


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

A PREPRINT

 **Esteban Ramos-Chavarriaga***

Department of Earth Sciences
EAFIT University
Medellín, Colombia
eramosc@eafit.edu.co

 **Juan F. Paniagua-Arroyave**

Department of Earth Sciences
EAFIT University
Medellín, Colombia
jpaniag2@eafit.edu.co

 **Jaap Nienhuis**

Department of Physical Geography
Utrecht University
The Netherlands
j.h.nienhuis@uu.nl

 **Juan D. Restrepo**

Department of Earth Sciences
EAFIT University
Medellín, Colombia
jdrestre@eafit.edu.co

September 15, 2021

ABSTRACT

Dominant dynamics of available wave energy dictate coastal gradients in alongshore sediment transport over wave-dominated continental shelves, where large-scale coastal evolution depend 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 Northern Andes.

Keywords Wave energy flux · Dominant dynamics · Data-driven architecture · Reduced-order model · Continental shelf · Coastal morphodynamics · Sediment transport

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 and Dean, 2003; Mortlock and Goodwin, 2015; Restrepo and Kjerfve, 2002; Salazar et al., 2018). Wave energy variability, driven by climate, wave-wave interactions, and multiple dissipation mechanisms, has been proven to relate significantly with seafloor topography (Ardhuin et al., 2009; Ardhuin et al., 2003). The stochastic interplay of oceanographic, hydroclimatic, and morphodynamic factors creates a complex dynamical system of nearshore wave climate, in which nonlinear processes regulate, balance, and produce the resultant wave behaviour and observed coastal impact.

*The corresponding author of this manuscript certifies that the contributors and conflicts of interest statements included in this paper are correct and have been approved by all co-authors.

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 long-term wave-induced morphodynamics and climate-driven patterns in coastal evolution.

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 robust data-driven techniques, these reanalysis parameters allow for unravelling wind-wave and wave-induced dynamic patterns (Brunton et al., 2020). Here, we analyse wave climate data from a coastal morphodynamics standpoint to develop a data-driven modeling architecture to assess, visualize and interpret the nonlinear dynamics of wave energy along continental shelves. We apply this architecture as a study case in the Pacific Coast of the Northern Andes compressional margin.

In particular, the problems we tackle in the present document are: i) the classification of P into low-rank regimes based on wave climate metrics, pioneering morphodynamical and geological characterisations, 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 gain morphodynamical insights and interpretability. 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, illuminating the importance of hidden leading-order variables in such complex systems.

2 Background and context

Ocean surface gravity waves exert decisive control on the morphodynamics of nearshore environments (Komar and Holman, 1986). "The shoreface" is 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 and 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 analysed 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 behaviour (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 and 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).

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 characterised by heavy rainfall, numerous rivers, and luxuriant vegetation (Correa and Morton, 2010; Morton et al., 2000) showing overall tropical meteorological conditions (Portilla-Yandún et al., 2019). The geological configuration of the littoral zone is characterised 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 the 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 and Gonzalez, 2000; Latrubesse and 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, characterised 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), characterised 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 and Kjerfve, 2002), with gently sloping platforms cut into Tertiary sedimentary rocks (Correa and Morton, 2010).

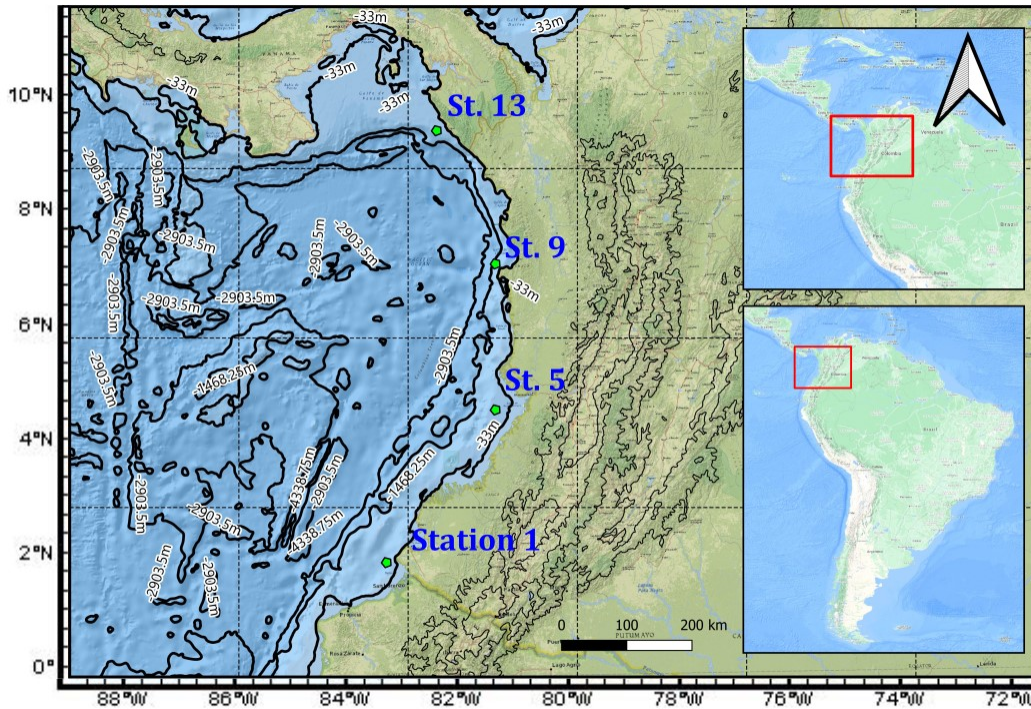


Figure 1: The Pacific oceanic basin and coastline of the Northern Andes at different spatial scales. Some of the study sites (red stroke points) shown in the figure represent significant results of each morphodynamically classified regime. The stations are located along the continental shelf around the 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).

3 Methods

We present the project data architecture proposed to assess wave energy by a variety of data-driven techniques (Figure 2). 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 ANOVAS, supervised and un-supervised learning algorithms, as well as wavelet and Fourier analysis. The data-driven predictive analytics consist 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 .

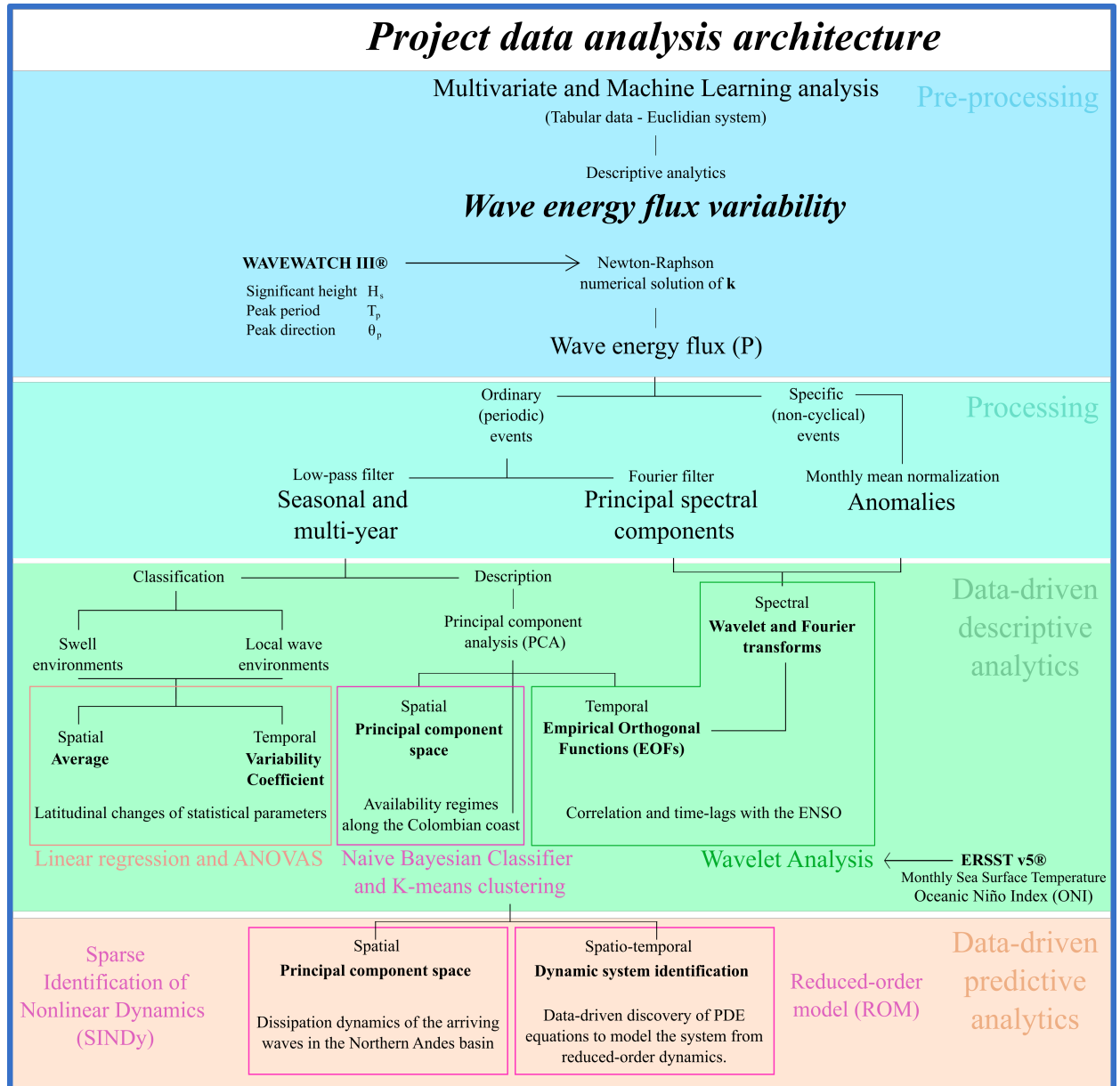


Figure 2: Data analysis architecture and modeling framework proposed to assess wave climate data, based on shoreface morphodynamics and coastal evolution. The architecture stages are pre-processing (cyan), processing (aquamarine), data-driven descriptive analytics (green), and data-driven predictive analytics (orange).

3.1 Wave energy flux from wave climate data

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. The modeling architecture proposed is graphically illustrated as a conceptual diagram in Figure 2. The assumptions we 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 $O(100 \text{ m})$, (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 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 castal 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 ($0.5^\circ \times 0.5^\circ$) and temporal resolution of 3 hours. Using the wave parameters from WW3 of H_s , T_p we calculated the wavenumber (k) to further compute the wave energy flux/transport (P in units of W/m) as (Holthuijsen, 2007 Eq. 5.5.12 therein),

$$P = Ec_g \quad (1)$$

where the energy is defined as $E = g\rho H_s/16$ (Lentz and Fewings, 2012), and the group velocity yields $c_g = 0.5\omega/k[1 + (2kh)/\sinh(2kh)]$, with the wavenumber (k) and intrinsic radian frequency (σ) following the well-known dispersion relation (i.e. Mei and Black, 1969 Eq. 1.4.11):

$$\sigma^2 = gk \tanh(kh) \quad (2)$$

where h is the water depth, σ is the intrinsic radian frequency. By linear wave theory $\omega = \sigma + kU$, where U is the time- and depth-averaged current velocity.

3.2 Data-driven analytics

3.2.1 Discovery of PDEs

We computed the measurement matrix of the system $X(i, n) = P(x, t)$ from time series of P , assuming the measured subsystem X (or sampled system) explains most of the defined system of interest. X contains a discrete collection of 87663 temporal realizations (n) of P at 13 coastal stations (i) along the continental shelf of the Pacific coast of the Northern Andes. By definition, the reanalysis data represent a high-fidelity statistical parametrization of climate-driven oceanic phenomena. Consequently, we do not have access to exact *partial differential equations* (PDEs) to describe the system. With these premises, we first enforce a general description of the dynamical system based on a nonlinear PDE, such as:

$$\frac{\delta X}{\delta t} = X_t = N(X, X_x, X_{xx}, XX_x, X_xX^2, X^3, \dots, x, t, \beta) \quad (3)$$

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 derivatives. 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 and 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 derivatives, and b is equal to the temporal derivatives 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:

$$X_t = \Theta(X, X_x)\alpha \quad (4)$$

with each column of Θ representing the candidate nonlinear terms, which are dependent on the systems state and its spatial derivatives. To solve the linear system, we apply a conventional least-squares regression ($A \setminus b$), as well as the convex relaxation of a sparse regression, referred to as the LASSO method (Rudy et al., 2017), which regularizes over α with the L1 and L2 norms.

Nevertheless, the optimization procedure is meaningless if the nonlinear terms defined as spatial numerical derivatives 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 derivatives of X (Tu et al., 2014). We can also compute the spatial derivatives (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 3).

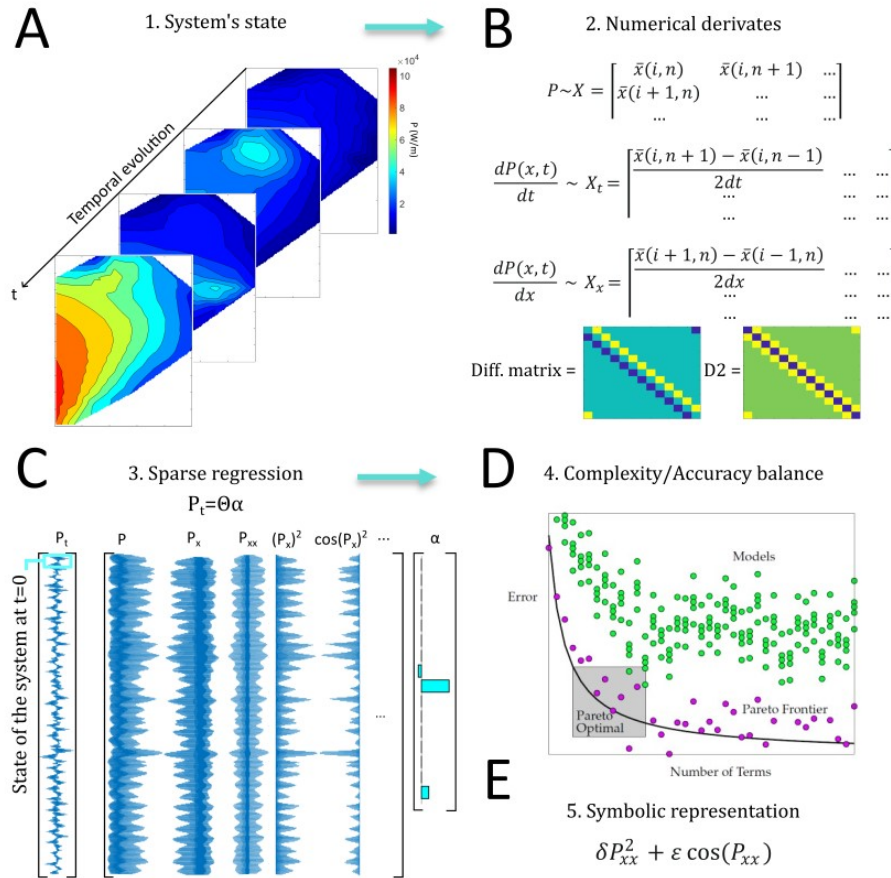


Figure 3: Data-driven discovery of PDEs, showing the model discovery and selection with the PDE-FIND algorithm (Brunton et al., 2016; Rudy et al., 2017). This methodology allows for a model selection that optimizes the mathematical description of the system's nonlinear behavior. Figure adapted from Brunton et al., 2016

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 Figure 3), 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 and 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 (monthly moving average 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.

3.2.2 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 and Kutz, 2017). 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 and Dean, 2007). 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), such that it can be expressed as the multiplication of three other matrices as:

$$X = U\Sigma V^T \quad (5)$$

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 (x, y and z) that best represents (in the statistical regression sense) the temporal variability of each continental shelf station. By multiplying each truncated temporal mode or eigen-process (V^T) by each station’s time series of X , we obtain a unique coordinate point representing each station in terms of the first three PCA temporal modes in the optimal coordinate system $x = V^T X(r = 1)$, $y = V^T X(r = 2)$, $z = V^T X(r = 3)$. This representation corresponds to the Principal Component Space or PCS (Figure 4), where we statistically model and classify each station’s unique “fingerprint”.

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 $X_t = AX$ (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.

3.2.3 Classification and clustering

To classify the spatial distribution of P in terms of the PCS, we applied the Naïve Bayes classifier (Langley and 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. 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 and 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

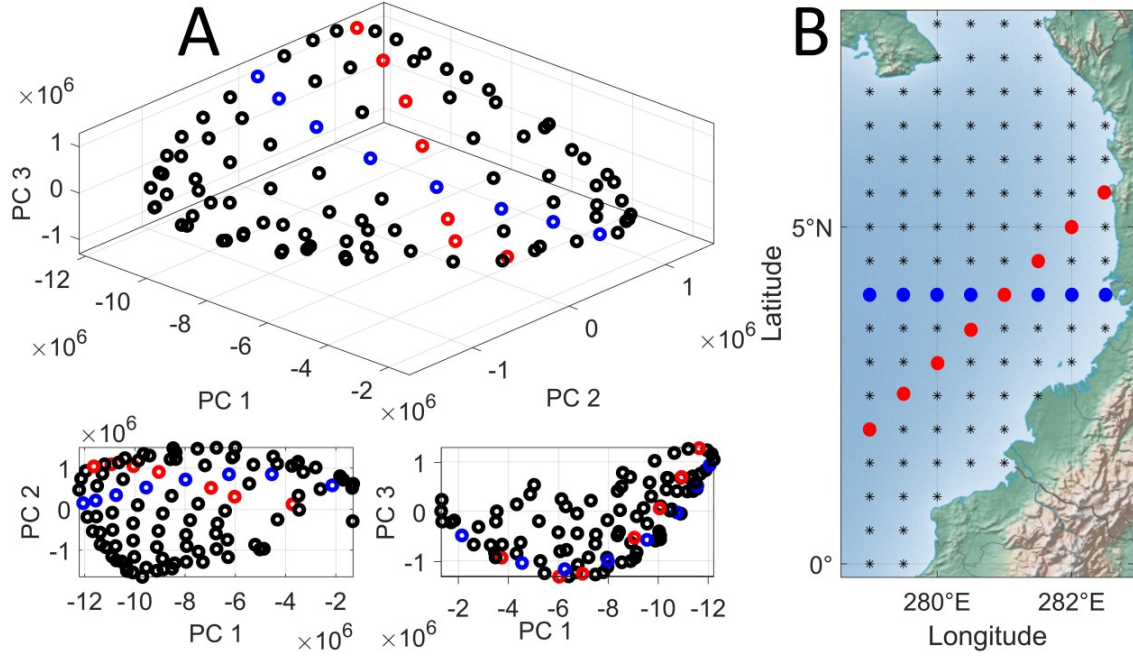


Figure 4: Principal Component Space of 101 stations of *wave energy flux* (P) variability along the Pacific oceanic basin of the Northern Andes. We highlight the longitudinal variability in blue, and the average incoming wave angle ($\phi_0 = 300$ Azimuth) in red. Panel A presents all 101 locations on 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 percent 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.

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 .

3.2.4 Wave-induced coastal morphodynamics

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

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

where

$$K_1 = 5.3e^{-6} K \left(\frac{1}{2n}\right)^{6/5} (\sqrt{g\gamma_b}/2\pi)^{1/5} \quad (7)$$

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), θ is the deep-water wave approach angle, and ϕ_0 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, Ashton, and Giosan, 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, Ashton, Nardin, et al., 2016 Suppl. Information):

$$E(\phi_0) = \frac{H_s^{12/5}(\phi_0)T_p^{1/5}(\phi_0)}{\sum_{\phi_0} H_s^{12/5}(\phi_0)T_p^{1/5}(\phi_0)} \quad (8)$$

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:

$$DoC_i = 2.28H_s - 68.5\left(\frac{H_s^2}{gT_p^2}\right) \quad (9)$$

$$DoC_o = (H_s - 0.3\sigma)T_p\left(\frac{g}{5000D}\right)^{0.5} \quad (10)$$

where g is the acceleration due to gravity, σ_s 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.

3.2.5 Climate-driven periodic variations

The time series analysis consist on 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 (Butterworth, 1930). We compare ordinary (or raw) data with the filtered time series to identify noise and frequencies of interest. Additionally, since the teleconnection indices (i.e., ONI for ENSO) represent anomalies in the 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 over the 30-year time span. 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 and seasonal time series, and the Oceanic Niño Index (ONI), to analyze the atmosphere-ocean coupling driving wave energy variability.

Wavelet transform (Torrence and Compo, 1998), or multiresolution spectral analysis, is the most robust technique to obtain the frequency content present in a signal (Kumar and 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 $f_0 = 0.033 \text{ cpy}$, 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 and Emery, 2014). Similarly, the Nyquist frequency is $f_N = 4 \text{ cpd}$. 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.

4 Results and discussion

4.1 Statistical and exploratory analysis

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 and 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 (DoC_i and DoC_o , respectively) 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 (S-N) 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 this results with the morphological parameters in Table 1, we assume that the wave climate present at the inner shoreface (DoC_i) is approximately the same as in the outer shoreface (DoC_o). Meaning wave hydrodynamics can be taken into account in our morphodynamic implementation.

Table 1: Wave and geomorphological parameters used in the present study, shown as latitudinal changes with a spatial resolution of $0.5 \times 0.5^\circ$. 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.

Stations		Mean H_s (m)	Mean T_p (s)	Mean P (W/m)	Depth (m)	Shelf width (m)
Regime	Coord. (lat, lon)					
Southern	1.5°N-79°W	0.7	11.1	4,600	223	96,000
Southern	2°N-79°W	0.9	12.2	9,200	494	87,000
Southern	2.5°N-78.5°W	0.9	10.6	6,200	80	67,000
Southern	3°N-78°W	0.7	10.8	5,700	127	116,000
Southern	3.5°N-77.5°W	0.7	10.2	4,500	62	67,000
Middle	4°N-77.5°W	0.8	12.2	6,800	267	47,000
Middle	4.5°N-77.5°W	0.8	13.1	8,200	618	40,000
Middle	5°N-77.5°W	0.9	13.8	10,700	924	31,000
Middle	5.5°N-77.5°W	0.9	14	11,300	1193	15,000
Middle	6°N-77.5°W	0.9	14.2	10,500	1759	47,000
Middle	6.5°N-77.5°W	0.9	14.4	12,900	651	29,000
Northern	7°N-78°W	1.1	14.5	18,900	2487	19,000
Northern	7°N-78.5°W	1.2	14.6	20,300	1374	10,000

From directional histograms, also known as wave roses of P (Figure 5-A), we observe the angular distributions of incoming waves from different directions (between 240° N and 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 (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.

We classified swell and local waves in terms of T_p (local waves: $0 < T_p < 10$ s and swell: $10 < T_p < 20$ s), 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 5-B). 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).

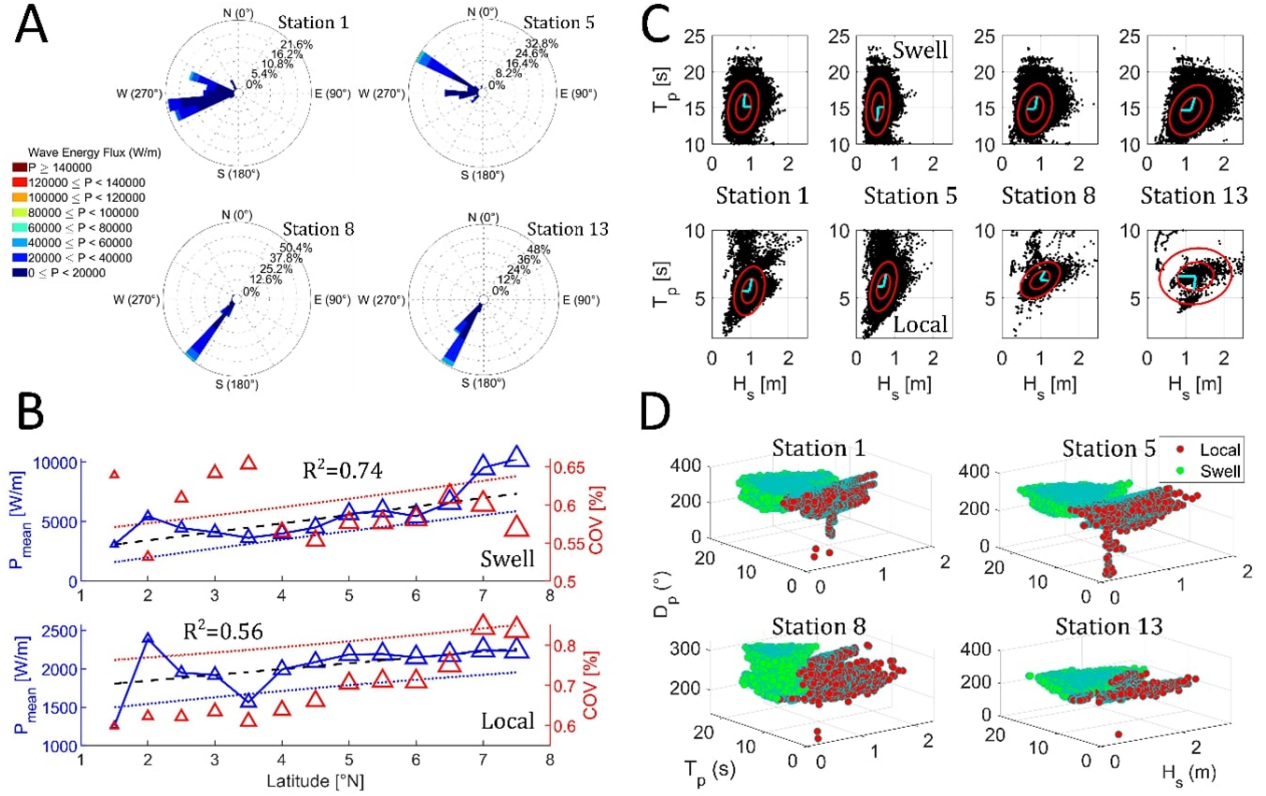


Figure 5: Exploratory analysis on wave climate reanalysis data at the Northern Andes Pacific coast. (A) Directional histograms of P angular distribution in stations 1, 5, 9, and 13 showing directional changes of the incoming waves. (B) Latitudinal changes of statistical parameters (mean and variability 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 characterisation 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.

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.

Variables	ANOVAS	
	Swell Energy vs Latitude	Local Energy vs Latitude
R^2	0.74	0.56
$RMSE$	1.296e+03	2.648e+02
SST	3.72e+07	3.68e+07
$pvalue$	0.13	0.03
t_{stat}	1.61	8.47
F_{Fisher}	22.13	5.24
$r_{Pearson}$	0.81	0.71
Equation	$y = 1510.17 + 904.3x$	$y = 1621.6 + 90x$
Degrees of freedom	12	12

We further applied a linear regression regularizing least squares to model P_{mean} in terms of latitudinal changes (Figure 5-B 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 (Table 2), 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 5-B) in the Northern Andes Pacific coast.

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 wavefronts with the undisturbed swell (Dimar, 2020). 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.

4.2 Model discovery and selection

To find the symbolic representation of our system, as a mathematical form of nonlinear terms, we must assume that the measurement matrix (X) is the solution to the PDE of interest. 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 6 presents the PDE discovery methodology 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.

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 term into the optimal basis to linearize each symbolic representation (ROM of each term). Then, we added all terms to obtain a 6-term PDE and a 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 critical dynamic evolution. By simulating each term and the complete PDE, we found the Pareto Front, defined as the curve's inflection point where the error starts increasing and the number of terms decreases (Figure 6-D). We find that the error using only the first term, P_x , is $NRMSE = 36.63\%$. whereas the error using four terms, $P_x + P_x^2 + PP_x + P_{xx}^2 + \cos(P_x)$, is $NRMSE = 37.16\%$. Thus, we select the PDE with one term and suggest that the error obtained from the model is almost entirely dependent on the basis set that embeds the system's dynamics, i.e., the ROM. Next, we perform an error analysis based on evaluating different training datasets (seasonal, ordinary, random, and uniform) on the reduced-order algorithms (PCA and DMD).

4.3 Reduced-order model

After discovering the ideal nonlinear term that best represents the data, we find the low-rank structure that allows linearizing the system. We computed the SVD algorithm applied to normalized training data to generate the PCA modes along the basin and the coastline (Figure 7-D). Further, we use the Koopman operator to produce the DMD modes (Figure 7-F). We contrast the different basis sets with three and five-mode expansions of the system to solve the reduced-order dynamics. We found the simplest and ideal model of the reduced-order dynamical system of interest to be a function of only one spatial derivative:

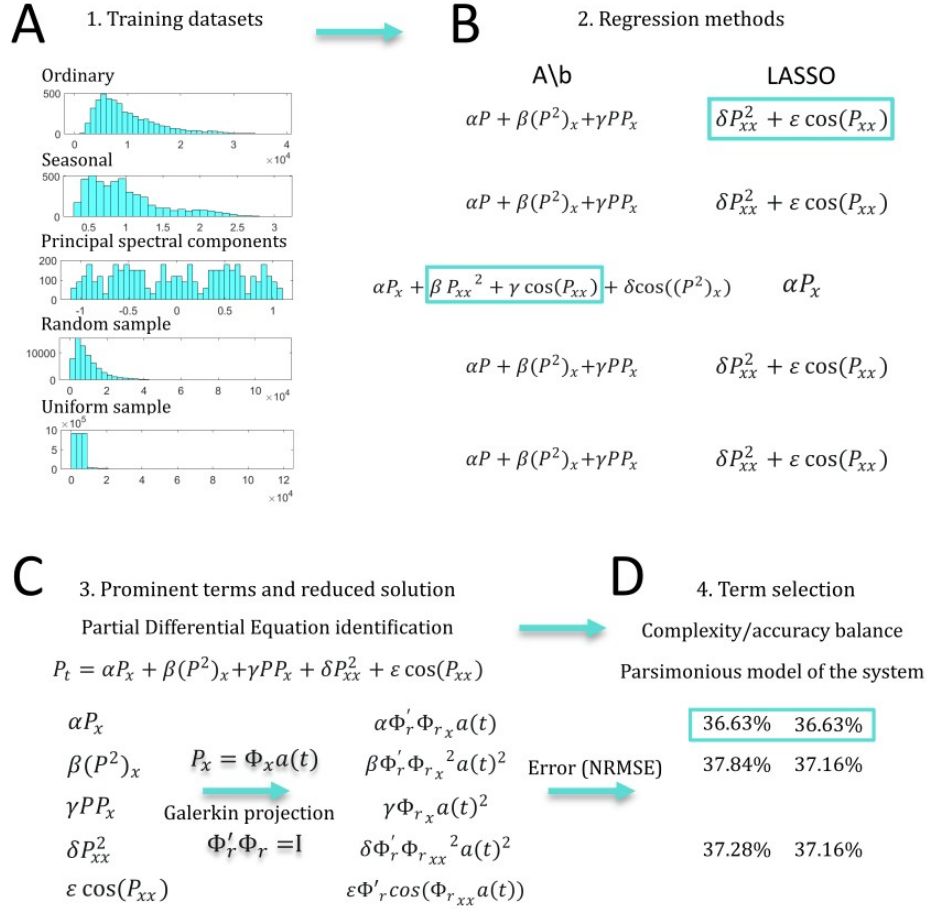


Figure 6: Methodological diagram proposed in the present study. We applied it to different training datasets (step 1) to discover the possible nonlinear terms (step 2), select the appropriate reduced model (step 3), and evaluate the simulated results versus actual data (step 4). This result illustrates 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.

$$P_t = X_t = \alpha X_x + \beta \quad (11)$$

where α is the coefficient for the nonlinear term, and β represents the bifurcation parameter space. By assuming the separation of variables $X(x, t) = \phi_r(x) a_r(t)$ 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 $\phi_r(x)$, such as:

$$X_t = \phi_r(x) \frac{\delta a_r(t)}{\delta t} = \alpha \left[\frac{\delta \phi_1(x)}{\delta x} a_1(t) + \frac{\delta \phi_2(x)}{\delta x} a_2(t) + \frac{\delta \phi_3(x)}{\delta x} a_3(t) \right] + \beta \quad (12)$$

Due to the orthogonality properties of the SVD algorithm, we assume $\phi_r' \phi_r = I$ by the Galerkin projection theory so we can multiply Eq. 12 by ϕ_r' to obtain a reduced-order dynamics of $a(t)$ as:

$$\frac{\delta a_r(t)}{\delta t} = \alpha \left[\phi_1' \frac{\delta \phi_1(x)}{\delta x} a_1(t) + \phi_2' \frac{\delta \phi_2(x)}{\delta x} a_2(t) + \phi_3' \frac{\delta \phi_3(x)}{\delta x} a_3(t) \right] + \beta \quad (13)$$

Eq. 13 represents the ODE of the reduced-order dynamics of wave energy flux P along the compressional continental shelf of the Northern Andes Pacific coast. We numerically solve the ROM using the ODE45 algorithm in MATLAB®, which outputs the dynamical solution of each $\frac{\delta a_r(t)}{\delta t}$. All the terms can be computed using numerical derivatives and matrix transformations, solved using numerical approximations. We finally reconstruct the system as:

$$X_{simulated} = \phi_r(x) a_r(t) \quad (14)$$

4.4 Model implementation

4.4.1 Morphodynamical 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 7 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 7-B right panel). Explaining 95% of the total variability, according to the cumulative percent of the variance of the singular values (Figure 7-B left panel).

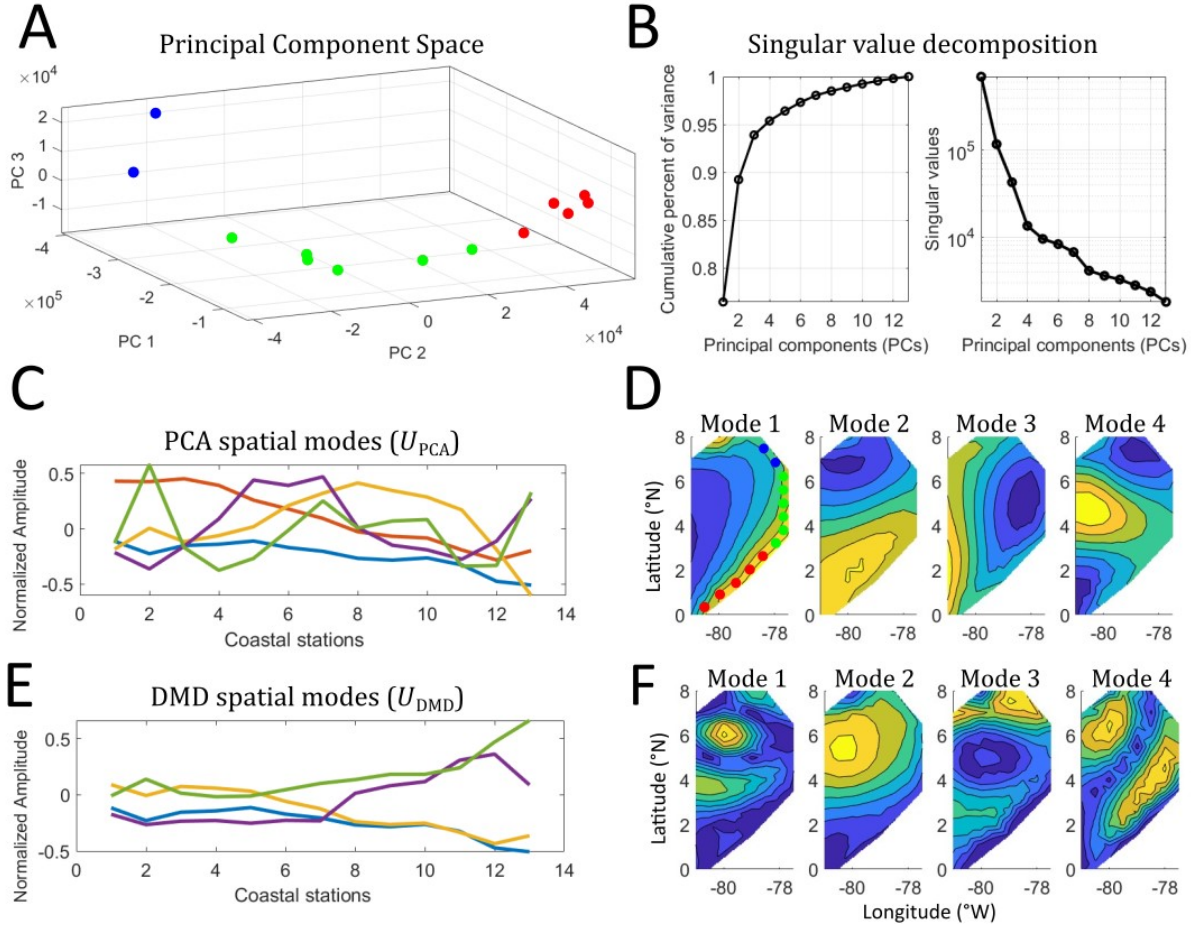


Figure 7: Data-driven modal decomposition of wave climate data at the Northern Andes Pacific basin and coastline. The singular value decomposition (SVD) from both PCA and DMD algorithms produces a dimensionality reduction space (A) to express the fundamental variability trends in the data. From the singular 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 7-A) 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 7-C and E, and Figure 7-D 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.

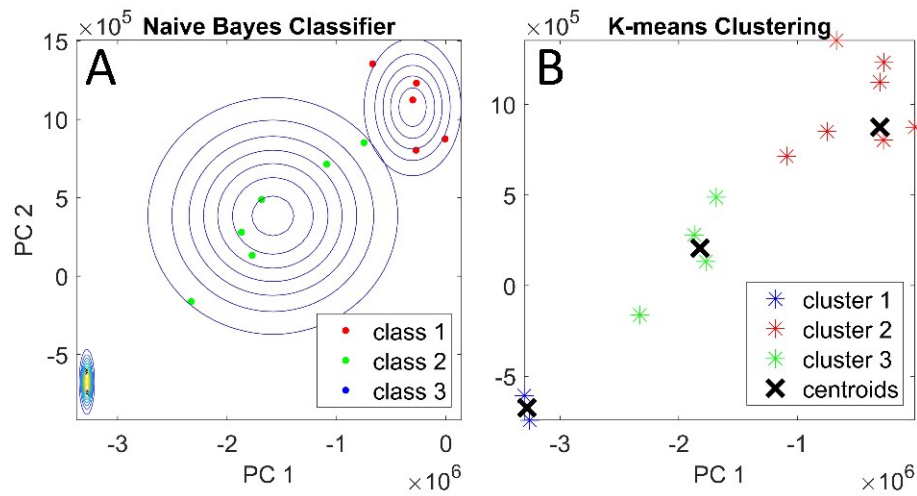


Figure 8: (A) Supervised and (B) unsupervised learning algorithms applied to the PCS of coastal P. The Naïve Bayes Classifier uses the previous expert classification of empirical coastal regimes provided as labels to obtain statistical significance on three distinct regimes. The K-means clustering algorithm 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 the PCS, we generated three modeling approaches for classifications of the morphodynamical coastal regimes: empirical (Figure 7-A), supervised, and unsupervised (Figure 8-A and B, respectively). From previous characterizations of wave-induced morphodynamics along the Pacific coast (Correa, 1996; Correa and Gonzalez, 2000; Correa and 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 and Morton, 2010). It does so by grouping the stations into morphodynamical regimes obtained from the spatio-temporal variability of P .

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.

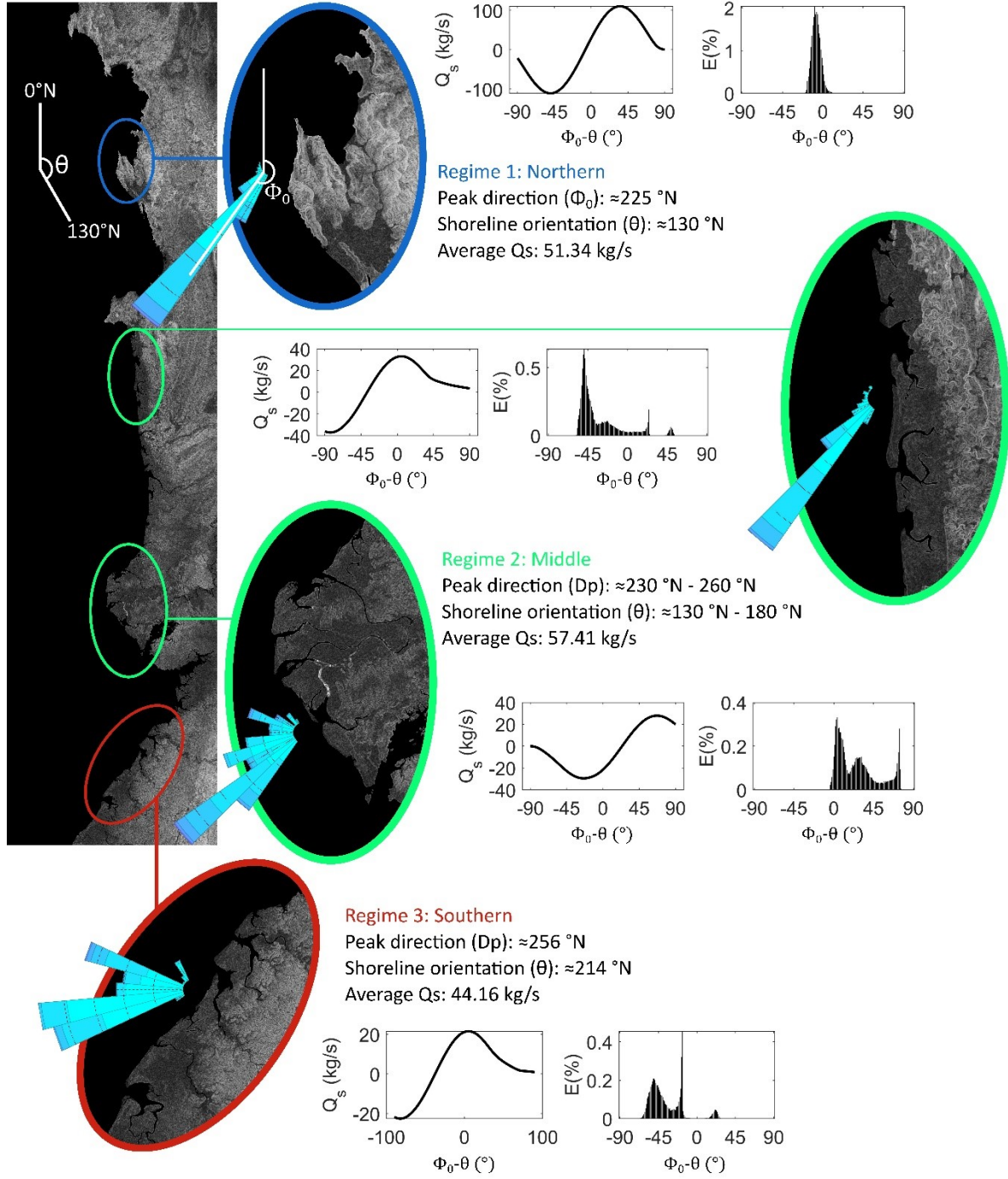


Figure 9: Morphodynamical classification of coastal regimes using wave climate metrics of alongshore sediment transport (Q_s) and directional energy (E) versus relative incoming wave angle ($\phi_0 - \theta$). In colors, we represent the distinct regimes classified to relate specific morphodynamics conditions. Stable morphodynamics (Northern regime in blue), transition between stable and unstable morphodynamics (Middle regime in green), and unstable morphodynamics (Southern regime in red).

We argue that this classification would serve as a basis for assessing future wave-induced coastal evolution scenarios. Figure 9 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 and 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 and López, 2008). For example, at the San Juan delta (Figure 9, 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 and Cantera, 2013; Restrepo and 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 9). 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.

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 ($\theta > 180$) and the high breaking wave angles ($\phi_0 - \theta > 45$) 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, generating the resultant swell dissipation trends and variability.

4.4.2 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 10 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 10-C 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 normalized root mean square error (NRMSE) analysis, and linear regression.

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

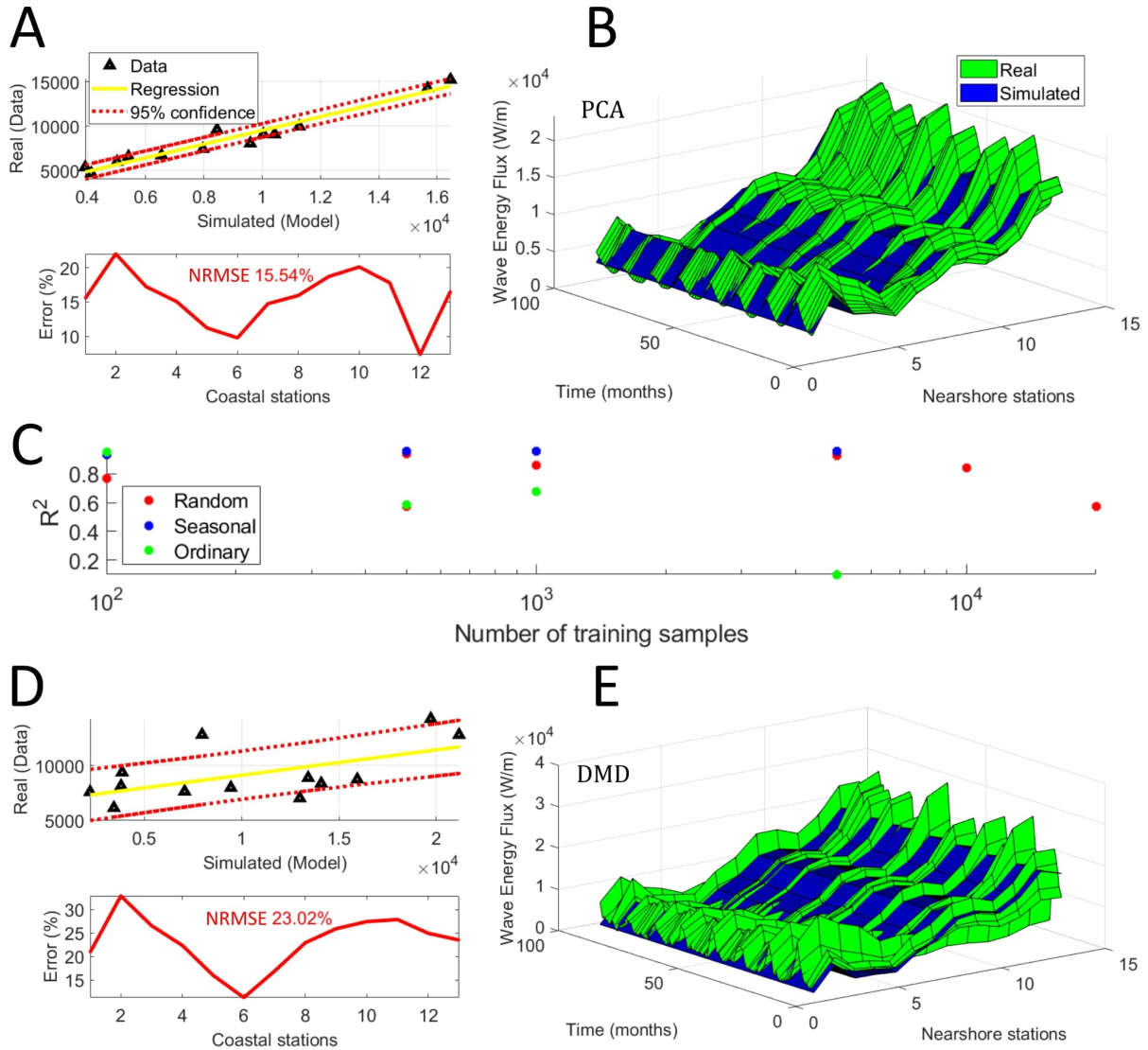


Figure 10: Model evaluation of the best-fitted ROMs using PCA (D and E) and DMD (A and B). (C) Pareto analysis, i.e., accuracy vs. complexity, as applied to the different training datasets.

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, i.e., αX_x (Whitman, 1999), we added the diffusion term of Burger's 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 derivatives to compute a linearization that captures the fundamental spatial variability of P along continental shelves.

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.

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.

5 Conclusions

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 ($R^2 = 0.86$ 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.

6 Acknowledgments

The authors would like to acknowledge the Vice presidency of Discovery and Creation at EAFIT University for funding this research project and master thesis (code 952-000015). This manuscript could not be possible without the guidance and support of Olga L. Quintero, José F. Duque, Marcela Jaramillo, Juan J. Restrepo B., and many others whose conversation enriched widely this document.

References

Almar, R., Ranasinghe, R., Bergsma, E. W. J., Diaz, H., Melet, A., Papa, F., Vousedoukas, 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), 125–153. [https://doi.org/10.1016/S0378-3839\(03\)00061-9](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>
- 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.
- 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>
- 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 cusped 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 [Publisher: Unpublished]. <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](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]*. Retrieved June 26, 2021, from <http://arxiv.org/abs/2006.11287>
- Comment: Accepted to NeurIPS 2020. 9 pages content + 16 pages appendix/references. Supporting code found at https://github.com/MilesCranmer/symbolic_deep_learning
- 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 classification. *John Wiley & Sons, Inc.*, 69.
- 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), 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](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>
- Holthuijsen, L. H. (2007). *Waves in oceanic and coastal waters*. Cambridge University Press.
- Ingeominas. (1996). Atlas geomorfología y aspectos erosivos del litoral caribe colombiano.
- Köhler, A., Ohnberger, M., & Scherbaum, F. (2010). Unsupervised pattern recognition in continuous seismic wave-field 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>
- Kumar, P., & Fofoula-Georgiou, E. (1997). Wavelet analysis for geophysical applications. *Reviews of Geophysics*, 35(4), 385–412. <https://doi.org/10.1029/97RG00427>
- Langley, P., & Sage, S. (1994). Induction of selective bayesian classifiers. *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>
- Lentz, S. J., & Fewings, M. R. (2012). The wind- and wave-driven inner-shelf circulation. *Annual Review of Marine Science*, 4(1), 317–343. <https://doi.org/10.1146/annurev-marine-120709-142745>
- 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., Osorio Arias, A., Ortiz Royero, J., & Ocampo-Torres, F. (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., Brichenon, 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., & Corraej, I. D. (2000). Frequent non-storm washover of barrier islands, pacific coast of colombia. *Journal of Coastal Research*, 16, 6.
- 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. (2016). Littoral steering of deltaic channels. *Earth and Planetary Science Letters*, 453, 204–214. <https://doi.org/10.1016/j.epsl.2016.08.018>
- 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., Nardin, W., Fagherazzi, S., & Giosan, L. (2016). Alongshore sediment bypassing as a control on river mouth morphodynamics. *Journal of Geophysical Research: Earth Surface*, 121(4), 664–683. <https://doi.org/10.1002/2015JF003780>
- 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-Aroyave, 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 cusped 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>
- 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>
- 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](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>
- 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. d. 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., Wagemann, 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\(245\)](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>
- 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>
- Veldhuizen, D. A. V., & Lamont, G. B. (1998). Evolutionary computation and convergence to a pareto front. *In Late breaking papers at the genetic programming*, 221–228.
- 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>
- Whitman, G. B. (1999). *Linear and nonlinear waves* (Vol. 42). John Wiley Sons.
- Zhang, J. M., Zou, L., Sun, T.-Z., Wen, Z.-H., & Yu, Z.-B. (2020). Experimental investigation on the propagation characteristics of internal solitary waves based on a developed piecewise dynamic mode decomposition method. *Physics of Fluids*, 32(8), 082102. <https://doi.org/10.1063/5.0015960>
- Zhang, Z., & Liu, Y. (2021). Robust data-driven discovery of partial differential equations under uncertainties. *arXiv:2102.06504 [cs, math]*. Retrieved June 8, 2021, from <http://arxiv.org/abs/2102.06504>
- 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>