Energy for Sustainable Development 93 (2026) 102004

Contents lists available at ScienceDirect

Energy for Sustainable Development

journal homepage: www.journals.elsevier.com/energy-for-sustainable-development

Assessment of hydrokinetic energy in the Cozumel ocean current: Turbine design considerations

Gerardo Cano-Perea [a], Ricardo Arévalo-Mejía [a], Humberto Salinas-Tapia [a], Carlos Díaz-Delgado [a], Edgar Mendoza-Baldwin [b], Boris Miguel López-Rebollar [c,*]

[a] Instituto Interamericano de Tecnología y Ciencias del Agua, Universidad Autónoma del Estado de México, Carretera Toluca-Atlacomulco km 14.5, 50200 Toluca, Estado de México, Mexico

[b] Instituto de Ingeniería de la Universidad Nacional Autónoma de México, Ciudad Universitaria, Edificio 17, 04510 México City, Mexico

[c] Department of Basic Science and Engineering, Universidad Autónoma Metropolitana, Unidad Lerma, Lerma de Villada 52005, Estado de México, Mexico

ARTICLE INFO

Keywords:
Ocean energy
Ocean currents
Hydrokinetic turbine
Clustering methods
Power curve modelling
Power potential
Capacity factor

ABSTRACT

In recent years, ocean currents have been identified as a promising resource for renewable energy generation. However, most conversion devices remain at an early stage of technical development, hindering progress towards large-scale commercial deployment. In this scenario, the resource assessment, with particular attention to hydrodynamic characterisation, re-emerges as an essential activity that demands comprehensive knowledge to meet the requirements of any generation project and to support the reduction of the cost of energy. The present research aims to evaluate the energetic potential of an ocean current located in the insular shelf of Cozumel Island, emphasising on the definition of the fundamental parameters needed for the hydrodynamic design of a medium-scale horizontal-axis hydrokinetic turbine with a radius of about 0.5 m. The evaluation was divided into two main sections. In the first section, 192 current speed time series were obtained from the HYbrid Ocean Coordinate Model (HYCOM), each associated with a geo-referenced node and a depth of 50 m. Subsequently, the K-Means++ algorithm was applied to cluster nodes based on their statistical features. Three specific approaches were considered for the feature extraction: statistical moments, principal component analysis, and deep autoencoding. The clusters were compared using the Silhouette, Davies–Bouldin and Calinski–Harabasz scores. While the first and second approaches produced quite similar and consistent results, the performance of the third approach was unexpectedly inferior. In the second section, four potential harnessing regions were chosen. For each region, the optimal power per node and cluster was computed using 25 generic power curves. It was assumed that nodes within the same cluster shared a single optimal curve. For the most promising region, the best-performing power curve suggests a cut-in speed of 0.69 m/s, a rated speed of 2.3 m/s and an effective speed for hydrodynamic design of 1.52 m/s. The maximum capacity factor observed was 0.325, with a power production of 6.55 MWh/yr m^2. Although reported results focus on the maximum yield, alternative scenarios for power production are also discussed. Finally, a preliminary set of parameters to describe the operative state of the hydrokinetic turbine is presented.

Introduction

In response to concerns about climate change and sustainable development, countries have gradually updated their renewable energy policies. For instance, in early 2022, the European Commission raised its 2030 targets for the share of renewable energy in final energy consumption to 45% (REN 21, 2022). Nevertheless, the current reliance on fossil fuels reveals an evident gap between aspiration and actual progress. In this scenario, ocean energy emerges as a promising alternative to diversify the energy sector. Compared to well-established solar and wind energy, ocean energy offers the advantage of being more stable and predictable over time, simplifying its integration into traditional power grids. Preliminary studies indicate that the potential for annual energy extraction could reach 331 exajoules (Lewis et al., 2011), a noteworthy magnitude relative to the 650 exajoules of global energy demand recorded in 2024 (IEA, 2025).

Hernández-Fontes et al. (2019) provided an extensive evaluation of the potential of ocean energy in Mexico. The research identified multiple locations suitable for harnessing energy from sources such as waves,

* Corresponding author.
E-mail addresses: gcanop120@alumno.uaemex.mx (G. Cano-Perea), rarevalom@uaemex.mx (R. Arévalo-Mejía), hsalinast@uaemex.mx (H. Salinas-Tapia), cdiazd@uaemex.mx (C. Díaz-Delgado), EMendozaB@iingen.unam.mx (E. Mendoza-Baldwin), b.lopez@correo.ler.uam.mx (B.M. López-Rebollar).

https://doi.org/10.1016/j.esd.2026.102004
Received 29 May 2025; Received in revised form 21 March 2026; Accepted 21 March 2026
Available online 20 April 2026

salinity and thermal gradients, and ocean currents. Regarding ocean currents, the eastern coast of the state of Quintana Roo was identified as the most promising site, with estimated power densities ranging from $32\ \mathrm{W/m^2}$ to $512\ \mathrm{W/m^2}$. It should be noted that reported values represent the theoretical hydrokinetic energy available and do not account for energy conversion losses. Prior research conducted by Cetina et al. (2006) examined the circulation patterns of the water currents in this region, identifying a predominant northeastward flow direction. In addition, some reversal periods were also observed in specific southern areas. Athié et al. (2011) investigated the Yucatan Channel and its correlation with the Cozumel Channel, a site where the current intensifies and the surface velocities increase up to 1.0 m/s. Although the core area of the Yucatan Current remained stable, the position of the maximum velocity varied throughout the research period.

In recent years, Alcérreca-Huerta et al. (2019) evaluated the hydrokinetic potential of the current near the Cozumel Channel. The study identifies the northern insular shelf-edge of Cozumel Island as the most suitable region, with an average velocity of 0.93 m/s and peak values in the range of 1.8–1.9 m/s. However, the reported velocities are lower than those required by the hydrokinetic turbines closest to commercial deployment, which are typically designed for operative velocities around 2.0 m/s. Despite this limitation, the current can be assumed to be viable for energy extraction as a result of a nearly continuous and unidirectional flow. Arredondo-Godínez (2020) provided an extensive guide to evaluate potential sites for the extraction of hydrokinetic energy, using the Cozumel Channel as a case study. This guide presents a comprehensive oceanographic characterisation of the current, providing valuable insights for subsequent turbine design activities. Notable findings include the spectral analysis of the current velocity and the detection of the principal oscillation modes, which are particularly relevant for the hydroelastic and structural design of the blades.

A techno-economic assessment was conducted by Bárcenas Graniel et al. (2021) at the same location, considering the operational parameters of four pre-commercial turbines. For the optimal device, the levelised cost of energy (LCOE) was estimated to be around 1.1458 USD/KWh, with an associated capacity factor of 12.5%. This value is not competitive compared to the energy generation costs of conventional renewable sources, which are typically around 0.24 USD/kWh (IRENA, 2022). To enhance competitiveness, a turbine specifically designed to account for the hydrodynamic characteristics of the Cozumel current is required. Achieving a capacity factor of at least 25% could reduce the LCOE to 0.685 USD/KWh.

In a similar research, Ring et al. (2022) performed an extreme value analysis for the velocity current data derived from the HYCOM numerical model. The author concludes that, while the adjusted HYCOM model fits well with empirical data, it exhibits an underestimation of the extreme values of approximately 22%. This underestimation has critical implications for engineering activities, particularly the design of turbine support structures and mooring systems. In a subsequent study, Maslo et al. (2023) investigated the potential impact on the ocean current in the Cozumel Channel resulting from the installation of multiple ideal turbines. The results suggest that, while the installation of a few turbines has a minor impact compared to the scale of the ocean current, if the number of turbines increases, severe flow disturbances can be expected, which could potentially modify the seabed morphology and ecosystems.

Although the Cozumel Channel has been identified as a potential site to harness energy, the current velocity tends to be lower compared to other locations, particularly those driven by tidal currents. This limitation represents a considerable challenge, as effective hydrokinetic energy extraction requires the development and optimisation of turbine models that have not been widely studied in the industry. As noted by Bárcenas Graniel et al. (2021), there are at least three other sites worldwide with similar low-energetic conditions: the Agulhas current on the southern coast of Africa (Gründlingh, 1983), the Kuroshio current along the Japanese coast (Sawada & Handa, 1998), and the Gulf Stream near the Florida coast (Duerr & Dhanak, 2012).

Chen (2010) proposed a design of a power plant to harness energy from the Kuroshio current, estimating that a pilot plant could generate approximately 30 MW. Shirasawa et al. (2016) conducted an experimental verification of a small-scale floating turbine for this current. The turbine, with a rotor diameter of 1.46 m, achieved a power coefficient of 0.43 for an operational speed of 1.27 m/s. Subsequently, a turbine prototype, known as Kairyu, was redesigned to reach a pre-commercial stage, but detailed information remains out of the literature. Similarly, for the Gulf Stream, the Ocean-Based Perpetual company has developed a prototype intended for deployment near the Florida coast. In a technical approach, Encarnacion et al. (2019) developed a turbine prototype specifically designed to address the limitations of less energetic flows. Research suggests that a low-solidity turbine (TSR = 7.75) can achieve a power coefficient of 0.405, with the added advantage of producing power through low-torque operation. This design strategy was employed to reduce the size and costs of the generator, potentially improving the viability of the turbine. Boretti (2020) reviewed several turbine projects at the MW-level generation scale, even though not all were explicitly developed for slow-moving ocean current applications.

Research previously reviewed illustrates different stages of the technical development of hydrokinetic conversion devices. However, compared to the wind energy industry, which benefits from well-established technical standards and regulatory frameworks, the ocean energy industry faces persistent knowledge gaps that hinder its progress towards commercial viability. Although this situation can be perceived as challenging, it also provides an opportunity to explore innovative methodologies to mitigate constraints.

The present study aims to estimate the hydrokinetic potential near the Cozumel Channel using current speed time series data from a set of 192 geo-referenced nodes. In contrast to previous assessments that primarily focused on resource quantification, this study integrates time-series clustering with power curve analysis to simultaneously characterise both the hydrokinetic resource and preliminary operational design parameters for a medium-scale turbine. The assessment was divided into two main sections. The first section addresses the mapping of the hydrokinetic resource by applying time-series clustering techniques to categorise nodes. Subsequently, the operational parameters for the candidate device were defined by conducting a power curve analysis in four regions of interest, which were pre-selected based on the results from the previous section.

Time-series clustering is a widely used methodology for identifying patterns in temporal data. In literature, notable examples related to the energy industry include the identification of electricity consumption patterns (Räsänen & Kolehmainen, 2009), the measurement of the impact of climate change on wind speed (Clifton & Lundquist, 2012), the improvement of wind power forecasting (Liu et al., 2014; Pravilovic et al., 2014), the evaluation of wind resources (Abreu et al., 2025; Azhar & Hashim, 2023), and the diagnosis and predictive maintenance of wind farms (Adedeji et al., 2022; Martins et al., 2024). While multiple techniques for time series clustering have been proposed (Aghabozorgi et al., 2015; Javed et al., 2020; Wang et al., 2006; Warren-Liao, 2005), for this research, a feature-based time series clustering is adopted as a fast-implementation approach to identify spatially coherent regions for power generation. Therefore, this constitutes an alternative technique for energy resource mapping. The primary assumption is that generated clusters should contain nodes with similar hydrodynamic conditions, such that a single turbine model should be able to operate efficiently in the surroundings. In the absence of performance data of commercial turbines to test different array configurations, the farm layout relies on the clusters computed from the statistical properties of the nodes and their corresponding time series.

In a similar scenario, several studies have been conducted to estimate the hydrokinetic potential of ocean currents (Alcérreca-Huerta et al., 2019; Duerr & Dhanak, 2012; Haas et al., 2017; Kabir et al., 2015; Ren et al., 2023; Yang et al., 2013). Moreover, advanced methodologies, many of them inherited from the wind energy industry, are capable of integrating economic, operational, and environmental factors (Charhouni et al., 2019; Haces-Fernandez et al., 2022; Herbert-Acero et al., 2014). Unfortunately, due to the relatively early stage of research on this topic, conducting a comprehensive assessment introduces significant challenges as data availability is limited. Given these constraints, a more pragmatic approach to deal with this problem is to assume an iterative process, where the assessment will be refined and complemented in successive iterations. For the present iteration, the assessment is primarily focused on the power output as the key performance metric, which is determined by a power curve analysis.

Lewis et al. (2019, 2021) provide a comprehensive review of the most relevant hydrokinetic turbine prototypes that are progressively approaching commercial deployment. While available data is predominantly focused on energy extraction from tidal currents, the operational principles are analogous to those employed for ocean current applications (International Renewable Energy Agency, 2014; Lewis et al., 2011). From the previous research, a provisional approach was adopted to formulate a generic power curve set, with operational parameters derived from the averaged specifications of the reported devices. Each power curve represents a candidate turbine, whose power output was estimated for each cluster. The optimal power curve was selected based on the maximum cumulated power production as a criterion, enabling the derivation of fundamental operational parameters, including the cut-in, cut-out, rated, and effective speeds. Subsequently, by including complementary parameters, such as rotor diameter, tip speed ratio, and density and viscosity of seawater, a robust preliminary framework can be established to support the hydrodynamic design of the device.

Materials and methods

In the context of ocean energy, while single descriptive parameters, such as the minimum, maximum, or mean current speed, are useful for preliminary resource mapping, they are not enough when the objective is the delimitation of operational regions. This restriction arises from the fact that similar hydrodynamic conditions can be associated with different flow patterns. For instance, two sites with nearly identical mean speeds or power generation might be derived from different data distributions, making the operation of a single turbine model impractical. Feature-based clustering deals with this problem by categorising nodes based on the similarities of a computed vector of statistical properties of the time series, hereafter referred to as feature-vector. The specific algorithm selected for clustering is the K-Means++, a variant of the K-Means algorithm that optimises the seeding technique for centroid initialisation (Arthur & Vassilvitskii, 2007).

Time series clustering

To define the optimal set of statistical parameters to compose the feature-vector, three different approaches were considered: statistical moments, principal component analysis (PCA) variables, and autoencoder-derived variables. Before describing the procedures and techniques for obtaining these parameters, the following section introduces the case study.

Case study: dataset

The case of study is the Cozumel Channel, a narrow passage in the Caribbean Sea bounded by the coast of Quintana Roo to the west and by the Cozumel Island to the east. As noted by Arredondo-Godínez (2020), this channel is connected to a large-scale system driven by the Yucatan Current, which flows northward along the Quintana Roo coastline. Near Cozumel Island, the flow accelerates due to the constriction of the channel and bathymetry. As it enters the Gulf of Mexico, the Yucatan Current transitions into the Loop Current, which subsequently becomes the Florida Current. According to Chávez et al. (2003), despite its relatively small dimensions, measuring approximately 50 km in length and 18 km in width, the Cozumel Channel can transport 5 Sv,[1] which represents around 20% of the mean transport of the Yucatan Current. Bárcenas Graniel et al. (2021) provide further information about the electricity consumption, generation, and current power grids on the island, while Muckelbauer (1990) describes the morphology and organisms.

The dataset used in the cluster analysis was obtained from a reanalysis of the HYbrid Ocean Coordinate Model (Chassignet et al., 2007). This product was selected as it offered relatively high spatial resolution and a suitable temporal range (Duerr & Dhanak, 2012; Kabir et al., 2015; Neary et al., 2012). Current velocities were calibrated using in situ measurements and the HYCOM + NCODA Gulf of Mexico 1/25° GOMl0.04 experiment. The in situ measurements were retrieved from a stationary long-range ADCP deployed near the central section of the channel, at a depth of 49 m and reporting at half-hour intervals during the period from April 2009 to May 2011. The overestimation of the zonal component was reduced from 19.3% to 10.2% in terms of the NMSE.[2] Numerical data for the selected region (see Fig. 1) were computed for a set of 192 nodes at a depth layer of 50 m, each providing hourly current velocity time series from 2 April 2009 to 31 March 2014. This activity partially adopted the methodology presented by Duerr and Dhanak (2012), and was carried out as part of the initiatives of the Mexican Centre for Innovation in Ocean Energy (CEMIE-Oceano) and the CANEK project.

The time series were examined for stationarity using the augmented Dickey–Fuller test, with a number of lags that were selected according to the AIC information criterion (Fuller, 1995). Moreover, a Hampel filter, applied considering a minimum window size of 6 h and a detection threshold of three standard deviations ($\sigma = 3$), demonstrates that the number of data entries that can be classified as outliers is not significant compared to the length of the time series. As the data were derived from a numerical model, no additional pre-processing steps were required. Nevertheless, a visual inspection of the frequency histograms revealed certain nodes where the distribution of the data suggests bimodality.

To validate the observations, a Hartigan-Dip test for unimodality was performed using R software (Hartigan & Hartigan, 1985; Maechler, 2003). From this analysis, it was possible to identify nine nodes with an evident bimodal distribution. However, certain nodes exhibited an excess mass in their data distributions, suggesting potential deviations from rigorous unimodality, even if they did not meet the criteria for being classified as bimodal. This situation implies that fitting continuous functions for each node and comparing their parameters (i.e., shape, scale, and location) is not a feasible method for feature extraction, and different alternatives must be explored.

Statistical moments

As a preliminary data exploration stage, ten descriptive statistical variables were estimated for each node in the dataset, including mean, standard deviation, quantiles (25%, 50% and 75%), coefficient of variation, skewness, kurtosis, and metrics to measure stationarity (Augmented Dickey–Fuller test) and homoskedasticity (Breusch–Pagan test). A correlation matrix was calculated, and it was possible to appreciate a relatively high correlation between variables. Breiman (2001) and Keogh and Mueen (2017) suggest that high dimensionality is usually a problem for clustering activities, including the K-Means algorithm that relies on the Euclidean distance (L2-distance) to define centroids. Despite not being an absolute rule and depending primarily on the nature of the data, dimensionality reduction was prioritised.

To ensure a more rigorous mathematical representation of the data distribution, the parameter set was reformulated using the first seven standardised raw moments as the initial guess. Raw moments were

[1] Sverdrup, unit of volumetric flow equivalent to 1×10^6 m^3/s.

[2] NMSE: Normalised Mean Squared Error.

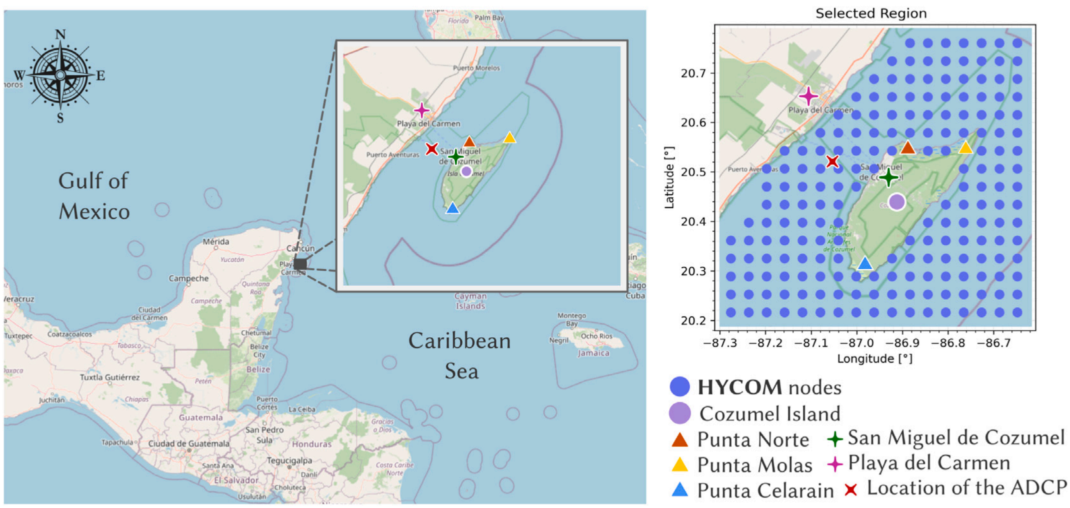

Fig. 1. Location of Cozumel Channel and HYCOM node distribution.

deliberately selected instead of central moments to reduce the dependency on the mean value. To detect and address multicollinearity issues, a variance inflation factor (VIF) analysis (Akinwande et al., 2015) was conducted. The results indicated that three raw moments could be excluded and confirmed that four raw moments were enough for clustering activities. For this first approach, the standardised (Z-score) raw moments selected to compose the feature-vector were the second, third, fourth, and fifth. Regarding the detection of bimodality in certain nodes, Pfister et al. (2013) and Kang and Noh (2019) discuss the *coefficient of bimodality* as a parameter that indirectly measures bimodality through the relationship between the third and fourth central moments. Although these specific moments were not included in the analysis, it is assumed that their effects are partially captured due to the inherent relationship between raw and central moments.

Principal component analysis

For the second approach, principal component analysis (PCA) was considered a suitable alternative to reduce dimensionality (Ayesha et al., 2020; Rastogi et al., 2023; Salem & Hussein, 2019). The initial set of parameters contained the same standardised seven raw moments computed previously. The analysis was performed using the open-source Python library scikit-learn. Given a relatively high correlation, the unsupervised Kernel-PCA analysis was used to transform a baseline feature space into a set of uncorrelated parameters, known as orthogonal principal components. The number of principal components was determined based on the explained variance criterion, leading to the conclusion that four principal components were enough to capture more than 90% of the variance from the original data. The high-dimensionality problems were expected to be mitigated, and the robustness of the clustering improved.

Deep autoencoder

Given the problem of bimodality, for this third approach, a different technique was considered for the feature-extraction process. Unlike the previous approaches, which relied on statistical parameters computed from time series, this approach was intended to extract features directly from the shape of the corresponding histograms by using a deep autoencoder model. In deep learning, a deep autoencoder is a neural network in which the input and output layers contain the same number of neurones. In an encoding section, the data is compressed as the number of neurones in layers is progressively reduced until a bottleneck layer is reached, after which the decoding section reconstructs the data by increasing the number of neurones in the subsequent layers (Chen & Guo, 2023; Li et al., 2023). By taking advantage of a compressed representation enhanced through deep learning, the features extracted from the bottleneck layer are expected to capture essential characteristics of the data distributions, including bimodality (Alqahtani et al., 2021).

Table 1
Summary of deep autoencoder network structure.

Layer	Type	Units	Activation
Input Layer	InputLayer	31	ReLU
Encoder	Dense	512	ReLU
Encoder	Dense	256	ReLU
...	...	...	...
Latent Space	Dense	4	Sigmoid
...	...	...	...
Decoder	Dense	256	ReLU
Decoder	Dense	512	ReLU
Output Layer	Dense	31	Sigmoid

For each node in the dataset, a training vector for the model was defined to store relative frequency values. Assuming a bin size of 0.1 m/s, each training vector contained 31 entries, corresponding to a flow speed in the range of 0 to 3.1 m/s. This range was chosen based on the minimum and maximum values observed in the time series. The deep autoencoder model was implemented using the high-level Python library Keras. Table 1 presents the general layout configuration of the model, assuming a symmetric structure. In the input and output layers, each neuron is associated with a bin of relative frequency. Therefore, the model included an *Input Layer*, seven encoding layers, one *Latent Space* layer, seven decoding layers, and an *Output Layer*, giving a total of 17 layers. As all layers were fully connected *Dense* layers, the number of trainable parameters reached 382,290.

It is important to note that neither regularisation nor dropout layers were included, as the primary objective is, even if some degree of overfitting might happen, the accurate encoding–decoding process rather than generalisation. The Rectified Linear Unit (ReLU) was selected as the pre-defined activation function, except for the Latent Space and output layer, where a Sigmoid function was applied to constrain the neurone weights in the range of 0–1, corresponding to the expected values of the feature-vector and the relative frequencies.

The model was compiled using the Nesterov-accelerated adaptive momentum (NAdam) optimiser and the mean squared error (MSE) as the loss function. Training was conducted for 2000 epochs, using a batch size of 32. Fig. 2a presents the training history based on the accuracy as a metric. For the last epoch, an accuracy of 0.875 was achieved, with a corresponding MSE loss of 4.2172 x $10^{-}5$. Whilst the accuracy metric is illustrative in indicating how well the autoencoder was trained to classify bins in the latent space based on the sigmoid activation threshold (0.5), the low MSE values are the metrics that

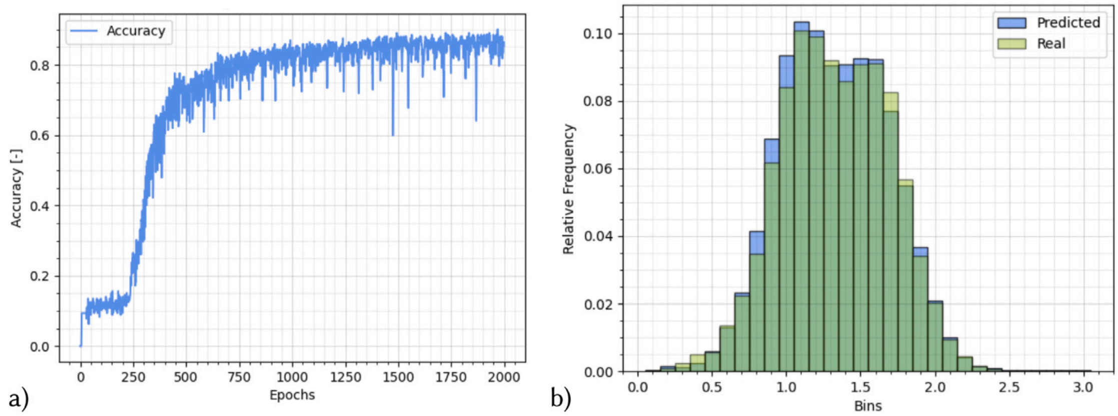

Fig. 2. Results for the Deep Autoencoder model. (a) Training history. (b) Predicted values for node HA454.

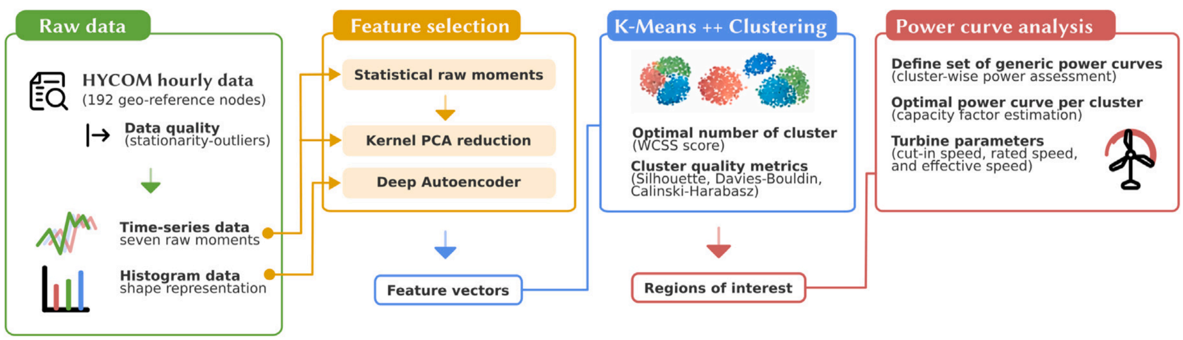

Fig. 3. Clustering activities: general workflow and relationship to power curve analysis.

suggest the global amplitude of the histogram is reasonably preserved and that bin values are reconstructed with small absolute deviations.

Fig. 2b provides a comparison between an actual histogram from the original dataset, which displays a remarkable bimodal distribution and its predicted counterpart generated by the deep autoencoder model. The feature-vector was computed for each node using the encoding sub-structure of the model. Given the properties of the activation function assigned to the Latent Space layer, no additional pre-processing activities were required before the clustering of the nodes based on the extracted feature-vectors.

In order to evaluate the quality of the clusters generated by the three proposed approaches, three different metrics were computed: Silhouette, Davies–Bouldin, and Calinski–Harabasz scores. These metrics were chosen based on the research presented by Shahapure and Nicholas (2020) and Javed et al. (2020).

Fig. 3 illustrates the general workflow for the clustering activities described previously and their relationship with the subsequent power curve analysis. Each approach generates different feature vectors that are subsequently processed using the K-Means++ algorithm. Once spatially coherent regions with similar hydrodynamic characteristics have been identified, the clustering results inform a power curve analysis in which the optimal operational regions are determined and from which the preliminary design specifications of the turbine candidate can potentially be derived.

Power curve modelling

As mentioned previously, the power assessment for each georeferenced node in the dataset will be performed using a simple set of generic power curves, which are introduced to simulate the operational behaviour of different turbine models. The hydrokinetic power output P of a single turbine model can be estimated from the following equation.

$$P = \frac{1}{2}\rho A U^3 C_P \tag{1}$$

where: ρ is the density of seawater, A the swept area of the rotor, U the velocity of the ocean current, and C_P the power coefficient. Based on the research conducted by Lewis et al. (2021), a cubic power curve modelling technique was adopted to define generic power curves. This technique was chosen due to the relatively good agreement with the 14 power curves of near-commercial hydrokinetic turbines evaluated in the study and its capacity to be fully defined by three essential conditions, which are detailed below.

$$U < 0.3 \cdot U_r \, : \, P = 0 \tag{2}$$

$$U_r > U > 0.3 \cdot U_r \, : \, P = \frac{1}{2}\rho A U^3 C_P \tag{3}$$

$$U > U_r \, : \, P = P_r = \frac{1}{2}\rho A U_r^3 C_P \tag{4}$$

where: U_r is the rated speed and P_r is the rated power. The first condition, expressed in Eq. (2), applies to the operational state below the cut-in speed, where the kinetic energy in the fluid flow is not enough to initiate power generation. The second condition, given by Eq. (3), describes a transition region where the turbine is expected to operate as close as possible to maximum efficiency, with power increasing cubically with flow speed. It is important to note that, even if the power coefficient is assumed to be constant, in practice, it depends on the rotor speed, pitch control, and geometrical features of the blades (Burton et al., 2011; Hachmann, 2019; Wang et al., 2012). The third condition represents the rated operational state for power generation. Despite increasing fluid flow speed, the power generation remains constant, limited not by the turbine but by the generator. Finally, a fourth condition, not particularly relevant for this analysis, describes the cut-off speed scenario when the turbine interrupts power generation to protect against structural damage.

In this research, 25 power curves were generated by varying the rated speed from 0.6 to 3.0 m/s. Based on the findings reported by Lewis et al. (2021), a mean power coefficient of $C_P = 0.37$ and a ratio between cut-in speed and rated speed $U_{cut-in} = 0.3 \cdot U_r$ were defined, as these values represent the mean performance of the near commercial tidal turbines reviewed. Furthermore, the capacity factor was computed using Eq. (5), defined as the ratio between the cumulated output power and the hypothetical output power assuming continuous operation at the rated speed. Power curves were evaluated for the hourly data.

$$CF = \frac{\sum \frac{1}{2}\rho A C_P \cdot U^3}{\sum \frac{1}{2}\rho A C_P \cdot U_r^3} \tag{5}$$

The power production is reported for each node as the mean annual power output, which takes into account the five-year data, and it is representative of a hypothetical turbine with a swept area of 1 square metre. While each individual node can be associated with an optimal power curve, the development of a turbine model for each location is not feasible. Therefore, four regions of interest were defined, and performance was evaluated by grouping internal nodes based on the results of the previous cluster analysis, reducing the number of potential turbine models to the number of clusters. In this specific approach, the optimal power curve is computed in terms of the cumulated power output of the nodes within each cluster. Then, it is important to emphasise that the optimal power curve of the cluster may be different compared to the optimal curve computed for individual nodes.

Results

Nodes clustering

The optimal number of clusters was defined using the Elbow method and the within-cluster sum of squares (WCSS) metric. This graphical method indicates that five clusters were sufficient to meet the slope-change criterion for the three feature-based approaches mentioned above. Fig. 4a illustrates the node distribution by cluster for each approach. It is important to note that, due to the random nature of centroid initialisation, the assignment of cluster colours does not follow any specific order.

While at first sight the visual inspection of clustering processes might seem challenging, a comprehensive analysis reveals certain recurrent patterns. In all three approaches, the clusters located in the southwestern regions are consistently associated with regions where suboptimal power production is expected due to relatively low current velocities. Moreover, nodes belonging to these clusters are also observed in coastal regions and in the lee of the island. In contrast, in the narrow passage between the island and the mainland, there is an evident fragmentation of clusters that is extended to the northeastward, and aligns with more favourable operational conditions as the velocity of the current intensifies. Nevertheless, the coexistence of multiple clusters in this region indicates complex hydrodynamic conditions that complicate energy harvesting. Consequently, if this region is selected for exploitation, the design and development of a single turbine model is deemed neither practical nor technically feasible.

Fig. 4b shows the Silhouette scores computed as a quantitative metric to evaluate the goodness of clusters. For both the Standard and Kernel PCA approaches, the highest scores are mainly found in regions where power production is low. In contrast, the number of nodes with the lowest scores, which suggests their placement on the decision boundary between clusters, is similar for both approaches, but their spatial distributions differ. In the case of the Deep Autoencoder approach, with the exception of the south-east region, there are no evident patterns in the score distribution.

Table 2
Clustering scores for the different approaches.

Approach	Silhouette	Davies–Bouldin	Calinski–Harabasz
Standard	0.5719	0.5149	724.8438
Kernel PCA	0.5537	0.6198	296.9134
Deep Autoencoder	0.5606	0.5383	219.4198

Table 2 presents the mean Silhouette score for each clustering approach, and the Davies–Bouldin and Calinski–Harabasz scores to complement the quantitative analysis. Based on the previous scores, the approach with the highest clustering quality was the Standard approach. The Kernel PCA approach ranked third according to Silhouette and Davies–Bouldin scores, and second according to the Calinski–Harabasz score. This quantitative analysis provides two relevant insights. First, despite the first and second approaches producing relatively similar clusters, their quality scores are different. Second, in contrast to initial expectations, the Deep Autoencoder approach did not outperform the other approaches, even if it was theoretically able to more effectively capture the complex features of the time series data, including the bimodality.

The observed disparity in scores may be attributed to the generation of clusters that are neither strictly convex nor well-separated in the feature spaces. As is noted by Javed et al. (2020), this condition suggests that some approaches create clusters with a reasonable internal cohesion, but they fail to maximise the between-cluster separation or attain an optimal separation-to-dispersion ratio. On the other hand, the reasons behind the underperformance of the Deep Autoencoder approach remain unclear. However, a possible explanation is that the encoded nature of the extracted features could be incompatible with the Euclidean assumptions of the K-Means algorithm. The latent space might exhibit anisotropic scaling or, even, individual latent features may encode multiple relevant features of the time series, resulting in an entangled representation.

Power assessment

Given the inherent challenges associated with clustering activities, the selection of a single optimal clustering approach is not entirely feasible. Therefore, the power assessment of the four pre-selected regions was performed based on the clusters generated by each of the previous approaches. As a supplementary illustration, Fig. 5 shows how the rated speed and power output are distributed, assuming that power curve analysis and optimal conditions were computed per each individual node. Notably, the two nodes with the maximum power output do not share the same optimal rated speed; indeed, certain nodes with lower rated speeds yield higher power outputs. This observation reinforces the notion that defining operating regions depends not only on a scalar magnitude, such as the mean speed, but on the entire velocity distribution. A node with consistently moderate velocities may outperform, in cumulative energy terms, a node whose higher speeds are episodic rather than sustained. Moreover, as is anticipated, the cubic dependence of velocity in Eq. (1) magnifies the difference in power output among nodes, placing the highest-potential region along the central section of the Cozumel Channel.

The selected regions are introduced in Fig. 6. The dashed line delineates a feasible region wherein nodes satisfy the minimum cut-in speed threshold for power production, as established in turbine specifications documented by Lewis et al. (2021). This threshold is approximately 0.5 m/s, which, when expressed in terms of the rated speed, is equivalent to around 1.6 m/s. Cut-in speeds lower than this value are not considered technically feasible for two primary reasons: first, there are no prototypes reported in the literature able to operate below this limit, and second, power production in this range is negligible.

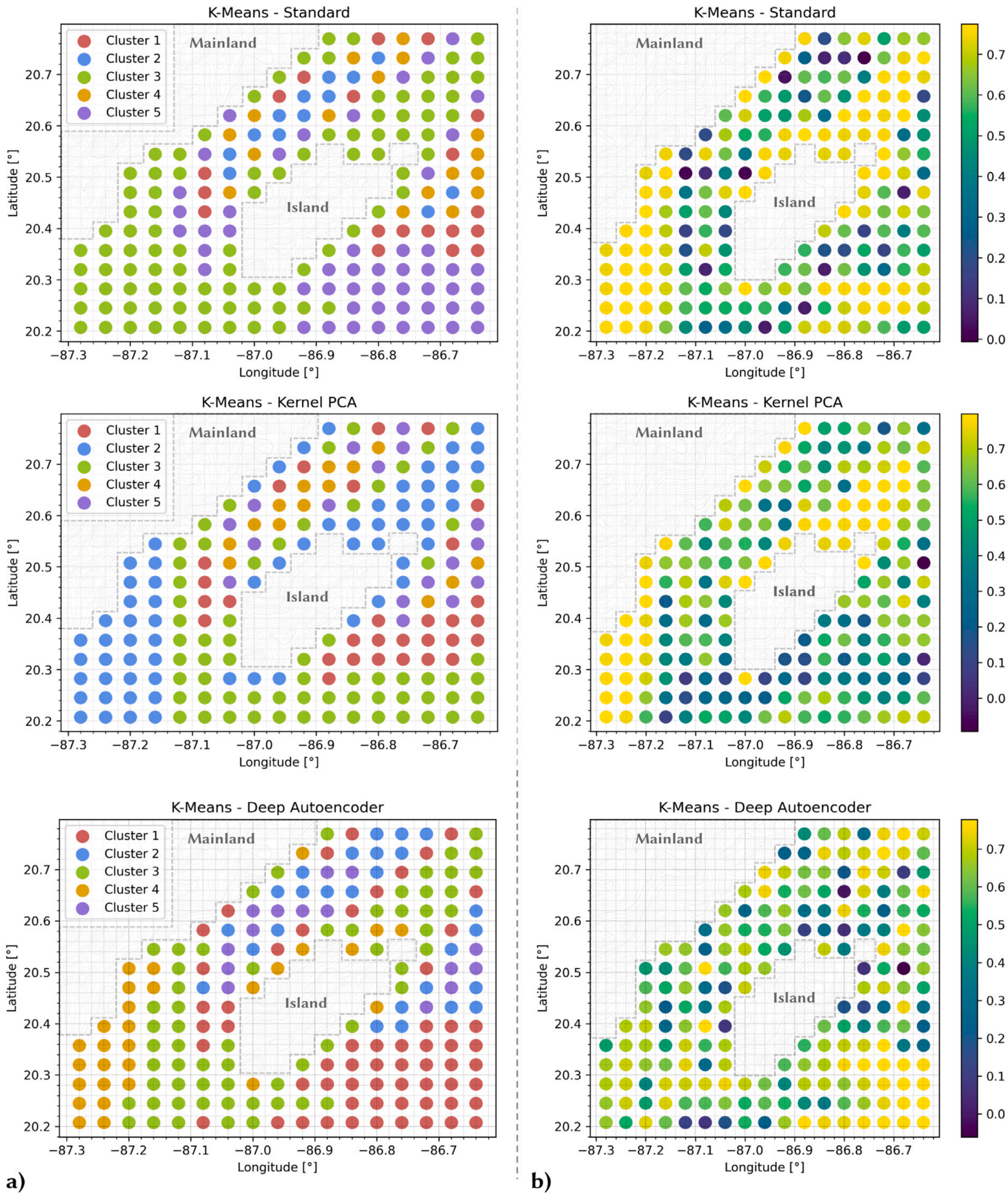

Fig. 4. Comparative analysis. (a) Cluster distributions. (b) Silhouette scores per node.

The selection of the three following regions was conducted based on an alternative criterion partially addressed by Encarnacion et al. (2019), which defines certain natural protected areas as external restrictions for power generation projects. After excluding these areas, the available regions for power production consist of two coastal regions, one adjacent to the mainland and the other to the island, and a third region located in the central section of the channel. Two potential mooring configurations can be associated with these regions: fixed seabed mooring for the onshore and floating mooring for the offshore. Bañuelos García et al. (2021) and Ouro et al. (2022) indicate that a depth in the range 30–50 m is recommended for the fixed mooring to ease the deployment and maintenance of turbines. On the opposite side, as floating mooring systems are complex, with greater installation and maintenance costs, achieving competitive generation costs requires a higher power output. Given this scenario, intermediate nodes located between coastal regions and the central channel were excluded as potential regions of interest because they benefit neither from favourable energy potential nor from feasible mooring conditions.

For the feasible region, it was found that the clusters produced by the Standard and Kernel PCA approaches exhibit identical node distributions. Fig. 7 shows the results of the power assessment. Four different clusters were identified in the region. The Cluster 4 exhibited

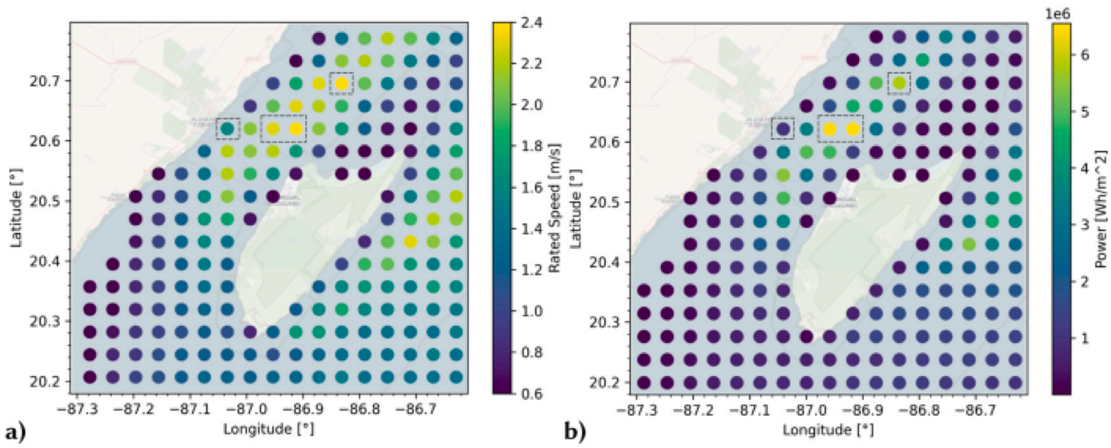

Fig. 5. Energy assessment per node. (a) Optimal rated speeds. (b) Optimal annual power production.

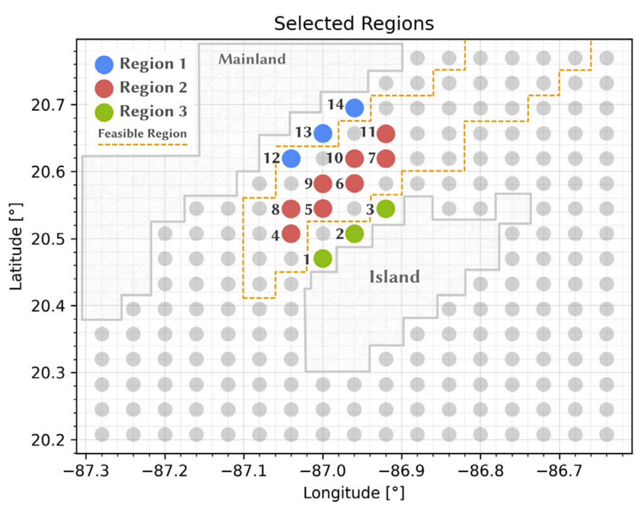

Fig. 6. Regions selected for the simulation energy extraction.

the most favourable scenario with a cluster-wise average of the mean annual power production of 5.18 MWh/yr m^2, followed by Cluster 5 with 3.09 MWh/yr m^2, Cluster 1 with 2.04 MWh/yr m^2, and Cluster 3 with 1.085 MWh/yr m^2. A similar pattern was observed for the cluster-wise average of the capacity factor with values of 0.25, 0.20, 0.17, and 0.15 for Cluster 4, 5, 1, and 3, respectively.

From the previous results, supporting the power production of Cluster 3 proves to be challenging, as nodes do not demonstrate favourable conditions either for power output or for the capacity factor. The low capacity factor observed suggests the manifestation of positively skewed data distributions. Therefore, it is quite possible that the power curve assigned to these nodes as optimal may not be entirely appropriate, and lower cut-in and rated speeds might be preferred. However, among the 25 power curves considered for this research, the power curve assigned already represents the minimal operational conditions. Consequently, further assessment assuming power curves with even lower speeds is not deemed meaningful for practical applications.

For the Deep Autoencoder approach, the node distribution changed as, aside from an isolated node, only three different clusters were identified, as is shown in Fig. 8. The highest cluster-wise averages of mean annual power production were observed in Cluster 5 and Cluster 2 with 4.91 MWh/yr m^2 and 3.01 MWh/yr m^2 respectively. Cluster 1, whose nodes are predominantly located in the onshore regions, presented the lowest average with 1.25 MWh/yr m^2. In the same order, the cluster-wise averages of capacity factor were 0.24, 0.19, and 0.18. It is worth noting that capacity factors in nodes within the same cluster are not entirely homogeneous. This suggests that, in terms of the capacity factor, the Standard and Kernel PCA approaches achieve a more effective node clustering. However, even if no fully uniform patterns were identified within individual clusters, a general trend towards higher capacity factors is evident for this approach based on the Deep Autoencoder.

When the cumulated powers of nodes from each approach were computed and compared, it was observed that the difference is minimal. Both the Standard and Kernel PCA approaches yielded approximately 110.04 MWh/yr m^2, while the Deep Autoencoder produced 109.91 MWh/yr m^2. The main differences arise when a pairwise comparison of nodes is performed. In Fig. 7, for example, certain nodes from Cluster 4 located in the central channel assumed an optimal rated speed of 2.3 m/s. In Fig. 8, these same nodes were associated with Cluster 2 and an optimal rated speed of 2.1 m/s. While the mean annual power of nodes in the first case tends to be slightly superior, the capacity factors in the second case are higher. As a consequence, the definition of a single farm layout or turbine array configuration proves intricate, as external conditions must be considered in the decision-making framework of the generation project.

For a base scenario, the objective is to maximise the power output without emphasising the storage capacities, end-user needs, or power grid conditions. Therefore, a predictable supply–demand balance is an essential requirement for effective energy distribution. Power assessments for Standard and Kernel PCA approaches appear to be more consistent with this scenario. In contrast, an alternative scenario, more closely aligned with the Deep Autoencoder approach, prioritises the capacity factor. The objective is firm power production, aiming to minimise fluctuations. This strategy is beneficial for sites where integration to power grids is limited or where energy demand and storage capacities are restricted.

Table 3 presents the estimated mean annual power production per node for the remaining pre-selected regions. These regions were initially chosen based on practical criteria, as potential sites for turbine deployment. In the short term, coastal regions were considered more viable due to their proximity to the shore, which facilitates the mooring and activities related to operation, maintenance, and monitoring. However, the results suggest, at least from the point of view of this assessment, that these regions are not an option for long-term power generation projects. Compared to the nodes located in the central channel of Region 2, the mean annual power production is minimal.

Table 4, which presents the cumulative mean annual power production, corroborates the trends identified earlier. Consistent with previous findings, power output estimations for the Deep Autoencoder were slightly lower compared to the remaining two approaches. It should be noted that this power estimation assumes a hypothetical configuration

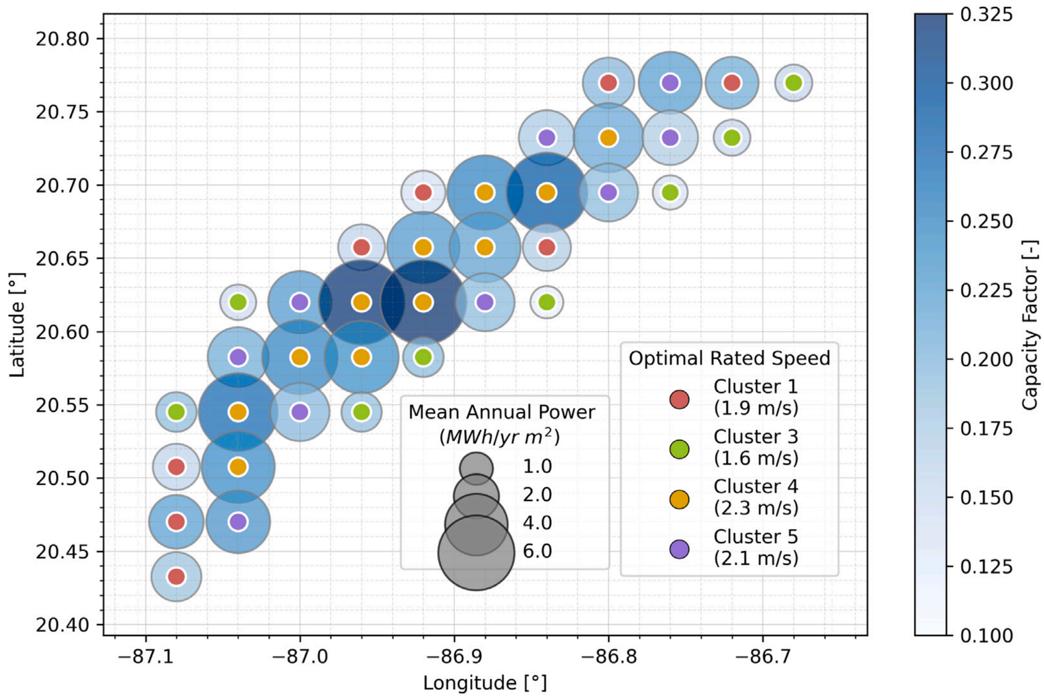

Fig. 7. Power assessment of the feasible region for Standard and Kernel PCA approaches.

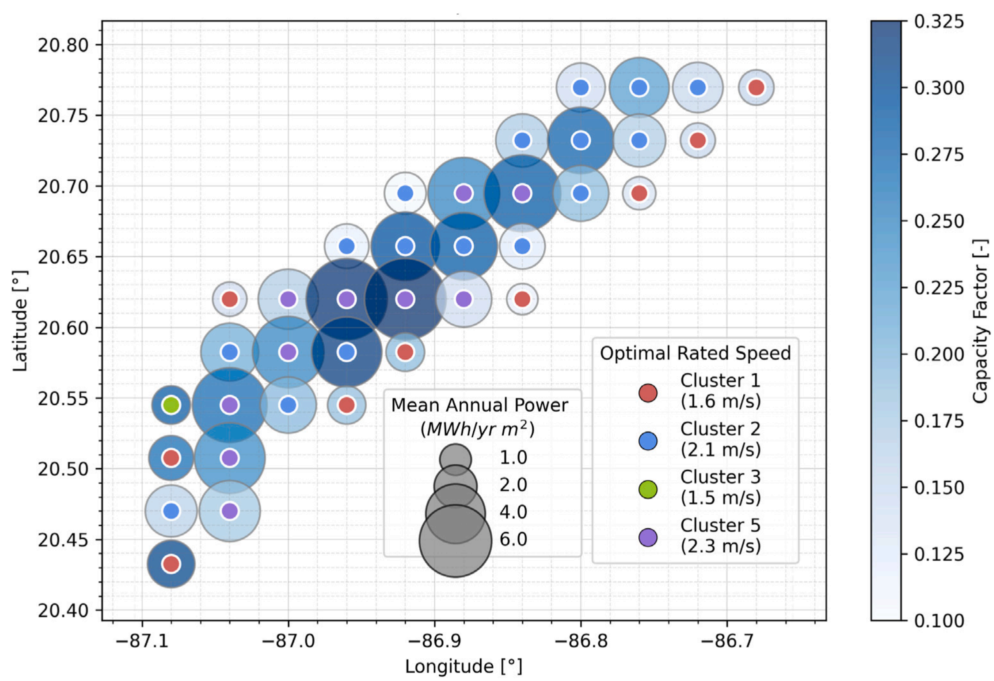

Fig. 8. Power assessment for the feasible region for the Deep Autoencoder approach.

Table 3
Theoretical mean annual power production per node in the pre-selected regions in MWh/yr m^2.

Approach	1	2	3	4	5	6	7	8	9	10	11	12	13	14
K-Means Standard/Kernel PCA	0.088	0.032	0.010	4.868	3.002	4.820	6.557	5.462	4.971	6.490	4.575	0.989	0.101	0.262
K-Means Deep Autoencoder	0.088	0.032	0.010	4.868	2.992	4.820	6.557	5.462	4.971	6.490	4.575	0.989	0.101	0.262

Table 4
Theoretical mean cumulated annual power production per region in MWh/yr m^2.

Approach	Region 1	Region 2	Region 3
K-Means Standard /Kernel PCA	0.1305	40.7491	1.3533
K-Means Deep Autoencoder	0.1305	40.7387	1.3533

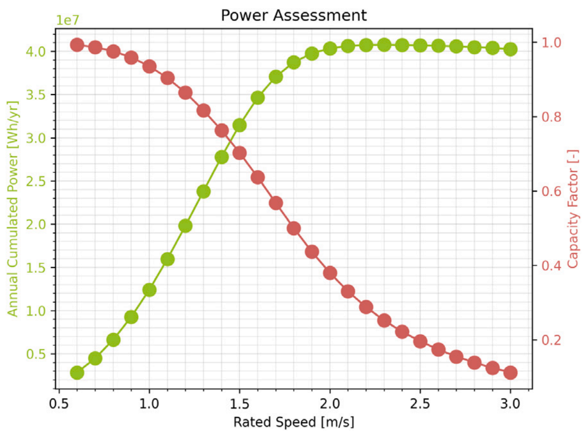

Fig. 9. Trade off comparison between capacity factor and annual cumulated power for Region 2.

in which a single turbine with a swept area of 1 m^2 is installed at each node. Therefore, while the values presented do not correspond to a fully realistic deployment, they serve as an effective metric to compare the relative potential of each cluster in terms of power production capability.

To complement these results, the optimal rated speed values per node are also presented. For the three nodes in Region 1, the optimal rated speed was computed to be around 0.7 m/s. In Region 2, based on the results obtained from both K-Means Standard and K-Means Kernel PCA clustering approaches, the optimal rated speed of the isolated node was 1.9 m/s, while the remaining nodes exhibited an optimal rated speed of 2.3 m/s. In contrast, when Region 2 was evaluated using the Deep Autoencoder Approach, all nodes converged to a uniform optimal rated speed of 2.3 m/s. Finally, in Region 3, an optimal rated speed of 1.6 m/s was associated with nodes 12 and 13, and a rated speed of 1.0 m/s to node 14.

Assuming Region 2 as the most feasible for power extraction, Fig. 9 shows the trade-off between annual cumulated power and capacity factor. As can be observed, the power curve that produces the maximum power output corresponds to a capacity factor that is not particularly high. However, increasing the capacity factor by adopting a lower rated speed is a worse option than reducing it by adopting a higher rated speed. In this particular region, due to the skewness and kurtosis of the current speed distributions, it is more suitable to harness energy from high velocities even though the optimal operative time is reduced.

Operational parameters

While the selection of the optimal power curve for a given node, or clustering region, is essential, this activity represents just the first hurdle of an extensive process required for the integral design of a hydrokinetic turbine. In terms of the rated speed, for example, lower values are commonly associated with turbine blades of high solidity (i.e. low aspect ratio). Due to restrictions in the rotational speed, the rotor generates power primarily through a high torque, which is essential to initiate movement when the devices operate at low cut-in speeds. On the opposite side, if the rated speed is high, a low solidity (i.e. high aspect ratio) for turbine blades is preferred. The torque applied to the rotor decreases, but is compensated by an increase in rotational speed (Burton et al., 2011; Hansen, 2023; Manwell et al., 2009). This criterion partially defines the operational behaviour and, along with the criteria discussed below, serves as a guideline rather than a general rule in a design process that depends not only on theoretical methods but also on empirical experience. For the present research, Region 2 was selected as the most promising region for power generation. The parameters needed for the hydrodynamic characterisation were defined based on the available site-specific data and the operation of a small turbine with a swept area of 1 m^2.

Assuming a rated speed of 2.3 m/s and a cut-in speed of 0.69 m/s, it is clear that these values cannot be attributed to an unusually slow current speed. Consequently, the Tip Speed Ratio, an dimensionless parameter that relates the rotational speed of the rotor to the free-stream speed, was set to a discrete value of $TSR = 5$. Previous research has shown that $TSR \leq 4$ may produce tip and root losses for similar hydrodynamic conditions (Cano-Perea et al., 2023). On the other hand, Encarnacion et al. (2019) suggests that values of $TSR \approx 7.75$ allow the development of thin turbine blades, reducing manufacturing costs. Nevertheless, although this proposal is feasible, high TSR configurations usually lead to side effects, such as an increase in turbulence and cavitation (European Commission and Directorate General for Research and Innovation and ECORYS and Fraunhofer, 2017; Fraenkel, 2002). Hence, for the turbine configuration of this research, a $TSR = 5$ was selected as an appropriate trade-off, avoiding the drawbacks associated with both extreme conditions.

The second essential parameter defined was the effective speed. Although power curves are theoretically formulated to maintain a constant power coefficient across the entire range between the cut-in speed and rated speed, this assumption does not always reflect real operational conditions. Even when control strategies, such as variable pitch or variable rotation speed, are implemented, fluctuations in the free-stream current velocity may cause deviations in the power coefficient. As a consequence, the blade geometry must be optimised for a specific speed associated with the optimal hydrodynamic performance, usually known as effective speed. As is indicated by Hachmann (2019) and Wang et al. (2006), this speed is representative of the most frequent or prevailing flow conditions. Based on this criterion, the most frequent velocities for each node in Region 2 were selected from the histograms and subsequently averaged, yielding an approximate effective speed of 1.52 m/s. This value represents approximately 62.5% of the rated speed.

The subsequent set of parameters was focused on the basic properties of the sea water at the site. The approximated values for density, salinity, and temperature were taken from the research presented by Alcérreca-Huerta et al. (2019), and they are consistent with mean values reported in the literature (Alemán et al., 2006; Millero et al., 2008; Qasem et al., 2021). Though no direct measurements were available, the kinematic viscosity was computed using the previous parameters and correlations described in the work of Sharqawy et al. (2010).

Table 5 presents a summary of the parameters discussed above.

The last parameter estimated was the well-known Reynolds number, a dimensionless parameter that expresses the ratio of inertial to viscous forces within a fluid flow. This parameter is particularly relevant, as it reflects the influence of hydrodynamic conditions, operational parameters, and geometrical features. According to Allmark et al. (2020) and Lloyd et al. (2021), the representative Reynolds number of a horizontal axis turbine can be associated with the flow conditions at approximately 70% of the blade radius; then a specific cross-sectional geometry is required. For this instance, the NACA 63815 was selected. The characteristic length is the blade chord at the corresponding radial position.

It is important to note that, in the absence of a fully defined blade geometry, the blade chord, as well as the Reynolds number, must be computed iteratively. An initial chord value is assumed, followed

Table 5
Design parameters proposed for a turbine operating in Region 2.

Parameter	Value	Units
Tip Speed Ratio	5	[–]
Angular Speed	15	[rad/s]
Rated Speed	2.3	[m/s]
Free-stream Speed	1.5	[m/s]
Cut-in Speed	0.69	[m/s]
Water Density	1025	[kg/m^3]
Kinematic Viscosity	1.005 $x10^{-6}$	[m^2/s]
Angular Speed	143.239	[rpm]
Initial Blade Chord	0.1533	[c/R]
Final Blade Chord	0.0981	[c/R]
Coefficient of Power	0.37	[–]
Operative No. Reynolds	416 429	[–]

by the estimation of lift and drag coefficients, here determined using computational fluid dynamics (CFD). With the assistance of the Blade Element Momentum (BEM) method, a new chord is computed by optimising the length to achieve the optimal blade solidity. Given the updated chord, hydrodynamic coefficients are recalculated, and the process is repeated until convergence is reached. As Reynolds number relies on the blade chord, its value changes every iteration. The convergence criteria adopted for this research was a 5%, and 6 iterations were required to meet this criterion. Resources related to empirical equations to initialise the blade chord and BEM method specifications are available in Burton et al. (2011), Hansen (2023), Manwell et al. (2009), Masters et al. (2011), while resources focused on the use and application of CFD simulation for tidal energy topics are available in Ferrer and Montlaur (2015).

From the previously defined set of parameters, it is now possible to proceed with a preliminary design of a horizontal-axis tidal turbine as a device for power generation. From this stage, the design may be supported by multiple analytical, numerical, or experimental resources. Even though the incorporation of AI tools offers the potential to enhance traditional methods and explore novel solutions and design configurations adapted to the oceanic environment. The primary challenge is the validation of the adopted assumptions, such as a power coefficient of $C_P = 0.37$ or the application of cubic power curves, both obtained from the literature. These assumptions should be refined or adjusted as necessary. Furthermore, external factors, including operational and economical constraints, should be incorporated in subsequent stages.

Conclusions

From three clustering approaches, the Standard approach, based on the implementation of statistical moments as features, demonstrated the most coherent segmentation of regions with potential for power generation. The Kernel PCA approach produced quite similar clusters. However, the Deep Autoencoder approach, which was initially expected to outperform the other approaches, did not exhibit satisfactory results either in terms of clustering coherence or performance metrics. As previously mentioned, it is probable that the encoded nature of features, extracted from the autoencoder's bottleneck, is not compatible with assumptions adopted in the K-Means algorithm. Deep autoencoders are usually trained to minimise reconstruction error, not to enforce geometric regularity in the latent space. Therefore, latent representations may be in a certain way anisotropic. Further research is required to explore alternative autoencoder configurations and to determine if additional pre-processing stages are needed prior to their use in clustering processes.

Regarding the K-Means algorithm, another fast-implementation algorithms were explored, including OPTICS (Ankerst et al., 1999) and DBSCAN (Schubert et al., 2017), to deal with the limitations of the hard clustering approach. In contrast to K-Means, these algorithms choose the optimal number of clusters based on the data properties. Even if some clustering regions align with those derived from K-Means, OPTICS and DBSCAN algorithms failed to cluster a significant number of nodes, which were labelled as noisy data points. Then nodes situated near the boundaries of clusters are representative of hydrodynamically complex regions for power production. As a proposal, soft-assignment clustering methods, such as fuzzy clustering (Yang, 1993), are recommended. Compared to traditional clustering, fuzzy clustering has the potential to assign a single node to different clusters by implementing scores of membership. From this perspective, it is expected to obtain a more accurate representation of ambiguous data distributions.

In terms of the power assessment, within the feasible region, multiple nodes were identified with relatively high power production, reaching values close to 6 MWh/yr m^2. However, many of these nodes are located near natural protected regions, which impose important restrictions that exclude potential regions. Notwithstanding these limitations, Region 2 remains a promising candidate, exhibiting a cluster-wise average of mean annual power production of approximately 5.09 MWh/yr m^2. These estimations are consistent with those reported by Bañuelos García et al. (2021), although some differences were observed in the central core region of the Cozumel channel. The main problem arises with the specific analysis of capacity factors.

According to the previous research, a capacity factor of approximately $CF \approx 0.4$ is required to achieve economically viable power generation. In contrast, the maximum capacity factor estimated in the present work was $CF = 0.325$. Although the disparity could be perceived as minimal, achieving a 25% increase in capacity represents an important technical challenge, particularly if it is considered that power curves were already selected for specific conditions of the channel. The design of a turbine able to operate at a lower rated speed could improve the capacity factor by increasing the operational time, but this also reduces the power output. In literature, alternative solutions to overcome this problem include the design of low solidity turbines to reduce the torque requirements (Encarnacion et al., 2019; Wang et al., 2023), hybrid floating systems that exploit the waves as a complementary energy source (Chen et al., 2016), and ducted turbines that induce an acceleration of the flow incident to the rotor (Belloni et al., 2017; Cano-Perea et al., 2023; Chang et al., 2016; Maduka & Li, 2021; Mei et al., 2025). Consequently, the objective to be pursued is the optimisation of the trade-off between maximising the power output and extending operation in terms of capacity factor, or alternatively, in reducing the cost of the energy generation.

Among turbine configurations reviewed by Lewis et al. (2011) and Boretti (2020), the three-bladed Nova Turbine (Encarnacion et al., 2019) and the two-bladed Atlantis SeaGen-S turbine were identified as the most compatible turbines with the hydrodynamic conditions of Region 2. The Nova turbine was designed to operate at even lower speed conditions, but to date, no empirical data have been published to confirm its performance under real-world conditions. In contrast, the information related to the SeaGen-S turbine is not extensive, but this device has been experimentally validated. Compared to the power curve selected for Region 2, the rated speed and cut-in speed increase by 8% and 44%, respectively. These values suggest that the SeaGen-S, like many other turbines, does not meet the general criteria established by Lewis et al. (2021),[3] and requires a minimum cut-in speed of approximately 1 m/s to initiate operation. While low cut-in speeds are technically viable, the deficit of momentum in fluid flow may prevent stable operation. Therefore, beyond the optimisation of the power curve, the design of novel turbine geometries is essential to address potential mechanical and hydrodynamic challenges associated with real-world deployment.

In addition, the data sources (i.e. HYCOM model) must be verified in order to effectively confirm or exclude potential regions for power

[3] $U_{cut-in} = 0.3 \cdot U_{rated}$.

production. Even with this preliminary assessment, certain regions, such as Region 1, were identified as unsuitable. From a power production perspective, onshore nodes are less competitive compared to nodes located in the central channel. Nevertheless, onshore locations may offer important advantages in terms of the mooring structures and operational and maintenance costs, which contribute to a reduction in the cost of energy. A further comprehensive assessment should therefore include these techno-economical restraints, along with additional operational restrictions that may be omitted.

Finally, the set of parameters estimated in this work represents the minimum required to facilitate the preliminary design of a hydrokinetic turbine. However, it is worth mentioning that effective speed, as one of the most fundamental operational parameters, still relies on a basic assumption inherited from the wind industry. Whilst not a rule of thumb, the effective speed of most wind sites aligns with the peak of the wind speed distribution, and notably, most of these distributions fit relatively well with Rayleigh or Weibull models. However, in ocean currents, the current speed distribution does not fit these models. In practice, even bimodal distributions have been identified; thus, the selection of the effective speed presents an opportunity for further investigation and methodological refinement.

For future works, the implementation of the BEM methods supported on CFD simulations remains a common and effective approach. However, it is essential to explore alternatives and empower the integration of novel techniques and tools, such as AI, to produce turbines capable of addressing the specific challenges discussed along this paper. The Cozumel Channel is a potential site for power generation, the challenge lies in adapting and enhancing the available technical resources and technologies to harness this energy effectively.

CRediT authorship contribution statement

Gerardo Cano-Perea: Writing – original draft, Software, Methodology, Conceptualization. **Ricardo Arévalo-Mejía:** Supervision, Software, Methodology, Conceptualization. **Humberto Salinas-Tapia:** Writing – review & editing, Supervision, Methodology. **Carlos Díaz-Delgado:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Edgar Mendoza-Baldwin:** Supervision, Methodology, Conceptualization. **Boris Miguel López-Rebollar:** Writing – review & editing, Supervision, Methodology, Investigation, Conceptualization.

Declaration of competing interest

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

Acknowledgements

The authors acknowledge the valuable contribution of Dr. Ismael Mariño Tapia and Dr. Miguel A. Cahuich for providing the adjusted HYCOM dataset.

Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.esd.2026.102004.

References

Abreu, S., Rodrigues, F., & Pereira, J. (2025). Clustering of renewable energy assets to optimize resource allocation and operational strategies. *Journal of Intelligent Information Systems*, *63*(3), 831–853. http://dx.doi.org/10.1007/s10844-024-00914-4.

Adedeji, P. A., Olatunji, O. O., Madushele, N., Akinlabi, S. A., & Adeyemo, J. A. (2022). Cluster-based wind turbine maintenance prioritization for a utility-scale wind farm. *Procedia Computer Science*, *200*, 1726–1735. http://dx.doi.org/10.1016/j.procs.2022.01.373.

Aghabozorgi, S., Seyed Shirkhorshidi, A., & Ying Wah, T. (2015). Time-series clustering – a decade review. *Information Systems*, *53*, 16–38. http://dx.doi.org/10.1016/j.is.2015.04.007.

Akinwande, M. O., Dikko, H. G., & Samson, A. (2015). Variance inflation factor: As a condition for the inclusion of suppressor variable(s) in regression analysis. *Open Journal of Statistics*, *05*(07), 754–767. http://dx.doi.org/10.4236/ojs.2015.57075.

Alcérreca-Huerta, J. C., Encarnacion, J. I., Ordoñez-Sánchez, S., Callejas-Jiménez, M., Gallegos Diez Barroso, G., Allmark, M., Mariño-Tapia, I., Silva Casarín, R., O'Doherty, T., Johnstone, C., & Carrillo, L. (2019). Energy yield assessment from ocean currents in the insular shelf of cozumel island. *Journal of Marine Science and Engineering*, *7*(5), 147. http://dx.doi.org/10.3390/jmse7050147.

Alemán, J., Pelegrí, J., & Sangrà, P. (2006). Ocean rheology. *Journal of Non-Newtonian Fluid Mechanics*, *133*(2–3), 121–131. http://dx.doi.org/10.1016/j.jnnfm.2005.12.002.

Allmark, M., Ellis, R., Lloyd, C., Ordonez-Sanchez, S., Johannesen, K., Byrne, C., Johnstone, C., O'Doherty, T., & Mason-Jones, A. (2020). The development, design and characterisation of a scale model horizontal axis tidal turbine for dynamic load quantification. *Renewable Energy*, *156*, 913–930. http://dx.doi.org/10.1016/j.renene.2020.04.060.

Alqahtani, A., Ali, M., Xie, X., & Jones, M. W. (2021). Deep time-series clustering: A review. *Electronics*, *10*(23), 3001. http://dx.doi.org/10.3390/electronics10233001.

Ankerst, M., Breunig, M. M., Kriegel, H.-P., & Sander, J. (1999). OPTICS: ordering points to identify the clustering structure. *ACM SIGMOD Record*, *28*(2), 49–60. http://dx.doi.org/10.1145/304181.304187.

Arredondo-Godínez, L. A. (2020). Guía para la evaluación de sitios con potencial energético de corrientes marinas: caso cozumel quintana roo méxico. (Master's thesis), Universidad Nacional Autonoma de México, url: https://hdl.handle.net/20.500.14330/TES01000804053.

Arthur, D., & Vassilvitskii, S. (2007). K-means++: the advantages of careful seeding. In *Proceedings of the eighteenth annual ACM-SIAM symposium on discrete algorithms* (pp. 1027–1035). Society for Industrial and Applied Mathematics, ISBN: 9780898716245.

Athié, G., Candela, J., Sheinbaum, J., Badan, A., & Ochoa, J. (2011). Yucatan current variability through the cozumel and yucatan channels. *Ciencias Marinas*, *37*(4A), 471–492. http://dx.doi.org/10.7773/cm.v37i4a.1794.

Ayesha, S., Hanif, M. K., & Talib, R. (2020). Overview and comparative study of dimensionality reduction techniques for high dimensional data. *Information Fusion*, *59*, 44–58. http://dx.doi.org/10.1016/j.inffus.2020.01.005.

Azhar, A., & Hashim, H. (2023). A review of wind clustering methods based on the wind speed and trend in Malaysia. *Energies*, *16*(8), 3388. http://dx.doi.org/10.3390/en16083388.

Bañuelos García, F. H., Ring, M., Mendoza Baldwin, E. G., & Silva Casarín, R. (2021). *Foundation design for marine turbines on rocky soils*. EPOMEX-Universida Autonoma de Colima: http://dx.doi.org/10.26359/epomex.cemie012021.

Bárcenas Graniel, J. F., Fontes, J. V. H., Garcia, H. F. G., & Silva, R. (2021). Assessing hydrokinetic energy in the mexican caribbean: A case study in the cozumel channel. *Energies*, *14*(15), 4411. http://dx.doi.org/10.3390/en14154411.

Belloni, C., Willden, R., & Houlsby, G. (2017). An investigation of ducted and open-centre tidal turbines employing CFD-embedded BEM. *Renewable Energy*, *108*, 622–634. http://dx.doi.org/10.1016/j.renene.2016.10.048.

Boretti, A. (2020). State-of-the-art of MW-level capacity oceanic current turbines. *Nonlinear Engineering*, *9*(1), 361–369. http://dx.doi.org/10.1515/nleng-2020-0022.

Breiman, L. (2001). Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical Science*, *16*(3), http://dx.doi.org/10.1214/ss/1009213726.

Burton, T., Jenkins, N., Sharpe, D., & Bossanyi, E. (2011). Wind energy handbook. Wiley: http://dx.doi.org/10.1002/9781119992714.

Cano-Perea, G., Mendoza, E., & López-Rebollar, B. M. (2023). Numerical analysis of a ducted water current turbine for low energetic flow conditions. *Journal of Ocean Engineering and Marine Energy*, *10*(1), 217–237. http://dx.doi.org/10.1007/s40722-023-00308-8.

Cetina, P., Candela, J., Sheinbaum, J., Ochoa, J., & Badan, A. (2006). Circulation along the mexican caribbean coast. *Journal of Geophysical Research: Oceans*, *111*(C8), http://dx.doi.org/10.1029/2005jc003056.

Chang, L.-Y., Chen, F., & Tseng, K.-T. (2016). Dynamics of a marine turbine for deep ocean currents. *Journal of Marine Science and Engineering*, *4*(3), 59. http://dx.doi.org/10.3390/jmse4030059.

Charhouni, N., Sallaou, M., & Mansouri, K. (2019). Realistic wind farm design layout optimization with different wind turbines types. *International Journal of Energy and Environmental Engineering*, *10*(3), 307–318. http://dx.doi.org/10.1007/s40095-019-0303-2.

Chassignet, E. P., Hurlburt, H. E., Smedstad, O. M., Halliwell, G. R., Hogan, P. J., Wallcraft, A. J., Baraille, R., & Bleck, R. (2007). The HYCOM (hybrid coordinate ocean model) data assimilative system. *Journal of Marine Systems*, *65*(1–4), 60–83. http://dx.doi.org/10.1016/j.jmarsys.2005.09.016.

Chávez, G., Candela, J., & Ochoa, J. (2003). Subinertial flows and transports in cozumel channel. *Journal of Geophysical Research: Oceans*, *108*(C2), http://dx.doi.org/10.1029/2002jc001456.

Chen, F. (2010). Kuroshio power plant development plan. *Renewable and Sustainable Energy Reviews*, *14*(9), 2655–2668. http://dx.doi.org/10.1016/j.rser.2010.07.070.

Chen, J.-H., Chiu, F.-C., Hsin, C.-Y., & Tsai, J.-F. (2016). Hydrodynamic consideration in ocean current turbine design. *Journal of Hydrodynamics*, *28*(6), 1037–1042. http://dx.doi.org/10.1016/s1001-6058(16)60708-4.

Chen, S., & Guo, W. (2023). Auto-encoders in deep learning—A review with new perspectives. *Mathematics*, *11*(8), 1777. http://dx.doi.org/10.3390/math11081777.

Clifton, A., & Lundquist, J. K. (2012). Data clustering reveals climate impacts on local wind phenomena. *Journal of Applied Meteorology and Climatology*, *51*(8), 1547–1557. http://dx.doi.org/10.1175/jamc-d-11-0227.1.

Duerr, A. E. S., & Dhanak, M. R. (2012). An assessment of the hydrokinetic energy resource of the florida current. *IEEE Journal of Oceanic Engineering*, *37*(2), 281–293. http://dx.doi.org/10.1109/joe.2012.2186347.

Encarnacion, J. I., Johnstone, C., & Ordonez-Sanchez, S. (2019). Design of a horizontal axis tidal turbine for less energetic current velocity profiles. *Journal of Marine Science and Engineering*, *7*(7), 197. http://dx.doi.org/10.3390/jmse7070197.

European Commission and Directorate General for Research and Innovation and ECORYS and Fraunhofer (2017). *Study on lessons for ocean energy development: Final report*. Publications Office: http://dx.doi.org/10.2777/389418.

Ferrer, E., & Montlaur, A. (2015). *CFD for wind and tidal offshore turbines*. Springer International Publishing, http://dx.doi.org/10.1007/978-3-319-16202-7.

Fraenkel, P. L. (2002). Power from marine currents. *Proceedings of the Institution of Mechanical Engineers, Part A: Journal of Power and Energy*, *216*(1), 1–14. http://dx.doi.org/10.1243/095765002760024782.

Fuller, W. A. (1995). *Introduction to statistical time series*. Wiley, http://dx.doi.org/10.1002/9780470316917.

Gründlingh, M. L. (1983). On the course of the agulhas current. *South African Geographical Journal*, *65*(1), 49–57. http://dx.doi.org/10.1080/03736245.1983.10559671.

Haas, K., Yang, X., Neary, V., & Gunawan, B. (2017). Ocean current energy resource assessment for the gulf stream system: The florida current. In *Marine renewable energy* (pp. 217–236). Springer International Publishing, http://dx.doi.org/10.1007/978-3-319-53536-4_9.

Haces-Fernandez, F., Cruz-Mendoza, M., & Li, H. (2022). Onshore wind farm development: Technologies and layouts. *Energies*, *15*(7), 2381. http://dx.doi.org/10.3390/en15072381.

Hachmann, C. (2019). *Power generation from large-scale tidal stream turbine arrays* (Ph.D. thesis), Faculty of Science and Engineering. The University of Manchester.

Hansen, M. O. L. (2023). Aerodynamics and the design of horizontal axis wind turbine. In *Wind energy engineering* (pp. 137–153). Elsevier, http://dx.doi.org/10.1016/b978-0-323-99353-1.00019-0.

Hartigan, J. A., & Hartigan, P. M. (1985). The dip test of unimodality. *The Annals of Statistics*, *13*(1), http://dx.doi.org/10.1214/aos/1176346577.

Herbert-Acero, J., Probst, O., Réthoré, P.-E., Larsen, G., & Castillo-Villar, K. (2014). A review of methodological approaches for the design and optimization of wind farms. *Energies*, *7*(11), 6930–7016. http://dx.doi.org/10.3390/en7116930.

Hernández-Fontes, J. V., Felix, A., Mendoza, E., Cueto, Y. R., & Silva, R. (2019). On the marine energy resources of Mexico. *Journal of Marine Science and Engineering*, *7*(6), 191. http://dx.doi.org/10.3390/jmse7060191.

IEA (2025). *Global energy review 2025*: *Technical report*, Paris: International Energy Agency, url: https://www.iea.org/reports/global-energy-review-2025 Licence: CC BY 4.0.

International Renewable Energy Agency (2014). *Ocean energy: Technology readiness, patents, deployment status and outlook*. International Renewable Energy Agency: ISBN: 9781523152193.

IRENA (2022). *Renewable power generation costs in 2022*: *Technical report*, International Renewable Energy Agency, url: https://www.irena.org.

Javed, A., Lee, B. S., & Rizzo, D. M. (2020). A benchmark study on time series clustering. *Machine Learning with Applications*, *1*, Article 100001. http://dx.doi.org/10.1016/j.mlwa.2020.100001.

Kabir, A., Lemongo-Tchamba, I., & Fernandez, A. (2015). An assessment of available ocean current hydrokinetic energy near the north carolina shore. *Renewable Energy*, *80*, 301–307. http://dx.doi.org/10.1016/j.renene.2015.02.011.

Kang, Y.-J., & Noh, Y. (2019). Development of hartigan's dip statistic with bimodality coefficient to assess multimodality of distributions. In A. J. Fernández (Ed.), *Mathematical Problems in Engineering*, *2019*(1), http://dx.doi.org/10.1155/2019/4819475.

Keogh, E., & Mueen, A. (2017). Curse of dimensionality. In *Encyclopedia of machine learning and data mining* (pp. 314–315). Springer US, http://dx.doi.org/10.1007/978-1-4899-7687-1_192.

Lewis, A., Estefen, S., Huckerby, W., Musial, W., Pontes, T., & Torres-Martínez, J. (2011). IPCC special report on renewable energy sources and climate change mitigation. In *Chapter ocean energy*. Cambridge University Press, url: http://www.ipcc.ch.

Lewis, M., McNaughton, J., Márquez-Dominguez, C., Todeschini, G., Togneri, M., Masters, I., Allmark, M., Stallard, T., Neill, S., Goward-Brown, A., & Robins, P. (2019). Power variability of tidal-stream energy and implications for electricity supply. *Energy*, *183*, 1061–1074. http://dx.doi.org/10.1016/j.energy.2019.06.181.

Lewis, M., O'Hara Murray, R., Fredriksson, S., Maskell, J., de Fockert, A., Neill, S. P., & Robins, P. E. (2021). A standardised tidal-stream power curve, optimised for the global resource. *Renewable Energy*, *170*, 1308–1323. http://dx.doi.org/10.1016/j.renene.2021.02.032.

Li, P., Pei, Y., & Li, J. (2023). A comprehensive survey on design and application of autoencoder in deep learning. *Applied Software Computing*, *138*, Article 110176. http://dx.doi.org/10.1016/j.asoc.2023.110176.

Liu, Y., Gao, X., Yan, J., Han, S., & Infield, D. G. (2014). Clustering methods of wind turbines and its application in short-term wind power forecasts. *Journal of Renewable and Sustainable Energy*, *6*(5), http://dx.doi.org/10.1063/1.4898361.

Lloyd, C., Allmark, M., Ordonez-Sanchez, S., Martinez, R., Johnstone, C., Germain, G., Gaurier, B., Mason-Jones, A., & O'Doherty, T. (2021). Validation of the dynamic load characteristics on a tidal stream turbine when subjected to wave and current interaction. *Ocean Engineering*, *222*, Article 108360. http://dx.doi.org/10.1016/j.oceaneng.2020.108360.

Maduka, M., & Li, C. (2021). Numerical study of ducted turbines in bi-directional tidal flows. *Engineering Applications of Computational Fluid Mechanics*, *15*(1), 194–209. http://dx.doi.org/10.1080/19942060.2021.1872706.

Maechler, M. (2003). Diptest: Hartigan's dip test statistic for unimodality [GPL software]. http://dx.doi.org/10.32614/cran.package.diptest.

Manwell, J. F., McGowan, J. G., & Rogers, A. L. (2009). *Wind energy explained: Theory, design and application*. Wiley: http://dx.doi.org/10.1002/9781119994367.

Martins, A. A., Vaz, D. C., Silva, T. A. N., Cardoso, M., & Carvalho, A. (2024). Clustering of wind speed time series as a tool for wind farm diagnosis. *Mathematical and Computational Applications*, *29*(3), 35. http://dx.doi.org/10.3390/mca29030035.

Maslo, A., Mariño-Tapia, I., Hernández, E. S., & Casarín, R. S. (2023). Modelling the impacts of a large marine turbine array in the cozumel channel. *Ocean Engineering*, *289*, Article 116153. http://dx.doi.org/10.1016/j.oceaneng.2023.116153.

Masters, I., Chapman, J. C., Willis, M. R., & Orme, J. A. C. (2011). A robust blade element momentum theory model for tidal stream turbines including tip and hub loss corrections. *Journal of Marine Engineering and Technology*, *10*(1), 25–35. http://dx.doi.org/10.1080/20464177.2011.11020241.

Mei, Y., Jing, F., Wang, X., Guo, B., Wang, S., & Lu, Q. (2025). Design and performance analysis of a tidal ducted turbine considering rotor-duct interaction. *Energy Conversion and Management*, *332*, Article 119758. http://dx.doi.org/10.1016/j.enconman.2025.119758.

Millero, F. J., Feistel, R., Wright, D. G., & McDougall, T. J. (2008). The composition of standard seawater and the definition of the reference-composition salinity scale. *Deep Sea Research Part I: Oceanographic Research Papers*, *55*(1), 50–72. http://dx.doi.org/10.1016/j.dsr.2007.10.001.

Muckelbauer, G. (1990). The shelf of cozumel, Mexico: Topography and organisms. *Facies*, *23*(1), 201–239. http://dx.doi.org/10.1007/bf02536714.

Neary, V., Gunawan, B., & Ryou, A. (2012). *Performance evaluation of HYCOM-GOM for hydrokinetic resource assessment in the florida strait*. Oak Ridge National Laboratory (ORNL), http://dx.doi.org/10.2172/1081936.

Ouro, P., Stansby, P. K., & Stallard, T. (2022). Analysis of the kinetic energy recovery behind a tidal stream turbine for various submergence levels. *International Marine Energy Journal*, *5*(3), 265–272. http://dx.doi.org/10.36688/imej.5.265-272.

Pfister, R., Schwarz, K. A., Janczyk, M., Dale, R., & Freeman, J. B. (2013). Good things peak in pairs: a note on the bimodality coefficient. *Frontiers in Psychology*, *4*, http://dx.doi.org/10.3389/fpsyg.2013.00700.

Pravilovic, S., Appice, A., Lanza, A., & Malerba, D. (2014). Wind power forecasting using time series cluster analysis. In *Discovery science* (pp. 276–287). Springer International Publishing, http://dx.doi.org/10.1007/978-3-319-11812-3_24.

Qasem, N. A. A., Generous, M. M., Qureshi, B. A., & Zubair, S. M. (2021). A comprehensive review of saline water correlations and data: Part II—Thermophysical properties. *Arabian Journal for Science and Engineering*, *46*(3), 1941–1979. http://dx.doi.org/10.1007/s13369-020-05020-5.

Räsänen, T., & Kolehmainen, M. (2009). Feature-based clustering for electricity use time series data. In *Adaptive and natural computing algorithms* (pp. 401–412). Springer Berlin Heidelberg, http://dx.doi.org/10.1007/978-3-642-04921-7_41.

Rastogi, A. K., Taterh, S., & Kumar, B. S. (2023). Dimensionality reduction algorithms in machine learning: A theoretical and experimental comparison. In *RAiSE-2023* (p. 82). MDPI, http://dx.doi.org/10.3390/engproc2023059082.

Ren, L., Pan, G., Yang, L., Wang, Y., Zheng, G., Yao, P., Zhu, Q., Zhu, Z., & Hartnett, M. (2023). Assessment of ocean circulation characteristics off the West Coast of Ireland using HF radar. *Remote Sensing*, *15*(22), 5395. http://dx.doi.org/10.3390/rs15225395.

REN 21 (2022). Renewables 2022 global status report. *Technical report*, REN21 Secretariat, url: http://www.ren21.net.

Ring, M., Rodríguez-Ocampo, P. E., Silva, R., & Mendoza, E. (2022). Extreme value analysis of ocean currents in the mexican caribbean based on HYCOM numerical model data. *Frontiers in Marine Science*, *9*, http://dx.doi.org/10.3389/fmars.2022.866874.

Salem, N., & Hussein, S. (2019). Data dimensional reduction and principal components analysis. *Procedia Computer Science*, *163*, 292–299. http://dx.doi.org/10.1016/j.procs.2019.12.111.

Sawada, K., & Handa, N. (1998). Variability of the path of the kuroshio ocean current over the past 25, 000 years. *Nature*, *392*(6676), 592–595. http://dx.doi.org/10.1038/33391.

Schubert, E., Sander, J., Ester, M., Kriegel, H. P., & Xu, X. (2017). DBSCAN revisited. Why and how you should (still) use DBSCAN. *ACM Transactions on Database Systems*, *42*(3), 1–21. http://dx.doi.org/10.1145/3068335.

Shahapure, K. R., & Nicholas, C. (2020). Cluster quality analysis using silhouette score. In *2020 IEEE 7th international conference on data science and advanced analytics*. IEEE, http://dx.doi.org/10.1109/dsaa49011.2020.00096.

Sharqawy, M. H., Lienhard, J. H., & Zubair, S. M. (2010). Thermophysical properties of seawater: a review of existing correlations and data. *Desalination and Water Treatment*, *16*(1–3), 354–380. http://dx.doi.org/10.5004/dwt.2010.1079.

Shirasawa, K., Tokunaga, K., Iwashita, H., & Shintake, T. (2016). Experimental verification of a floating ocean-current turbine with a single rotor for use in kuroshio currents. *Renewable Energy*, *91*, 189–195. http://dx.doi.org/10.1016/j.renene.2016.01.035.

Wang, X., Li, H., Chen, J., Jiang, C., & Bao, L. (2023). Research on solidity of horizontal-axis tidal current turbine. *Energies*, *16*(8), 3467. http://dx.doi.org/10.3390/en16083467.

Wang, X., Smith, K., & Hyndman, R. (2006). Characteristic-based clustering for time series data. *Data Mining and Knowledge Discovery*, *13*(3), 335–364. http://dx.doi.org/10.1007/s10618-005-0039-x.

Wang, L., Tang, X., & Liu, X. (2012). Blade design optimisation for fixed-pitch fixed-speed wind turbines. *ISRN Renewable Energy*, *2012*, 1–8. http://dx.doi.org/10.5402/2012/682859.

Warren-Liao, T. (2005). Clustering of time series data—a survey. *Pattern Recognition*, *38*(11), 1857–1874. http://dx.doi.org/10.1016/j.patcog.2005.01.025.

Yang, M.-S. (1993). A survey of fuzzy clustering. *Mathematical and Computer Modelling*, *18*(11), 1–16. http://dx.doi.org/10.1016/0895-7177(93)90202-a.

Yang, X., Haas, K. A., & Fritz, H. M. (2013). Theoretical assessment of ocean current energy potential for the gulf stream system. *Marine Technology Society Journal*, *47*(4), 101–112. http://dx.doi.org/10.4031/mtsj.47.4.3.