Applied Ocean Research 137 (2023) 103596

Contents lists available at ScienceDirect

Applied Ocean Research

journal homepage: www.elsevier.com/locate/apor

A hybrid model for online short-term tidal energy forecasting

Thomas Monahan *, Tianning Tang, Thomas A.A. Adcock

Department of Engineering Science, University of Oxford, Oxford, OX1 3LZ, United Kingdom

ARTICLE INFO

Keywords:
Hybrid machine learning
Tidal energy
Short-term power forecasting
Renewable energy
Singular spectrum analysis decomposition

ABSTRACT

A hybrid model is proposed for the short-term online prediction of tidal currents. The harmonic residual analysis (HRA) model is designed to augment the numerical schemes employed by tidal energy installations by forecasting the residual error of existing methods. Using a combination of techniques from Information and Fractal Theory, a novel component selection criterion for singular spectrum analysis (SSA) is used to remove true noise from the residual time series and to decompose the signal into components that are appropriate for linear-recurrent forecasting (LRF) and high order fuzzy time series (HOFTS) respectively. The performance of the HRA method is evaluated using a combination of simulated and real data from sites in the United Kingdom and the United States. Results demonstrate the model's viability for 6-minute and 1-hour forecast horizons across sites exhibiting variable degrees of non-linearity. Empirical analysis of the resultant tidal energy forecast verifies the superior accuracy and reliability of the HRA method when compared with existing numerical schemes. Simulated data from three sites at the Pentland Firth, UK is also provided to facilitate further study of the site's power generation characteristics and to allow for direct model performance comparisons.

1. Introduction

Tidal energy is a promising form of renewable energy. Unlike alternative modes of renewable power, it is highly predictable as it is primarily driven by the relative movements of the earth, moon, and sun which can be known indefinitely far in advance. This predictability has significant advantages for grid integration (Frost, 2022) as well as helping to bridge periods of low production from other renewables such as wind and solar. Tidal energy extraction may be done using either an impoundment (tidal barrage) or free standing turbines (tidal stream) (Adcock et al., 2021). The present paper focuses on the second of these.

While tidal currents are highly predictable in the long-term, short-term predictions (real-time to 1 h) suffer from the stochastic and non-linear behavior introduced by turbulence, waves, bathymetric interactions, and meteorological forcing (wind stress, atmospheric pressure, etc.) (Neill et al., 2014). The resultant variation in power production and forecast accuracy imposes severe challenges to the efficient dispatching of grid resources. Accurate short-term forecasts are critical to facilitate the economical implementation of tidal energy systems and can reduce the reliance on expensive energy storage systems (Beaudin et al., 2010). Due to the cubic relationship between current velocity and the power produced by tidal stream turbines, small errors in current predictions can cause significant issues for tidal energy practitioners. Hence, it is imperative that high-accuracy forecasting tools are developed. Due to the significant physical and economic implications of power production, tidal energy installations primarily rely on less accurate but reliable physics-based forecasting. As outlined in Section 2, while many improvements have been made to tidal current forecasting using machine learning, the lack of a physical basis has inhibited the widespread adoption of these methods. Furthermore, prior models, as discussed in Section 2, have been developed in an offline fashion, meaning the model is pretrained on historical data and deployed. Online machine learning models continually update as data becomes available and can therefore adapt to localized changes in the underlying data distribution (Saber and Khandelwal, 2017). This is critical to account for both the aperiodic behavior of fast-moving tidal currents and seasonal changes in flow characteristics. The aim of this paper is to develop an online forecasting method to enhance the predictive capabilities of the physical schemes currently employed by tidal energy practitioners.

2. Review of tidal forecasting literature

Numerous studies have proposed methods for tidal current forecasting across short, medium, and long-term forecast horizons. We define short-term forecasts as (real-time to 1-h), medium-term forecasts as (1-h to 1 day), and long-term forecasts as (>1 day). The two most pervasive themes in the growing tidal forecasting literature are data-preprocessing methods and hybrid/combination models. Due to the quasi-deterministic nature of tidal currents, data-preprocessing

* Corresponding author.
E-mail address: thomas.monahan@eng.ox.ac.uk (T. Monahan).

https://doi.org/10.1016/j.apor.2023.103596
Received 22 February 2023; Received in revised form 2 May 2023; Accepted 7 May 2023
Available online 26 May 2023

Table 1
Summary of tidal forecasting literature. m- denotes ensemble/aggregate forecasting methods. The associated reference number is provided before the names of each author.

Author	Method	Location	Forecast horizon	Type	Online	Train time
Kavousi-Fard and Su (2017)	DWT + m-SVR	Bay of Fundy, Canada	10 min	Data-Preprocessing + Kernel (aggregate)	No	Minutes-Hours
Safari et al. (2017)	EEMD + LSSVM	Florida, United States	6 min	Data-Preprocessing + Kernel	No	Minutes-Hours
Sarkar et al. (2019)	GP	Pentland Firth, UK (SIM)	1 h to 1 week	Probabilistic	No	Hours
Sarkar et al. (2018)	GP	United States[a]	12 h	Probabilistic	No	Hours
Safari et al. (2018)	AAQR + NPI	North America	6, 10 min, 1 h (single-step)	Probabilistic	No	Hours
Zhang et al. (2022)	DA-LSSVM	Straits of North America	6 min, 1 h (single-step)	Optimization + Kernel	No	Minutes-Hours
Lee and Jeng (2002)[b]	ANN	Taiwan[a]	1 h, 1 Year	Deep Learning	No	Hours
Chen et al. (2007)[b]	DB-WT + ANN	Taiwan + South China Sea	1–5 year(s)	Data Preprocessing + Deep Learning	No	Minutes-Hours
Qiao et al. (2020)	GA -BPNN	Zhoushan, China	10 min	Optimization + Deep Learning	No	Minutes-Hours
Aly and El Hawary (2013)	ANN + FLSM	Bay of Fundy, Canada	10 min	Statistical + Deep Learning	No	Minutes-Hours
Aly (2020)	WNN + ANN + FLSM	Bay of Fundy, Canada	10 min	Deep Learning	No	Minutes-Hours
Remya et al. (2012)	EOF + GA	Numerical (SIM)	1 h–24 h	Data Preprocessing + Genetic	No	Minutes-Hours
Kavousi-Fard (2016)	ARIMA + SVR	Bay of Funday, Canada	10 min	Optimization + Linear + Kernel	No	Minutes-Hours
Qian et al. (2022)	H-ELM + LSTM	Zhejiang, China	5 min	Deep Learning	No	24 min[c]
Yin and Wang (2016)	OS-ELM + IGG-Fuzzy-SA	United States[a]	1, 3, 6, 12 h (single-step)	Data Preprocessing + Statistical	Yes	.01 s[c].

[a]Indicates multiple ports in the same country.
[b]Indicates forecasts for tidal heights.
[c]Indicates training/prediction time was provided.

methods are applied to filter noise and improve model performance. Hybrid/combination models have been shown to improve predictive accuracy and reliability across forecasting horizons (Chang, 2014; Hajirahimi and Khashei, 2019). These models seek to maximize the available information by leveraging the advantages of each model to produce a globally optimized forecast. While the following review exclusively looks at tidal literature, the methods used and trends presented are widely used in wind, solar, and energy load forecasting (Wu and Hong, 2007; Soman et al., 2010).

A summary of each of the following methods can be found in Table 1. After decomposing the signal into sub-harmonic components using the discrete wavelet transform (DWT), Kavousi-Fard and Abdollah forecast each sub-harmonic component using an aggregate of forecasts made by several Support Vector Machines (SVM) (Kavousi-Fard and Su, 2017). Due to their superior performance over wavelet transforms for nonlinear and non-stationary time series, Empirical Mode Decomposition (EMD) methods have been successfully applied to tidal current prediction. Additionally, localized training was shown by Yang et al. (2009) to improve predictive performance for non-linear and short-term forecasts. Using an ensemble of EMDs (EEMD), Safari et al. (2017) was able to enhance the predictive accuracy of a Least-Square Support Vector Machine through improved localized learning. The authors of Zhang et al. (2022) use the fractal characteristics of tidal current data to determine optimal training timescales. They combine this insight with a dragonfly optimization algorithm to optimize the parameter selection of the LSSVM. While SVM methods demonstrate superior generalizability over standard deep learning methods, training times are prohibitive for online applications.

Probabilistic machine learning methods have shown promise in handling the uncertainty and noise inherent to real tidal current measurements. The authors of Sarkar et al. (2019) provide a method for spatiotemporal distribution modeling using a Gaussian Process (GP). Additionally, Sarkar et al. (2018) demonstrate the predictive accuracy of GPs to be superior to traditional HA methods in the presence of noise and non-linearities. Furthermore, the ability to forecast probabilistic confidence intervals can help to reduce the cost of power system operation. Safari et al. develop a non-parametric interval (NPI) model with a bi-level optimization formula for extreme learning machine (BOF-ELM) to forecast confidence intervals for tidal currents and levels (Safari et al., 2018).

Due to the remarkable ability of Neural Networks to model exclusive and non-linear relationships, considerable work has gone into their application for tidal levels and currents forecasting. In 2002, Lee and Tang demonstrated that Artificial Neural Networks (ANN) could produce higher accuracy forecasts than HA (Lee and Jeng, 2002). These deep learning methods suffer from a number of drawbacks; training is computationally intensive, data-hungry, and highly prone to overfitting or converging to local optima. A variety of techniques have therefore been applied to combat these shortcomings. Chen et al. employed the Daubechies Wavelet transform (DB-WT) to remove the high-frequency 'noise' components before using Artificial Neural Networks (ANN) to predict tidal heights (Chen et al., 2007). In order to prevent local optima convergence, Remyra et al. use a Genetic Algorithm (GA) to determine optimal Back Propagation Neural Network (BPNN) parameters for short-term tidal current prediction (Remya et al., 2012). In Aly and El Hawary (2013), tidal currents are first forecasted using a Least Squares Fourier Series Model (FLSM). The associated prediction error referred to as 'innovations' is then used as input to the ANN. The final forecast is the sum of the FLSM prediction and the ANN forecast. Long-Short-Term Memory (LSTM) models have also been used for tidal current forecasting due to their ability to capture both long and short-term dependencies. The authors of Qian et al. (2022) develop a Hybrid Hierarchical Extreme Learning Machine (H-ELM) and LSTM method to predict the random and periodic components of multi-layer tidal currents.

Online tidal level predictions were realized in Yin and Wang (2016) using a sample pool selection scheme to optimize an online sequential extreme learning machine (OS-ELM) in conjunction with the improved Gath-Geva Fuzzy Segmentation algorithm (IGG-Fuzzy-SA) and harmonic analysis (HA). To the best of our knowledge, this is the only study that has realized online tidal prediction and only considers tidal levels. Hence, there is a clear need for an online tidal current prediction method.

Outside of the tidal literature, multiple works have looked to combine the interpretability and reliability of physical methods with machine learning for short-term forecasting tasks. The authors of Dolara et al. (2015) combined a theoretical model of solar radiation with an

ANN for the task of photovoltaic power plant production forecasting. The PH-ANN model was able to learn the non-linear cloud cover dynamics missed by the solar radiation model and outperformed a stand-alone ANN. In Shen et al. (2021), the predictive capabilities of a modified gravity model were improved using a CNN encoder for the task of metro transit passenger flow forecasting. Daily precipitation forecasts were bolstered in Dong et al. (2023) by using XGBoost for bias correction of a numerical weather prediction model. Similarly, the authors of Cui et al. (2021) showed that daily evapotranspiration estimation could be improved by using a particle swarm optimization algorithm (PSO) to combine and optimize two surface temperature-vegetation index (Ts-VI) triangle models.

In order to combat the limitations faced by traditional statistical and deep learning models we propose a Harmonic Residual Analysis scheme (HRA) for online short-term tidal current forecasting. The HRA is designed to be easily modified to fit into the forecasting methods currently used by practitioners. With cumulative training and forecasting times of less than 10 seconds, our algorithm is capable of real-time online forecasting. Furthermore, the SSA-LRF and HOFTS only require 3 and 5 days of consecutive training data respectively. This form of localized training allows for rapid adaptability to variable conditions. The automated component selection scheme for the SSA-LRF is able to remove noise and identify components that are best suited for the LRF and Fuzzy forecasts respectively. By incorporating the stability of physical methods with the local adaptability of the combined SSA-LRF and HOFTS methods, HRA provides reliable and high-accuracy short-term online forecasts.

3. Overview of methods

3.1. Harmonic analysis

Tidal currents are primarily driven by the gravitational potentials induced by movements of the Earth, Sun, and Moon and can be described by Potential Field Theory. As such, a tidal prediction model can be derived from the characteristic harmonic frequencies of the elliptical orbits of these astronomical bodies. The resultant model is defined by tidal constituents which correspond to a particular harmonic frequency. This fundamental idea is the basis of the Harmonic Analysis method whose basic procedure is as follows:

For a time series of current velocity observations $y(t)$ with time steps t at fixed intervals Δt we can predict the tidal heights using N constituents such that

$$y(t) = \sum_{k=1}^{N} (a_k^+ e^{i\omega_k(t-t_0)+i\beta_k}) + (a_k^- e^{-i\omega_k(t-t_0)-i\beta_k}) + x_0 + x_1(t-t_0), \tag{1}$$

where a_k^+ and a_k^- correspond to the unknown complex amplitudes for the kth components, ω_k is the angular frequency and β_k is the corresponding astronomical argument for the kth components. The inclusion of the additional terms x_0 and x_1 is necessary to model the offset and trend respectively. Given that we are interested in modeling currents, $y(t)$ will have the form $y(t) = U(t) + iV(t)$ with the real and imaginary parts corresponding to the East–West and North–South Velocity components respectively. Fitting of this equation can be carried out using an Iteratively Reweighted Least-Squares Fitting procedure (IRLS). Let $\phi_m^k = w_k(t-t_0) + i\beta_k$. We can now rewrite Eq. (1) in matrix form such that

$$\begin{bmatrix} e^{i\phi_1^1} & \dots & e^{i\phi_1^N} & e^{-i\phi_1^1} & \dots & e^{i\phi_N^1} & 1 & (t_1-t_0) \\ e^{i\phi_2^1} & \dots & e^{i\phi_2^N} & e^{-i\phi_2^1} & \dots & e^{-i\phi_2^N} & 1 & (t_2-t_0) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ e^{i\phi_M^1} & \dots & e^{i\phi_M^N} & e^{-i\phi_M^1} & \dots & e^{-i\phi_M^N} & 1 & (t_N-t_0) \end{bmatrix} \begin{bmatrix} a_1^+ \\ \vdots \\ a_N^+ \\ a_1^- \\ \vdots \\ a_N^- \\ c_0 \\ c_1 \end{bmatrix}$$

$$= \begin{bmatrix} u_1 + iv_1 \\ u_2 + iv_2 \\ \vdots \\ u_N + iv_N \end{bmatrix}. \tag{2}$$

In order to reduce the sensitivity to non-tidal variations the IRLS method makes use of an iteratively determined weighting matrix W. A detailed description of the weighting matrix W and how to calculate it is given in Franco (1973). The final solution is obtained using

$$a = (S^* W S)^{-1} S^* W y, \tag{3}$$

where S is a matrix of sinusoidal basis functions evaluated at t_i. Given a sufficiently long time series (18.6 years) it is possible to resolve more than 500 tidal constituents using Harmonic Analysis. However, within the context of tidal current analysis, continuous time series longer than 60 days are rare. As such, it is important that we impose an automated criterion for tidal constituent selection. The Rayleigh criterion defines the minimum resolvable frequency between two constituents to be T^{-1} for a time series of length T (Franco, 1973). The Harmonic Analysis described in this paper is carried out using a Python adaptation of the MATLAB Harmonic Analysis program UTide (Codiga, 2011) with a Rayleigh threshold $R = 1$.

3.2. Singular Spectrum Analysis (SSA)

SSA is a powerful non-parametric tool for time-series analysis and filtering (Elsner and Tsonis, 1996). Similar to Principal Component Analysis (PCA) but for one-dimensional series, SSA decomposes a signal into independent oscillatory components. These components can then be grouped together and reconstructed to produce a filtered version of the input series. Common applications of SSA are noise removal, trend/periodicity extraction, and of particular interest to us: forecasting. We outline the 4 primary steps of SSA below:

1. **Embedding** The embedding procedure transforms the input time series into a sequence of lagged vectors using a sliding window. Once again consider an arbitrary time series $y(t)$, for convenience we let $y_i = y(t_i)$. Our time series can be written as $\{y_0, y_1, \dots, y_T\}$ where T is the length of the time series. The embedding procedure is defined in terms of the window length k, where $k < \frac{T}{2}$. The number of lagged vectors X_j produced is therefore given by $L = T - K + 1$ where $X_j = (y_j, \dots, y_{j+L-1})^T$ for $j = 1, 2, \dots, K$. The lagged vectors are then transformed into a 2-d Hankel Matrix of the form

$$X = \begin{bmatrix} x_1 & \dots & x_K \\ \vdots & \ddots & \vdots \\ x_L & \dots & x_T \end{bmatrix} \tag{4}$$

2. **Decomposition** The decomposition of our Hankel Matrix is carried out using Singular Value Decomposition (SVD). The Eigenvectors and Eigenvalues are calculated for $S = XX^T$. Since X is a Hankel matrix, we recognize XX^T is symmetric and diagonal. The eigenvectors U_i and corresponding eigenvectors λ_i for $i = 1, 2, \dots, d$ where $d = rank(X)$ are then arranged in descending order. Note $L = d$ if all eigenvalues are non-zero. For each eigenvector U_i, we compute the corresponding vector $V_i = X^T \times \frac{U_i}{\sqrt{\lambda_i}}$. The final SVD is the sum of all V_i with principal components given by elementary matrices E_i of the form

$$E_i = \sqrt{\lambda_i} U_i \times V_i^T. \tag{5}$$

3. **Grouping** The Grouping stage is critical to the success of the SSA and subsequent Linear Recurrent Forecast and consists of grouping the eigentriples (λ_i, U_i, V_i) into disjoint subsets. The eigenvalues λ_i are sorted in descending order with their magnitudes defined by their principal components' contribution to

the overall signal variance S^2. Typical eigentriple grouping procedures partition the components into disjoint sets based on their relative contributions. Some commonly used methods are grouping the first n components that account for 95% of the variance, correlation-matrix inspection, mean-based selection, and clustering (Hassani and Thomakos, 2010). We demonstrate the need for an alternative method in Section 3.4 and present our unsupervised algorithm for component selection in Section 3.5.

4. **Reconstruction** Once we have grouped the elementary matrices from the SVD we recover the corresponding time series using Diagonal Averaging. Let $L^* = min(L, K)$, and $K^* = max(L, K)$. The reconstructed series $y'(t)$ is given by

$$y'(t) = \begin{cases} \frac{1}{k} & \sum_{m=1}^{k} y_{m,k-m+1} \quad 0 \le k \le L^* \\ \frac{1}{L^*} & \sum_{m=1}^{L^*} y_{m,k-m+1} \quad L^* \le k \le K^* \\ \frac{1}{T-k} & \sum_{m=k-K^*+1}^{T-k+1} y_{m,k-m+1} \quad K^* \le k \le T \end{cases} \tag{6}$$

3.2.1. Linear Recurrent Forecasting (LRF)

Linear Recurrent Forecasting relies on the assumption that our newly reconstructed series satisfies a Linear Recurrent Relation (LRR). Through careful selection of SSA components, we are able to extract a signal that satisfies this condition. That is our time series $y'(t) = \{y'_0, y'_1, \dots, y'_T\}$ can be expressed as a LRR of order d such that

$$y'_{i+d} = \sum_{k=1}^{d} c_k y'_{i+d-k} \quad \text{for} \quad 1 \le i \le T - d \tag{7}$$

where c_k are the LRR coefficients which can be estimated using the method outlined in Afshar and Bigdeli (2011), and order $d < T$. Using this relation the Linear Recurrent Forecasting prediction $\tilde{y}$ is defined by

$$\tilde{y}(t) = \begin{cases} y'_i & i = 1, 2, \dots, T \\ \sum_{j=1}^{N-1} c_j \tilde{y}_{i-j} & i = T+1, \dots, T+H \end{cases} \tag{8}$$

where H is the number of steps ahead to forecast. It can be clearly seen that the Linear Recurrent Forecast is simply a weighted sum of the previous values of the reconstructed signal. Improper SSA component selection will therefore lead to inaccurate forecasts due to our assumption that y' satisfies a LRR. As such, we develop a robust method for SSA component selection in the following two sections.

3.3. Permutation Entropy (PE)

PE measures the complexity of a time series through the comparison of adjacent values. The partitioning of our time series into a set of 'ordinal patterns' and subsequent comparisons allow for the discrimination between chaotic, deterministic, and stochastic processes (Bandt and Pompe, 2002). PE has been shown to be robust to noise and invariant to other non-linear transformations. Furthermore, PE is applicable to arbitrary time series and is efficient to calculate making it an ideal candidate for the real-time selection of SSA components.

The complete time series can be partitioned into the set of ordinal patterns $\Pi = \{\pi_1, \pi_2, \dots \pi_{n!}\}$ defined in terms of the order n and the time lag τ. The order n determines the number of values contained within a given permutation. It follows that the set of all permutations Π will have cardinality $n!$. The time lag τ determines the number of time steps spanned by each segment of a permutation. Due to its temporal significance, τ plays a critical role in determining the frequency characteristics of the PE (Olofsen et al., 2008). The relationship between the order n and the time lag τ is illustrated in Fig. B.10. The procedure for calculating the PE can be defined in three steps:

1. Partition the input series into a sequence of sets whose length is equal to n.
2. Sort each set according to the relative magnitude of its components and identify its corresponding $\pi_i \in \Pi$.
3. Sum the number of sets corresponding to each π_i and compute the probability $p(\pi_i)$ for each π_i such that

$$p(\pi_i) = \frac{\text{Number of Occurrences of Corresponding Sets}}{\text{Total Number of Sets}}. \tag{9}$$

The final permutation entropy $H(n)$ is given by

$$H(n) = \sum_{i=1}^{n!} p(\pi_i) \log(p(\pi_i)). \tag{10}$$

It is clear the range of values $H(n)$ can take on is $0 \le H(n) \le \log(n)$. The normalized PE can therefore be written as $N(PE) = H(n)/\log(n)$. For the remainder of this paper, when referring to PE we will be referring to the normalized PE (i.e. $H(n)/\log(n)$).

Given $H \in [0, 1]$ it remains to be seen what the physical significance of these values are within the context of our time series. There are three cases:

1. $H(n) \approx 0$: The time series is deterministic. This corresponds to a singular π_i being more dominant.
2. $0 < H(n) < 1$: The time series is characterized by both deterministic and stochastic processes. This corresponds to a varied distribution of π_is.
3. $H(n) \approx 1$: The time series is stochastic. The distribution of π_i's is uniform.

We elect to calculate PE using the Python package Ordpy described in detail in Pessa and Ribeiro (2021).

3.4. Complexity analysis of SSA components

To establish the motivation for our automated component selection procedure we look at the SSA Decomposition of the UTide Residual of data collected from Estes Head (EPT0003) located in Eastport, Maine. The Eastport data was collected from 01 July, 2000, to 30 August, 2000, with a sampling rate of 6 min. The PE of the components of the SSA decomposition and their relationship to their respective 'significance' levels are shown in Fig. 1. The decomposition was carried out for a 7-day subset (1680 Samples) of the original series using a window length of $k = 540$.

It can be clearly seen that the magnitude of the SSA component's contribution to the overall variance is not necessarily an indication of its association with noise. As outlined in Section 3.2 and shown in Fig. 1, typical SSA grouping procedures fail to account for this behavior and include true noise in our reconstruction whilst simultaneously discarding deterministic information. Hence, a principled method is needed in order to ensure the proper selection of SSA components.

3.5. Unsupervised binary grouping of SSA components

Due to site-dependent temporal and frequency characteristics as well as the added degrees of freedom introduced by PE parameters n and τ it is impossible to impose a general PE threshold with which to describe a signal as being either noisier or more deterministic. Thus, our approach is to empirically determine this threshold using the decomposed components' PE characteristics. We adopt a method similar to Colonna and Nakamura (2018), using a simple Unsupervised Binary Grouping algorithm (UBG). Given two sets, UBG seeks to maximize the distance between the means of each set. The eigenvalues of the SSA decomposition do not contain any information relating to the structure or complexity of its signal. As such, we must first reconstruct each component to determine its PE. After computing the set $\{H_1, H_2, \dots, H_{k-1}\}$ of PEs for the reconstructed signals, we calculate the threshold H_{T1} using the following Binary Grouping Procedure:

1. Let $H_{T0} = mean(\{H_1, H_2, \dots, H_{k-1}\})$
2. Form two sets, A, B with A being the set of components with $H(n) < H_{T0}$ and B the set of components with $H(n) \ge H_{T0}$ and compute the mean values μ_A, μ_B for each.

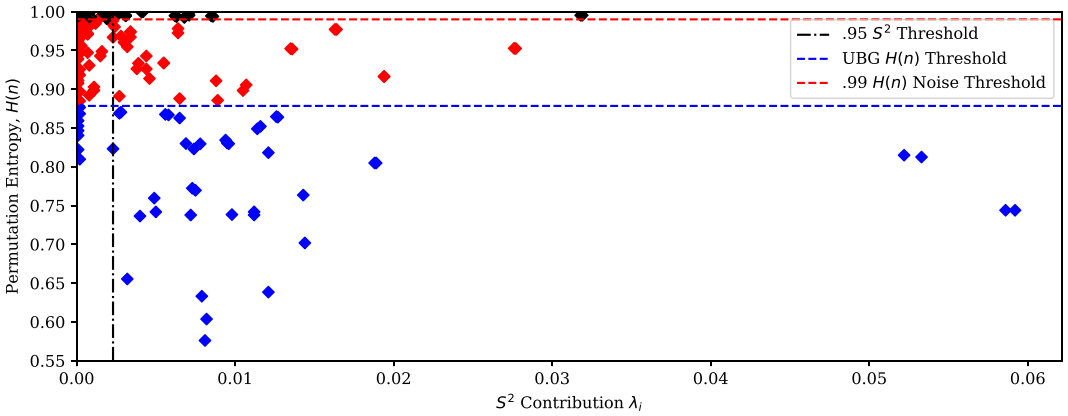


Fig. 1. Permutation Entropy of Reconstructed SSA Components for 7-day Eastport UTide Residuals. S^2 Contribution refers to the eigenvector λ_i's contribution to the overall signal variance. PE is calculated for each reconstructed signal for the length of the original series (1680 Samples). Points are colored according to their membership to either the LRF signal, the HOFTS signal, or true noise signal denoted by blue, red, and black points respectively. The standard .95 S^2 threshold is shown in black. The Unsupervised Binary Grouping $H(n)$ threshold is shown in blue. The true noise threshold for components with $H(n) > .99$ is shown in red.

3. $H_{T1} = (\mu_A + \mu_B)/2$.
4. Continue until $|H_{T0} - H_{T1}| < \epsilon$, where ϵ is the termination condition.

During our testing, $\epsilon = .001$ proved to be more than sufficient for our application. While noise adaptive thresholding is effective at separating the low and high-frequency components from the reconstructed signal it does not guarantee the signal exhibits sufficient self-similarity for LRF. As such, we adopt the Hurst Exponent as a measure of the series' self-similarity. A complete explanation of the Hurst exponent and how it is calculated is given in Hurst (1951). If a series has a Hurst exponent greater than 0.5 then it contains self-similarity and can therefore be used for LRF. If a series has a Hurst index of 0.5 or less then it is non-independent or randomly independent and cannot be used for LRF. In this case, the SSA is simply used to remove noise from the signal and the combined low and high-frequency signals are provided to the Fuzzy Time Series method outlined in the next section.

3.6. Fuzzy time-series forecasting

Interest in Fuzzy Time Series (FTS) has grown dramatically in recent years due to the advantages FTS provides over conventional methods. Namely, FTS requires significantly less data than statistical methods (Singh, 2007). FTS makes use of linguistic variables to form simple logical relationships and generate subsequent forecasts. As such, FTS forecasting is efficient, interpretable, and adaptable. FTS is built on fuzzy set logic. Traditional set logic states that elements must belong to single sets, membership x is therefore defined by $x \in \{0, 1\}$. Fuzzy sets allow for simultaneous set membership with membership being defined for the range $x \in [0, 1]$. The basic procedure for FTS Forecasting can be defined as follows:

1. **Universe of Discourse** We define the Universe of Discourse, D, for our time series $y(t)$ as $D = [\min(y(t)), \max(y(t))]$ where $D = \{u_1, u_2, \ldots, u_n\}$.
2. **Partition** D We can partition the Universe of Discourse into a set of overlapping intervals defined by the fuzzy sets f_j. Let f_A be an arbitrary fuzzy set membership function such that $f_A \rightarrow [0, 1]$. The fuzzy set A is given by
$$A = \frac{f_A(u_1)}{u_1} + \frac{f_A(u_2)}{u_2} + \cdots + \frac{f_A(u_n)}{u_n}. \tag{11}$$
The resultant FTS $F(t)$ is defined by the set of all fuzzy sets f_j for $j = A, B, \ldots$, defined on the Universe of Discourse D. There are two primary types of membership functions: even spacing, and uneven spacing. [2] provides a more detailed description of partitioning methods and their applications.
3. **Create Fuzzy Logical Relationships** We assume that there exists a relationship $R(t-1, t)$ between consecutive values $F(t-1)$ and $F(t)$. This implies that $F(t) = F(t-1) \circ R(t-1, 1)$ where $\circ$ is an arbitrary composition operator. Thus the Fuzzy Logical Relationship (FLR) is given by $F(t-1) \rightarrow f(t)$. We can extend this treatment to include multiple FLRs by assuming $F(t)$ is caused by the N preceding terms. It follows that $F(t-N), \ldots, F(t-1) \rightarrow F(t)$.
4. **Forecast** Using the FLRs defined in the previous section we can compute future values by identifying their corresponding FLR. Forecasted fuzzy values are de-fuzzified based on the fuzzy set membership functions defined in Step 2.

While FTS can be applied to arbitrary time series, we found it necessary to difference our time series prior to applying FTS in order to remove non-stationarity and improve forecasting accuracy. A number of methods have been proposed for FTS, we make use of Chen's method which leverages High Order Fuzzy Time Series (HOFTS). A full description of Chen's method is given in Chen (1996). FTS Forecasts in this paper are implemented using the pyFTS toolbox (Silva et al., 2018).

4. Harmonic residual analysis forecasting scheme

Based on the methods above, we present an Online Harmonic Residual Analysis Forecasting scheme that utilizes HA-SSA-UBG-LRF-FTS which we refer to as (HRA) (see Figs. 2 and 3).

The basic procedure for Real-Time Online prediction can be summarized as follows:

1. Perform HA on the input time series, generate a forecast, and compute the residuals between the HA and the observed data.
2. HA residuals for the previous 3 days of observations are fed into the SSA. The series is embedded into the Hankel Trajectory Matrix and decomposed using SVD.
3. Individual signals are reconstructed for each eigentriple. The PE of the resultant signals is computed and stored. UBG is applied to determine the optimal PE threshold with which to define our signal and noise sets.
4. Components with $PE > .99$ are removed. The remaining components are partitioned into two sets corresponding to low and high PE.
5. The high PE set is reconstructed and set aside to be predicted using HOFTS.
6. The components of the low PE set are reconstructed and the moving Hurst exponent is calculated. If the Hurst exponent is greater than 0.6 then the signal demonstrates sufficient autocorrelation and is forecasted using LRF. If the Hurst exponent is less

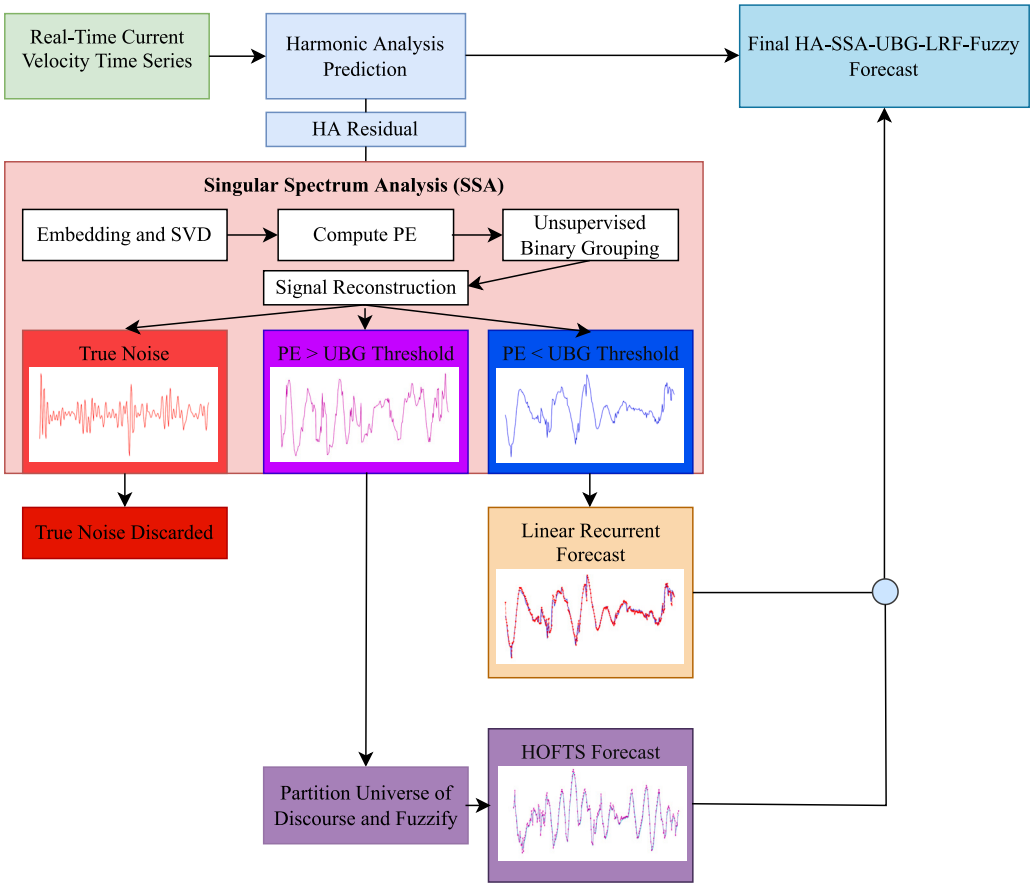


Fig. 2. HRA model architecture. HA = Harmonic Analysis, SVD = Singular Value Decomposition, PE = Permutation Entropy, UBG = Unsupervised Binary Grouping, LRF = LRF, HOFTS = High-Order Fuzzy Time Series.

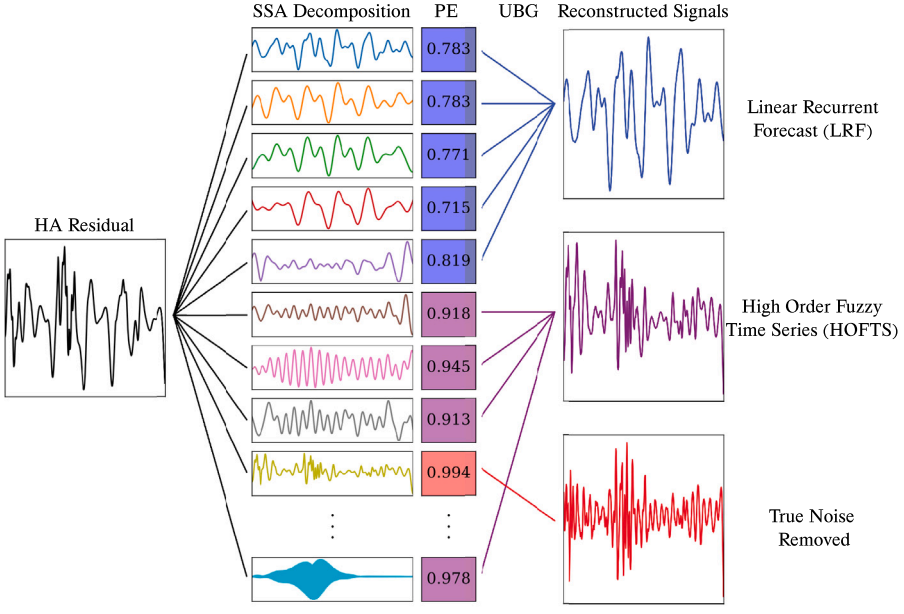


Fig. 3. SSA decomposition and automated grouping explained. Components are shown in descending order of their contribution to the signal S^2 and scaled for visualization.

than 0.6 then the signal is inappropriate for use in LRF and is combined with the high PE signal in the next step.
7. The high PE signal is then presented to the HOFTS. The data is differenced, partitioned using an entropy partitioner, and then predicted using the Chen Method whose procedure is outlined in Section 3.6. Due to the low computational cost of HOFTS, we make use of the entire set of historical residuals. As such, the performance of the HOFTS will improve as the duration of the Online implementation increases.
8. The final HRA Forecast is the ensemble of predictions from the Harmonic Analysis, Linear Recurrent Forecast, and Fuzzy Time Series.

Due to the residual-based construction of our prediction scheme, two weeks of continuous observations are necessary in order to realize real-time Online predictions. This is the first time SSA-LRF and HOFTS have been used to forecast tidal currents. The proposed model is similar in structure to the model outlined in Sulandari et al. (2020), however, we have made a number of critical changes to their proposed architecture. First, the automated component selection for SSA ensures that the reconstructed signal is suitable for LRF. Using the novel selection criterion based on information and fractal theory, our model is able to recognize and adapt to local anomalies. In this way, the HRA method is insulated from the erroneous and unpredictable errors faced by other ML methods. Second, we find it to be inappropriate to use the residuals of the LRF in our HOFTS prediction. Due to the fact that component grouping is not necessarily uniform between decompositions, LRF residuals may not be consistent between decompositions and can lead to inaccurate forecasts. Additionally, by using ML methods to bolster the performance of physical schemes, the resultant HRA model retains the reliability demanded by tidal energy practitioners and combines it with the high accuracy of ML.

4.0.1. A note on data leakage

Data leakage is a critical consideration when decomposing time series and can result in erroneous predictions as well as model testing not being rigorous (Niyogi, 2023). A brief overview of the two main types of leakages and how HRA avoids them is given below.

1. ***Feature Leakage*** *results from the inclusion of data that will not be available to the model in testing. HRA predictions are computed using only the past values of the univariate reference series and are therefore uncontaminated by feature leakage.*
2. ***Train-Test Leakage*** *occurs when there is an overlap between the train and test sets for model training or decomposition. This can occur as a result of random sampling, or if decomposition is carried out on the entire time series prior to separation (one-time decomposition)* (Niyogi, 2023). HRA avoids train-test leakage through temporal partitioning (partitioning before and after a given timestamp) prior to the HRA algorithm being run. In this way, HA and the SSA decomposition are only run on the train set and receive no information or overlap with the test series.

5. Results

5.1. Pentland simulation

The Pentland Firth, which connects the Atlantic Ocean to the North Sea, is characterized by exceptionally fast tidal currents, thus making it a high-interest site for tidal energy development. We have adopted the numerical scheme proposed and validated in Adcock et al. (2013) to simulate the currents at the Pentland Firth. We first benchmark our model using this simulated data in order to test our model's ability to capture severe non-linear current interactions without the presence of measurement noise or meteorological effects. The simulation models the tidal flows in the Pentland Firth by solving the shallow water equations using the Discontinuous Galerkin (DG) method (Kubatko et al., 2006). The depth-averaged model is forced with the tidal constituents K1, K2, M2, MU2, N2, NU2, O1, and S2 and shows reasonable agreement with field data at various locations.

Table 2
Model performance for simulated 6 min online one-step forecasts for 72 h of Pentland simulation data.

Node	Prediction data	RMSE	MSE	MAE	Max error	r^2
A	U	0.00059	0.0	0.00047	0.00199	0.99999
	V	0.00107	0.0	0.00085	0.00336	0.99983
B	U	0.04269	0.00182	0.02769	0.2899	0.99933
	V	0.04933	0.00243	0.03188	0.29052	0.98641
C	U	0.0273	0.00075	0.02014	0.12429	0.99485
	V	0.06088	0.00371	0.03971	0.32678	0.98903

A plot of the Pentland Firth, node locations, and the simulated currents is shown in Fig. 4. The time series consists of simulated observations from January 1st, 2014 through October 31st, 2014 at a sampling rate of 5 min. We present a case study of three nodes characterized by increasing levels of non-linearity. C.5 provides the coordinates for each node and numerical designations.

5.1.1. Experimental setup

To test our model's forecasting performance we benchmark using a combination of simulated and real data. We provide case studies of sites characterized by increasing degrees of non-linearity. Model performance is tested for different sampling intervals and the power predictive capabilities are evaluated. We assess multiple aspects of our model's predictive capabilities using Root-Mean Squared Error (RMSE), Mean Squared Error (MSE), Mean Absolute Error (MAE), Maximum Absolute Error (Max Error), and R-Squared (R^2). The definition of each statistical measure is given in Appendix A. Additionally, the SSA parameters used for each site can be found in Appendix E, Table E.8.

Due to the fact that tidal energy site stakeholders demand a physical basis for forecasting, the focus of this paper is not on pure Machine learning methods. A comprehensive set of comparisons between the HRA method and current state-of-the-art ML methods (RF, SVR, ANN, and LSTM) can be found in Appendix D for both real (Table D.7) and simulated data (Table D.6). It should be noted that HRA outperformed or tied the best-performing ML methods for 26/30 and 19/20 performance metrics for the simulated and real data respectively. Furthermore, it can also be seen that the pure ML methods have significant variability in performance accuracy across sites. A description of the model architectures and hyperparameters for each method can also be found in Appendix D.0.1.

5.1.2. HRA model performance (Simulated Data)

HA predictions were computed using 15 days of reference data. SSA-LRFs were then run for 7 days using the computed HA residuals and a training-set size of 72 h (720 observations). The Hurst exponents for the 72-h training sets are greater than .7 indicating the decomposed signal has a high level of long-range correlation and is thus suitable for LRF. The high-PE signal was then provided to the HOFTS. The subsequent 3-day Online forecasts are displayed in Fig. 5. It should be noted that the "spin-up" period for full online operation can be completed using only 15 days of reference data. Given most tidal energy sites will have ongoing data streams, this should not be prohibitive. It is clearly seen that despite increasing non-linearity between nodes, the model demonstrates a high level of agreement with the test data. Table 2 presents forecasting results. Forecasting accuracy decreases as the tidal signal is muddled by increased non-linearity and non-stationarity as expected.

Node A has minimal non-linearity and is easily modeled by HRA. The HA is able to account for 99.99% of the total signal variance, and the subsequent LRF and HOFTS are able to account for the remaining signal variance to an extremely high level of precision. There are

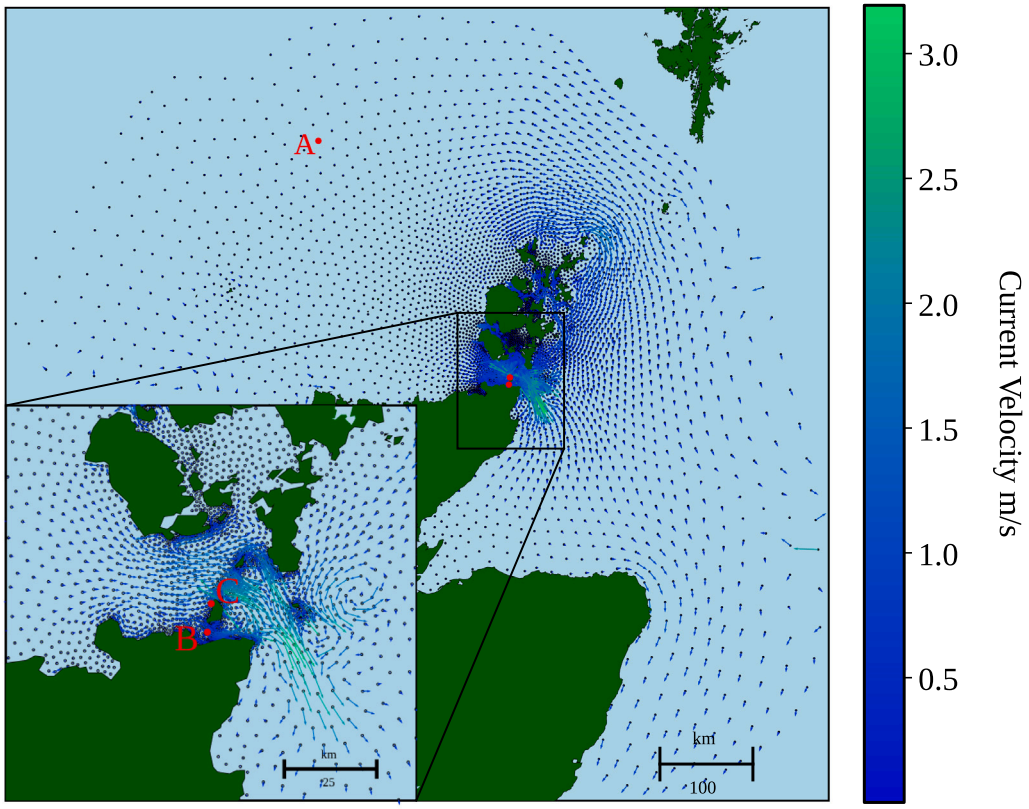


Fig. 4. Map of the Pentland Firth with node locations and velocity field overlaid. The depth-averaged simulation contains 5200 nodes with grid resolution varying from 150 m in the Pentland Firth and 20 km beyond the continental shelf.

increased levels of non-linearity and non-stationarity present at Node B resulting in steep fluctuations in current velocity as the tide direction changes. The HRA shows improved tracking ability through these sharp changes. While there is a tendency to over-predict the sharp changes in North–South velocity during the flood tide, the model is able to quickly return to agreement with the data. The increased accuracy around and after these turn points is reflected in the low RMSE values of .043 and .049 for the East–West and North–South currents respectively.

Node C is characterized by extreme non-linear tidal flow interactions. As a result, we observe sharp changes in velocity during the slack-tide periods as well as a shift in dominant velocity components from East–West to North–South. Despite these extreme characteristics, the proposed model is able to account for over 99.4% of the signal variance. While typical tidal energy sites are not characterized by such severe non-linearity, these results demonstrate the ability of the residual analysis method to capture non-linearities and have potential applications in archipelago regions where fast-moving tidal current mixing is prevalent.

5.2. Real sites

In this section, we evaluate our model's performance on real data by simulating Online forecasts for two potential tidal energy sites in the United States. The Cook Inlet in Alaska is a high-interest tidal energy site that contains nearly a third of the nation's potential tidal energy resource (Wang and Yang, 2020). Data is taken from the East Foreland (COI0503) from 18 May 2005, to 29 June 2005, at an approximate depth of 30.24 m, with measurements taken at a sampling rate of 6 min. The second set of data was taken from Estes Head (EPT0003) located in Eastport, Maine. The Eastport data was collected from 01 July 2000, to 30 August 2000, with a sampling rate of 6 min.

Table 3
Summary statistics for Eastport and Cook Inlet 6 min predictions.

Node	Prediction data	RMSE	MSE	MAE	Max error	r^2
Cook Inlet	U	0.06834	0.00467	0.05429	0.23987	0.99792
	V	0.0697	0.00486	0.0543	0.25361	0.84652
Eastport	U	0.02099	0.00044	0.01548	0.19906	0.95441
	V	0.03111	0.00097	0.02349	0.17139	0.99836

5.2.1. HRA model performance (Real Data)

As in the previous section, HA predictions were generated using 15 days of reference data. Linear Recurrent forecasts were run for seven days using only 36 h (360 observations) as a training set size. The reduced training set size yielded increased performance for the sharp oscillations at the site at Estes Head. This was validated by the Hurst exponent which showed values less than 0.5 for training sets greater than 36 h and a Hurst exponent of ≈.6 for 36 h. The performance statistics for the final 3-day combined forecasts are shown in Table 3.

The East Foreland is characterized by a dominant East–West current and a relatively slow-moving North–South current exhibiting a high degree of noise. It can be seen in Fig. 6 that the HRA method demonstrates a high level of agreement for the East–West component through the sharp changes present during peak current velocities. This observation is confirmed by the r^2 value of .998. Due to the high level of noise, the North–South component exhibits less self-similarity which is confirmed by the reconstruction's Hurst exponents being less than 0.5. The adaptive construction of the HRA method is able to accommodate these signal characteristics by foregoing the LRF and predicting the "noise-free" residual using the HOFTS. It can be seen that the model prediction is suitable for a majority of test points and demonstrates reasonable agreement even with sharp velocity fluctuations.

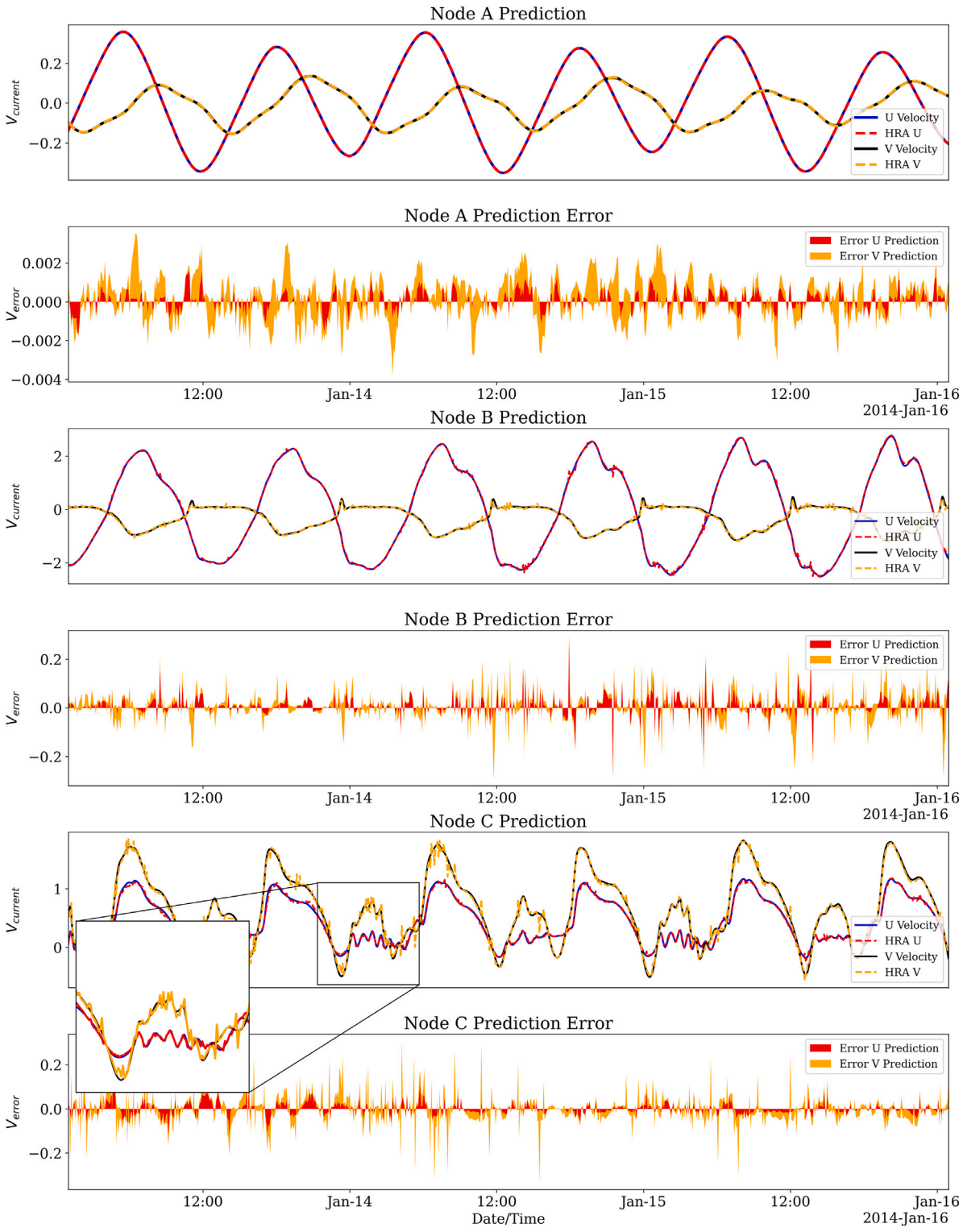


Fig. 5. 72 h of 6 min online forecasts and associated prediction error for Pentland Simulation Data. Note the current velocity error axes are scaled for each node.

The Eastport site is comparatively balanced, with both velocity components contaminated by severe non-linearity due to the site's location in a chain of archipelagos. As we demonstrated using the Pentland simulation, the HRA method is able to capture severe non-linear current interactions. As a result, the HRA method boasts Mean Average Errors of 0.015 m/s and 0.023 m/s for the East–West and North–South components respectively. The increased accuracy and high stability of the HRA method can be seen in Fig. 7.

6. Power predictive performance

Accurate tidal current predictions 1 h ahead are critical for the grid integration of tidal turbines. Typical acoustic doppler current

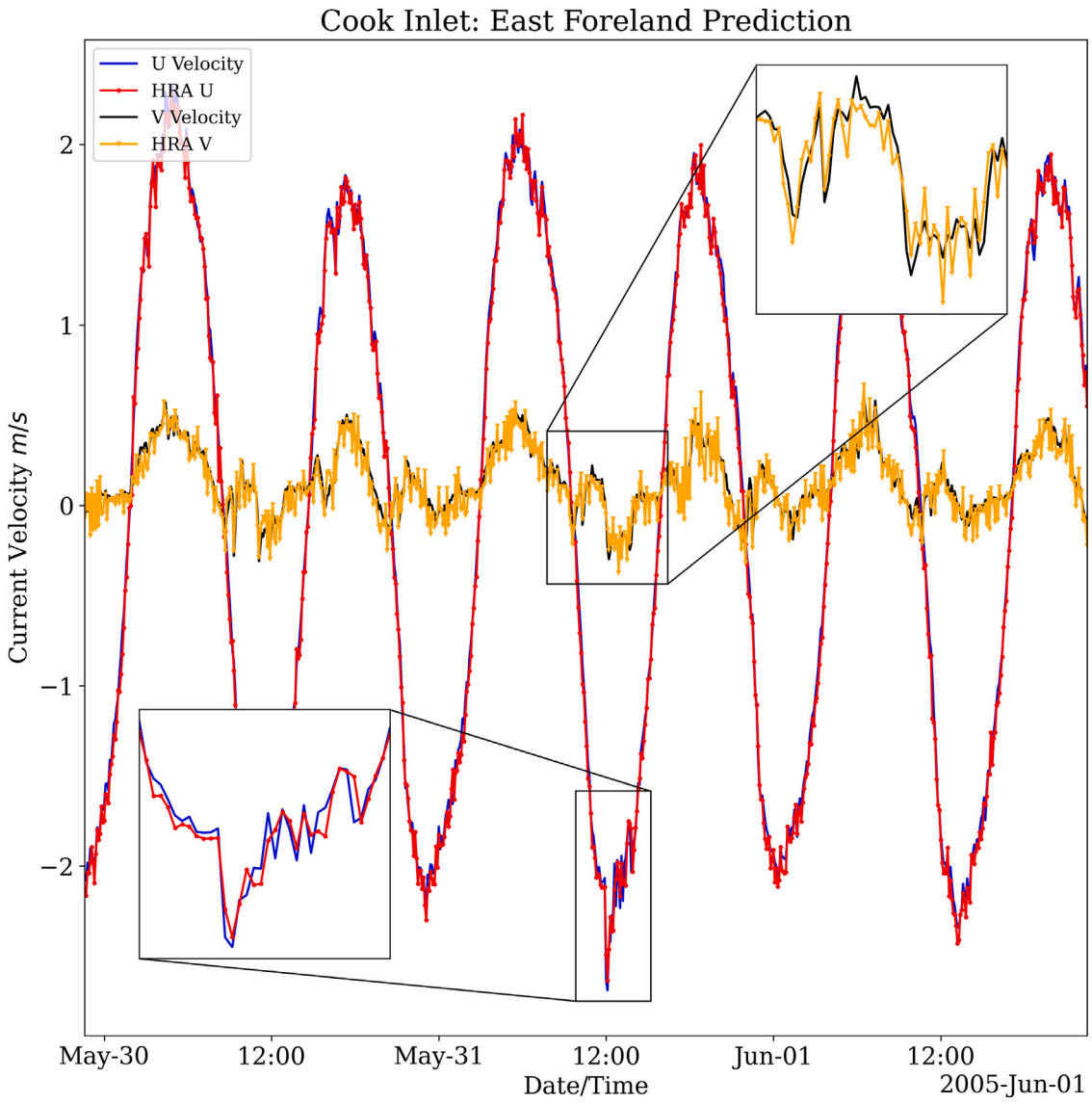


Fig. 6. 72 h of 6-min online predictions for the ADCP at the East Foreland in the Cook Inlet, AK. Observed North–South and East–West velocities are shown in black and blue. HRA predictions for the North–South and East–West components are shown in orange and red.

profilers (ADCP) provide current measurements at a sampling rate of 6 min. As such, forecasting can either be carried out using multi-step or single-step forecasting. When we discuss time-series forecasting we typically only consider one-step forecasting; that is forecasting a single time step ahead. Multi-step time series forecasting is a difficult task that is performed using either a recursive or single-shot forecasting method (Kline, 2004). The recursive forecasting method makes multi-step predictions by recursively generating predictions one step ahead. As a result, recursive multi-step forecasting suffers from severe error propagation and rapid divergence. Single-shot multi-step forecasting predicts the entire multi-step horizon in one step. As such, the single-shot method is unable to consider the dependencies between observations (Taieb et al., 2012). Due to these shortcomings and the relatively slow-changing nature of tidal currents, single-step 1-h predictions are preferred for energy forecasts and are performed using down-sampled data.

It is important to understand the relationship between tidal currents and tidal energy generation in order to understand the importance of high-accuracy forecasts. Turbines are defined by a cut-in and rated velocity, V_{cut-in} and V_{rated}. Below V_{cut-in} the turbine is not operational and above V_{rated} power output is constant. The power equation for a standard tidal turbine is given in Eq. (12). A complete description of Tidal Power Generation systems is given in Liu et al. (2015). Due to the cubic and discontinuous nature of the power equation, the forecasted power output is highly sensitive to errors in tidal current predictions.

This cubic relationship is clearly illustrated in Fig. 8 Panel (a). Despite high average accuracies, small errors in current velocity forecasts can result in significantly different power outputs. To illustrate our model's power predictive performance results for 1-h tidal current predictions and the associated power forecasts are provided.

6.0.1. HRA model performance (1-h Horizon)

The effectiveness of HRA is evaluated for 1-h predictions using downsampled data from the Pentland Firth simulation, Eastport, and the Cook Inlet. HA predictions are generated using 30 days of reference data from each site (720 observations). SSA-LRF forecasts are run for 15 days on the HA residuals using a training set length of 15 days (360 observations) and a window length of 150. The window length was determined based on the recommendations given in Golyandina (2010) and validated using the Hurst exponent. The SSA reconstructed high-PE signal is presented to the HOFTS and the final prediction is aggregated. Due to the smoothing nature of downsampling, the behavior of the

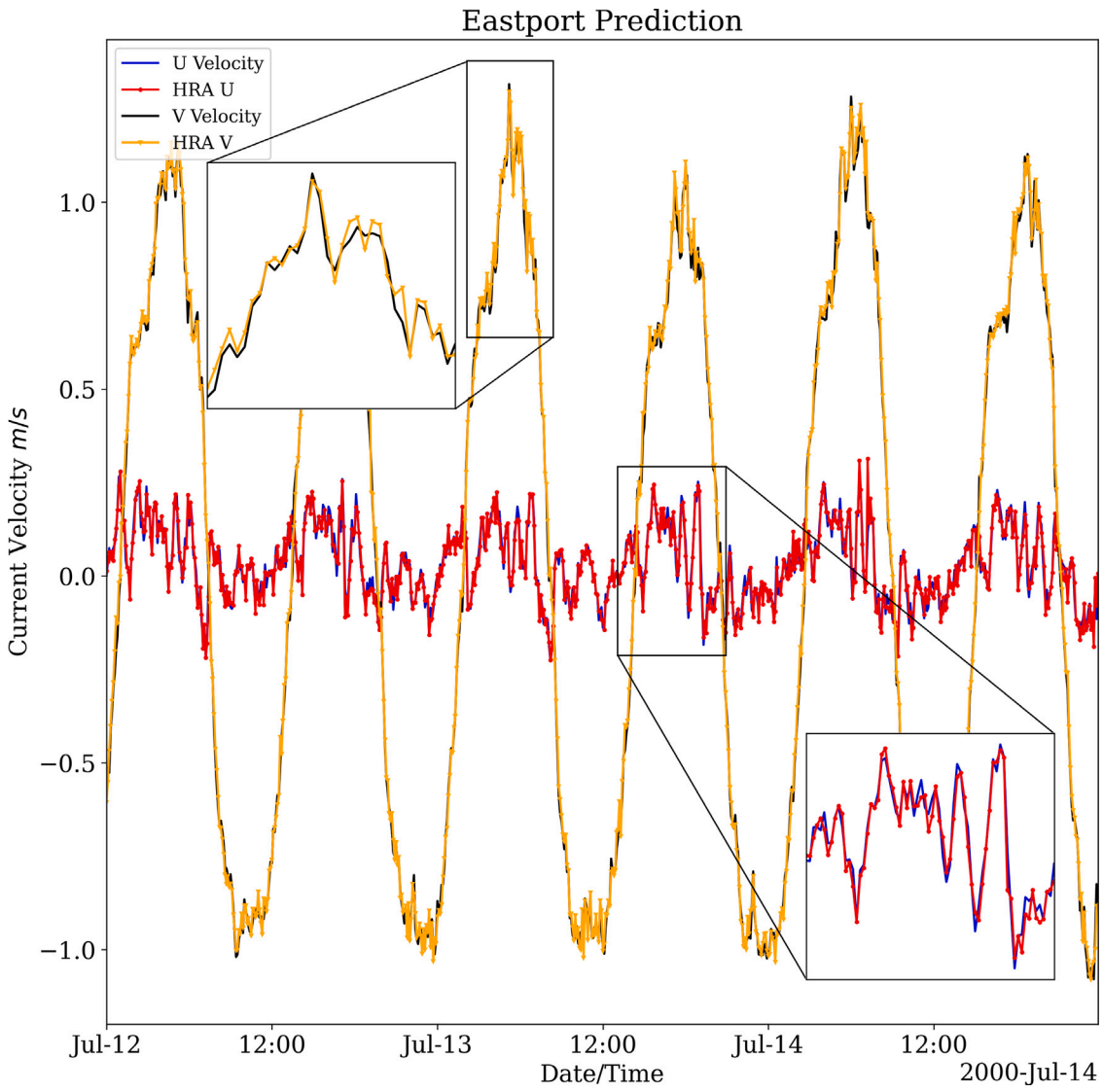
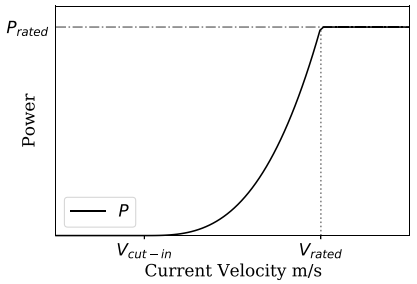


Fig. 7. 72 h of 6-min online predictions for the ADCP at Estes Head in Eastport, ME. Observed North–South and East–West velocities are shown in black and blue. HRA predictions for the North–South and East–West components are shown in orange and red.



(a)

$$P = \begin{cases} 0 & 0 \le V_t < V_{cut-in} \\ 0.5 C_p \rho A V_t^3 & V_{cut-in} \le V_t < V_{rated} \\ P_{rated} & V_t \ge V_{rated} \end{cases} \quad (12)$$

(b)

Fig. 8. Panel (a) shows the theoretical power curve P for a Tidal Current Turbine (TCT). Panel (b) shows the Power Equation where C_p is the power capture coefficient (0.4 – 0.5), ρ is the water density, A is the swept area of the blades (81π), V_t is the tidal current speed, V_{cut-in} is the minimum current speed for turbine operation (1 m/s) and V_{rated} is the rated operational speed (3 m/s) (Liu et al., 2015).

resultant time series and computed HA residuals is different from that of the 6-min data. For the 6-min data, the reconstructed SSA low-PE signal's Hurst exponents were all found to be suitable for the LRF forecasts with the exception of the Cook Inlet's North–South component. Conversely, for the downsampled 1-h data, all reconstructed signals were suitable for LRF with the exception of the Eastport East–West component. As such, after removing true noise components with $PE > .99$ the combined low and high-PE signals for the Eastport East–West component were predicted by the HOFTS. Table 4 presents the performance indicators for the 15 days of online forecasts.

Table 4
Pentland, Eastport, and Cook Inlet 1 h performance results for 15-days of simulated online forecasts.

Node	Prediction data	RMSE	MSE	MAE	Max error	r^2
n4768	U	0.05287	0.00279	0.03388	0.23079	0.99953
	V	0.06564	0.00431	0.0519	0.15373	0.98797
n4496	U	0.09211	0.00848	0.07223	0.2879	0.98107
	V	0.15409	0.02374	0.12071	0.38046	0.97955
Cook Inlet	U	0.05389	0.0029	0.03706	0.25675	0.99885
	V	0.02507	0.00063	0.01677	0.08573	0.97461
Eastport	U	0.03073	0.00094	0.02083	0.12205	0.85397
	V	0.05301	0.00281	0.04074	0.22691	0.99547

It can be clearly seen in Table 4 that the HRA method yields similar performance for 1-h predictions as it did for 6-min predictions. Furthermore, the model demonstrates high performance across sites exhibiting significantly different characteristics and degrees of non-linearity. The high performance across assessment metrics over a full Spring-Neap cycle verifies that the HRA method is suitable for 1-h forecasts.

Due to the low self-similarity in the SSA reconstructed signal for the Eastport East–West component, the increased reliance on the HOFTS prediction results in a decreased prediction accuracy. Despite this, the prediction retains a low MAE of 0.02 and a maximum error of 0.122. Hence, we have shown that by combining the stability of numerical methods with the flexibility of ML methods, HRA provides increased reliability in the presence of unknown conditions and results in a reduction of maximum error compared to conventional ML models.

Power predictions are computed using Eq. (12). Fig. 9 shows a 10-day subset of the 15-day forecast as well as the associated Absolute Error distributions comparing the HRA method with traditional HA. It is clear that HRA can significantly reduce both the maximum and mean prediction error for real-time power forecasting when compared to numerical schemes. As such, we present the HRA method as a reliable tool to augment the numerical schemes currently employed by tidal energy practitioners.

7. Conclusion

To the best of our knowledge, this paper presents the first online tidal current prediction method. The hybrid ML approach combines the reliability of classical HA with the improved localization and flexibility of ML methods. The primary developments from this paper are as follows:

1. The experimental results verify the HRA method's viability for both 6-min and 1-h online tidal current forecasting. The method is robust to non-linearity and non-stationarity and has been shown to work across a diverse set of sites.
2. This is the first time an automated component selection criteria has been used for SSA-LRF. The unsupervised binary grouping approach, augmented by the moving Hurst Exponent, results in improved LRF accuracy and reliability. The method shows promise as a general tool for SSA-LRF component selection, however, further study is necessary.
3. We have demonstrated that the HRA approach can achieve better accuracy over traditional HA and can therefore be used to augment the forecasting schemes at existing tidal energy sites. Moreover, the proposed model realizes accurate online predictions for both tidal currents and the resultant power generation. Comparison with state-of-the-art online prediction methods remains to be determined.

Future work will look to expand the HRA method to incorporate multiple data streams, including adjacent ADCPs and meteorological forcing variables (wind stress, atmospheric pressure, etc.). The inclusion of these parameters can enhance the model's predictive capabilities and increase tolerance to severe weather. Data for the three nodes (n500, n4768, and n4496) from the Pentland simulation will be made available and can be used to study the behavior of fast-moving tidal currents as well as act as a standard benchmark for future models.

CRediT authorship contribution statement

Thomas Monahan: Conceptualization, Methodology, Software, Writing – original draft. **Tianning Tang:** Writing – review & editing, Supervision. **Thomas A.A. Adcock:** Writing – review & editing, Supervision, Data curation.

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.

Data availability

Data will be made available on request.

Acknowledgment

TT acknowledges support by the Eric and Wendy Schmidt AI in Science Postdoctoral Fellowship, a Schmidt Futures program.

Appendix A. Performance measures

Root Mean Squared Error (RMSE):

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{A.1}$$

Mean Squared Error (MSE):

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{A.2}$$

Mean Absolute Error (MAE):

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{A.3}$$

Maximum Absolute Error:

$$Max\ \ Error = \max(|y_i - \hat{y}_i|) \tag{A.4}$$

R-Squared (r^2):

$$r^2 = 1 - \frac{\frac{1}{n}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2}{\frac{1}{n}\sum_{i=1}^{m}(y_i - \bar{y}_i)^2} \tag{A.5}$$

Note n is the number of points sampled, y_i is the actual value, $\hat{y}_i$ is the predicted value, and $\bar{y}_i$ is the mean value.

Appendix B. Visual summary of permutation entropy

We present herein a visual explanation of Permutation Entropy in Fig. B.10. Panel (a) illustrates the relationship between the order n and time lag τ with the timestep shown on the x-axis and the corresponding unitless observation values shown on the y-axis. The importance of τ in determining the frequency characteristics analyzed can be clearly seen. Panel (b) is the corresponding normalized histogram of permutation occurrences for (a). From this, it can be seen how the partitioned segments are classified.

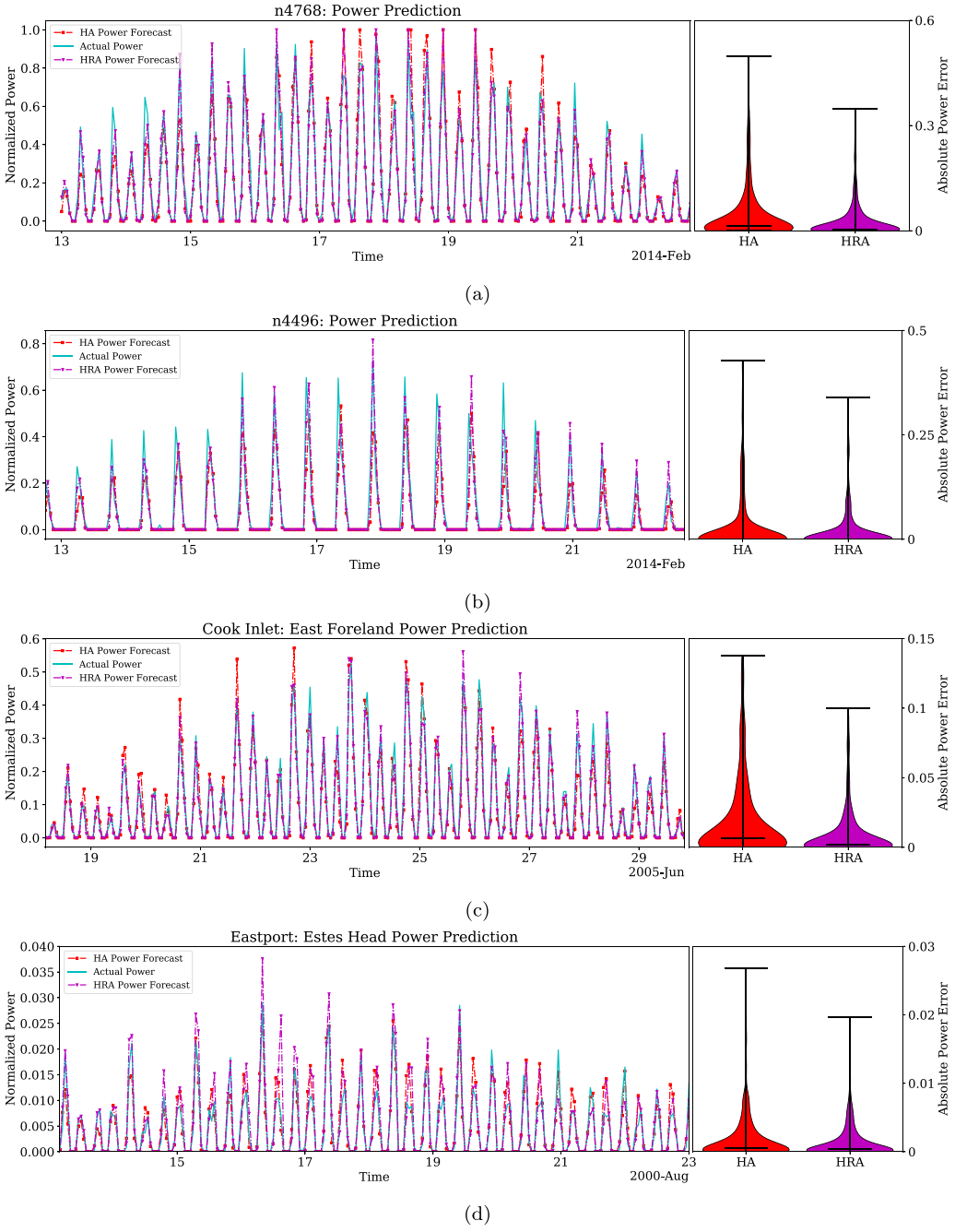


Fig. 9. Power prediction comparisons between Harmonic Analysis and the HRA method for 1 h Downsampled data over a Spring-Neap Cycle. (a) Pentland n4768 (b) Pentland n4496 (c) Cook Inlet, East Foreland (d) Eastport, Estes Head. The corresponding violin plot shows the distribution of absolute error values with whiskers indicating the mean and maximum error values. Power is normalized such that 1 corresponds to currents above V_{rated} and 0 below V_{cut-in}.

Table C.5
Location of case study nodes at Pentland Firth.

Node	Number	Latitude	Longitude
A	500	59.8599068915N°	4.9797111830W°
B	4768	58.6557103353N°	3.1391667722W°
C	4496	58.6918591258N°	3.1296765346W°

Appendix C. Node locations and simulation designations

See Table C.5.

Appendix D. Machine learning benchmarks

See Tables D.6 and D.7.

D.0.1. ML hyperparameters and architecture

An overview of the model architecture and hyperparameters for each ML algorithm used for benchmarking is provided below.

- **RF:** RF regression was implemented using the sci-kit-learn RandomForestRegressor module (Pedregosa et al., 2011). Predicting based off of the previous 8 values yielded the best performance in our testing.
- **SVM:** SVM was implemented using the sci-kit-learn SVM module. A radial basis function (RBF) kernel was selected based on the

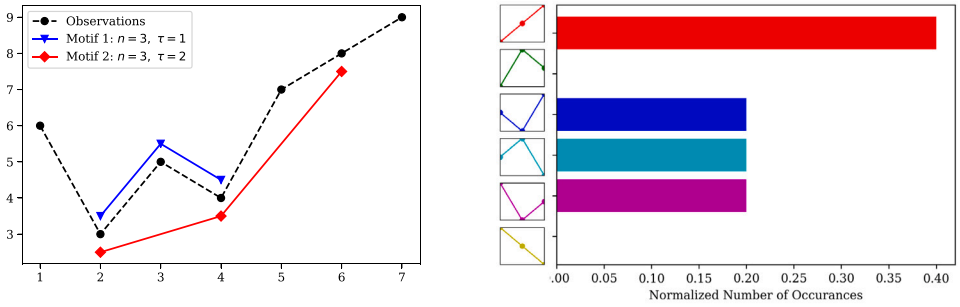


Fig. B.10. Simplified example of the relationship between n and τ for PE.

Table D.6
Machine learning comparisons for Pentland Firth Simulation sites. Model predictions are evaluated over 3 days on data sampled at 6 min intervals. Note, the best values are highlighted in bold for each performance measure and current direction respectively.

Site	Method	Prediction data	RMSE	MSE	MAE	Max error	r^2
Node A	UTide	U	0.0412	0.0017	0.0364	0.0749	0.9494
		V	0.035	0.0012	0.0321	0.0544	0.7107
Node A	RF	U	0.0059	**0.0**[b]	0.0036	0.0321	0.9992
		V	0.0012	**0.0**[b]	0.0009	0.0043	**0.9998**[b]
Node A	SVR	U	0.069	0.0048	0.0581	0.1411	0.7933
		V	0.0618	0.0038	0.0558	0.0983	−5.9303
Node A	ANN	U	0.0012	**0.0**[b]	0.0008	0.0051	**1.0**[a]
		V	**0.0011**[a]	**0.0**[b]	0.0009	0.0046	**0.9998**[b]
Node A	LSTM	U	0.0036	**0.0**[b]	0.0021	0.0207	0.9997
		V	0.0021	**0.0**[b]	0.0019	0.0054	0.9994
Node A	HRA	U	**0.0006**	**0.0**[b]	**0.0005**	**0.002**	**1.0**[a]
		V	**0.0011**[a]	**0.0**[b]	**0.0008**	**0.0034**	**0.9998**[b]
Node B	UTide	U	0.2118	0.0449	0.1734	0.7643	0.8801
		V	0.3725	0.1388	0.3181	0.957	0.839
Node B	RF	U	0.0861	0.0074	0.0542	0.4807	0.9836
		V	0.1627	0.0265	0.0974	0.7945	0.9779
Node B	SVR	U	0.0822	0.0068	0.065	0.4287	0.9846
		V	0.1676	0.0281	0.1115	0.8274	0.9757
Node B	ANN	U	0.0685	0.0047	0.042	0.3891	0.9897
		V	0.2393	0.0573	0.1466	1.0067	0.946
Node B	LSTM	U	0.0722	0.0052	0.045	0.5516	0.9886
		V	0.1517	0.023	0.0821	0.7839	0.9806
Node B	HRA	U	**0.0427**	**0.0018**	**0.0277**	**0.2899**	**0.9993**
		V	**0.0493**	**0.0024**	**0.0319**	**0.2905**	**0.9864**
Node C	UTide	U	0.3942	0.1554	0.33	1.076	0.9676
		V	0.1473	0.0217	0.1123	0.4348	0.9286
Node C	RF	U	0.1293	0.0167	0.1013	0.52	0.995
		V	**0.0497**	**0.0025**	**0.0318**	0.3411	0.9888
Node C	SVR	U	0.0914	0.0084	0.0725	0.4204	**0.9975**
		V	0.0563	0.0032	0.0483	0.3649	0.9871
Node C	ANN	U	0.13	0.0169	0.0772	0.7508	0.9949
		V	0.0724	0.0052	0.0427	0.6064	0.9865
Node C	LSTM	U	0.2049	0.042	0.1563	0.6391	0.988
		V	0.0785	0.0062	0.0568	0.3549	0.9753
Node C	HRA	U	**0.0273**	**0.0008**	**0.0201**	**0.1243**	0.9949
		V	0.0609	0.0037	0.0397	**0.3268**	**0.989**

[a]Denotes values for which two models are tied for best performance.
[b]Denotes values for which multiple values are tied for best performance.

Table D.7
Machine learning comparisons for real sites. Model predictions are evaluated over 3 days on data sampled at 6 min intervals. Note, the best values are highlighted in bold for each performance measure and current direction respectively.

Site	Method	Prediction data	RMSE	MSE	MAE	Max error	r^2
Eastport	UTide	U	0.0741	0.0055	0.0588	0.2527	−0.2413
		V	0.3003	0.0902	0.255	0.7226	0.8621
Eastport	RF	U	0.0479	0.0023	0.0369	0.1904	0.6622
		V	0.057	0.0033	0.0437	0.2026	0.9939
Eastport	SVR	U	0.0468	0.0022	0.0366	0.1897	0.6429
		V	0.0537	0.0029	0.0412	0.2075	0.9946
Eastport	ANN	U	0.046	0.0021	0.0351	0.1819	0.7024
		V	0.5586	0.312	0.4759	1.358	−4.9505
Eastport	LSTM	U	0.0462	0.0021	0.0356	**0.1761**	0.6942
		V	0.0549	0.003	0.0425	0.2115	0.9945
Eastport	HRA	U	**0.021**	**0.0004**	**0.0155**	0.1991	**0.9544**
		V	**0.0311**	**0.001**	**0.0235**	**0.1714**	**0.9984**
Cook Inlet	UTide	U	0.4408	0.1943	0.3901	1.0072	0.899
		V	0.1265	0.016	0.0995	0.3494	−0.1896
Cook Inlet	RF	U	0.0867	0.0075	0.065	0.3339	0.997
		V	0.074	0.0055	0.0555	0.2807	0.7846
Cook Inlet	SVR	U	0.0857	0.0073	0.065	0.3258	0.9971
		V	0.0842	0.0071	0.0635	0.305	0.6719
Cook Inlet	ANN	U	0.0915	0.0084	0.0668	0.3988	0.9967
		V	0.1044	0.0109	0.0805	0.3203	0.785
Cook Inlet	LSTM	U	0.0961	0.0092	0.0747	0.3806	0.9963
		V	0.0746	0.0056	0.056	0.3105	0.7728
Cook Inlet	HRA	U	**0.0683**	**0.0047**	**0.0543**	**0.2399**	**0.9979**
		V	**0.0697**	**0.0049**	**0.0543**	**0.2536**	**0.8465**

recommendations from Zhang et al. (2022). The SVM produced the highest performance when training to predict off of the 4 previous values.

- **ANN:** The backpropagation ANN was implemented using Keras (Chollet et al., 2015). The model used an input layer with 8 neurons and two hidden layers with sizes 16 and 8 respectively. The ADAM optimizer was used with a learning rate of 0.0001 and was trained for 500 epochs. We found that the best performance resulted from training the model to predict based off of the 8 previous values.
- **LSTM:** The LSTM was implemented using the Keras LSTM module. The model used an input layer with 8 neurons and one hidden layer with size 8. The ADAM optimizer was used with a learning rate of 0.00001 and was trained for 500 epochs. We found that the best performance resulted from training the model to predict based on the 8 previous values.

It should be noted that training times for the LSTM and ANN models exceeded 10 min and can therefore not be used in an Online manner. In our testing, we found that training separate models to predict the U and V components respectively yielded superior results for all models.

Appendix E. Model parameters

The SSA parameters used for each site and sampling rate are reported in Table E.8. It should be noted that grid partitioning is used for the HOFTS on all sites and is set to 15 partitions.

Table E.8
Model parameters for SSA decomposition.

Site	Sampling rate	Train set length (samples)	Window length (samples)
Pentland Firth	6 min	720	350
	1 h	360	150
Cook Inlet	6 min	360	150
	1 h	360	150
Eastport	6 min	360	150
	1 h	360	150

References

Adcock, T.A.A., Draper, S., Houlsby, G.T., Borthwick, A.G., Serhadl Serhadlıoğlu, S., 2013. The available power from tidal stream turbines in the pentland firth. Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci. 469 (2157), 20130072.

Adcock, T.A.A., Draper, S., Willden, R.H.J., Vogel, C.R., 2021. The fluid mechanics of tidal stream energy conversion. Annu. Rev. Fluid Mech. 53, 287–310.

Afshar, K., Bigdeli, N., 2011. Data analysis and short term load forecasting in Iran electricity market using singular spectral analysis (SSA). Energy 36 (5), 2620–2627.

Aly, H.H., 2020. A novel approach for harmonic tidal currents constitutions forecasting using hybrid intelligent models based on clustering methodologies. Renew. Energy 147, 1554–1564.

Aly, H.H.H., El Hawary, M.E., 2013. A proposed ANN and FLSM hybrid model for tidal current magnitude and direction forecasting. IEEE J. Ocean. Eng. 39 (1), 26–31.

Bandt, C., Pompe, B., 2002. Permutation entropy: a natural complexity measure for time series. Phys. Rev. Lett. 88 (17), 174102.

Beaudin, M., Zareipour, H., Schellenberglabe, A., Rosehart, W., 2010. Energy storage for mitigating the variability of renewable electricity sources: An updated review. Energy Sustain. Dev. 14 (4), 302–314.

Chang, W.Y., 2014. A literature review of wind forecasting methods. J. Power Energy Eng. 2 (04), 161.

Chen, S.M., 1996. Forecasting enrollments based on fuzzy time series. Fuzzy Sets and Systems 81 (3), 311–319.

Chen, B.F., Wang, H.D., Chu, C.C., 2007. Wavelet and artificial neural network analyses of tide forecasting and supplement of tides around Taiwan and south China sea. Ocean Eng. 34 (16), 2161–2175.

Chollet, F., et al., 2015. Keras. https://keras.io.

Codiga, D.L., 2011. Unified tidal analysis and prediction using the Utide matlab functions.

Colonna, J.G., Nakamura, E.F., 2018. Unsupervised selection of the singular spectrum components based on information theory for bioacoustic signal filtering. Digit. Signal Process. 82, 64–79.

Cui, Y., Jia, L., Fan, W., 2021. Estimation of actual evapotranspiration and its components in an irrigated area by integrating the shuttleworth-wallace and surface temperature-vegetation index schemes using the particle swarm optimization algorithm. Agricult. Forest Meteorol. 307, 108488.

Dolara, A., Grimaccia, F., Leva, S., Mussetta, M., Ogliari, E., 2015. A physical hybrid artificial neural network for short term forecasting of PV plant power output. Energies 8 (2), 1138–1153.

Dong, J., Zeng, W., Wu, L., Huang, J., Gaiser, T., Srivastava, A.K., 2023. Enhancing short-term forecasting of daily precipitation using numerical weather prediction bias correcting with xgboost in different regions of China. Eng. Appl. Artif. Intell. 117, 105579.

Elsner, J.B., Tsonis, A.A., 1996. Singular Spectrum Analysis: A New Tool in Time Series Analysis. Springer Science & Business Media.

Franco, A., 1973. Tidal mathematics: The analysis of tides. Science 180 (4087), 729–730.

Frost, C., 2022. Cost reduction pathway of tidal stream energy in the UK and France.

Golyandina, N., 2010. On the choice of parameters in singular spectrum analysis and related subspace-based methods. arXiv preprint arXiv:1005.4374.

Hajirahimi, Z., Khashei, M., 2019. Hybrid structures in time series modeling and forecasting: A review. Eng. Appl. Artif. Intell. 86, 83–106.

Hassani, H., Thomakos, D., 2010. A review on singular spectrum analysis for economic and financial time series. Stat. Interface 3 (3), 377–397.

Hurst, H.E., 1951. Long-term storage capacity of reservoirs. Trans. Am. Soc. Civ. Eng. 116 (1), 770–799.

Kavousi-Fard, A., 2016. A hybrid accurate model for tidal current prediction. IEEE Trans. Geosci. Remote Sens. 55 (1), 112–118.

Kavousi-Fard, A., Su, W., 2017. A combined prognostic model based on machine learning for tidal current prediction. IEEE Trans. Geosci. Remote Sens. 55 (6), 3108–3114.

Kline, D.M., 2004. Methods for multi-step time series forecasting neural networks. In: Neural Networks in Business Forecasting. IGI Global, pp. 226–250.

Kubatko, E.J., Westerink, J.J., Dawson, C., 2006. Hp discontinuous Galerkin methods for advection dominated problems in shallow water flow. Comput. Methods Appl. Mech. Engrg. 196 (1–3), 437–451.

Lee, T., Jeng, D., 2002. Application of artificial neural networks in tide-forecasting. Ocean Eng. 29 (9), 1003–1022.

Liu, M., Li, W., Wang, C., Billinton, R., Yu, J., 2015. Reliability evaluation of a tidal power generation system considering tidal current speeds. IEEE Trans. Power Syst. 31 (4), 3179–3188.

Neill, S.P., Hashemi, M.R., Lewis, M.J., 2014. The role of tidal asymmetry in characterizing the tidal energy resource of Orkney. Renew. Energy 68, 337–350.

Niyogi, D., 2023. A novel method combines moving fronts, data decomposition and deep learning to forecast intricate time series. arXiv preprint arXiv:2303.06394.

Olofsen, E., Sleigh, J., Dahan, A., 2008. Permutation entropy of the electroencephalogram: a measure of anaesthetic drug effect. Br. J. Anaesth. 101 (6), 810–821.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: Machine learning in python. J. Mach. Learn. Res. 12, 2825–2830.

Pessa, A.A., Ribeiro, H.V., 2021. Ordpy: A python package for data analysis with permutation entropy and ordinal network methods. Chaos 31 (6), 063110.

Qian, P., Feng, B., Liu, X., Zhang, D., Yang, J., Ying, Y., Liu, C., Si, Y., 2022. Tidal current prediction based on a hybrid machine learning method. Ocean Eng. 260, 111985.

Qiao, X., Guo, F., Zhang, R., 2020. Short-term tidal current prediction based on GA-bp neural network. In: IOP Conference Series: Earth and Environmental Science, Vol. 513. IOP Publishing, 012061.

Remya, P., Kumar, R., Basu, S., 2012. Forecasting tidal currents from tidal levels using genetic algorithm. Ocean Eng. 40, 62–68.

Saber, A.Y., Khandelwal, T., 2017. IoT based online load forecasting. In: 2017 Ninth Annual IEEE Green Technologies Conference (GreenTech). IEEE, pp. 189–194.

Safari, N., Ansari, O.A., Zare, A., Chung, C., 2017. A novel decomposition-based localized short-term tidal current speed and direction prediction model. In: 2017 IEEE Power & Energy Society General Meeting. IEEE, pp. 1–5.

Safari, N., Mazhari, S.M., Khorramdel, B., Chung, C., 2018. Tidal current and level uncertainty prediction via adaptive linear programming. IEEE Trans. Sustain. Energy 10 (2), 748–758.

Sarkar, D., Osborne, M.A., Adcock, T.A.A., 2018. Prediction of tidal currents using Bayesian machine learning. Ocean Eng. 158, 221–231.

Sarkar, D., Osborne, M.A., Adcock, T.A.A., 2019. Spatiotemporal prediction of tidal currents using Gaussian processes. J. Geophys. Res.: Oceans 124 (4), 2697–2715.

Shen, L., Shao, Z., Yu, Y., Chen, X., 2021. Hybrid approach combining modified gravity model and deep learning for short-term forecasting of metro transit passenger flows. Transp. Res. Rec. 2675 (1), 25–38.

Silva, P.C., Lucas, P., Sadaei, H., Guimarães, F., 2018. Pyfts: Fuzzy time series for python. Belo Horizonte.

Singh, S., 2007. A simple method of forecasting based on fuzzy time series. Appl. Math. Comput. 186 (1), 330–339.

Soman, S.S., Zareipour, H., Malik, O., Mandal, P., 2010. A review of wind power and wind speed forecasting methods with different time horizons. In: North American Power Symposium 2010. IEEE, pp. 1–8.

Sulandari, W., Subanar, S., Lee, M.H., Rodrigues, P.C., 2020. Time series forecasting using singular spectrum analysis, fuzzy systems and neural networks. MethodsX 7, 101015.

Taieb, S.B., Bontempi, G., Atiya, A.F., Sorjamaa, A., 2012. A review and comparison of strategies for multi-step ahead time series forecasting based on the NN5 forecasting competition. Expert Syst. Appl. 39 (8), 7067–7083.

Wang, T., Yang, Z., 2020. A tidal hydrodynamic model for cook inlet, alaska, to support tidal energy resource characterization. J. Mar. Sci. Eng. 8 (4), 254.

Wu, Y.K., Hong, J.S., 2007. A literature review of wind forecasting technology in the world. In: 2007 IEEE Lausanne Power Tech. IEEE, pp. 504–509.

Yang, H., Huang, K., King, I., Lyu, M.R., 2009. Localized support vector regression for time series prediction. Neurocomputing 72 (10–12), 2659–2669.

Yin, J., Wang, N., 2016. An online sequential extreme learning machine for tidal prediction based on improved Gath–Geva fuzzy segmentation. Neurocomputing 174, 85–98.

Zhang, A., Lin, Y., Sun, Y., Yuan, H., Wang, M., Liu, G., Hu, Y., 2022. Tidal current prediction based on fractal theory and improved least squares support vector machine. IET Renew. Power Gener. 16 (2), 389–401.