



On detrending stream velocity time series for robust tidal flow turbulence characterization

Shyuan Cheng^a, Vincent S. Neary^e, Leonardo P. Chamorro^{a,b,c,d,*}

^a Mechanical Science and Engineering, University of Illinois, Urbana, 61801, IL, USA

^b Sandia National Laboratories, Albuquerque, 87123, NM, USA

^c Civil and Environmental Engineering, University of Illinois, Urbana, 61801, IL, USA

^d Aerospace Engineering, University of Illinois, Urbana, 61801, IL, USA

^e Geology, University of Illinois, Urbana, 61801, IL, USA

ARTICLE INFO

Keywords:

Detrend
Empirical mode decomposition
Sifting
Riverine flow

ABSTRACT

We investigated the impact of detrending techniques on turbulence quantities from tidal stream flow data, focusing on the autocorrelation function, ρ_{uu} , and velocity spectrum, $\Phi(f)$. Standard detrending methods, including high-pass frequency-based and polynomial-based techniques, are examined, alongside a proposed alternative method, the empirical mode decomposition (EMD). Our results highlight that intervals of flow acceleration and deceleration, typical in tidal and riverine flows, significantly affect the estimation of turbulence quantities using high-pass frequency filtering and polynomial detrending of varying orders. These methods can strongly influence ρ_{uu} and $\Phi(f)$, thereby affecting the accurate estimation of derived quantities. We examine two variations of detrending data using EMD; the first removes only the EMD residue, and the second removes both the residue and the largest scale intrinsic mode function (IMF). By comparing the detrended spectra with the modeled von Kármán spectra, we demonstrate that the second variation (i.e., removing the residue and the largest scale IMF) successfully removed the large-scale trend of the data while retaining the energy of other scales.

1. Introduction

Dozens of successful deployments of tidal energy converters (TECs) around the globe demonstrate the potential of tidal energy to contribute to an expanding global renewable power generation portfolio, including a wide range of distributed energy needs for isolated coastal and island communities (Huckerby et al., 2012; Chowdhury et al., 2021; OES, 2023). The maximum extractable global mean annual tidal power along continental shelves is estimated to be some fraction of the 2.5 terawatts (TW) tidal energy dissipation rate (Egbert and Ray, 2003), with estimates ranging significantly from 0.06 TW (Hammons, 1993) to 1 TW (Kempener and Neumann, 2014).

The current global installed capacity for tidal energy is approximately 15 MW, primarily consisting of 1st-stage pilot farm projects. However, there are new projects totaling blue an additional 109 MW of installed capacity, representing a seven-fold increase, under development (IRENA and OEE, 2023b). Full-scale tidal energy conversion (TEC) technologies have been demonstrated with rated capacities of up to 2 MW per device (O. Orbital Marine Power, 2023). With this projected increase in global installed capacity in the coming decades,

reductions in levelized costs of energy (LCOE) are expected to occur through economies of scale, the streamlining of supply chains, and technology improvements; further details can be found in IRENA and OEE (2023a).

Thorough characterization of the tidal energy resource and turbulent inflow conditions is crucial for advancing tidal energy development. It provides key inputs, like the hydrodynamic loads, hydrokinetic power densities, and turbulence statistics for siting, project development, and designing and certifying TEC devices (IEC, 2015; Neary et al., 2013). Classification and type-certification of TEC devices require assessment of the maximum mean current speed at hub height, e.g., during the maximum perigean tide, along with the median value of the turbulence intensity occurring at 1.5 m/s (Neary et al., 2019; Bittencourt et al., 2014; Deb et al., 2023).

Estimations of tidal energy potential often combine field measurements with numerical models (Iglesias et al., 2021; Chen et al., 2013; Wang and Yang, 2017; Burić et al., 2021; Yang et al., 2020, 2021; Maldar et al., 2022). For instance, Chen et al. (2013) used a 3D finite-element model to evaluate tidal current energy potential at Kinmen

* Corresponding author at: Mechanical Science and Engineering, University of Illinois, Urbana, 61801, IL, USA.

E-mail addresses: vsneary@sandia.gov (V.S. Neary), lpchamo@illinois.edu (L.P. Chamorro).

Nomenclature

Φ_u	streamwise velocity spectrum (m^2/s).
ρ_{uu}	autocorrelation of the streamwise velocity component.
σ_u^2	streamwise velocity variance (m^2/s^2).
τ	time lag (s).
f_s	sampling frequency (Hz).
L_u	integral length scale (m).
T_u	integral time scale (s).
U	time-averaged streamwise velocity (m/s).
u	streamwise velocity (m/s).
u'	streamwise velocity fluctuation (m/s).
EMD	empirical mode decomposition.
IMF	intrinsic mode function.
WSS	wide-sense stationary.

Island, Taiwan, aiming to identify an optimal location for a tidal power plant. Similarly, Wang and Yang (2017) numerically assessed the power potential of minor tidal channels in Puget Sound, USA. They employed a 3D hydrodynamic model (FVCOM) to investigate the impact of a power plant on tidal circulation, revealing increased turbulence levels in the turbine wake. Yang et al. (2021) employed a high-resolution 3D tidal hydrodynamic model to characterize the energy resource of Salish Sea tidal streams, highlighting the significance of evaluating both current magnitude and kinetic energy flux for optimal turbine placement. However, while these studies have primarily focused on hydrodynamic characteristics, estimating tidal-turbine interaction often neglects tidal turbulence parameters due to the challenges of obtaining high-resolution turbulence measurements. Recently, Deb et al. (2023) incorporated macro-scale turbulence quantities into a hydrostatic primitive equation (HPE) model to analyze the tidal energy resource in the Western Passage, Maine, USA. Their assessment of turbulent kinetic energy (TKE) and turbulence intensity across various channel depths facilitated the evaluation of suitable cross-sections for tidal energy converter (TEC) deployment.

The interaction between turbines and tidal currents may be particularly strong; tidal turbulence significantly affects the fatigue load on hydrokinetic turbines, requiring modeling and methods to assess turbulence levels (Yadegari et al., 2023). A specific turbulence model is required to simulate these types of flows. Such a model relies on input parameters from field measurements and returns tidal turbulence characteristics like those used for wind turbulence models, such as the IEC Kaimal Model and the IEC von Kármán Isotropic Model. These models generate time series of velocity fluctuations based on turbulence spectral representation and account for coherent structures, as demonstrated in the TurbSim stochastic unsteady turbulent inflow simulator (Kelley and Jonkman, 2005).

To utilize the classical velocity spectrum model, it is necessary to have turbulence statistics such as the variance of velocity fluctuations (σ_u) and integral length scale (L_u) as input parameters. However, to estimate these statistical quantities and perform spectrum analysis correctly, the turbulence signal must be wide-sense stationary (WSS) and pass the stationarity test, i.e., $\langle X(t) \rangle = \mu$, and their covariance function, C , follows $C(t_1, t_2) \equiv C(t_2 - t_1) = C(\tau)$, where $\tau = t_2 - t_1 \geq 0$ is the time lag (Grami, 2016). Unfortunately, most tidal currents exhibit large-scale patterns, changes in direction, and acceleration/deceleration over the measurement period, resulting in a non-stationary process. This leads to an inability to converge on mean quantities and higher-order statistics using the raw signal. Consequently, turbulence statistics at tidal sites being reported only over short time intervals, typically on the order of several minutes in the literature (Gunawan et al., 2014; Thomson et al., 2012; Milne et al., 2017). However, such short time periods may not

capture the effects of large-scale eddies and can restrict the development of spectrum models as part of a multiple-criteria decision-making model for selecting tidal energy converter sites. This emphasizes the critical need for the development of detrending techniques capable of efficiently removing dominant trends in the raw signal and generating a wide-sense stationary (WSS) signal suitable for calculating turbulence statistics.

Several detrending methods have been suggested to eliminate large-scale trends in raw velocity fluctuation data and obtain a WSS signal. However, conventional methods such as frequency-based filtering (low-pass, high-pass, and band-pass filters) and polynomial-based filtering are vulnerable to failure or are not suitable for turbulence data due to intermittent events. Determining the correct parameters for these detrending methods is challenging, and there are no established guidelines for designing filters for turbulence data processing. Furthermore, most detrending methods are limited to one-dimensional and single-variant data.

Here, we present a robust method for detrending multi-dimensional and multi-variant data using the so-called empirical mode decomposition (EMD) sifting process to remove large-scale trends. We compare the performance of EMD with other common detrending methods, such as high-pass filters and polynomial-based detrending, when applied to tidal flow data. We present a methodology for empirical mode decomposition (EMD) with minimal parameter tuning required, which yields superior results in spectrum and autocorrelation analyses when compared to other commonly used detrending methods. This has been demonstrated through its application to experimental turbulence tidal signals.

2. Tidal current data

We examined the effectiveness of detrend methods for two tidal sites: Nodule Point, WA, and East River, NY. The first dataset was obtained from an Acoustic Doppler Velocimeter (ADV) during the spring tide in 2011, with a sampling frequency of $f_s = 32$ Hz, over a period of 4.3 days. The ADV was positioned at the apex of the Tidal Turbulence Tripod, which was situated 4.7 m above the seabed that was 22 m deep at Nodule Point, located on the eastern side of Marrowstone Island. For more information on the sampling parameters and Doppler noise, refer to Thomson et al. (2012).

The second dataset was collected using a three-axis ADV for a duration of 39 days between June 9th to July 17th, 2011, with a sampling frequency of $f_s = 20$ Hz. The upward-looking Sontek ADV was deployed at the Roosevelt Island Tidal Energy (RITE) site, owned by Verdant Power. The RITE Project site is located on the east side of Roosevelt Island, between the boroughs of Manhattan and Queens, New York City. For details on the site characteristics, refer to Gunawan et al. (2014).

2.1. Data processing

The unprocessed Acoustic Doppler Velocimeter (ADV) data underwent the phase-space-thresholding (PST) test (Goring and Nikora, 2002) to detect spikes and outliers. Detected spikes and outliers were replaced with time-averaged values as detailed in Gunawan et al. (2011). We focus on intervals characterized by sufficiently high fluid velocities relevant to energy conversion. Fig. 1a depicts a time series of tidal flow, highlighting specific intervals that meet the criterion.

3. Methodology

Tidal flows exhibit non-stationarity statistical features; they fail stationarity tests such as the one described by Foken and Wichura (1996); see example in Fig. 1. Time series from these streams are not in a state of wide-sense stationarity (WSS), which means they lack a constant mean, or their covariance function shifts over time (Florescu, 2014).

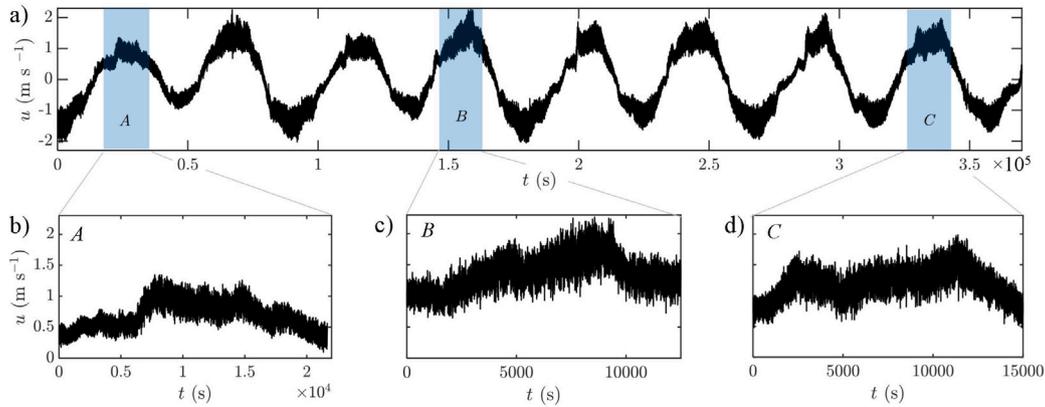


Fig. 1. (a) Sample streamwise velocity time series of a selected tidal resource. (b-d) Close view of the signal within selected A – C regions.

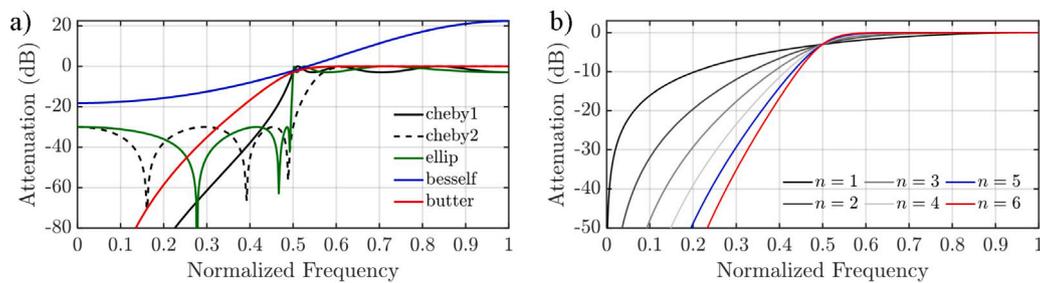


Fig. 2. Comparison of the frequency response generated by (a) various 6th order high-pass filters, and (b) Butterworth filter for filter order $n = 1$ to 6.

Such non-stationarity presents challenges for using turbulence formulations (Monin and Yaglom, 1999; Durbin and Reif, 2011; Stanisci, 2012) since statistical quantities do not converge and are dependent on the extent of the time series. Conversely, dividing the continuous time series into shorter time windows, where each window may exhibit a stable mean and variance, could lead to the omission of large-scale eddies (Thomson et al., 2012). Therefore, it becomes essential to implement effective detrending techniques when aiming to develop turbulence models that span extended durations characterized by high-momentum tides with time scales on the order of hours (as illustrated in Fig. 1 b-d). Standard detrending methods can be categorized into frequency-based filters (e.g., high-pass, low-pass, and band-pass filters) and polynomial-based approaches. These two methods have advantages and disadvantages, which are discussed briefly below.

While detrending techniques such as regression-based and frequency-based filtering are widely used in the signal processing community, for instance, in processing Electroencephalography (EEG) and Magnetoencephalography (MEG) data (de Cheveigné and Arzounian, 2018; Eleuteri et al., 2012), they suffer from limitations such as significant autocorrelation and spectra artifacts or sensitivity to various detrending parameters. These parameters lack robust tuning criteria and often require trial and error. Furthermore, the aforementioned methods are limited to one-dimensional data with a single variant and cannot be easily extended to multi-dimensional or multi-variant data.

We propose an alternative detrending method using empirical mode decomposition (EMD) to address these issues. EMD is a powerful non-stationary modal decomposition tool that does not require parameter tuning and is robust to highly intermittent turbulence data. Furthermore, EMD applies to multi-dimensional and multi-variant data (Rilling et al., 2007; Rehman and Mandic, 2009, 2010), making it an ideal tool for detrending tidal turbulence signals.

3.1. Frequency-based filtering

High-pass filters are commonly employed to eliminate slow large-scale patterns in signals (van Driel et al., 2021). However, the frequency-based nature of these filters alters the behavior of the turbulence spectrum, rendering them challenging. The frequency response of standard high-pass filters is shown in Fig. 2a, with the cutoff frequency set to half the sampling frequency. This figure illustrates the common artifacts of frequency-based filterings, such as the ripple effect in the passband (i.e., normalized frequency ≥ 0.5) of Chebyshev Type I and elliptic filters. Such issues are absent in Butterworth and Chebyshev Type II filters, which exhibit flat passband responses. However, these filters have wide transition bands and suffer from slow roll-off. Based on the frequency response characteristic, using the maximally flat passband Butterworth filter response is recommended to minimize artifacts in the passband spectrum. However, this smoothness comes at the cost of decreased roll-off steepness, so a higher filter order is typically necessary to improve the roll-off performance, as demonstrated in Fig. 2b. Other limitations have also been reported while dealing with frequency-based filter design, including the distortion of transient features (de Cheveigné and Arzounian, 2018; Acunzo et al., 2012) and the ringing effect (Tanner et al., 2015; Widmann and Schröger, 2012). These are particularly problematic when dealing with turbulence data, as they can alter the magnitude of intermittent events. In addition, low cutoff or high-order filter designs often exhibit transient fluctuations that take time to attenuate (Vanierschot et al., 2009).

3.2. Regression-based detrending

Large-scale signal trends can be identified using different techniques, including linear and nonlinear regression analysis, which involves solving a least-squares fitting problem. The methods used to identify these trends can significantly affect the inferred statistics.

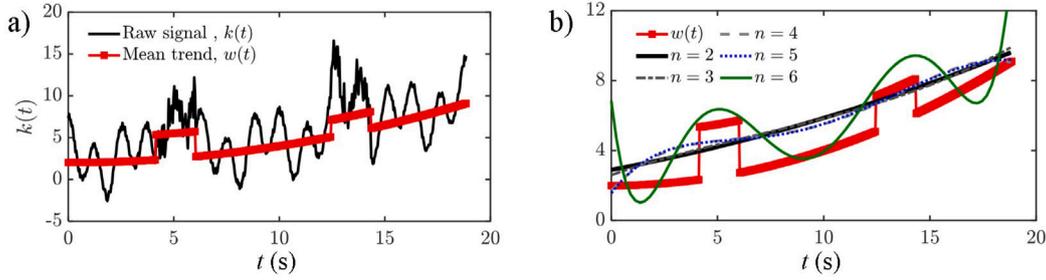


Fig. 3. Polynomial regression on the sample data with added noise and glitches mimicking distinct features of flow data; (a) raw signal $k(t)$ (black) superimposed with the large mean pattern and glitches $w(t)$ (red), (b) fitted curves using n -th order polynomial regression for various n values.

Linear and nonlinear regression analyses involve minimizing problems with linear or nonlinear constraints and bounds (Freedman, 2009; Aldrich, 2005). Here, we utilize linear regression for data fitting. The problem is defined as follows (Devore, 2008):

$$\min_x \sum_{i=1}^m (y_i - f(x_i, \beta))^2(x), \quad (1)$$

$$f(x, \beta) = \sum \beta_j \varphi_j(x).$$

The fitted curve y_i is a function of the independent variable x_i and the fitting coefficient β , represented as $y_i = f(x_i, \beta)$. The base function $\varphi(x)$, such as $\varphi_j = x^j$ for polynomial fitting or $\varphi_j = e^{jx}$ for exponential fitting, is also involved in the fitting process. Polynomial regression is of particular interest as the order of the polynomial determines the removed signal scales, potentially providing an alternative to frequency-based filtering. However, the least-square regression method minimizes the variance of the unbiased coefficient estimators under the Gauss–Markov theorem (Harville, 1976; Hallin, 2014), which assumes uncorrelated errors in the regression model. This assumption does not align with the nature of turbulence data.

Minimizing the least square regression variance is highly susceptible to glitches and intermittent events, as Fig. 3 exemplifies. We illustrate this using a sample signal $k(t)$, consisting of a pattern $w(t)$ and a summation of two modes. The signals $w(t)$ and $k(t)$ are shown in Fig. 3a, and are defined as follows:

$$w(t) = t^2/50 + 2 + g(t),$$

$$g(t) = \begin{cases} 3 + 5r_1(t) & 4.15 < t < 6 \\ 2 + 5r_1(t) & 12.45 < t < 14.3 \\ r_1(t) & \text{otherwise} \end{cases} \quad (2)$$

$$k(t) = 2 \cos(2t) + 3 \cos(5t) + w(t).$$

The quadratic trend $w(t)$ is added to the random noise distributed as $r_1(t)$, which comprises two sections of higher mean and fluctuation, mimicking the peak flood behavior in tidal turbulence (Milne et al., 2013; Villarini and Smith, 2010). Polynomial regression is applied to the sample signal $k(t)$, which consists of the large pattern and two modes. Fig. 3b illustrates the polynomial regression results, demonstrating that even with the correct polynomial order ($n = 2$), the regression method fails to capture the large pattern due to the influence of outliers. The results highlight that selecting proper fitting parameters heavily affects the regression-based filtering outcomes. Moreover, increasing the polynomial order does not necessarily yield a better fit as it may lead to overfitting, which is evident in the $n = 6$ fit.

There have been various attempts to prevent the issues of underfitting or overfitting in polynomial regression; however, there is currently no widely agreed-upon approach for selecting the most suitable fitting parameters (Araujo, 2018; Bu and Zhang, 2021; Lever et al., 2016; Chatterjee and Hadi, 1986).

3.3. Empirical Mode Decomposition (EMD) for data detrending

Empirical Mode Decomposition (EMD) is a sifting process developed by Huang et al. (1998) to analyze non-stationary and nonlinear signals. EMD is designed to extract local features and the time-frequency

distribution of a signal. The EMD process comprises two steps. First, the signal is decomposed into multiple intrinsic mode functions (IMFs). Second, the instantaneous time-frequency distribution of the data is extracted from each IMF using Hilbert-Huang transform (HHT) (Huang et al., 1999). The sifting process extracts IMFs based on local frequency or wavenumber information and produces IMFs in ascending order of local frequencies or spatial scales. The first IMF contains the highest local frequencies or spatial scales, and the final IMF contains the lowest local frequencies or scales of the oscillation. The residual component of the decomposition contains the trend of the signal or data. Despite being an effective method for analyzing non-stationary and nonlinear signals, there is still ongoing research to improve the accuracy and efficiency of EMD.

The fast and adaptive EMD algorithm is based on a sifting process that resembles the one proposed by Thirumalaisamy & Ansell (Thirumalaisamy and Ansell, 2018) for extracting intrinsic mode functions (IMFs) from a time-varying signal $x(t)$. This procedure involves identifying the local extrema of the signal, which are used to construct upper and lower envelopes using cubic splines. The average of these envelopes is then computed as $m_1(t)$, and the difference between the signal and its envelope, denoted as $h_1(t)$, is obtained as $h_1(t) = x(t) - m_1(t)$.

To obtain the first intrinsic mode function (IMF) more accurately, the difference signal $h_1(t)$ is considered as a new signal, and its upper and lower envelopes are computed along with their new mean $m_2(t)$. Subsequently, a new difference signal $h_2(t) = h_1(t) - m_2(t)$ is determined, and this iteration is repeated until a stopping criterion is satisfied. The sifting process continues until the last residue shows no significant variation. An example of the one-dimensional, single variant EMD applied on a clean signal $x(t)$ is presented in Fig. 4, where $x(t)$ comprises two modes, a nonlinear quadratic pattern and an offset, as given by:

$$x(t) = 2 \cos(2t) + 3 \cos(5t) + t^2/50 + 2. \quad (3)$$

Fig. 4 demonstrates the efficacy of the EMD algorithm in accurately identifying and separating the different modes present in the raw signal. The extracted intrinsic mode functions (IMFs) are given alongside the modes used to construct the raw signal in Fig. 4b,c, showing that the IMFs accurately capture the amplitude of the modes with an error of less than 5% in the peak magnitude difference. Furthermore, the overall nonlinear pattern of the signal is captured in the residue, as shown in Fig. 4d. These results highlight the ability of EMD to robustly sift out multi-scale trends without heavy parameter tuning or prior knowledge of the raw signal.

4. Results

In this section, we evaluate the efficacy of the aforementioned filters, i.e., frequency-based, regression-based, and EMD approaches, in detrending raw data. To assess the performance of these methods, we compute velocity spectra and autocorrelation functions of the detrended signals. The objective is to retain the intermediate frequencies while eliminating low-frequency, large-scale patterns. Additionally, we use autocorrelation functions to compare the performance of the different filters. A successful filter should preserve essential frequency components of the signal while effectively removing the trend.

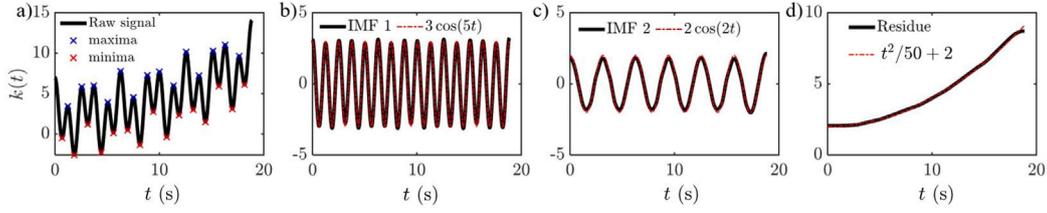


Fig. 4. Demonstration of fast and adaptive EMD. (a) Raw signal, $k(t)$, (b) first IMF, (c) second IMF, and (d) residue.

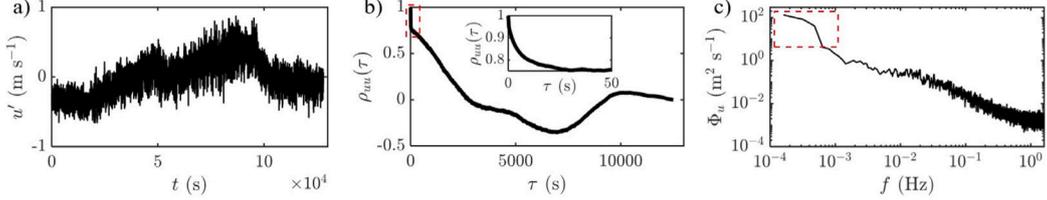


Fig. 5. (a) Raw streamwise velocity fluctuations in interval section B of Fig. 1, (b) autocorrelation function, ρ_{uu} , and (c) corresponding streamwise velocity spectrum of the unfiltered signal.

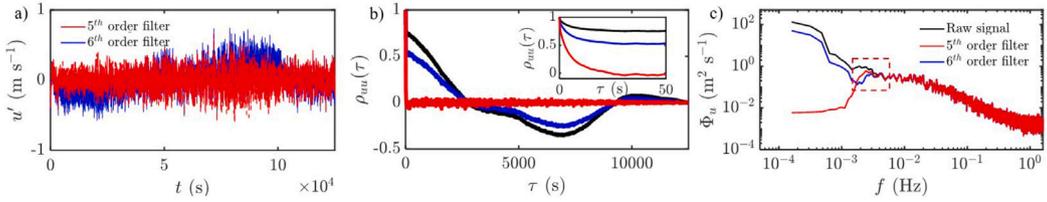


Fig. 6. (a) Time series of high-pass filtered streamwise velocity fluctuations in interval A of Fig. 1, (b) autocorrelation function, ρ_{uu} , and (c) the corresponding streamwise velocity spectrum.

4.1. Raw data

Fig. 5a illustrates the velocity fluctuation data during time interval B, as shown in Fig. 1a, which exhibits periods of acceleration and deceleration (see Fig. 1b), displaying substantial fluctuations at a large scale in the signal. The non-stationary nature of the signal can also be inferred from the autocorrelation function of the streamwise velocity, ρ_{uu} , defined as (Hong et al., 2022):

$$\rho(\tau) = \overline{u'(t)u'(t-\tau)} / \sigma_u^2. \quad (4)$$

Here, $u' = u - U$ represents the velocity fluctuations, where $U = \langle u \rangle$ corresponds to the mean velocity, τ is the time lag and σ_u^2 is the variance of the signal. The decay of correlation in ρ_{uu} appears very slow, where it does not approach zero until a time lag of approximately 3000 s. A closer examination of the highlighted region for $\tau \leq 50$ seconds in the inset reveals that ρ_{uu} reaches a plateau value of approximately 0.75 and decays very slowly afterward. This plateau behavior indicates non-ergodic and non-stationary behavior (Papoulis and Unnikrishna Pillai, 2002; Park and Park, 2018). The general large-scale variability in Fig. 1b affects the velocity spectrum leading to a sharp increase in energy level at the largest scale for $f \leq 10^{-3}$ Hz, as highlighted by the red-dashed box in Fig. 5c.

4.2. Spectral high-pass filtering

A high-pass Butterworth filter (Butterworth et al., 1930) with a cutoff frequency of 2.5×10^{-5} Hz and filter orders of 5 and 6 are used to illustrate the effectiveness of frequency-based detrending. The filtered signal, its corresponding velocity spectrum, and autocorrelation function are presented in Fig. 6. The velocity fluctuations, u' , in Fig. 6a indicates that the 5th-order filter successfully removes the large-scale patterns, resulting in a significant decrease in energy at low frequencies; however, the 6th-order filter with a different roll-off performance

fails to filter all the low-frequency energy, resulting in the plateau of $\rho_{uu} \approx 0.55$ at $\tau = 50$ s. Although the fifth-order filter performs well in autocorrelation, an artifact appears in the spectrum of Fig. 6c near the cutoff frequency, demonstrating that frequency-based filters are not optimal for filtering turbulence data.

4.3. Polynomial regression filtering

Fig. 7 illustrates a least square regression filtering method. Fig. 7a shows the raw velocity fluctuations overlaid with polynomial regression trends. Figs. 7b-e display the corresponding detrended signals, demonstrating how the filtered structures are affected by changing the polynomial order. This figure also highlights the difficulty of using regression-based filtering methods in selecting the appropriate parameters. Lower polynomial orders can result in underfitting, where the slow drift cannot be removed, as shown in Figs. 7b-c. Overfitting can occur with larger polynomial orders, as seen in the case of $n = 15$ at the beginning of the signal (highlighted by a red box) when comparing Figs. 7d,e. However, the $n = 15$ case showed better filtering performance at the end of the signal (red-dashed box), whereas the $n = 9$ case demonstrated underfitting. This example highlights the challenges associated with regression-based methods, where determining the correct regression parameters might require trial and error. There is no guarantee for an n that produce the right polynomial fit.

Fig. 8 displays the autocorrelation function and the velocity spectrum Φ_u for the detrended signals presented in Fig. 7. Underfitting polynomials are expected to produce a large-scale trend and slow convergence in autocorrelation. Although the higher-order polynomial fits ($n = 9$ and 15), exhibit better performance in ρ_{uu} , the nature of least square regression does not consider the frequency domain property. As a result, the selection of the filtered scales is ambiguous. This is evidenced by comparing the corresponding spectrum filtered by 9-th and 15-th order polynomials. Note that the 15-th order polynomial does

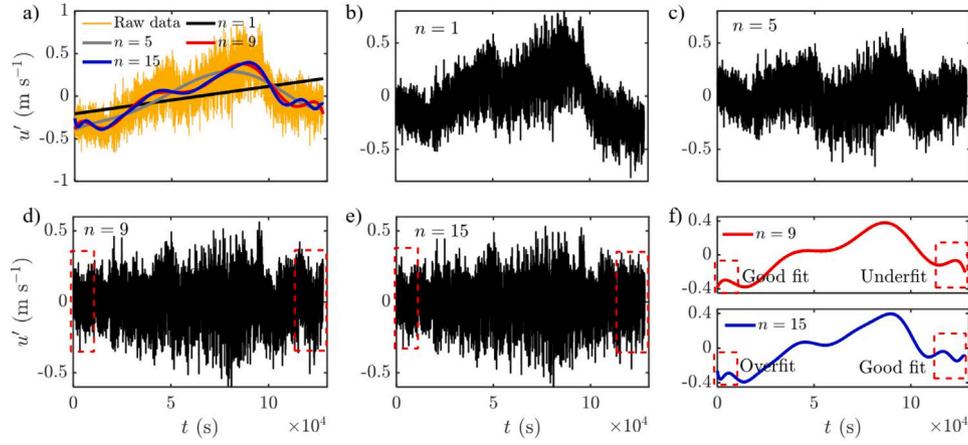


Fig. 7. (a) Raw signal in interval B (Fig. 1c) superimposed with polynomials, (b-e) detrended data with n -th order polynomial regression, (f) the 9-th and 15-th order polynomial showing the under and overfitting.

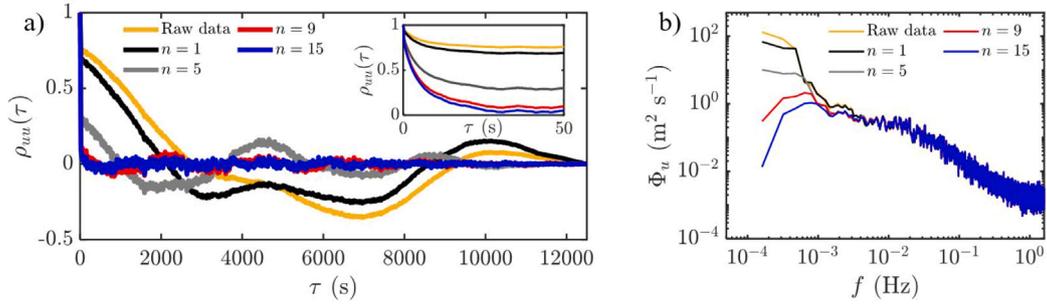


Fig. 8. (a) Autocorrelation function, ρ_{uu} , of the detrended data using polynomial regression, and (b) corresponding velocity spectra.

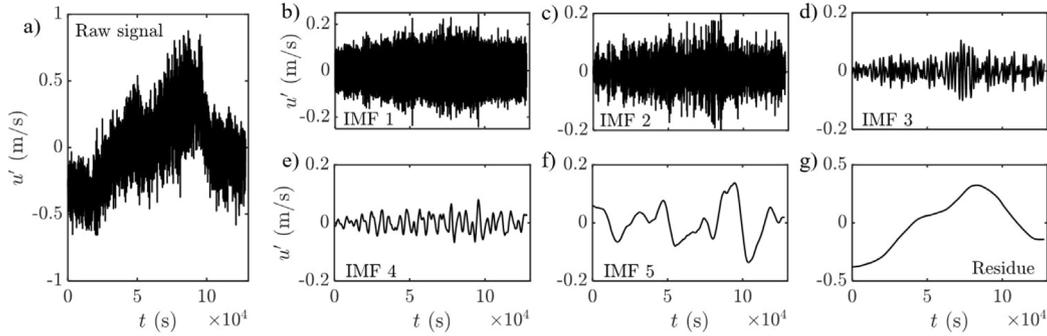


Fig. 9. IMFs obtained from fast and adaptive EMD for streamwise velocity fluctuations in interval B (Fig. 1c).

not filter out higher frequency (i.e., smaller scale) structures; instead, an unphysical sharp drop in the lowest frequency band is observed. This again highlights that regression-based filtering is suboptimal for detrending tidal turbulence data.

4.4. Single-variant EMD filtering

Fig. 9 presents the IMFs generated through EMD in section B of the data (Fig. 1b), using a sifting tolerance set at 0.1% of the signal amplitude. Fig. 10 displays two versions of detrended data. The first version subtracts only the residue from the raw signal, marked with a green dashed line in Fig. 10a. In contrast, the second version involves the addition of the last IMF and residue, represented by the red line in Fig. 10a. The first detrended signal still displays sufficiently large-scale fluctuations (Fig. 10b), while the second version exhibits a signal with acceleration and deceleration periods removed correctly (Fig. 10c).

A comparison of the autocorrelation function, ρ_{uu} , and velocity spectrum, Φ_u , between the raw signal and the two detrended signals are given in Figs. 11 and 12 for the three intervals selected in Fig. 1. The blue lines in Fig. 11 indicate signals where both the residue and the last IMF were subtracted, showing $\rho_{uu} \approx 0$ at $\tau = 50$ s. In contrast, the red lines indicate signals where only the residue was subtracted, demonstrating $\rho_{uu} > 0.2$ at $\tau = 50$ s with large-scale oscillations observed for a sufficiently high time lag ($\tau > 1000$ s). This finding confirms that residue and last IMF subtraction may be appropriate for riverine and tidal turbulence data where glitch exists. Moreover, the velocity spectrum in Fig. 12 supports the effectiveness of EMD in detrending riverine and tidal turbulence data. The signals with the residual and last IMF removed (blue lines) overlap with the raw signal at the high-frequency band while effectively removing the lowest frequency artifact, resulting in negligible slopes in the energy-containing range for one-dimensional spectra. This is consistent with previous studies (Pope, 2000; Saddoughi and Veeravalli, 1994).

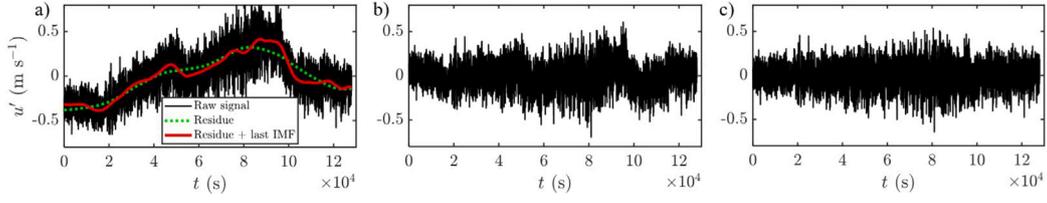


Fig. 10. (a) Raw streamwise velocity fluctuations at interval B (Fig. 1c) superimposed with trends obtained by one-dimensional one-variant EMD; the green dotted line shows the EMD residue and the red line is the residue with the addition of the last IMF, (b) signal with residue subtracted, and (c) signal with residue and last IMF subtracted.

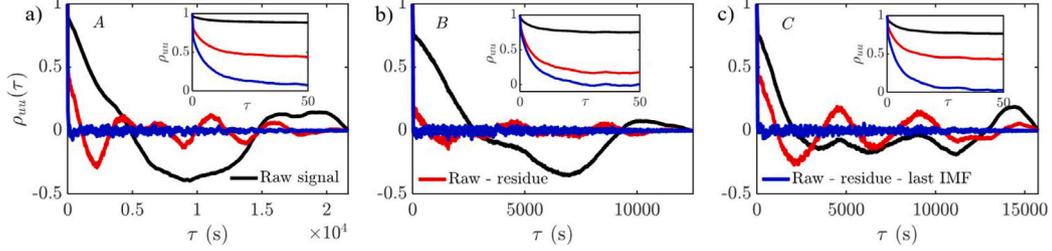


Fig. 11. Autocorrelation function, $\rho_{uu}(\tau)$, comparison between raw signal and signals filtered by EMD at the interval (a) A (Fig. 1b), (b) B (Fig. 1c), and (c) C (Fig. 1d). Insets show details for $\tau \leq 50$ s.

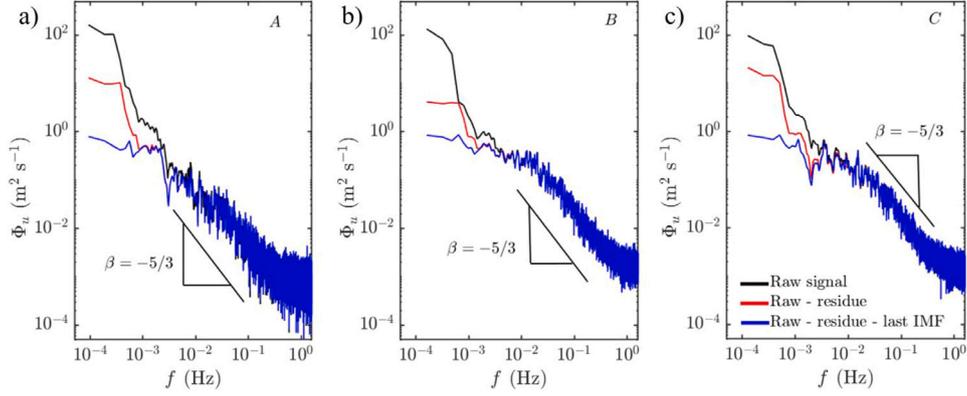


Fig. 12. Streamwise velocity spectra, Φ_u , comparison between raw signal and signals filtered by EMD at various selected the interval (a) A (Fig. 1b), (b) B (Fig. 1c), and (c) C (Fig. 1d).

With the detrended wide-sense stationary data obtained from EMD, we can calculate the input parameters required for the von Kármán spectrum model. The integral length scale, L_u , is determined as (Pope, 2001)

$$L_u = T_u \times U. \quad (5)$$

where $U = \langle u_{raw} \rangle$ represents the convection velocity with $\langle \rangle$ representing the time-average operator (Fig. 1b-d), and T_u denoting the integral time scale. It is defined as follows (Chamorro et al., 2013):

$$T_u = \int_0^{t_0} \rho_{uu}(\tau) d\tau, \quad (6)$$

where t_0 is time lag where the correlation is sufficiently low (Chamorro et al., 2013; Cheng et al., 2022); here, we set this low limit to $\rho_{uu}(t_0) \leq 0.05$. The integral length scale, L_u , and streamwise velocity fluctuations, σ_u , from the raw data and those from the detrended data by the EMD method are provided in Fig. 13. The analysis reveals a significant overestimation of σ_u and L_u when using the raw signal. This overestimation is caused by imposing a mean on a non-stationary signal, leading to erroneous large-scale fluctuations, as shown in Fig. 13c. By subtracting both EMD residue and the last IMF from the signal, a more accurate estimate of turbulence intensity ($\sim 10\%$) and integral length scale ($L_u \sim O(10^1)$ m) was obtained, which is consistent with similar velocity range riverine/tidal data (Milne et al., 2013; Thiébaud et al., 2020; Thomson et al., 2012).

4.5. Justification on subtracting both EMD residue and last IMF

Fig. 14 illustrates the outcome of applying the sifting algorithm to the signal $k(t)$, as depicted in Fig. 3 and defined by Eq. (2). This demonstrates that when the time series is noisy and contains large-scale outliers, the combination of both EMD residue and the last IMF successfully captures the true background trend, whereas using EMD residue alone fails to do so. Although there were larger errors in peak amplitude in the sections where outliers are present, the modes were well captured in the first three IMFs. The fourth IMF showed a smooth spline fit approximation of the piecewise function $g(t)$, which consists of two outlier sections. It is important to note that due to the spline-fit nature of the EMD algorithm, the fourth IMF and residue in Fig. 14d,e could not accurately capture the piecewise function $g(t)$ and quadratic pattern. Instead, it was necessary to combine the last IMF with the residue to obtain a smooth approximation for $w(t)$ in Eq. (2), as evidenced in Fig. 14f. This further stresses the robustness of EMD detrending when compared to regression-based detrending methods.

Lastly, the comparison between model spectra using the IEC von Kármán model (I.E. Commission et al., 2005) given in Eq. (7)

$$\Phi_u(f) = \frac{4\sigma_u^2 L_u / U}{(1 + 71(f L_u / U)^2)^{5/6}} \quad (7)$$

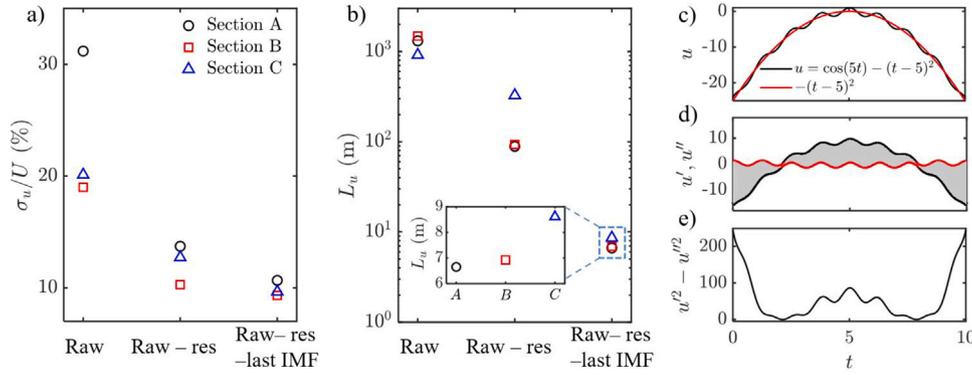


Fig. 13. Comparison between turbulence statistics of the raw signal, signal with EMD residue substrated, and signal with EMD residue and last IMF substrated. (a) Normalized turbulence intensity σ_u/U , and (b) integral length scale L_u ; (c-e) simple signal illustrating the overestimation of σ_u for non-stationary data.

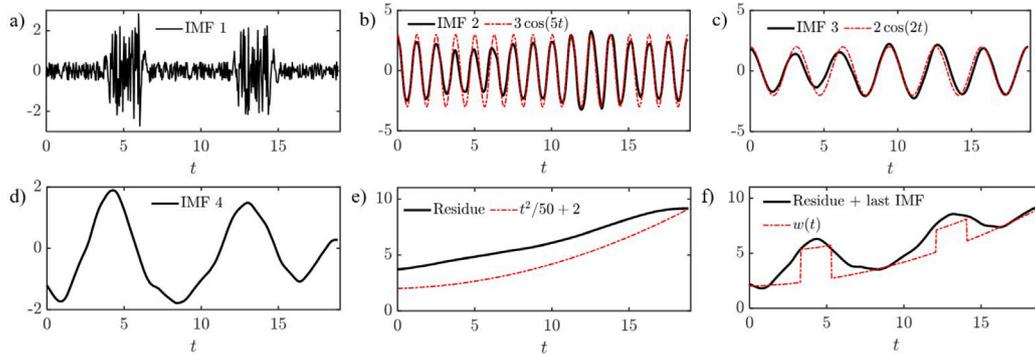


Fig. 14. Demonstration of fast and adaptive EMD applied on the signal of Fig. 3a. (a) First IMF, (b) second IMF, (c) third IMF, (d) fourth IMF, (e) residue, and (f) last IMF added residue.

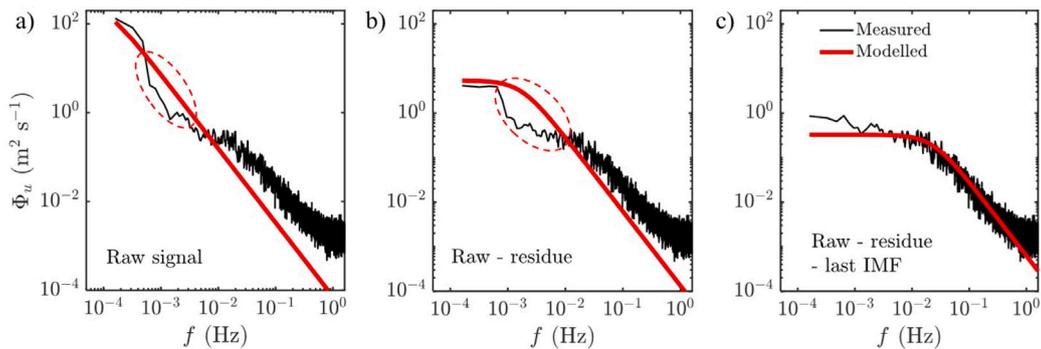


Fig. 15. Comparison between measured and estimated spectra from classic von Kármán model using interval B (Fig. 1c). (a) Unfiltered raw signal, (b) signal with residue subtracted, and (c) signal with residue and last IMF subtracted.

and the spectrum of the measured signal for the raw u' signal and the EMD-filtered data are shown in Fig. 15. The results indicate a good match between the modeled and measured spectra within the inertial subrange for the signal after removing the last IMF and residue obtained from EMD, as shown in Fig. 15c. However, the non-stationary signals produce unrealistically large σ_u and L_u values, resulting in large-scale fluctuations and a sharp increase in energy at the lowest frequency band (Fig. 15a,b, red-dashed ellipse). The incorrect estimation of σ_u and L_u results in overestimating the integral scale energy level and correlation length scale and underestimating the inertial subrange energy.

These results emphasizes the importance of the EMD detrend process for riverine and tidal turbulence data.

5. Discussion

Examining distinct detrending methods, frequency-based filtering, regression-based techniques, and empirical mode decomposition (EMD), reveals their effects on non-stationary tidal turbulence data. Each technique offers a distinct perspective in addressing detrending challenges.

Frequency-based filters manipulate signal frequency content to isolate or remove specific fluctuation scales. While they can eliminate

low-frequency trends effectively, they often introduce unwanted artifacts in autocorrelation (ρ_{uu}) and velocity spectrum (Φ_u). This occurs due to abrupt frequency component removal, causing ringing and transient effects that distort turbulence signal spectral characteristics. Such distortions could affect the analysis and modeling. Hence, while useful in specific contexts, frequency-based filters' susceptibility to introducing artificial features hampers their use with non-stationary tidal turbulence data.

Regression-based detrending, though conceptually simple, can pose implementation challenges. Determining the correct polynomial order requires a balance between capturing trends and avoiding overfitting or underfitting. This method might struggle to handle certain phenomena in tidal data. While offering flexibility, the success of regression-based approaches heavily relies on precise parameter selection and understanding of data characteristics.

In contrast, empirical mode decomposition (EMD) uses the sifting process to automatically obtain various scales of intrinsic mode functions (IMFs), providing an adaptive and robust solution for detrending non-stationary tidal turbulence data and avoids the complex filter parameter tuning process in regression-based detrending. Its adaptability ensures that local, intermittent-like events and glitches minimally affect decomposition, treating them as intrinsic components rather than artifacts. By removing both residue and the largest scale IMF, EMD-detrended signals eliminate large-scale trends while preserving other scale dynamics, demonstrating a smooth spectrum without artifacts observed in frequency-based filtered signals. This aligns with the inherent variability and intermittency of tidal turbulence data.

6. Conclusion and remarks

Our research underscores the significance of detrending non-stationary riverine and tidal current data before estimating turbulence statistics required for classical turbulence spectrum models. This is especially crucial for tidal flows with high velocities and substantial energy conversion potential, where the high-momentum tides often exceed 10 min in duration and exhibit pronounced non-stationary behavior. We assessed two detrending filters: frequency-based (like low-pass filters) and regression-based (such as polynomial regression detrending). Our findings highlight the challenges associated with these methods, including the need for intricate parameter tuning in filter design, the vulnerabilities of regression-based detrending to issues like under- and over-fitting, and sensitivity to glitches. Also, frequency-based filters introduce artifacts in autocorrelation (ρ_{uu}) and velocity spectrum (Φ_u), manifested as ringing and transient effects, rendering them less suitable for handling this type of turbulence data.

We propose a robust detrending process using empirical mode decomposition (EMD) to address these limitations. This approach effectively extracts various scales of fluctuations from the raw signal, categorizing them as intrinsic mode functions (IMFs), while the large-scale trend is captured in the form of residue. We tested two EMD-detrended signals on tidal turbulence data. One subtracted only the residue, while the other removed both the residue and the IMF, representing the largest scale fluctuation. Our outcomes reveal that the latter approach eliminates the large-scale trend within riverine and tidal data. As a result, improved estimations of turbulence statistics, such as turbulence intensity and integral length scale L_u , were achieved. Also, this method provided a velocity spectrum Φ_u that exhibited enhanced compatibility with the classical von Kármán spectrum model. The proposed EMD detrending method is flexible and can be applied to real-world non-stationary data, such as riverine flow, tidal flows, and atmospheric boundary layer wind flow. This method allows for accurate estimation of turbulence quantities and may be instrumental for future turbulence model construction.

CRediT authorship contribution statement

Shyuan Cheng: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Vincent S. Neary:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Funding acquisition, Formal analysis, Conceptualization. **Leonardo P. Chamorro:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Vincent Neary and Leonardo P. Chamorro reports financial support was provided by Department of Energy. If there are other authors, they 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.

Acknowledgments

Sandia National Laboratories is a multi-mission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration, United States under contract DE-NA0003525. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

References

- Acunzo, D., MacKenzie, G., van Rossum, M., 2012. Systematic biases in early ERP and ERF components as a result of high-pass filtering. *J. Neurosci. Methods* 209 (1), 212–218.
- Aldrich, J., 2005. Fisher and regression. *Stat. Sci.* 401–417.
- Araujo, A., 2018. Polynomial regression with reduced over-fitting—The PALS technique. *Measurement* 124, 515–521.
- Bittencourt, C., Argyriadis, K., Steiniger, M., 2014. Type classification for horizontal axis tidal turbine (HATT): An important step towards commercialisation. In: *International Conference on Offshore Mechanics and Arctic Engineering*, Vol. 45547. American Society of Mechanical Engineers, V09BT09A005.
- Bu, C., Zhang, Z., 2021. Research on curve fitting and overfitting based on Bayesian method. In: *2021 6th International Symposium on Computer and Information Processing Technology*. ISCIPT, IEEE, pp. 141–144.
- Burić, M., Grgurić, S., Mikulčić, H., Wang, X., 2021. A numerical investigation of tidal current energy resource potential in a sea strait. *Energy* 234, 121241.
- Butterworth, S., et al., 1930. On the theory of filter amplifiers. *Wirel. Eng.* 7 (6), 536–541.
- Chamorro, L., Hill, C., Morton, S., Ellis, C., Arndt, R., Sotiropoulos, F., 2013. On the interaction between a turbulent open channel flow and an axial-flow turbine. *J. Fluid. Mech.* 716, 658–670.
- Chatterjee, S., Hadi, A., 1986. Influential observations, high leverage points, and outliers in linear regression. *Stat. Sci.* 379–393.
- Chen, W.-B., Liu, W.-C., Hsu, M.-H., 2013. Modeling assessment of tidal current energy at Kinmen Island, Taiwan. *Renew. Energy* 50, 1073–1082.
- Cheng, S., Kaufman, S., Tipnis, V., Best, J., Chamorro, L., 2022. Effects of low clay concentrations on nearly isotropic turbulence. *Phys. Rev. Fluids* 7 (7), 073801.
- Chowdhury, M., Rahman, K., Selvanathan, V., Nuthammachot, N., Suklueng, M., Mostafaeipour, A., Habib, A., Akhtaruzzaman, M., Amin, N., Techato, K., 2021. Current trends and prospects of tidal energy technology. *Environ. Dev. Sustain.* 23, 8179–8194.
- de Cheveigné, A., Arzouanian, D., 2018. Robust detrending, rereferencing, outlier detection, and inpainting for multichannel data. *Neuroimage* 172, 903–912.

- Deb, M., Yang, Z., Wang, T., Kilcher, L., 2023. Turbulence modeling to aid tidal energy resource characterization in the Western Passage, Maine, USA. *Renew. Energy*.
- Devore, J.L., 2008. *Probability and statistics for engineering and the sciences*.
- Durbin, P., Reif, B., 2011. *Statistical Theory and Modeling for Turbulent Flows*. John Wiley & Sons.
- Egbert, G.D., Ray, R.D., 2003. Semi-diurnal and diurnal tidal dissipation from topex/poseidon altimetry. *Geophys. Res. Lett.* 30 (17).
- Eleuteri, A., Fisher, A., Groves, D., Dewhurst, C., 2012. An efficient time-varying filter for detrending and bandwidth limiting the heart rate variability tachogram without resampling: MATLAB open-source code and internet web-based implementation. *Comput. Math. Methods Med.* 2012.
- Florescu, I., 2014. *Probability and Stochastic Processes*. John Wiley & Sons.
- Foken, T., Wichura, B., 1996. Tools for quality assessment of surface-based flux measurements. *Agricult. Meteorol.* 78 (1–2), 83–105.
- Freedman, D., 2009. *Statistical Models: Theory and Practice*. Cambridge University Press.
- Goring, D., Nikora, V., 2002. Despiking acoustic doppler velocimeter data. *J. Hydraul. Eng.* 128 (1), 117–126.
- Grami, A., 2016. Probability random variables and random processes. In: *Introduction to Digital Communications*. Academic Press, pp. 151–216.
- Gunawan, B., Neary, V., Colby, J., 2014. Tidal energy site resource assessment in the East River tidal strait, near Roosevelt Island, New York. *Renew. Energy* 71, 509–517.
- Gunawan, B., Neary, V., McNutt, J., 2011. *Ornl Adv Post-Processing Guide and MATLAB Algorithms for MHK Site Flow and Turbulence Analysis*. ORNL/TML-2011/338.
- Hallin, M., 2014. *Gauss–Markov Theorem in Statistics*. Wiley StatsRef: Statistics Reference Online.
- Hammons, T.J., 1993. Tidal power. *Proc. IEEE* 81 (3), 419–433.
- Harville, D., 1976. Extension of the Gauss–Markov theorem to include the estimation of random effects. *Ann. Statist.* 4 (2), 384–395.
- Hong, L., Cheng, S., Houseago, R.C., Parsons, D.R., Best, J.L., Chamorro, L.P., 2022. On the submerged low-cauchy-number canopy dynamics under unidirectional flows. *J. Fluids Struct.* 113, 103646.
- Huang, N., Shen, Z., Long, S., 1999. A new view of nonlinear water waves: the Hilbert spectrum. *Annu. Rev. Fluid Mech.* 31 (1), 417–457.
- Huang, N., Shen, Z., Long, S., Wu, M., Shih, H., Zheng, Q., Yen, N., Tung, C., Liu, H., 1998. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* 454 (1971), 903–995.
- Huckerby, J., Jeffrey, H., Sedgwick, J., Jay, B., Finlay, L., 2012. An international vision for ocean energy-version II. *Ocean Energy Syst. Implement. Agree.* 20.
- I.E. Commission, et al., 2005. *Wind turbines-part 1: design requirements*. IEC61400-1.
- IEC, 2015. *Tidal and Other Water Current Converters—Part 101: Wave Energy Resource Assessment and Characterization*. Energy—Wave, IEC Marine.
- Iglesias, I., Bio, A., Bastos, L., Avilez-Valente, P., 2021. Estuarine hydrodynamic patterns and hydrokinetic energy production: The Douro Estuary case study. *Energy* 222, 119972.
- IRENA, OEE, 2023a. *Renewable Capacity Statistics 2023*. International Renewable Energy Agency (IRENA) and Ocean Energy Europe (OEE).
- IRENA, OEE, 2023b. *Scaling Up Investments in Ocean Energy Technologies*. International Renewable Energy Agency (IRENA) and Ocean Energy Europe (OEE).
- Kelley, N., Jonkman, B., 2005. *Overview of the Turbsim Stochastic Inflow Turbulence Simulator*. Tech. Rep., National Renewable Energy Lab.(NREL), Golden, CO.
- Kempener, R., Neumann, F., 2014. *Tidal Energy Technology Brief*. International Renewable Energy Agency (IRENA), pp. 1–34.
- Lever, J., Krzywinski, M., Altman, N., 2016. Points of significance: model selection and overfitting. *Nat. Methods* 13 (9), 703–705.
- Maldar, N.R., Ng, C.Y., Patel, M.S., Oguz, E., 2022. Potential and prospects of hydrokinetic energy in malaysia: A review. *Sustain. Energy Technol. Assess.* 52, 102265.
- Milne, I., Sharma, R., Flay, R., 2017. The structure of turbulence in a rapid tidal flow. *Proc. Math. Phys. Eng. Sci.* 473 (2204), 20170295.
- Milne, I., Sharma, R., Flay, R., Bickerton, S., 2013. Characteristics of the turbulence in the flow at a tidal stream power site. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* 371 (1985), 20120196.
- Monin, A., Yaglom, A., 1999. *Statistical Fluid Mechanics: The Mechanics of Turbulence*. Tech. Rep., Massachusetts Inst of Tech, Cambridge.
- Neary, V.S., Gunawan, B., Sale, D., 2013. Turbulent inflow characteristics for hydrokinetic energy conversion in rivers. *Renew. Sustain. Energy Rev.* 26, 437–445.
- Neary, V., Haas, K., Colby, J., 2019. *Marine Energy Classification Systems: Tools for Resource Assessment and Design*. Tech. Rep., Sandia National Lab.(SNL-NM), Albuquerque, NM.
- O. **Orbital Marine Power, 2023. Orbital marine power unveils new 30 MW tidal energy project in orkney waters.** <https://orbitalmarine.com/westray-tidal-energy-project>.
- OES, 2023. *OES Annual Report: An Overview of Ocean Energy Activities in 2022*. International Energy Agency (IEA).
- Papoulis, A., Unnikrishna Pillai, S., 2002. *Probability, random variables and stochastic processes*.
- Park, K., Park, M., 2018. *Fundamentals of Probability and Stochastic Processes with Applications to Communications*. Springer.
- Pope, S., 2000. *Turbulent Flows*. Cambridge University Press.
- Pope, S.B., 2001. Turbulent flows. *Meas. Sci. Technol.* 12 (11), 2020–2021.
- Rehman, N., Mandic, D., 2009. Empirical mode decomposition for trivariate signals. *IEEE Trans. Signal Process.* 58 (3), 1059–1068.
- Rehman, N., Mandic, D., 2010. Multivariate empirical mode decomposition. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* 466 (2117), 1291–1302.
- Rilling, G., Flandrin, P., Gonçalves, P., Lilly, J., 2007. Bivariate empirical mode decomposition. *IEEE Signal Process. Lett.* 14 (12), 936–939.
- Saddoughi, S., Veeravalli, S., 1994. Local isotropy in turbulent boundary layers at high reynolds number. *J. Fluid. Mech.* 268, 333–372.
- Stanisic, M., 2012. *The Mathematical Theory of Turbulence*. Springer Science & Business Media.
- Tanner, D., Morgan-Short, K., Luck, S., 2015. How inappropriate high-pass filters can produce artifactual effects and incorrect conclusions in ERP studies of language and cognition. *Psychophysiology* 52 (8), 997–1009.
- Thiébaud, M., Filipot, J., Maisondieu, C., Dambans, G., Duarte, R., Droniou, E., Chaplain, N., Guillou, S., 2020. A comprehensive assessment of turbulence at a tidal-stream energy site influenced by wind-generated ocean waves. *Energy* 191, 116550.
- Thirumalaisamy, M., Ansell, P., 2018. Fast and adaptive empirical mode decomposition for multidimensional, multivariate signals. *IEEE Signal Process. Lett.* 25 (10), 1550–1554.
- Thomson, J., Polagye, B., Durgesh, V., Richmond, M., 2012. Measurements of turbulence at two tidal energy sites in Puget Sound, WA. *IEEE J. Ocean Eng.* 37 (3), 363–374.
- van Driel, J., Olivers, C., Fahrenfort, J., 2021. High-pass filtering artifacts in multivariate classification of neural time series data. *J. Neurosci. Methods* 352, 109080.
- Vanierschot, M., Persoons, T., Van den Bulck, E., 2009. An eulerian time filtering technique to study large-scale transient flow phenomena. *Exp. Fluids* 47, 613–626.
- Villarini, G., Smith, J., 2010. Flood peak distributions for the eastern united states. *Water Resour. Res.* 46 (6).
- Wang, T., Yang, Z., 2017. A modeling study of tidal energy extraction and the associated impact on tidal circulation in a multi-inlet bay system of puget sound. *Renew. Energy* 114, 204–214.
- Widmann, A., Schröger, E., 2012. Filter effects and filter artifacts in the analysis of electrophysiological data. *Front Psychol.* 3, 233.
- Yadegari, M., Bak Khoshnevis, A., Boloki, M., 2023. An experimental investigation of the effects of helical strakes on the characteristics of the wake around the circular cylinder. *IJST-Trans. Mech. Eng.* 47 (1), 67–80.
- Yang, Z., Wang, T., Branch, R., Xiao, Z., Deb, M., 2021. Tidal stream energy resource characterization in the Salish sea. *Renew. Energy* 172, 188–208.
- Yang, Z., Wang, T., Xiao, Z., Kilcher, L., Haas, K., Xue, H., Feng, X., 2020. Modeling assessment of tidal energy extraction in the Western passage. *J. Mar. Sci. Eng.* 8 (6), 411.