Decomposition of Turbulent Flow Structures using EMD

Hemanth Sarabu

4^{th} May, 2018

1 Introduction

This report outlines an approach for using Empirical Mode Decomposition (EMD) to better visualize and understand the multi-scale dynamics associated with fully turbulent flow. Relevant previous work on EMD and its applications is covered briefly in Section 2, and the EMD algorithm itself is described and elaborated upon in Section 3. The code written for this project is validated against published work in Section 4 and, Section 5 highlights the results obtained on applying EMD to a turbulent flow field dataset.

2 Previous Work

The EMD method was formally introduced in 1998 [4]. The adaptive technique was stated to have the ability to decompose complicated nonlinear and non-stationary processes into a finite number of intrinsic mode functions (IMFs) that can yield well-behaved Hilbert transforms. This is achieved by basing the decomposition on local characteristic time scales; this allows the process to be represented in the frequency-time domain where the extracted instantaneous frequencies are devoid of erroneous harmonics such as those generated from a Fast Fourier Transform. In this work, the IMFs are defined as a class of functions that satisfy the following necessary conditions to define a meaningful instantaneous frequency:

1. The functions are symmetric with respect to the local zero mean, and the number of extrema and the number of zero crossings must either equal or differ at most by one in the whole dataset.

2. At any point, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero.

In [9] , EMD is used to study the characteristics of IMFs produced from white noise. It is indicated that the IMFs generated by EMD are distributed normally and the Fourier spectra of the components look identical and cover the same area on the semi logarithmic period scale. This result is significant in the context of this project as (seen later in this report) similar behavior is observed from IMFs produced from the turbulent flow dataset. Flandrin et al., reaffirm EMD's adaptive capabilities in stochastic situations involving broadband noise in [3] and report that when used on datasets consisting of fractional Gaussian noise, it acts as a dyadic filter bank similar to those seen in wavelet decompositions. In [7], Rilling et al., discuss issues with the original implementation of

EMD in [9] and propose algorithmic variations involving stopping criteria, end conditioning and online implementation. This work describes the different components of the EMD process in a concise manner allowing the reader to develop an intuition for their functions.

Despite its fairly recent introduction as a data analytical tool and its limited mathematical understanding, EMD has been used in a wide range of applications involving time-frequency domain analyses. It has been used to insight in studies such as passenger flow forecasting [8], two-phase debris flow analysis in landslides [6], melanoma classification optimization [5] amongst many others in audio engineering, speech analysis, climate studies and medicine.

3 EMD Algorithm

The EMD algorithm recursively detects all local minima/maxima in the given signal, estimates lower/upper envelopes by fitting a cubic spline through all extrema, subtracts the mean of the envelopes as a low-pass centerline, thus segregating the high-frequency oscillations as a potential IMF. This is continued recursively on the extracted low-pass centerline until a stopping criterion is reached.

1. Identify local maxima and minima in the signal
2. Deduce an upper and a lower envelope by interpolation (cubic splines)
3. Subtract the mean envelope from the signal
4. Iterate until SD criterion is achieved
5. Subtract the so-obtained intrinsic mode function (IMF) from the signal
6. Iterate on the residual

The images that are shown in this section of the report are from Patrick Flandrin's slides in [2]. The steps enumerated above are explained in more detail using images. The signal used to perform EMD on is synthetic and the X axis can be thought of as time and Y as amplitude.

The first step is to identify all local extrema and approximate two splines that define the envelope of the signal. The maxima (shown in blue) are found and a cubic spline is fit to obtain the upper bound of the envelope. Similarly, the minima (shown in red) and the respective lower bound for the envelope are found.

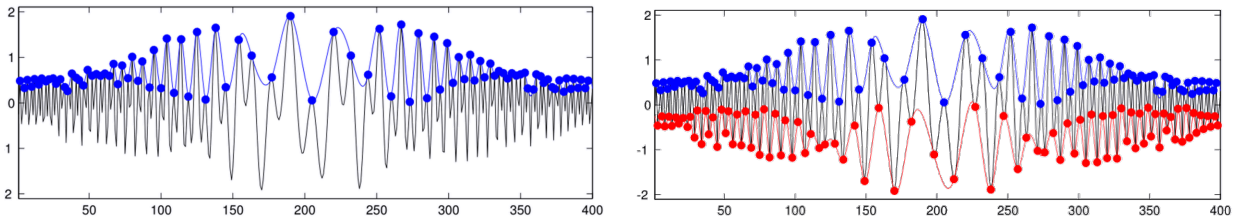

Figure 1: Identifying extrema and fitting splines to obtain envelope

The mean of the envelope is computed by simple averaging of the two curves. Physically, this mean (shown in Magenta) is meant to capture the low frequency components and trends on which the higher frequencies are riding. This is removed from the original signal allowing higher frequencies to be filtered out. Each such step can be likened to a low-pass filtering process. The residue is shown in the right of Figure 2.

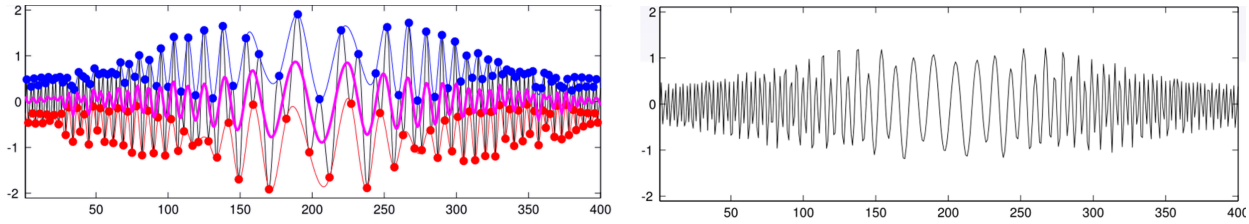

Figure 2: Subtracting mean from original signal yields residue

Obtaining the first residue (residue 1) marks the completion of the first iteration to obtain IMF1. Residue 1 is then treated through the same steps as the original signal to yield a new residue (residue 2); the extrema in residue 1 are found, curves are fit and the mean is calculated and subtracted from residue 1.

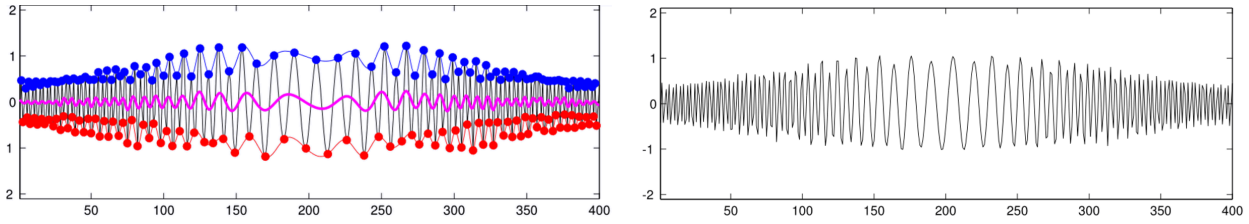

Figure 3: Sifting to filter out low frequencies

This recursive low-pass filtering process approaches a polynomial approximation of the lowest frequencies in a given signal. The same allows the algorithm to isolate the highest local frequencies. A stopping criterion is required to dictate when the sifting process can be halted. The stopping criterion introduced in the original formulation [4] is:

$$SD \leq SD_{crit} \tag{1}$$

$$SD = \frac{1}{T}\sum_{t=0}^{T}\frac{|r_{(k-1)}(t) - r_{(k)}(t)|^2}{r_{k-1}^2(t)} \tag{2}$$

Here SD is the measure of deviation between two consecutive residues in the sifting process, r is the data point at time t and k is the iteration.

The residue obtained form the sifting process after the stopping criterion is satisfied is IMF1 (Figure 4, left). IMF1 is then subtracted form the original signal and the steps described till this point treating the subtracted signal as input. The new IMF obtained is IMF2 (Figure 4, right) and contains the highest local frequencies from the original signal save those captured in IMF1.

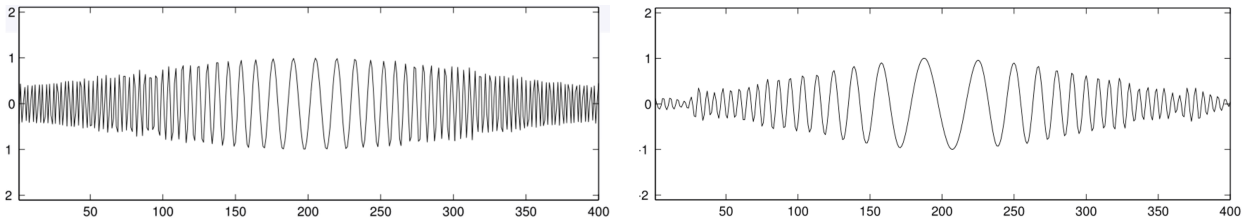

Figure 4: IMF1 (left) and IMF2 (right)

In the same manner, the IMF2 is subtracted from the signal it was generated from to obtain a new signal ($originalsignal - IMF1 - IMF2$) from which IMF3 is extracted. This is continued recursively until all IMFs are extracted. Figure 5 shows IMFs 1 - 8 and the final residue.

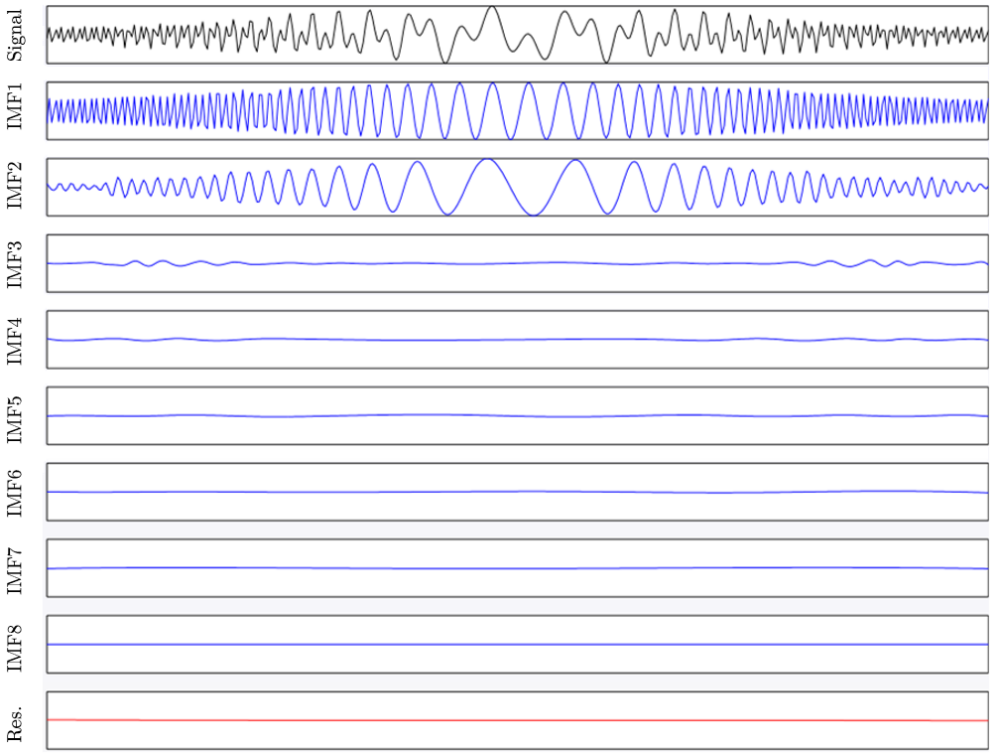

Figure 5: IMFs 1-8 and Residue

The sifting process is explained in [4] although the paper fails to illuminate on the physical significance of over-sifting and 'under-sifting'. It is important to note that EMD attempts to maximize the capture of the highest local frequencies at each time instance in an IMFs. A drawback of the original algorithm is that the IMFs produced are prone to mode-mixing when spectral components are closely spaced or due to the presence of intermittence in the components. EEMD [10] and CEEMD [12] are two noise-assisted variations of EMD that contend to resolve this issue. The turbulent flow data that is being analyzed in this report records no interruptions in any time scale and naturally, mode mixing is not observed.

4 Code Validation

The EMD program used in this project was written from the ground up in MATLAB following the algorithm described in [4]. The code is submitted along with this report. The program is validated against the EMD code and test cases made available by Patrick Flandrin in [1]. The results obtained using both the codes are compared for two synthetically generated datasets; the first is a nonlinear triangular signal with a riding sin wave and the second is a FM signal.

The IMFs produced by both codes are in reasonable agreement, enough for the plots to overlap. The RMS deviation was found to be 2.16% for the first dataset and 5.65% for the second. Figures 5 and 6 indicate the IMFs obtained from both codes and the absolute value of local deviation to visualize where the codes do not agree. It was found that for both test cases, the deviations are noticeable at the edges in IMFs with significant activity at the beginning and end of the record. Future work to understand and minimize the deviations in IMFs produced should include investigating sophisticated end-effect conditioning techniques. Despite the deviations in the IMFs produced, due to the recursive subtractive nature of the algorithm, the reconstruction error is virtually zero in both cases.

4.1 Triangular Signal

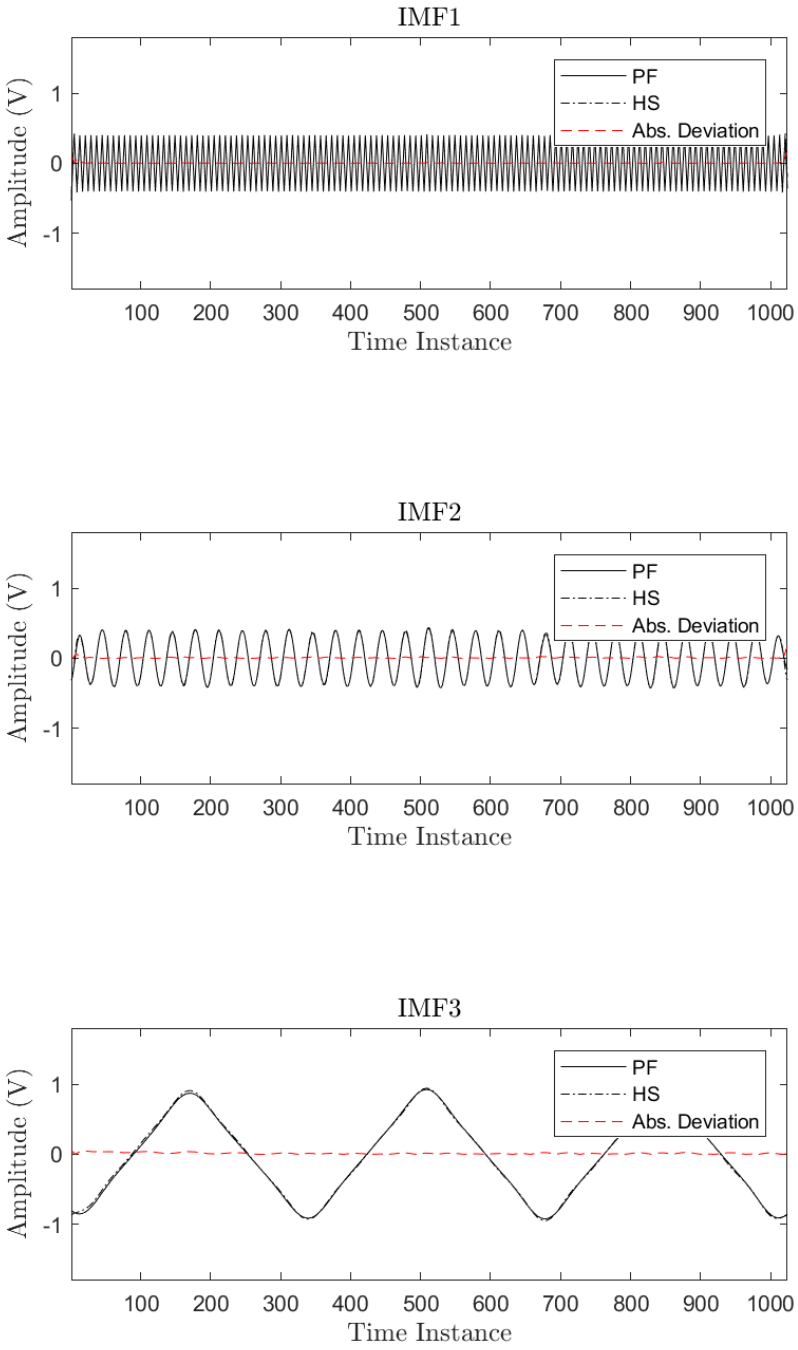

Figure 5: Comparisons for IMFs 1-3

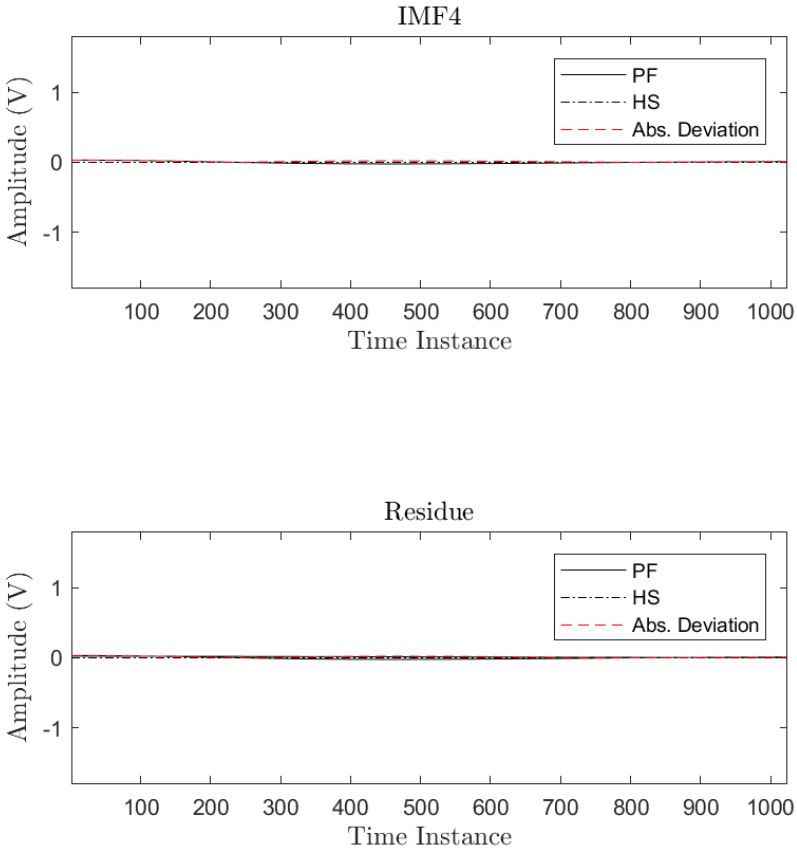

Figure 6: Comparisons for IMF4 and residue

4.2 fmsignal

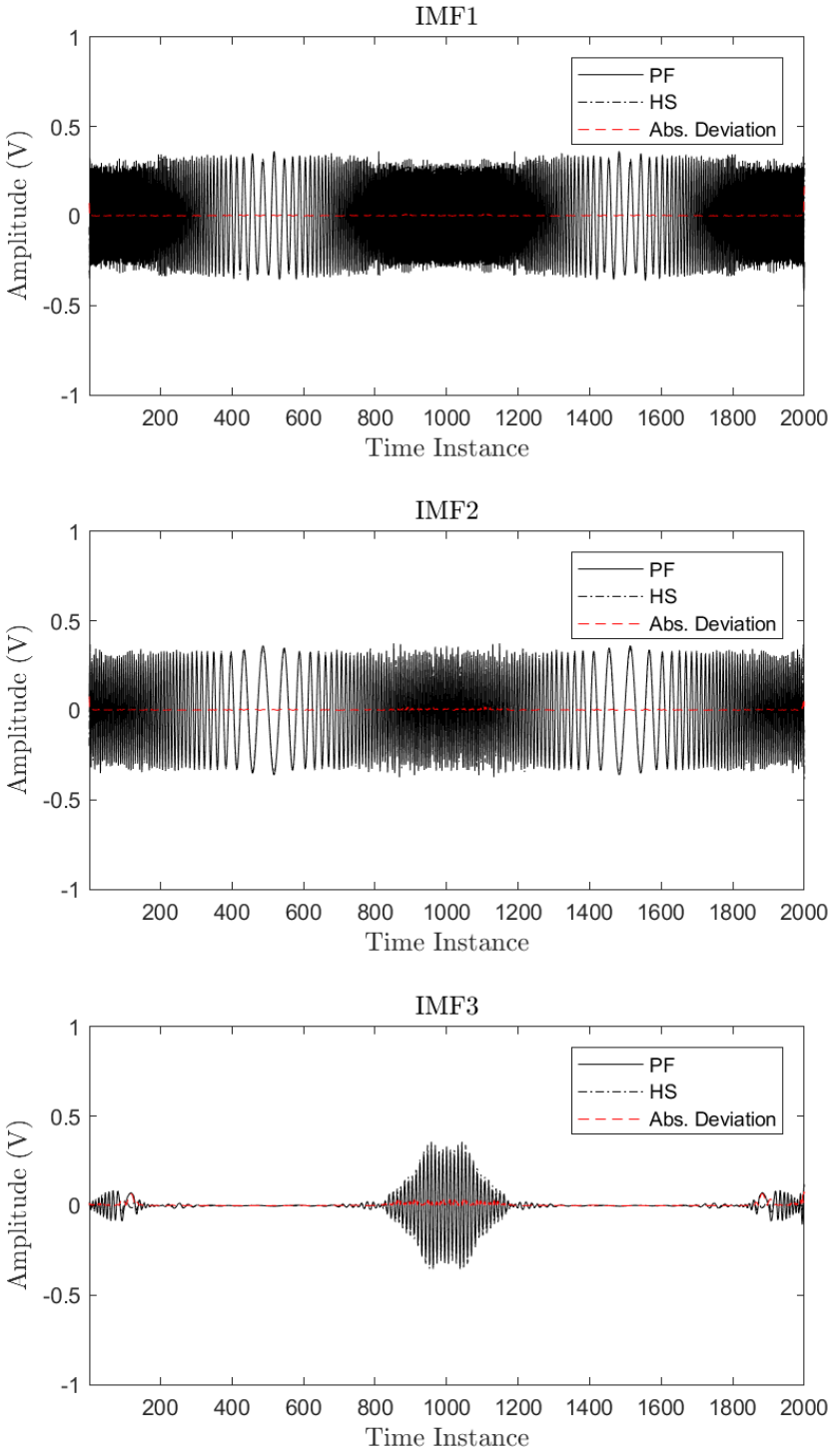

Figure 7: Comparisons for IMFs 1-3

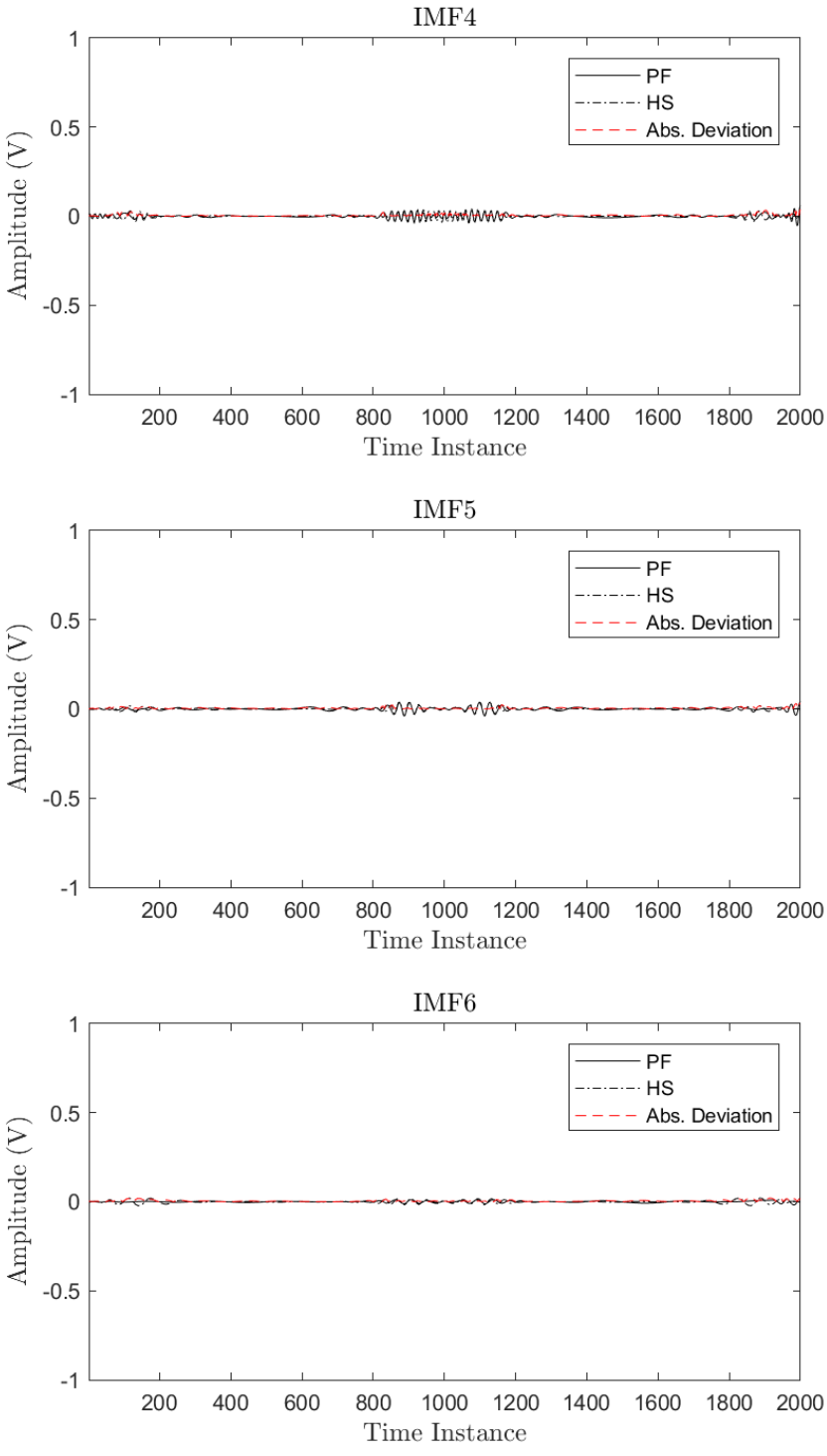

Figure 8: Comparisons for IMFs 4-6

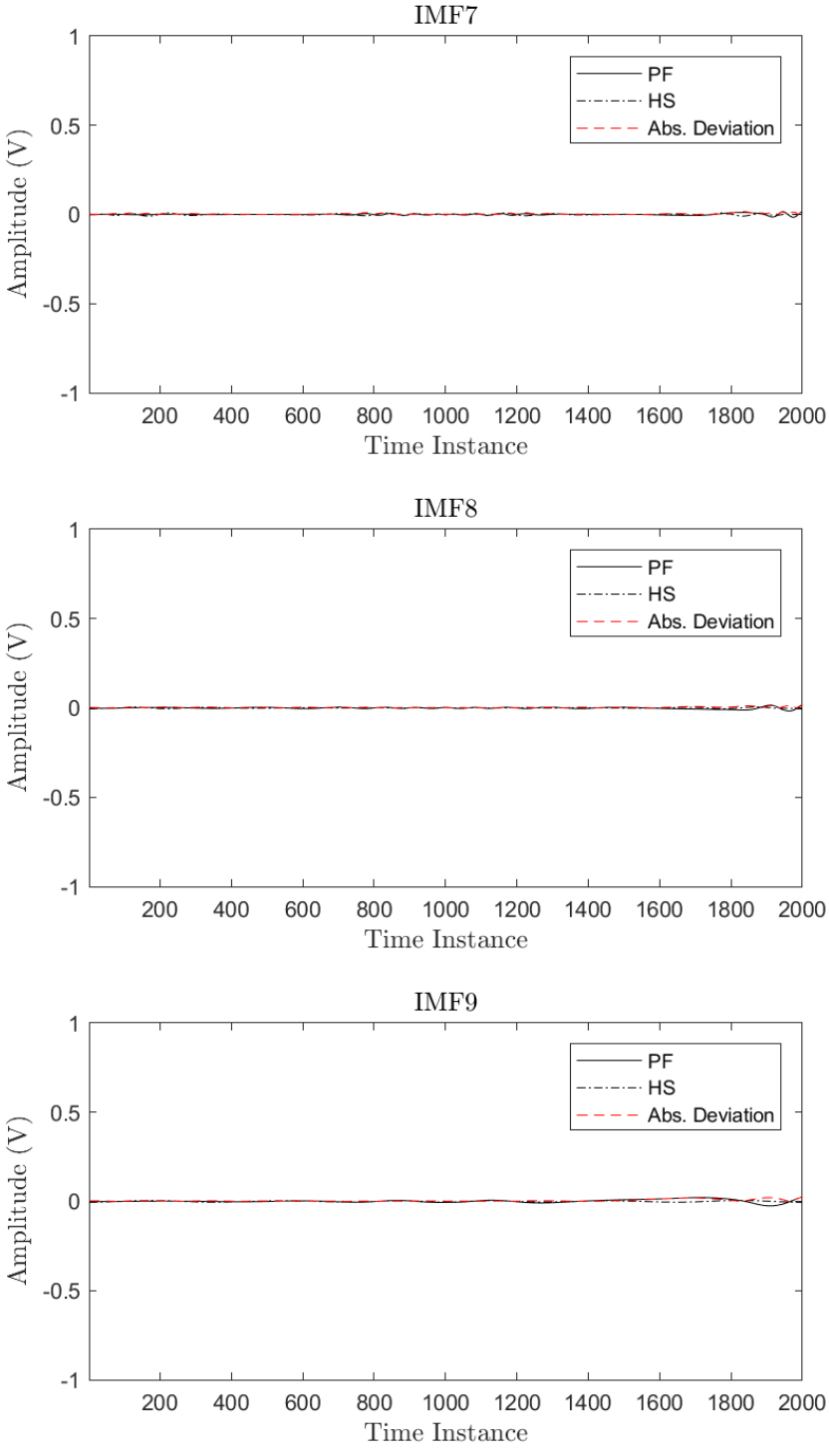

Figure 9: Comparisons for IMFs 7-9

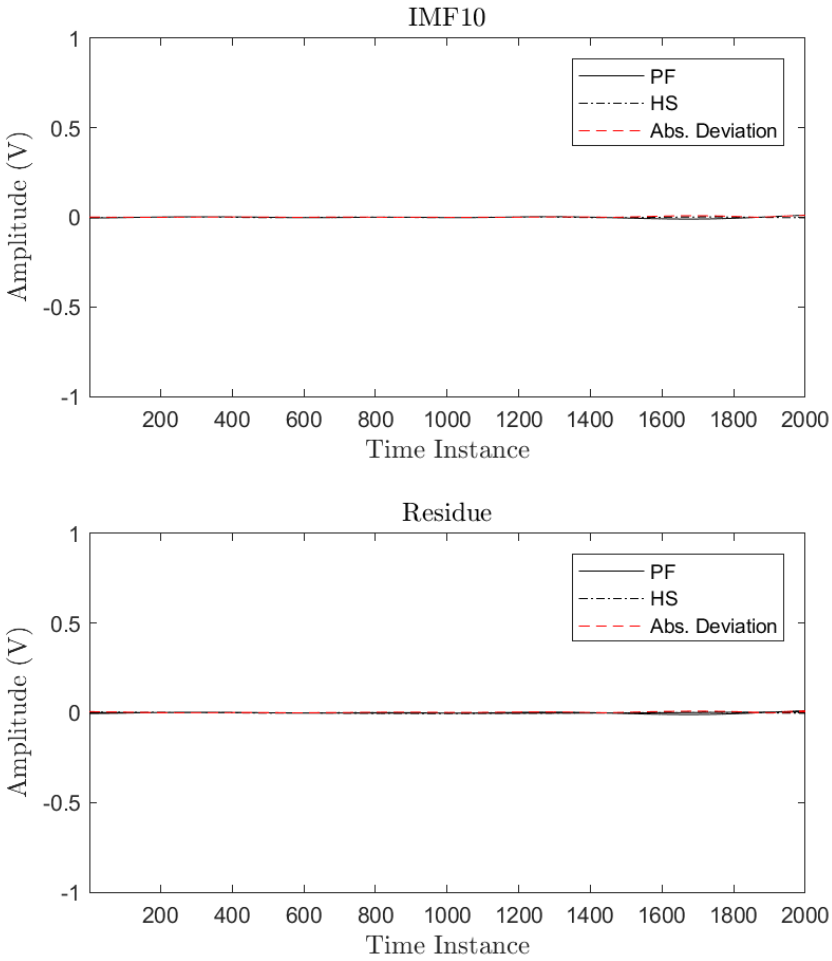

Figure 10: Comparisons for IMFs 10 and residue

5 Application to Turbulent Flow PIV Data

This section discusses how the EMD algorithm is applied to PIV data collected in a fully turbulent shear flow. The data of interest spans a spatial grid of dimensions (17,17) and is recorded for a length of 1 second at a sample rate of 5khz. The trace at each grid point is treated as a 1 dimensional dataset and IMFs are derived for each location. Referred to as the Pseudo-Bi-Dimensional Empirical Mode Decomposition [11], each time trace is decomposed independently, without the incorporation of information from other dimensions (spatial in this case). The stopping criterion shown in Section 2 is used with $SD_{crit} = 0.01$. For each grid location, 14 IMFs and residue were extracted. The animations of the IMFs can be found here: https://tinyurl.com/yblxv9qy

An example where a time trace is decomposed into its IMFs is presented using Figures 11 and 12. Axes labels are omitted in the interest of saving space as these plots are only used to qualitatively convey the characteristics of IMFs. A general trend that is easily recognized is the tendency for local frequencies to reduce with IMF number. In this example, it is difficult to find a pattern in the signals for IMFs 1 - 3, purely from a qualitative standpoint. From IMF4, however, the signals seem to behave "nicely" owing to the low frequency structures that appear coherent and less erratic.

Figure 12 consists of the power spectra of the parent signal (data) in black and the power spectra of the IMFs in color. Each IMF captures a band in the power spectrum of the raw signal, traversing from high to low frequencies.

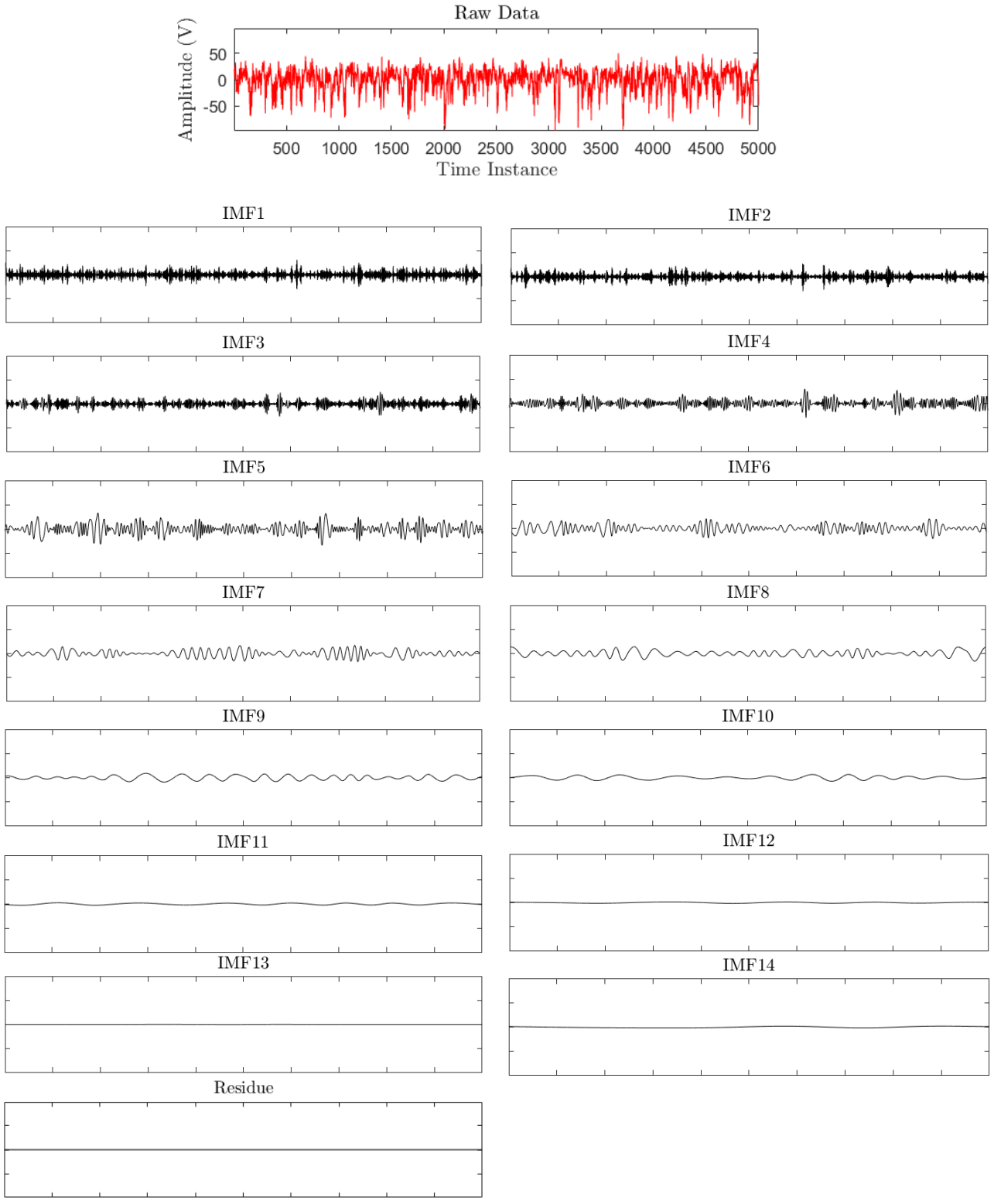

Figure 11: IMFs extracted from a single trace

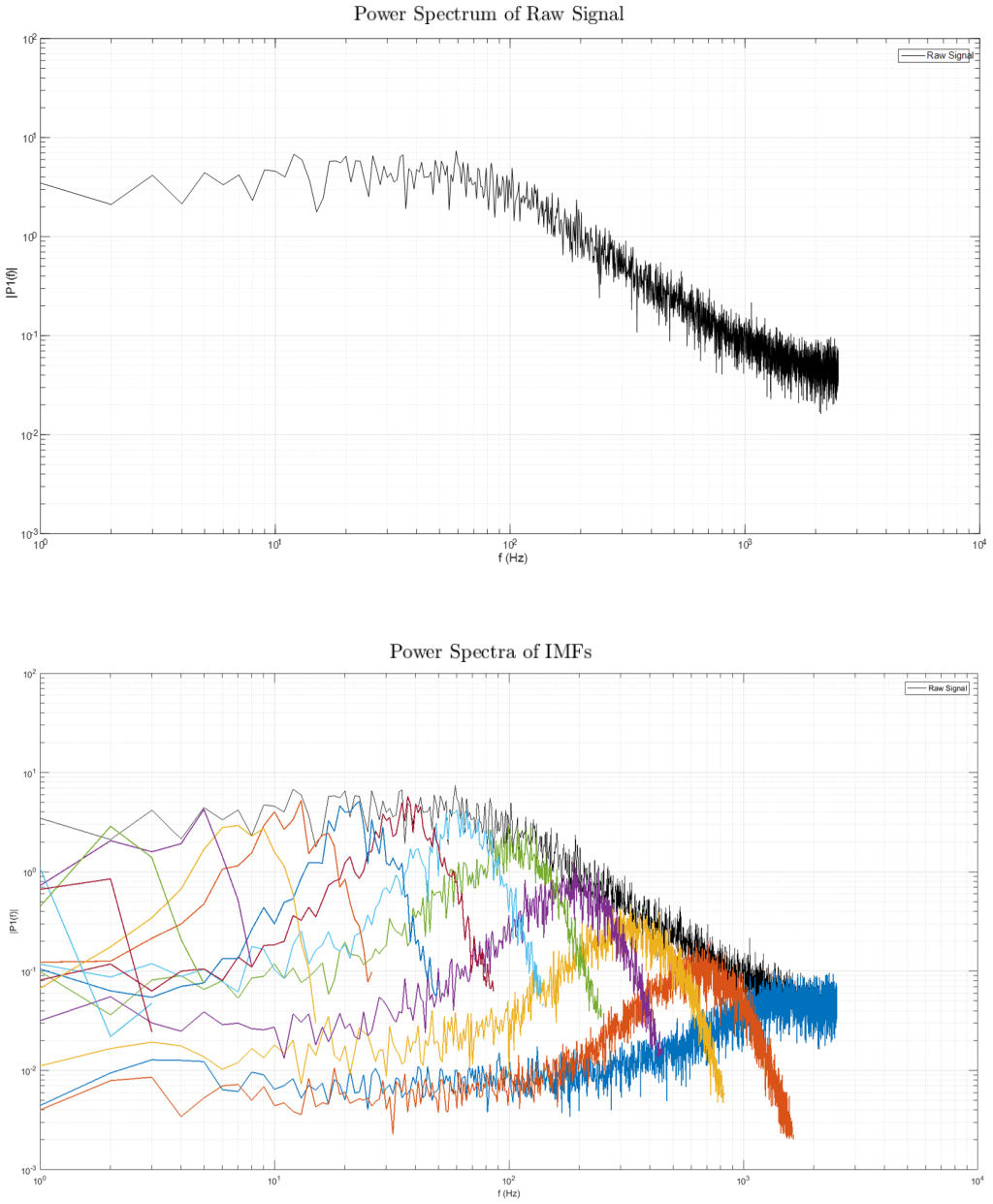

Figure 12: Power spectrum of raw signal (top) and power spectra of IMFs(bottom)

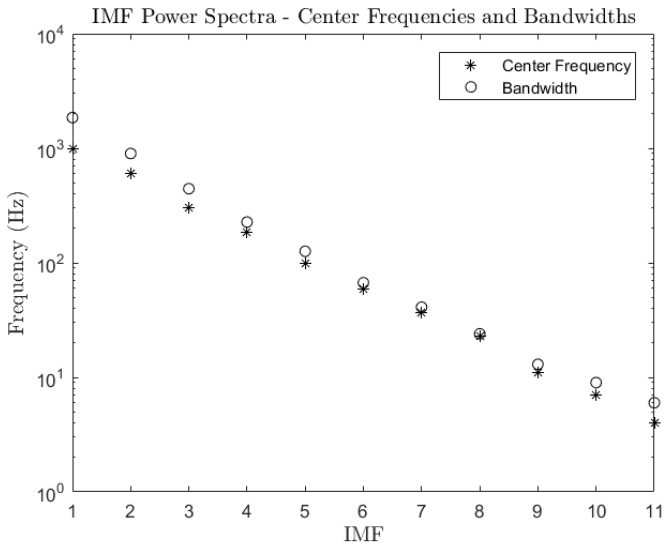

Figure 13: IMF bandwidths and center frequencies

The bands occupied by each IMFS in the power band is visualized in Figure 13 (left) using their respective bandwidths and center frequencies. The center frequencies of the IMFs appear to lie along a a linear path in the log scale. It is seen that the bandwidths follow the same convention.

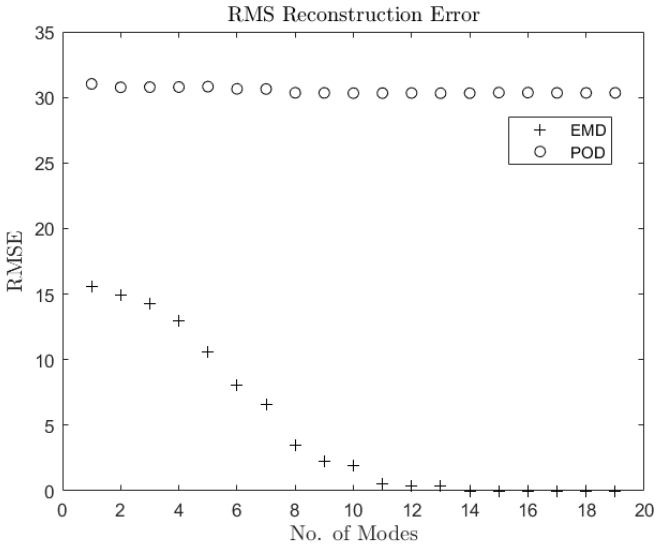

Figure 14: Reconstruction error vs modes used for EMD and POD (right)

The RMS reconstruction error is calculated and plotted as a function of the modes used to approximate the original signal. In Figure 14, the RMSE from IMFs and modes obtained using POD are compared. It is evident that EMD is able to produce higher fidelity reconstruction using fewer components.

6 Conclusion and Future Work

One objective of the project was to understand the EMD algorithm and write a program capable of producing IMFs. While the MATLAB code has been validated against state-of-the-art implementations and checked for robustness, there is scope for algorithmic improvements pertaining to end conditioning and parallelization.

The studies conducted on applying EMD to turbulent flow data reveal the algorithm's capability of systematic decomposition based on scale. Due to the homogeneous nature of the dataset, the implementation of point-wise EMD may be adequate to decompose the 3 dimensional process (2 in space and 1 in time). However, the next step would be to implement a bi-dimensional EMD algorithm that does take into account the spatial dimension in order to produce IMFs. This would allow for the analyses of less homogeneous and more interesting flow fields and help build towards the goal of this project i.e to be able to isolate and study multi-scale phenomena using EMD. Future work would also involve analyzing each IMF window produced in this study using 2D Fourier transforms and alternative techniques to evaluate the content captured.

References

[1] Patrick Flandrin. Matlab/c codes for emd and eemd with examples.

[2] Patrick Flandrin. The way emd works, 2012. [Online; accessed 19-March-2018].

[3] Patrick Flandrin, Paulo Gonçalves, and Gabriel Rilling. Emd equivalent filter banks, from interpretation to applications. In *Hilbert–Huang transform and its applications*, pages 99–116. World Scientific, 2014.

[4] Norden E Huang, Zheng Shen, Steven R Long, Manli C Wu, Hsing H Shih, Quanan Zheng, Nai-Chyuan Yen, Chi Chao Tung, and Henry H Liu. The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. In *Proceedings of the Royal Society of London A: mathematical, physical and engineering sciences*, volume 454, pages 903–995. The Royal Society, 1998.

[5] Piyapat Kosapan and Suparerk Janjarasjitt. Empirical mode decomposition of blood flow data for melanoma classification. In *Biomedical Engineering International Conference (BMEiCON), 2016 9th*, pages 1–4. IEEE, 2016.

[6] Yu Lei, Peng Cui, Zeng Chao, and Yayong Guo. An empirical mode decomposition-based signal process method for two-phase debris flow impact. 15, 08 2017.

[7] Gabriel Rilling, Patrick Flandrin, Paulo Goncalves, et al. On empirical mode decomposition and its algorithms. In *IEEE-EURASIP workshop on nonlinear signal and image processing*, volume 3, pages 8–11. NSIP-03, Grado (I), 2003.

[8] Yu Wei and Mu-Chen Chen. Forecasting the short-term metro passenger flow with empirical mode decomposition and neural networks. *Transportation Research Part C: Emerging Technologies*, 21(1):148–162, 2012.

[9] Zhaohua Wu and Norden E Huang. A study of the characteristics of white noise using the empirical mode decomposition method. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, volume 460, pages 1597–1611. The Royal Society, 2004.

[10] Zhaohua Wu and Norden E Huang. Ensemble empirical mode decomposition: a noise-assisted data analysis method. *Advances in adaptive data analysis*, 1(01):1–41, 2009.

[11] Zhaohua Wu, Norden E Huang, and Xianyao Chen. The multi-dimensional ensemble empirical mode decomposition method. *Advances in Adaptive Data Analysis*, 1(03):339–372, 2009.

[12] Jia-Rong Yeh, Jiann-Shing Shieh, and Norden E Huang. Complementary ensemble empirical mode decomposition: A novel noise enhanced data analysis method. *Advances in adaptive data analysis*, 2(02):135–156, 2010.