Arctic Sea Temperature Data: Comparative Decomposition Analysis with EEMD, CEEMDAN & VMD

Anthony Luby[1], Dhouha Kbaier[1] and Mark Brandon[2]
Open University (UK), [1]School of Computing & Communications, [2]School of Environment, Earth & Ecosystem Sciences
Email: Anthony.Luby@open.ac.uk

***Abstract*—Evaluating components of environmental time series, such as diurnals, seasonal and trends of Arctic Sea water temperatures are key to providing an understanding of any pattern shifts in such a sensitive environment. This analysis requires a robust decomposition method, given the challenges posed by the presence of noisy, missing, non-stationary and non-linear signal characteristics. Traditional frequency analysis techniques, such as Fourier techniques, are compromised in the context of these data types. Two variants of the Empirical Mode Decomposition method and a Variational Mode Decomposition technique are comparatively assessed as prime candidates, given the data context. A baseline application of the three techniques to a synthesised signal is established, followed by direct application to an annual sea water temperature dataset from the Gascoyne Bay inlet in the Arctic.**

Keywords—Arctic, temperature, EMD, decomposition

I. Introduction

Arctic sea water temperatures are increasing at a rate of 2 to 3 times that of global sea water temperatures, a phenomenon known as Arctic amplification as emphasised by Rantanen et. al. [1], with an increased rate over the last decade. Garcia-Soto et al. [2] examines seven key indicators of climate change, including ocean temperature and the direct effects on the environment consequently, such as melting ice sheets. A tool that would aid the understanding of the behaviour of such systems would be of benefit, such as a robust component analysis method. This paper seeks to examine the performance of three decomposition techniques applied to a temperature time series interval from an instrument stationed in the Gascoyne Inlet, Nunavut, North Canada. This data was accessed through the Canadian Ocean Networks data portal.

From observed datasets, we seek to make inferences about unobserved parameters of a stochastic process, such as means and variances. A structural decomposition of a time series further provides a framework of components derived from the overall process namely, diurnal, seasonal and trend components. Detailing descriptive statistics of these individual components can lead to a more informed understanding of the system.

Traditional component analysis techniques such as Fourier analysis perform sub optimally in the presence of non-linearity and other signal characteristics already mentioned. Ensemble Empirical Mode Decomposition (EEMD), Complete Ensemble Mode Decomposition (CEEMD) and Variable Mode Decomposition (VAR) are more favourable comparable decomposition robust techniques against such signal attributes. These are documented and compared by Liu et. al. [3]

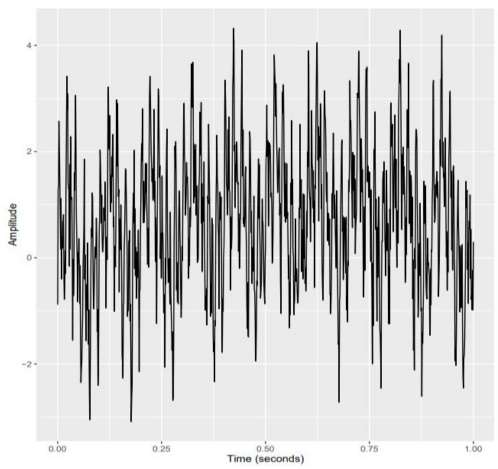

Fig. 1. Synthetic signal, composed of 10, 50, & 100 Hz sine

II. The Data

A. Synthetic Data

As an aid to understanding the performance of the techniques, the EEMD method is applied to a constructed signal, a combined signal of 10, 50 and 100Hz sine waves with the same amplitude and with a nonlinear trend.

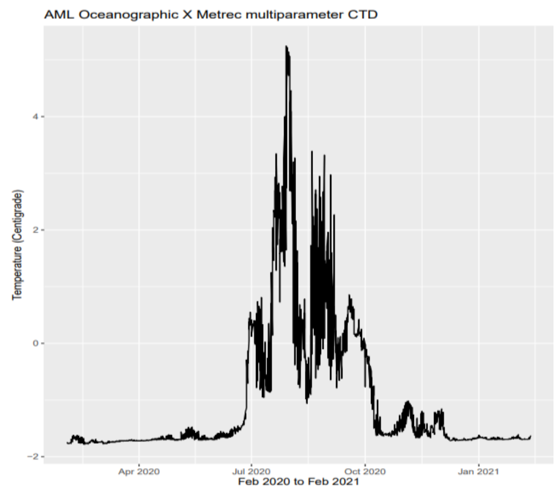

Fig. 2. Gascoyne Temperature Data (Feb. 2019 -Feb. 2020)

Subsequently, the signal is contaminated with a bounded noise component (Fig. 1) and the application is repeated to examine the effect of the noise component on all the methods' performance. The rationale behind this approach is to expose the methods to different frequencies, and to have a direct control of comparative quantitative measurement of recovered frequencies with input frequencies. Observed environmental data have invariably an element of noise and non-linearity, providing a real dimension to the test of the methods.

B. Observed Gascoyne Inlet Temperature Data

A representative annual data interval was selected for the Gascoyne observed temperature dataset, see Fig. 2. This data was obtained from Ocean Networks Canada from a deployed sonde in the Gascoyne Inlet Bay.

A data interval between Feb. 2019 and Feb. 2020 was extracted and down sampled to every 5 hours. This annual interval is intended to capture all the component cycles expected, such as the diurnal, seasonal cycle and trend. It should be noted that very little variation in the temperature is observed during the winter months, as expected in an Arctic environment.

III. METHODOLOGY

EEMD [5] and CEEMD [6] are improved variants of the Empirical Mode Decomposition method as introduced by Huang and Norden [4]. The EMD technique decomposes a signal into individual components called Intrinsic Mode Functions (IMFs) by an iterative 'sifting' process, without transformation into the frequency domain. This sifting process proceeds by fitting a smooth power envelope to the extrema of the signal, calculating the average of the fitted envelope and subtracting from the original signal, to form the first IMF, the highest frequency, which is the resultant signal applied to the next iteration. What follows are two variants of the EMD method.

A. Ensemble Empirical Mode Decomposition

EEMD is a variant and extension of the EMD process. EMD iteratively 'sifts' or extracts out progressively lower frequencies, until some stopping criteria is fulfilled. Wu and Huang [5] proposed EEMD to address mode mixing introduced by intermittent frequencies in the original signal. This is achieved through averaging of ensembles of IMFs with the addition of varying levels of white noise.

B. Complete Ensemble Empirical Mode Decomposition

As an attempt to deal with an introduction of a residue from the addition of white noise in EEMD, Torres et. al. [6] proposed Complete EEMD. A mixture of pairs of negative and positive white noise are added to the signal to cancel the noise addition effect.

C. Variable Mode Decomposition

As an alternative to EMD variants, Dragomiretskiy and Zosso [7] introduced VMD, a non-iterative method. Component frequencies are extracted by a Hilbert transform, mixing with a tuned exponential signal, optimised with a quadratic penalty, to determine the modes concurrently and adaptively.

IV. RESULTS- SYNTHETIC DATA

A. Ensemble Empirical Mode Decomposition

Application of the EEMD algorithm to the synthetic signal without noise and the resultant IMF's (1-3) can be viewed in Fig. 3.

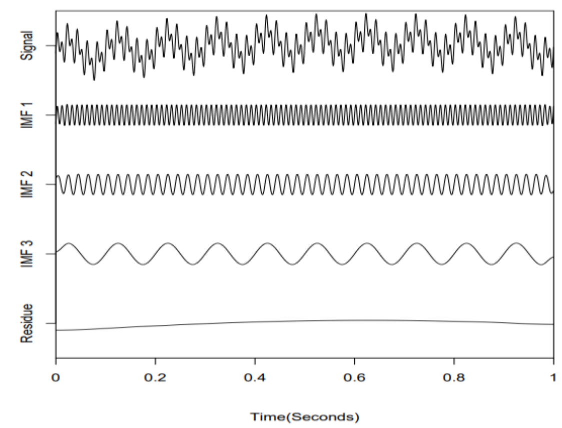

Fig. 3 EEMD decomposition of noise excluded synthetic signal

Here the individual component frequencies are derived effectively from the progressively higher (100Hz) to the lowest frequency (10Hz). The residue-r, is identified as a trend [9] and derived from the following formula [8]:

$$x(t) = \sum_{J=1}^{n} c_J(t) + r_n(t) \qquad (1)$$

This is consistent with the trend applied to the synthesised signal. The addition of noise results in a much less effective decomposition (Fig. 4). The expectation of EMD and its variants, is a progressive 'sifting' of discrete components.

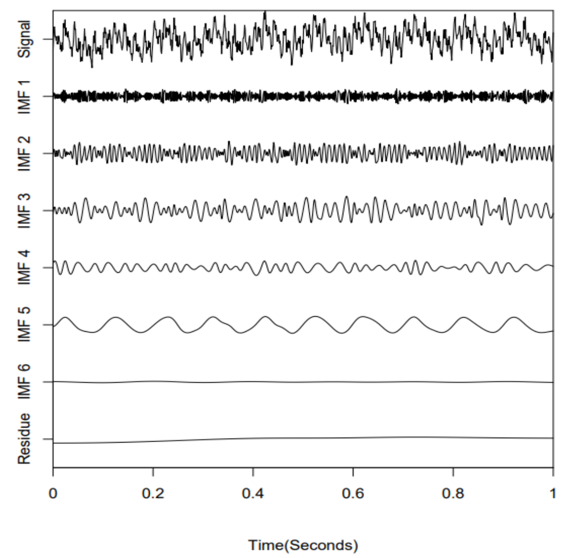

Fig. 4. EEMD results from Noise contaminated synthetic signal

Mode mixing appears evident from IMF1 to IMF3 with standard deviations of 104, 58 and 23, while IMF 4 to 6 have standard deviations of 10, 5 and 2 respectively.

The energy content of the decomposed signal can be viewed in Table I. Here we observe that the power content is uniform, consistent with the input signals of 500 each. The energy is computed for each IMF using:

$$E(y) = \sum_{n=-\infty}^{\infty} y^2 \qquad (2)$$

This is applied to all the EMD variant methods, the derived frequencies can be observed directly from the model results. These frequencies are in cycles per week, based on the sampling frequency of 0.02976169, applied to the EMD algorithms as the down sampled observed signal was every 5 hours. (5 hours/168(hours in a week).

B. Complete Ensemble Empirical Mode Decomposition

Torres et. al. [6] proposed Complete EEMD. Pairs of negative and positive white noise are added to the signal. This is used in an attempt to negate the the noise addition effect.

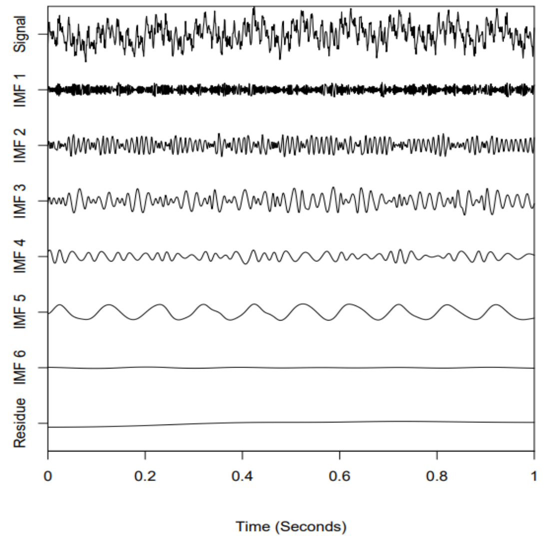

Fig. 5. CCEMD results from noisy synthetic signal.

In Fig. 5, above, a total of 6 resulting IMFs is displayed and a residue reflective of the trend, from the CEEMD algorithm. IMF-5 demonstrates the most regular period, which is the 10 Hz sine wave input.

The energy content is the same as the EEMD results, with the three input frequencies approximated with consistent uniform power distributions (Table I).

C. Variable Mode Decomposition

Fig. 6 displays the significant IMFs out of a total of 18. The VMD method is compromised in the same way that Fourier techniques are compromised by non-stationarity and non-linearity.

Initially, a function is run to estimate the number of modes, and this determines the mode number parameter setting prior to application of the VMD algorithm. Comparatively, the VMD returns 3 times the number of modes, as the EMD techniques, but does similarly resolve the key input components.

The energy content of the VMD modes were calculated using:

$$p_{dft} = |df_t|^2 \;/\; N \text{ where N is the number of samples} \quad (2)$$

The results (in Table 1) depict 3 main components, consistent with the input signal with uniform energy distribution. With VMD, the 10 Hz Mode has most energy content at the expense of the 100 Hz component, taken with Mode 2. Two 100Hz components are identified with contrasting energy.

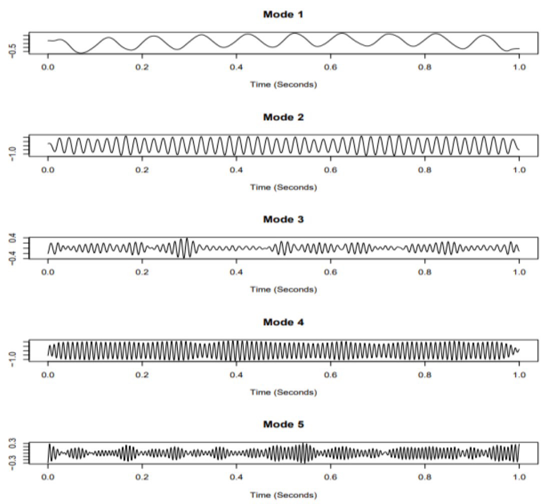

Fig. 6. VMD results from noisy synthetic signal (Shown 5 out of 18 IMF's)

The contribution of the remaining modes, IMF 6 -15 contain relatively very low power input and are thus not displayed.

Table I IMF & Mode Energy Decomposition of the Synthesised Data

	EEMD		**CEEMD**		**VMD Modes**	
	Energy	***Freq.***	***Energy***	***Freq.***	***Energy***	***Freq.***
IMF-1	199	365	199	365	943	10
IMF-2	479	105	479	105	531	50
IMF-3	553	52	553	52	36	98
IMF-4	174	32	174	31	385	99
IMF-5	500	10	500	10	15	134
IMF-6	38	6	38	6	15	169

V. RESULTS- GASCOYNE OBSERVED DATA

Having assessed the performance of three techniques on the synthetic signal, the next step was to apply the same methodology to the observed Gascoyne Inlet data.

A. *EEMD Decompostion Of Observed data*

Very low frequency content is observed from the results of the EEMD decomposition. In Fig. 7 we observe IMFs 1- 8, with the first 4 IMFs displaying high frequency content, atthat scale.

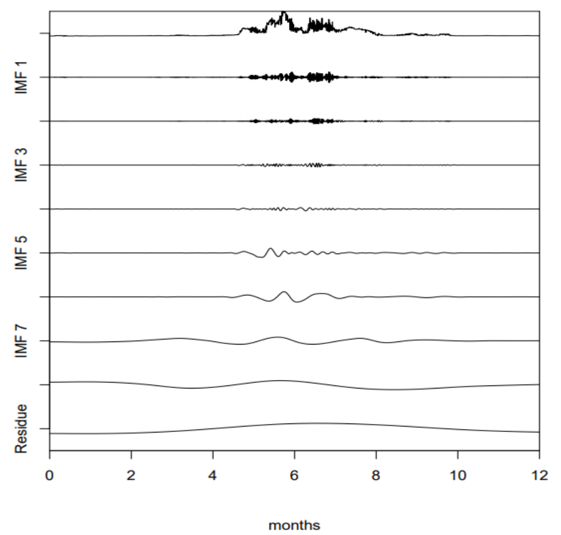

Fig. 7. EEMD results of Gascoyne data.

In the power content table (Table II), 3 main higher frequency components are evident, IMF 1 (13 cycles/week), IMF 2 (6 cycles/week) and IMF 3 (3 cycles/week). IMF 6 to 8 contribute most to the total energy.

Table II Analysis of IMFs after EEMD Decomposition of Gascoyne data

EMMD	Freq.	Energy	Variance	% of Total IMF Energy
IMF 1	13.14	121.66	13	5.9
IMF 2	6.04	29.21	5	1.4
IMF 3	2.73	14.92	2	0.7
IMF 4	1.45	12.62	0.7	0.6
IMF 5	0.44	109.55	0.17	5.3
IMF 6	0.24	245.13	0.3	11.8
IMF 7	0.08	394.81	0.002	19.0
IMF 8	0.037	1148.9	0.0003	55.3

B. *CEEMD Decompostion Of Observed Data*

The CCEMD results are like EEMD as can be observed in Fig.8.

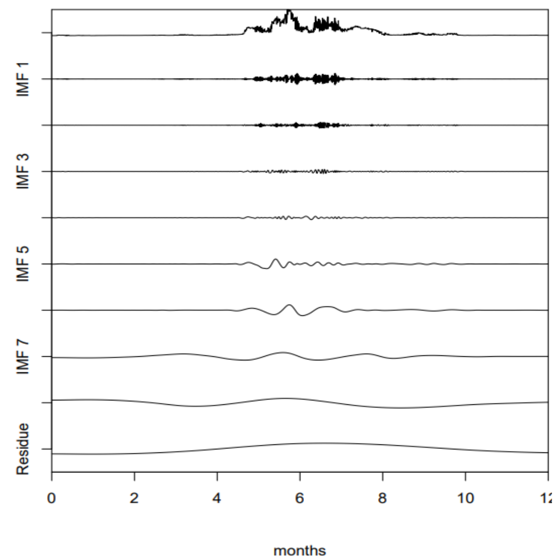

Fig. 8. CCEMD results of Gascoyne data.

C. *VMD Decompostion Of Observed Data*

The results of the VMD decomposition are shown in Fig.9, resulting with only four components. Modes 2 to 4 could be consistent with a Diurnal cycle or semi diurnal.

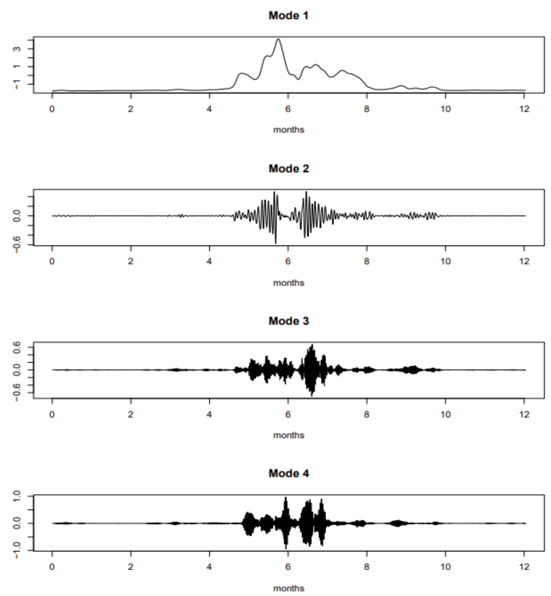

Fig. 9. VMD results of Gascoyne data.

Results are shown in Table III. Most of the energy is represented in a low frequency (Mode 1) of 0.018. Mode 2 & 3 display similar frequencies to the EMD methods, although mode 4 (13 cycles/week has a power content of approx. 50% of IMF-1 from the EEMD and CEEMD application.

VI. Conclusions

In this paper, we have applied three decomposition techniques to both synthetic and observed data and examined the resultant frequency energy metrics, in addition, in the synthesized case, the ability to discriminate the input frequencies.

Table III IMF & Mode Energy from the Observed Data

	EEMD		CEEMD		VMD Modes	
	Ener gy	*Freq.*	*Energy*	*Freq.*	*Energy*	*Freq.*
IMF-1	121	13	121	13	7652	0.02
IMF-2	29	6	29	6	20	3
IMF-3	15	3	15	3	25	7
IMF-4	13	1	13	1	62	13
IMF-5	110	0.4	110	0.4	-	-
IMF-6	245	0.2	245	0.2	-	-
IMF-7	394	0.08	394	0.08		
IMF-8	1148	0.04	1148	0.04		

All three techniques do resolve the original frequencies in the synthetic signal, albeit the EMD variants are the most consistent with the original uniform energy input to all frequencies. VMD estimates excessive modes compared to the EMD types in the synthetic case.

When all three are applied to the Gascoyne data, the results are consistent to a point, where three comparative frequencies are resolved. IMF 6 (EEMD & EEMD) and Mode 7 (VMD) cycles per week would be consistent with a diurnal. The first mode with VMD, assigns a very high-power contribution to the overall signal, this frequency is indicative of a type of low order trend. The EMD methods appear to spread this over IMF's 5 to 8.

Suggested future work could include investigating parameter settings for the techniques, instead of the 'out of the box' approach taken here, reconstructing the signal to assess fidelity with the original, and finally applying the Hilbert Huang transform to the resultant frequencies, compared to the traditional Fourier methodology.

Acknowledgment

The authors would like to express their gratitude to the owners and contributors to the various R package used, including hht and VMDecomp. and to Ocean networks Canada for the data.

References

[1] Rantanen, M., Karpechko, A.Y., Lipponen, A. et al. The Arctic has warmed nearly four times faster than the globe since 1979. Commun Earth Environ 3, 168 (2022). https://doi.org/10.1038/s43247-022-00498-3

[2] Garcia-Soto Carlos, Cheng Lijing, Caesar Levke, Schmidtko S., Jewett Elizabeth B., Cheripka Alicia, Rigor Ignatius, Caballero Ainhoa, Chiba Sanae, Báez Jose Carlos, Zielinski Tymon, Abraham John Patrick,An Overview of Ocean Climate Change Indicators: Sea Surface Temperature, Ocean Heat Content, Ocean pH, Dissolved Oxygen Concentration, Arctic Sea Ice Extent, Thickness and Volume, Sea Level and Strength of the AMOC (Atlantic Meridional Overturning Circulation),journal: Frontiers in Marine Science,vol:8,year:2021,url:https://www.frontiersin.org/articles/10.3389/fmars.2021.642372,doi:10.3389/fmars.2021.642372,ISSN=2296-7745

[3] Liu T, Luo Z, Huang J, Yan S. A Comparative Study of Four Kinds of Adaptive Decomposition Algorithms and Their Applications. Sensors (Basel). 2018 Jul 2;18(7):2120. doi: 10.3390/s18072120. PMID: 30004429; PMCID: PMC6068995

[4] Huang, Norden E., et al. "The Empirical Mode Decomposition and the Hilbert Spectrum for Nonlinear and Non-Stationary Time Series Analysis." Proceedings: Mathematical, Physical and Engineering Sciences, vol. 454, no. 1971, 1998, pp. 903–95. JSTOR, http://www.jstor.org/stable/53161.

[5] Wu, Zhaohua & Huang, Norden. (2009). Ensemble Empirical Mode Decomposition: a Noise-Assisted Data Analysis Method. Advances in Adaptive Data Analysis. 1. 1-41. 10.1142/S1793536909000047.

[6] Torres, M. E., Colominas, M. A., Schlotthauer, G., Flandrin, P. (2011). A complete ensemble empirical mode decomposition with adaptive noise. 2011 IEEE International Conference on Acoustics, Speech, and Signal Processing, pp.4144-4147, doi: 10.1109/ICASSP.2011.5947265.

[7] Zosso, Dominique & Dragomiretskiy, Konstantin. (2013). Variational Mode Decomposition. IEEE Transactions on Signal Processing. 62. 10.1109/TSP.2013.2288675.

[8] Kbaier Ben Ismail, Dhouha & Lazure, Pascal & Puillat, Ingrid. (2016). Application of Hilbert-Huang decomposition to temperature and currents data in the Réunion island. 10.1109/OCEANS.2016.7761460.

[9] Moghtaderi, Azadeh et al. "Trend Filtering: Empirical Mode Decompositions versus ℓ1 and Hodrick-Prescott." *Adv. Data Sci. Adapt. Anal.* 3 (2011): 41-61.