DUJASE Vol. 4(2) 109-116, 2018 (January & July)

Complexity Analysis of the Multivariate Wind Measurements: Renewable Energy Applications

Mosabber Uddin Ahmed

Department of Electrical and Electronic Engineering, University of Dhaka, Dhaka-1000, Bangladesh

E-mail: mosabber.ahmed@du.ac.bd

Received on 13.08.18. Accepted for publication on 04.02.19

ABSTRACT

Complexity analysis of real world multivariate wind data is addressed using the recently proposed multivariate multiscale entropy (MMSE) analysis. Both the original (univariate) MSE and the multivariate MSE methods are shown to perform better than traditional complexity analysis techniques, since they operate on multiple temporal scales of the signals and are, thus, able to extract information regarding inherent long range correlations in the data, signatures of structural complexity. The MMSE method, in addition, can also quantify inter-channel correlations in multivariate data and is perfectly suited for the analysis of multichannel data, where the channels exhibit different dynamical properties, such as three-dimensional wind speed. To cater for the non-stationarity of wind recordings, a novel scheme is presented for obtaining data-driven scales from input data using multivariate extension of empirical mode decomposition (MEMD), in order to obtain robust estimates. Our method can thus characterise different wind dynamics regimes and cloud-cover conditions in complexity domain. Finally, we illuminate how the different dynamic complexities associated with different wind regimes, and their connection with atmospheric parameters, such as temperature, or cloud cover, can be used as baseline knowledge in several important settings in renewable energy.

Keywords: Multivariate sample entropy, Multivariate empirical mode decomposition, Multivariate multiscale entropy, Complexity, Wind speed data, Long-range correlations.

1. Introduction

Electrical power generation using wind farms has emerged as one of the solutions to our increased energy demand; however, power quality from wind farms is subject to variations due to the intermittent nature of wind. Wind speed variability causes vibrations in the mechanical structures of wind turbines and also results in significant voltage variations at the output terminals [1], [2]. Short term wind speed forecasting helps to mitigate these problems by improving control of wind farms through accurate prediction of the generated wind power. However, existing statistical methods address the wind forecasting problem using conventional linear statistical methods [3], [4] which may be suboptimal given the nonlinear, non-stationary, complex and intermittent nature of wind. It is therefore imperative to first obtain an idea about nonlinear and auto- and cross-channel correlation properties among the wind speed channels at multiple scales before building forecasting models. To that end, we propose a framework for nonlinear multivariate complexity analysis, which assesses the stochastic complexity of 3D wind speed and helps to enhance existing short term wind speed/power forecasting methods.

Most real world signals including the wind signal exhibit complex dynamical behaviour and a number of nonlinear descriptors have been proposed to characterize the underlying signal generating mechanisms. These nonlinear measures include dynamical complexity [5], local predictability [6], correlation dimension [7], Lyapunov exponents [8] and phase synchrony [9]. Complexity[1] measures, in particular, reveal nonlinear structures or patterns in the data, typically manifested by nonlinear correlations at multiple scales in a time series. These measures have been extensively employed to distinguish between physiological time series originating from different physical systems or to identify different dynamic regimes coming from the same system [5], [10]. However, their usefulness is yet to be explored in environmental real world applications such as the classification of wind regimes based on their dynamical complexity and nonlinearity, a main focus of this paper. Employing nonlinear complexity measures on multivariate wind data is justified, and previous studies on different univariate wind speed data have shown the evidence of power law decay characteristics, which are typically associated with long range temporal correlations [1] and, hence, increased dynamical complexity [8].

The notion of entropy is commonly used to define signal complexity by effectively evaluating the amount of structure in a time series by assessing its degree of regularity/irregularity [11], [12]. There are many established measures of complexity that are based on different versions of entropy: Pincus introduced approximate entropy [11] as a complexity measure which

[1] The term `complexity' has many different meanings. In this paper, we are concerned with structural complexity which is maximized for data with long-term correlations, unlike Kolmogorov entropy measure which is maximized for random sequences.

was successfully used to distinguish different stages of sleep from electroencephalogram (EEG) and respiratory motion data [13]. Another complexity measure, known as sample entropy (SampEn) [12], was proposed by Richman and Moorman and is a modification of approximate entropy that makes it possible to operate on shorter data segments. Costa *et al.* noticed a discrepancy in the sample entropy method for physiological data sets and attributed it to the fact that sample entropy estimates were only defined for a single temporal scale. They argued that the dynamics of a complex nonlinear system manifests in multiple inherent scales of the observed time series and, thus, sample entropy estimates calculated on a single scale are not sufficient descriptors.

This led to the multiscale entropy (MSE) method in which multiple scales of input data are first extracted using so-called `coarse graining' method and sample entropy estimates are subsequently calculated for each scale separately [5], [10].

While the MSE measure has been successfully applied to distinguish between different physiological time series [5], [10], the method is not a perfect match for processing real world non-stationary data due to its deterministic way of generating scales via coarse graining of input data. The coarse graining process is based on low-pass filtering and is, therefore, not suited for extracting high frequency components of a signal. More critically, it reduces the input data length by the scale factor for each successive data scale, thereby, imposing a limit on the length of input data which can be effectively processed via MSE. To cater for the above issues regarding non-stationarity and the minimum length of input data, it was proposed to replace the coarse graining process in the original MSE method with the empirical mode decomposition (EMD) algorithm [14]. The resulting EMD-based MSE method [15] is highly suitable for the complexity analysis of non-stationary data sets owing to its data driven nature. Another motivation of using EMD in conjunction with the MSE method is that the standard MSE fails to cater for signals containing one or more pronounced trends; little can be inferred from entropy estimates as the trends tend to dominate other interesting features. From the statistical perspective, it is therefore imperative that any trends be removed before meaningful interpretation can be made from sample entropy values. Since EMD decomposes data into narrow-band *quasi-stationary* signals, calculating sample entropy estimates on EMD-based output components, rather than coarse-grained components, improves the original MSE method. For instance, EMD naturally captures a trend in the data within its residue (last component), which can then be removed prior to the MSE analysis.

Since real world wind data is inherently multivariate[2], we require multichannel extensions of existing algorithms for their processing. In the context of complexity analysis of wind speed data, direct multichannel processing of the data is a prerequisite since complexity of the data cannot be solely attributed to long range temporal (auto-)correlations within each channel; the inter-dependence (or cross-correlation) among multiple channels also critically contributes to the complexity of the signal [8]. Likewise, for a fair and meaningful comparison between sample entropy estimates obtained from different channels of multivariate data, a fully multivariate scheme must be used to obtain multiple scales. This would ensure that the scales generated from each channel are same in number and similar in terms of spectral properties (belonging to same frequency bands), thus, making their comparison physically meaningful.

To this end, we develop a robust framework for the structural complexity analysis of wind using the recently developed multichannel extensions of both MSE [16], [17] and EMD [18] algorithms. These extensions have been shown to outperform their univariate counterparts in the analysis of real world multivariate signals as they consider correlation among multiple data channels, which are ignored if those channels are processed separately using univariate algorithms [19]. More specifically, the proposed framework for multivariate wind speed data analysis generates data-driven scales from the multivariate extension of EMD (MEMD) which are subsequently analyzed by using the multivariate sample entropy (MSampEn) estimate. Besides generating comparable scales from multiple data channels, owing to its mode alignment property [20] the proposed MEMD-based MMSE method produces scales of same length as the length of input signal, thus, removing the limitation on input data size in the original MMSE method. The resulting method is also suitable for non-stationary multivariate wind data analysis owing to the data-driven nature of MEMD algorithm as opposed to deterministic coarse graining process used in MMSE.

The paper is organized as follows: Section 2 describes Multivariate Sample Entropy (MSampEn) estimation; this is followed by an introduction to the standard and proposed MEMD-enhanced MMSE frameworks for the complexity analysis of multivariate data in Section 3. In Section 4, simulation results regarding the complexity analysis of multivariate wind data are presented and discussed and finally conclusion is drawn.

2. Multivariate Sample Entropy

In many applications involving different observables from a same system, data often come in the form of multivariate signals. Such data typically exhibits both temporal and cross-channel correlations which cannot be catered for by using univariate algorithms applied to multiple data channels separately. For that cause, a multivariate extension of sample entropy (MSampEn) has been proposed recently which enables entropy calculation for multichannel data, by taking into account both within- and cross-channel dependencies [16]. The method introduces a general multivariate embedding process based on combining observations from multiple channels in a single composite

[2] Wind speed data is inherently trivariate with the speed components in north-south, east-west and vertical directions.

delay vector. To calculate multivariate sample entropy (MSampEn), recall from multivariate embedding theory [26], that for a p-variate time series $\{x_{k,i}\}_{i=1}^{N}, k = 1,2,\dots,p$, observed through p measurement functions $h_k(y_i)$, the composite delay vector can be defined as

$$X_m(i) = [x_{1,i}, x_{1,i+\tau_1}, \dots, x_{1,i+(m_1-1)\tau_1}, x_{2,i}, x_{2,i+\tau_2}, \dots, x_{2,i+(m_2-1)\tau_2}, \dots, x_{p,i}, x_{p,i+\tau_p}, \dots, x_{p,i+(m_p-1)\tau_p}], \quad (1)$$

where $M = [m_1, m_2, \dots, m_p] \in \mathbb{R}^p$ is the embedding vector, $\tau = [\tau_1, \tau_2, \dots, \tau_p]$ is the time lag vector, and the composite delay vector $X_m(i) \in \mathbb{R}^m$ (where $m = \sum_{k=1}^{p} m_k$).

For a p-variate time series $\{x_{k,i}\}_{i=1}^{N}, k = 1,2,\dots,p$, the MSampEn is outlined in Algorithm 1.

3. Multivariate Complexity Analysis

The MSE method in its original formulation is only limited to the analysis of scalar time series, however, multivariate time series are routinely measured in experimental, biological and meteorological systems. Although such multivariate time series can be treated as a set of individual time series by considering each variable separately, this is only efficient if all the variables are statistically independent or uncorrelated at the very least, which is rarely the case in real world recordings. Moreover, the univariate MSE method cannot model cross-channel correlations present among multiple channels of the input data. To that cause, Ahmed and Mandic [16] [17] have recently extended the univariate MSE to the multivariate case; they used coarse graining to generate multiple scales and subsequently calculated multivariate sample entropy (MSampEn) on each scale to define a multivariate complexity measure known as Multivariate Multi-scale Entropy (MMSE). In this manuscript, in addition to the standard version of MMSE which uses coarse graining to define scales in the input data, an MEMD enhanced MMSE method is also presented which employs MEMD to generate intrinsic scales of the data, better suited to the properties of wind.

Algorithm 1: Multivariate sample entropy (MSampEn)

1: Form $(N-\delta)$ composite delay vectors $X_m(i) \in \mathbb{R}^m$, where $i = 1,2,\dots,N-\delta$ and $\delta = \max\{M\} \times \max\{\tau\}$.

2: Define the distance between any two composite delay vectors $X_m(i)$ and $X_m(j)$ as the maximum norm, that is, $d[X_m(i), X_m(j)] = \max_{l=1,\dots,m}\{|x(i+l-1) - x(j+l-1)|\}$.

3: For a given composite delay vector $X_m(i)$ and a threshold r, count the number of instances P_i for which $d[X_m(i), X_m(j)] \le r$, $j \ne i$, then calculate the frequency of occurrence, $B_i^m(r) = \frac{1}{N-\delta-1} P_i$, and define

$$B^m(r) = \frac{1}{N-\delta} \sum_{i=1}^{N-\delta} B_i^m(r). \quad (2)$$

4: Extend the dimensionality of the multivariate delay vector in (2) from m to $(m+1)$. This can be performed in p different ways, as from a space with embedding vector $M = [m_1, m_2, \dots, m_k, \dots, m_p]$, the system can evolve to any space for which the embedding vector is $[m_1, m_2, \dots, m_k + 1, \dots, m_p]$ (k=1, 2, ..., p). Thus, a total of $p \times (N-\delta)$ vectors $X_{m+1}(i)$ in $\mathbb{R}^{m+1}$ are obtained, where $X_{m+1}(i)$ denotes any embedded vector upon increasing the embedding dimension from m_k to (m_k+1) for a specific variable k. In the process, the embedding dimension of the other data channels ($\ne k$) is kept unchanged, so that the overall embedding dimension of the system undergoes the change from m to $(m+1)$.

5: For a given $X_{m+1}(i)$, calculate the number of vectors Q_i, such that $d[X_{m+1}(i), X_{m+1}(j)] \le r$, where $j \ne i$, then calculate the frequency of occurrence, $B_i^{m+1}(r) = \frac{1}{p(N-\delta)-1} Q_i$, and define $B^{m+1}(r) = \frac{1}{p(N-\delta)} \sum_{i=1}^{p(N-\delta)} B_i^{m+1}(r)$. (3)

6: Finally, for a tolerance level r, estimate $MSampEn$ as

$$MSampEn(M, \tau, r, N) = -ln\left[\frac{B^{m+1}(r)}{B^m(r)}\right]. \quad (4)$$

Algorithm 2: Standard multivariate multiscale entropy

1: Multiple coarse-grained time series are generated from the original time series $\{x_{k,i}\}_{i=1}^{N}, k = 1,2,\dots,p$, where p denotes the number of variates (channels) and N the number of samples in each variate**.**

2: The elements of the coarse-grained time series of scale factor ϵ are calculated as:

$$y_{k,j}^{\epsilon} = \frac{1}{\epsilon} \sum_{i=(j-1)\epsilon+1}^{j\epsilon} x_{k,i}, \quad (5)$$

where $1 \le j \le \frac{N}{\epsilon}$ and $k = 1, \dots, p$.

3: Calculate the multivariate sample entropy, MSampEn (described in Algorithm1) for each coarse-grained multivariate $y_{k,j}^{\epsilon}$, and plot MSampEn as a function of the scale factor ϵ.

The following subsections explain both methods in some detail and give interpretation of their results for input multichannel white Gaussian noise as well as 1/f fractal noise.

A. Standard multivariate multiscale entropy

The standard multivariate multiscale entropy (MMSE) method is described in detail in Algorithm 2. The coarse graining based MMSE method assesses relative complexity of normalized multi-channel temporal data by plotting multivariate sample entropy as a function of the scale factor whereby:

a) A multivariate time series is considered more structurally complex than another if, for the majority of time scales, its multivariate entropy values are higher than those of the other time series;

b) A monotonic decrease in multivariate entropy values with the scale factor reveals that the signal in hand only contains information at the smallest scale, and is thus not dynamically complex;

c) A constant MSE curve over all the scales indicates long term correlations in the data, a signature of truly complex systems.

For illustration, Fig. 1(a) shows the standard multivariate MSE analysis [16], [17] for trivariate random white noise (uncorrelated), conforming to the interpretation that the MSampEn values monotonically decrease with scale, whereas for the $1/f$ fractal noise (long-range correlated) the MSampEn remains constant over multiple scales. This has a physical justification, as by design $1/f$ noise is structurally more complex than uncorrelated white noise.

B. MEMD-enhanced multivariate multiscale entropy

In standard multivariate multiscale entropy (MMSE), multiple data scales are generated by applying the same coarse graining process used for the univariate MSE to each input channel in parallel (see listing of Algorithm 2). Performing coarse graining separately for each channel is inherently not a data-adaptive process and also lacks direct multivariate approach; therefore, it fails to generate `aligned' and `intrinsic' temporal scales from the data - a prerequisite for high fidelity multiscale analysis. Moreover, the output of the coarse graining process reduces the length of each subsequent scale to the length of the original time series divided by the corresponding scale factor, ϵ. The method, thus, imposes the constraint that the highest scale should have enough data points to be able to calculate valid entropy estimates. This somewhat limits the applicability of the coarse graining based MMSE method for short real-world data.

To alleviate the above problems, we propose to use multivariate empirical mode decomposition (MEMD) to generate multiple scales of a given multivariate data and subsequently perform multivariate entropy analysis on cumulative[3] IMFs (scales). The proposed method is outlined in Algorithm 3.

To illustrate the performance of the proposed MEMD based MMSE method, it was applied to synthetically generated trivariate white noise and trivariate $1/f$ noise. The $1/f$ noise possesses long-range correlations and its standard entropy (at scale 1) is lower than that of white noise, however, the $1/f$ noise is structurally complex whereas the trivariate white noise is not, and any complexity measure should be higher for $1/f$ noise at increasing scales. Observe from Fig. 1(b) that though trivariate white noise has higher complexity than $1/f$ noise for the first scale, the complexity becomes lower than $1/f$ noise for higher scales. This example on synthetic data illustrates, that by design, $1/f$ noise is structurally more complex than uncorrelated random noise, a result consistent with standard MSE/MMSE [5], [16], [17] as shown in Fig. 1(a).

[3] Due to their narrow-band nature, an alternative option is to additionally apply coarse graining to the IMF-scales themselves, with minimal risk of aliasing.

Algorithm 3: MEMD-enhanced multivariate multiscale entropy

1: Generate multiple scales from J IMFs obtained by applying MEMD to a given multivariate time series $\{x_{k,i}\}_{i=1}^{N}$ $for \;\; k = 1,2,\dots,p$, where p denotes the total number of variates (channels) and N represents the total number of samples in each variate which does not change across MEMD-based scales.

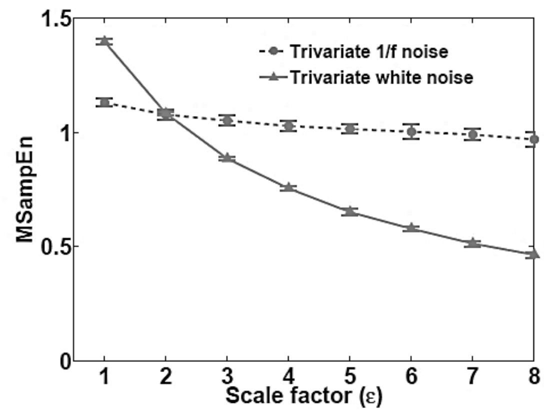
2: Define data-driven 'scales' of x as the cumulative sum of IMFs either by $\mathrm{c}_n = \sum_{j=n}^{J} \mathrm{c}_j$ (Approach 1) or by $\mathrm{c}_n = \sum_{j=1}^{J-n+1} \mathrm{c}_j$ (Approach 2), where $n \in [1, J]$ denotes the cumulative IMF index, and c_j denotes the jth IMF. *Only Approach 1 is used in the sequel*.

3: Calculate and plot multivariate sample entropy measure, given in (4), for each scale n.

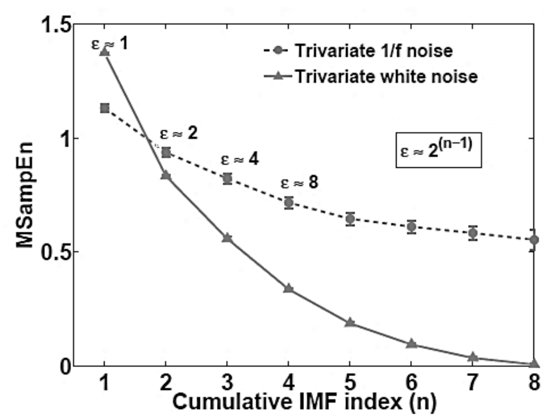
---

4. Applied Multi-Scale Complexity Analysis On Meteorological Data

In this section, we apply multiscale complexity analysis on several multivariate meteorological data sets obtained under different environmental conditions.



(a) Multivariate MSE



(b) MEMD-enhanced Multivariate MSE

Fig. 1: Multivariate multiscale entropy (MMSE) analysis for trivariate white and 1/f noise, each with 10,000 data points using: (a) coarse graining based standard multivariate MSE, and (b) MEMD-enhanced multivariate MSE. The curves represent an average of 20 independent realizations and error bars the standard deviation (SD).

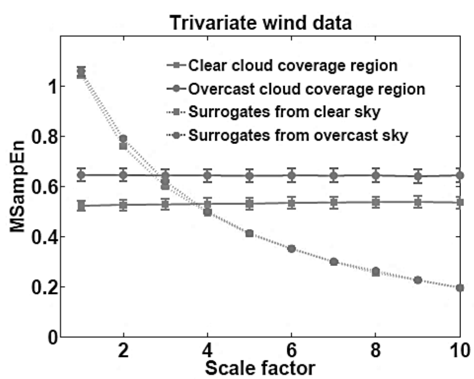
The data sets included:

a) wind speed and temperature data collected under different cloud-cover conditions;

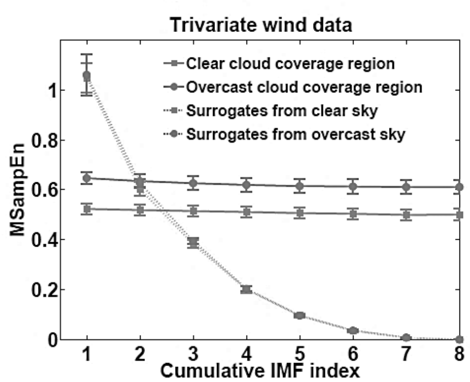
b) wind speed data corresponding to `high', `medium' and `low' wind regimes which were classified based on their speed variations;

In this manuscript, we refer to them as Cloud-cover data and Variance data, respectively. Further detail of the data sets is provided in the respective subsections where the results of complexity analysis performed via original MMSE and MEMD-enhanced MMSE on those data is given.

For rigor, we also performed the complexity analysis on a set of multivariate surrogates generated from the above data sets via random shuffling; this provided a reference for a suitable comparison of complexity estimates obtained from different physical systems. Randomized shuffling of the input data channels effectively destroyed temporal and cross-channel correlations among their samples, while preserving their first and second order statistical properties. This way, significant difference between observed complexity estimates from input data sets and their respective (randomly shuffled) surrogates, over a range of scales, would reject the null hypothesis of both temporal and cross-channel independence, implying a higher complexity and nonlinear coupling in the considered data sets.



(a) Multivariate MSE



(b) MEMD-enhanced Multivariate MSE

Fig. 2: Multivariate multiscale entropy (MMSE) analysis of 3D wind data using: (a) coarse graining based standard multivariate MSE, and (b) MEMD-enhanced multivariate MSE. The curves represent an average of 20 trials and error bars standard error (SE).

A. Complexity analysis of wind during different cloud covering

The energy balance of the Earth's climate is greatly influenced by clouds because of the cooling effect of albedo (as it reflects solar radiations) and the greenhouse warming effect (as it absorbs and re-emits terrestrial radiations back to the earth surface). This also depends on a number of factors, including the size of the droplets, the density of the clouds, their thickness, altitude and temperature, among others. The interaction of clouds with radiation thus alters the surface-atmosphere heating distribution, which in turn drives atmospheric motion that is responsible for the redistribution of clouds. Due to the complexity of the multiscale nature of cloud formation and cloud-radiation interactions, its total affect on the climate system is still unclear and thus provides one of the major uncertainties in climate modelling and prediction [29].

To analyze the underlying complexity of different cloud coverage regimes, we examined air temperature, wind speed and direction data taken from the Iowa Environmental Mesonet (IEM), Iowa State University Department of Agronomy's website.[4] The data contained two groups, one where more than 90% of the sky was covered with cloud (termed as overcast) and another where there were no clouds below 12000 feet (termed as clear). The wind speed, direction and air temperature data were all collected in one minute intervals from the Washington station of the Automated Weather Observing System (AWOS) stations network. Each group consisted of 20 trials with 3000 samples. We also generated 500 random shuffled surrogates of the above data to provide a suitable reference for performing a comparison between complexity curves from different systems. The values of the parameters used to calculate MSampEn were $m = 2$, $\tau = 1$ and $r = 0.20\,\times$(standard deviation of the normalized time series), for all the three data channels.

The data and the set of random surrogates were first analyzed using the standard multivariate multiscale entropy method and the results are shown in Fig. 2(a). Next, the proposed MEMD-enhanced multivariate multiscale entropy method was applied, and Fig. 2(b) shows the results. It can be noticed that both the methods clearly differentiated between the clear and overcast cloud coverage regimes and detected higher complexity in the overcast cloud coverage. This confirms that the formation of cloud increases the complexity of the underlying environmental system. This is intuitive since the interaction between temperature and wind increases during cloud formation which is absent in clear sky conditions [29].

Moreover, it is evident that although surrogates showed greater complexity than the considered data at higher (lower indexed) scales (ϵ) and cumulative IMF index (n); their complexity decreased at lower (higher indexed) scales implying an absence of significant `complex' structures (correlations) at multiple scales. The Cloud-cover data set,

[4] http://mesonet.agron.iastate.edu/request/awos/1min.php

on the other hand, exhibited higher complexity on all scales as evident by approximately flat and high MMSE curves for cloud-cover and clear-sky regimes, using both coarse graining and MEMD-enhanced MMSE methods.

B. Complexity analysis of 'high', 'low', and 'medium' wind dynamics

We shall now show that the MMSE method allows us to characterize different wind regimes based on the variance of their speeds. For this purpose, we identified and analyzed \`high', \`medium' and \`low' wind regimes corresponding to high, medium and low variance of their respective speeds. The data set used in our simulation was recorded using a 3D ultrasonic anemometer (measurements taken in the north-south, east-west and vertical direction) at a sampling frequency of 50Hz in the courtyard of Institute of Industrial Science (IIS) of the University of Tokyo. Fig. 3 shows a plot of the magnitude of wind speed recordings used in the simulation; it can be noticed that the wind dynamics was changing with time allowing us to define three wind regimes, labelled as \`low', \`medium', and \`high'.

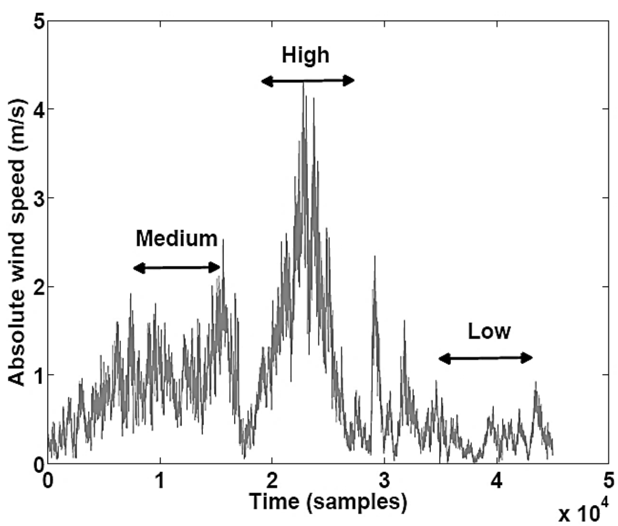
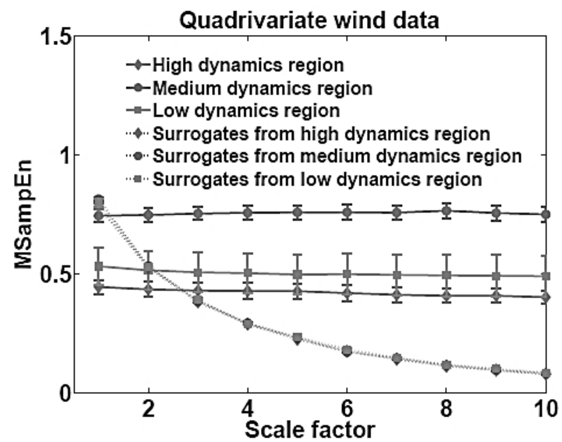


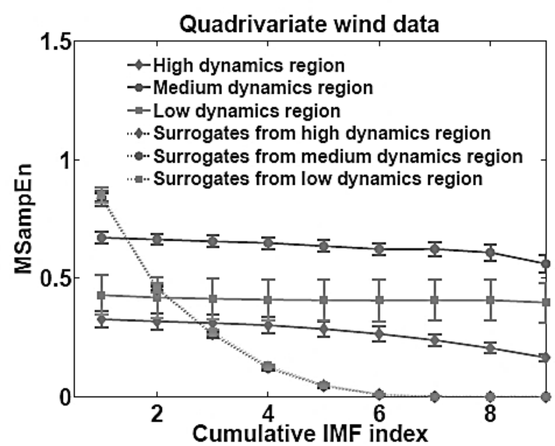
Fig. 3: Magnitude of the 3D wind signal. The wind dynamics regimes are identified as 'low', 'medium', and 'high'.

To generate the mean and variance (error bars) of the calculated sample entropy estimates, we divided the data set into 18 segments, with 6 segments each containing high, medium and low dynamics wind speed data. To reduce the effects of high frequency noise, the data was preprocessed by a moving average filter. The values of the parameters used to calculate MSampEn were $m = 2$, $\tau = 1$ and $r = 0.20\,\times$(standard deviation of the normalized time series), for each of the four data channels. We also performed the complexity analysis of corresponding random-shuffled surrogate time series, generated for each 18 data segments, to provide reference complexity curves for a suitable comparison with the original data.

Fig. 4(a) shows the standard multivariate multiscale entropy analysis, performed by considering the three wind directions as variables in a trivariate model. Observe that the multivariate approach was capable of detecting long-range correlations in the wind speed for all the wind regimes as the MMSE curve was similar to that of $1/f$ noise (cf. Fig. 1(a)), conforming with the existing results [27],[1],[28]. Fig. 4(a) also shows that, as desired, the medium dynamics regime had higher complexity than either high or low dynamics regime. This is also intuitively clear as medium wind dynamics has fewest constraints, and is thus most complex as mild winds come from a wide range of directions [25][28]. Surrogate series, as expected, showed higher MMSE values than that of input data only at highest temporal scales (lower index scales), their complexity decreasing with increasing scale indexes; this behaviour is similar to that of random uncorrelated multivariate noise (see Fig. 1).



(a) Multivariate MSE



(b) MEMD-enhanced Multivariate MSE

Fig. 4: Multivariate multiscale entropy (MMSE) analysis of 3D wind data using: (a) coarse graining based standard multivariate MSE, and (b) MEMD-enhanced multivariate MSE. The curves represent an average of 6 trials and error bars the standard error (SE).

Fig. 4(b): shows the corresponding MEMD-enhanced multivariate multiscale entropy analyses. Observe that MEMD-enhanced MMSE for the trivariate model not only showed a comparatively higher complexity in the medium wind dynamics regime, but was able to differentiate among the low, medium and high dynamics regimes at higher scales, as the error bars did not overlap, hence, a clear improvement over the standard MMSE method. Moreover, as we can consider the wind with medium dynamics as the least constrained system, as opposed to the high or low dynamics regimes which are constrained [25], this interpretation is also in agreement with the general complexity loss theory with constraints [30].

5. Discussion and scope of the work

The results presented in the previous section suggest the presence of long range (auto-) and cross-channel correlations in the chosen data as MMSE curves for those data were relatively flat; note that relatively higher and constant MMSE curves imply similar complexity in multiple scales of the data, a characteristic common to data sets containing long range correlations such as $1/f$ (fractal) noise time-series. It was also observed that the analyzed data sets collected under different environmental conditions were easily distinguished based on their complexity curves, suggesting a direct influence of environmental conditions, e.g cloud covering, on meteorological data characteristics.

The study provides the conclusive evidence, which has great potential in helping to accurately model wind speed for short-term wind power forecasting[5] which is extremely important for wind turbine operation and efficient energy harvesting. In the literature, different techniques have been used to forecast wind speed: a) methods employing numerical weather prediction models which incorporate physical properties of environment such as pressure, terrain etc; b) statistical methods using autoregressive integrated moving average (ARIMA) based models [3], least mean squares (LMS) filters and their respective multivariate extensions e.g. vector autoregressive moving average (VARMA) models [4] and quaternion LMS [31]; c) methods using artificial neural networks [32]; d) spatio-temporal methods integrating information of wind speed at neighbouring sites [33].

The methods and results presented in this paper could be directly integrated into the statistical methods employed for wind forecasting. For instance, higher and constant complexity (MMSE) values at multiple scales of wind data would suggest the use of *multi-scale* statistical ARIMA, VARMA and quaternion LMS models for wind speed data forecasting, that is, applying statistical models to each scale separately and then performing prediction. For data sets exhibiting different complexity values at multiple scales (varying complexity or MMSE curves), different model parameters could be tuned for each scale depending on their complexity values, for improved speed forecasting. MEMD based MMSE could be of great significance in this multi-scale analysis framework as MEMD inherently generates quasi-stationary scales (components) which could qualify for the stationarity requirements of these statistical methods.

Moreover, the generic multivariate nature of the algorithms and analysis presented in the paper makes it possible to be easily combined with existing spatial-temporal methods for wind forecasting. For this cause, wind speed information from neighbouring sites could be combined with the original speed data to create a multivariate signal. Multivariate complexity analysis on the resulting signal would then exploit the speed information from neighbouring sites and is expected to yield better forecasting estimates.

6. Conclusion

A data adaptive multivariate framework for dynamical complexity analysis of real-world wind data has been introduced. The proposed method has been shown to alleviate the stationarity requirements of the current multiscale entropy methods by defining data-adaptive scales through multivariate empirical mode decomposition (MEMD), thus making full use of cross-channel information based upon multivariate sample entropy estimate and MEMD. As a result, the proposed methodology has the ability to produce robust and physically meaningful complexity estimates for real-world systems, which are typically multivariate, finite in duration, and of noisy and heterogeneous natures. The method has been validated on several case studies based on real-world wind data and can be used for analysing underlying dynamical complexity of climatic variables.

References

1 R. G. Kavasseri and R. Nagarajan, “Evidence of crossover phenomena in wind-speed data,” *IEEE Transaction on Circuits and Systems*, vol. 51, pp. 2255–2262, 2004.

2 Z. Jiang K. Xie and W. Li, “Effect of wind speed on wind turbine power converter reliability,” *IEEE Transaction on Energy Conversion*, vol. 27, pp. 96–104, 2012.

3 E. Erdem and J. Shi, “ARMA based approaches for forecasting the tuple of wind speed and direction,” *Applied Energy*, vol. 88, pp. 1405–1414, 2011.

4 B. T. Ewing, J. B. Kruse, J. L. Schroeder, and D. A. Smith, “ Time series analysis of wind speed using VAR and the generalised impulse response technique,” *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 95, pp. 209–219, 2007.

5 M. Costa, A. L. Goldberger, and C. -K. Peng, “Multiscale entropy analysis of complex physiologic time series,” *Physical Review Letters*, vol. 89, no. 6, pp. 068102, 2002.

6 T. Gautama, D. P. Mandic, and M. M. Van Hulle, “Indications of nonlinear structures in brain electrical activity,” *Physical Review E*, vol. 67, no. 4, pp. 046204, 2003.

7 P. Grassberger, “Generalised dimension of strange attractors,” *Physics Letters A*, vol. 97, pp. 227–231, 1983.

8 H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis*, Cambridge University Press, Cambridge, UK, 2000.

[5]Accurate long term forecasting would require more complex and computationally expensive numerical weather prediction (NWP) models employing several environmental variables in a single complex model.

9 D. Looney, C. Park, P. Kidmose, M. Ungstrup, and D. P. Mandic, "Measuring phase synchrony using complex extensions of EMD ," in *Proceedings of the IEEE Statistical Signal Processing Symposium*, pp. 49–52, 2009.

10 M. Costa, A. L. Goldberger, and C. -K. Peng, "Multiscale entropy analysis of biological signals," *Physical Review E*, vol. 71, pp. 021906, 2005.

11 S. M. Pincus, "Approximate entropy as a measure of system complexity," *Proceedings of National Academy of Sciences USA*, vol. 88, no. 6, pp. 2297–2301, 1991.

12 J. S. Richman and J. R. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy," *American Journal of Physiology - Heart and Circulatory Physiology*, vol. 278, no. 6, pp. H2039–2049, 2000.

13 N. Burioka, C. Germaine, F. Halberg, D. T. Kaplan, H. Suyama, T. Sako, and E. Shimizu, "Approximate entropy of human respiratory movement during eye-closed waking and different sleep stages," *Chest*, vol. 123, pp. 80–86, 2003.

14 N. E. Huang, Z. Shen, S. Long, M. Wu, H. Shih, Q. Zheng, N. Yen, C. Tung, and H. Liu, "The empirical mode decomposition and Hilbert spectrum for non-linear and non-stationary time series analysis," *Pro-ceedings of the Royal Society A*, vol. 454, pp. 903–995, 1998.

15 H. Amoud, H. Snoussi, D. Hewson, M. Doussot, and J. Duchene, "Intrinsic mode entropy for nonlinear discriminant analysis," *IEEE Signal Processing Letters*, vol. 14, pp. 297–300, 2007.

16 M. U. Ahmed and D. P. Mandic, "Multivariate multiscale entropy: A tool for complexity analysis of multichannel data," *Physical Review E*, vol. 84, pp. 061918, 2011.

17 M. U. Ahmed and D. P. Mandic, "Multivariate multiscale entropy analysis," *IEEE Signal Processing Letters*, vol. 19, no. 2, pp. 91–94, 2012.

18 N. Rehman and D. P. Mandic, "Multivariate empirical mode decomposition," *Proceedings of the Royal Society A*, vol. 466, pp. 1291–1302, 2010.

19 D. Looney and D. P. Mandic, "Multi-scale image fusion using complex extensions of EMD," *IEEE Transactions in Signal Processing*, vol. 57, no. 4, pp. 1626–1630, 2009.

20 N. Rehman and D. P. Mandic, "Filter bank property of multivariate empirical mode decomposition," *IEEE Transactions on Signal Processing*, vol. 59, no. 5, pp. 2421 – 2426, 2011.

21 L. Huixia, Q. Meng, Y. Wang, and M. Zeng;, "Multi-scale entropy analysis of single-point wind speed in outdoor near-surface environments,"in *International Conference on Electrical and Control Engineering (ICECE)*, 2011.

22 M. Hu and H. Liang, "Adaptive multiscale entropy analysis of multivariate neural data," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 1, pp. 12–15, 2012.

23 C. Vincent, G. Giebel, P. Pinson, and H. Madsen, "Resolving nonstationary spectral information in wind speed time series using the Hilbert-Huang transform," *American Meteorological Society*, vol. 49, pp. 253– 267, 2010.

24 N. Rehman and D. P. Mandic, "Empirical mode decomposition for trivariate signals," *IEEE Transactions in Signal Processing*, vol. 58, pp. 1059–1068, 2010.

25 Danilo P. Mandic and Vanessa Su Lee Goh, *Complex Valued Non-linear Adaptive Filters: Noncircularity, Widely Linear Neural Models*, John Wiley & Sons, Ltd, UK, 2009.

26 L. Cao, A. Mees, and K. Judd, "Dynamics from multivariate time series," *Physica D: Nonlinear Phenomena*, vol. 121, no. 1-2, pp. 75 – 88, 1998.

27 L. Telesca and M. Lovallo, "Analysis of the time dynamics in wind records by means of multifractal detrended fluctuation analysis and theFisher-shannon information plane," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2011, no. 07, pp. P07001, 2011.

28 R. B. Govindan and H. Kantz, "Long-term correlations and multifrac-tality in surface wind speed," *EPL (Europhysics Letters)*, vol. 68, no. 2, pp. 184, 2004.

29 B. Sun, P. Y. Groisman, R. S. Bradley, and F. T. Keimig, "Temporal changes in the observed relationship between cloud cover and surface air temperature," *Journal of Climate*, vol. 13, no. 24, pp. 43414357, 2000.

30 A. L. Goldberger, L. A. N. Amaral, J. M. Hausdorff, P. C. Ivanov, C. -K. Peng, and H. E. Stanley, "Fractal dynamics in physiology: Alterations with disease and aging," *Proceedings of National Academy of Sciences USA*, vol. 99, no. Suppl 1, pp. 2466–2472, 2002.

31 C. C. Took, G. Strbac, K. Aihara, and D. P. Mandic, "Quaternion-valued short term forecasting of three dimensional wind and atmospheric paremeters," *Renewable Energy*, vol. 36, pp. 1754–1760, 2011.

32 M. Mabel and E. Fernandez, "Analysis of wind power generation and prediction using ANN: A case study," *Renewable Energy*, vol. 33, pp. 986–1078, 2008.

33 D. Villanueva, A. Feijo, and J. L. Pazos, "Simulation of correlated wind speed data for economic dispatch evaluation," *IEEE Transactions on Sustainable Energy*, vol. 3, pp. 142–149, 2011.