FORECASTING WIND SPEEDS AT TALL TOWER HEIGHTS WITHIN MISSOURI

A Dissertation presented to
the Faculty of the Graduate School
at the University of Missouri

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

---

by
SARAH SHARLENE BALKISSOON
Dr. Neil Fox, Dr. Anthony Lupo
MAY 2022

The undersigned, appointed by the Dean of the Graduate School, have examined the dissertation entitled:

FORECASTING WIND SPEEDS AT TALL TOWER HEIGHTS
WITHIN MISSOURI

presented by Sarah Sharlene Balkissoon ,
a candidate for the degree of Doctor of Philosophy and hereby certify that, in their opinion, it is worthy of acceptance.

Dr. Neil Fox

Dr. Anthony Lupo

Dr. Sue Ellen Haupt

Dr. Charles Li

Dr. Patrick Market

Dr. Samuel Walsh

ACKNOWLEDGMENTS

I would like to thank my advisors Dr. Neil Fox and Dr. Anthony Lupo for their guidance and support throughout my PhD program. For my committee members, I am grateful for all your suggestions, advise, time and help. Without your invaluable contribution to this chapter of my life, it would not have been completed and come to fruition. I would like to thank Dr. Bahtiyar Efe, Dr. Samniqueka Halsey and Dr. Stephen Penny for their support and technical expertise shared throughout the span this degree. It would be amiss to not acknowledge the contributions of my colleagues, friends and family for their moral support which served as nothing short of a pillar in my academic journey.

Thank you!,
Sarah

Contents

List of Tables

List of Figures

ABSTRACT

Forecasting of wind speeds is necessary for the planning and operations of the wind power generating plants. This research investigates the short term forecasting of wind speeds at tall tower heights for stations within Missouri: Columbia, Neosho and Blanchard. The first objective was to characterize the chaotic nature of this parameter using mono and multi fractal analysis using the Rescale Range Analysis (R/S Analysis) and the Multifractal Detrended Fluctuation Analysis respectively (MF-DFA). It was determined that the system was fractal and there were no trends indicative of increasing fractality and complexity with increasing height. The second objective was the qualitative and quantitative chaotic characterization of the wind speeds using phase-space portraits and the Largest Lyapunov Exponent (LLE) respectively. The methods confirm the results of the fractal analyses. A simple non-linear prediction algorithm, Empirical Dynamical Modeling (EDM) was then used to forecast the wind speeds using a moving window. It was determined that the EDM was comparable to persistence. It beats this benchmark model in the very short term range of one time step or 10 minutes. The third objective was to cluster the data using Self-Organizing Maps (SOMs), having identified the optimum number of clusters as 4 using the Elbow and Silhouette Methods, among others. Three continuous intervals belonging to a particular cluster, which represented approximately 50% and over of the input vectors or rows from the data frame were identified. These intervals were then used as inputs into a Long Short-Term Memory Network (LSTM) with variables, pressure and wind speeds, as well as a lagged series LSTM with embedding dimension, d, and time delay τ . These were compared to the Moving window Auto Regressive Integrated Moving Average (ARIMA) and to persistence. It was determined that the lagged series LSTM improved on the LSTM with wind speed and pressure series inputs, and all models beat persistence. The lagged LSTM beats the Moving ARIMA for at least 2 of the forecasting times of 60 and 120 minutes for all intervals.

Chapter 1

Introduction

Clean energy resources are needed to mitigate the emission of green house gases in the atmosphere due to human activities. One such greenhouse gas is carbon dioxide, which accounted for approximately eighty percent of such emissions by human activities in the U.S. for the year 2019 (United State Environmental Protection Agency, 2021). The main source for the release of carbon dioxide in the United States, in the year 2019, with approximately 35 percent of the total carbon dioxide emissions and 28 percent of the total green house gas emissions, is the combustion of fossil fuels namely gasoline and diesel for the purpose of transportation (United State Environmental Protection Agency, 2021). The second contributor is electricity generation, which accounts for 31 percent of the total carbon dioxide emissions and 24 percent of the total green house gas emissions (United State Environmental Protection Agency, 2021). The reduction of these emissions can come from fuel switching (United State Environmental Protection Agency, 2021) which entails producing more energy from renewable sources.

One such clean energy source is wind energy. Wind is caused by the uneven heating of Earth's surface and atmosphere from solar energy together with the planetary rotation. This causes pressure gradients (which is the difference in this variable over a given distance). Wind results from the flow of air from high to low pressure values. If the high and low pressures are closer together, then the pressure gradients are stronger and thus so too the winds. Wind can slow down near the surface or in areas of wind shear due to friction. Friction is caused from impediments such as buildings, forests, mountains and hills. Friction not only slows the wind but makes it traverse in different directions with varying speeds thus creating turbulence; defined as the rapid changes in wind speed velocities. Turbulence can also be caused by the rising and sinking motions caused by the changes in temperature (thermal gradients) during day time heating.

Missouri has a range of wind potentials which is above the required wind speeds for operational turbines. Out of the total wind power capacity of 122,465 MW installed in the US, Missouri has, as of the last quarter in 2020, 1987 MW (Office of Energy Efficiency & Renewable Energy, 2020). This source is deemed feasible if the power density is greater than 500 Wm^2 . At a height greater than or equal to 50 m above the ground, this implies having a wind speed of 7-8 $\mathrm{m\,s^{-1}}$. This requirement is satisfied by thirteen percent of the Earth's surface (Ferreira, Santos, and Lucio, 2019).

To harness this renewable energy source, forecasting of tall tower winds speeds is necessary. The accurate prediction of this source implies improvements to planning the wind power generating plants (Ferreira, Santos, and Lucio, 2019). Short term prediction of this variable is needed for the operation of the wind turbine (Ferreira, Santos, and Lucio, 2019). This dissertation thus examines the short term forecasting of wind speeds within Missouri using tall tower data. The second chapter deals with characterizing the fractal nature of the wind speeds; looking at its mono- and multi-fractality. The third chapter also establishes the chaotic characteristics of the wind speeds via the construction of phase space and the determination of the Largest Lyapunov Exponent. Thereafter a simple non-linear algorithm, Empirical Dynamical Modeling was conducted. The fourth chapter then investigates the clustering and subsequent forecasting of wind speeds using Neural Networks. The final chapter gives a summary of the previous chapters.

References

Ferreira, Moniki, Alexandre Santos, and Paulo Lucio (2019). “Short-term forecast of wind speed through mathematical models”. In: Energy Reports 5, pp. 1172–1184.

Office of Energy Efficiency & Renewable Energy (2020). U.S. Installed and Potential Wind Power Capacity and Generation. Tech. rep. Energy.gov.

United State Environmental Protection Agency (2021). Overview of Greenhouse gases. Tech. rep. EPA.

Chapter 2

Fractal Characteristics of Tall Tower Wind Speeds in Missouri

2.1 Abstract

The Hurst exponent H is used to determine the measure of predictability of a time series. The value between 0 and 1 with 0.5 representative of a random or uncorrelated series, $H > 0.5$ and $H < 0.5$ reflect a data set which is persistent and anti-persistent respectively. The fractal dimension can be given from the Hurst exponent. The fractal dimension is a factor of the complexity of which the system is being repeated at various scales. If the fractal dimension does not change with scale it is deemed monofractal if not, multifractal. The Hurst exponents were determined in this study using the Rescale Range Analysis (R/S Analysis) and Multifractal Detrended Fluctuation Analysis (MF-DFA) for monofractal and multifractal investigations respectively. These methods were applied to daily 10 minute wind speed time series data for the year 2009 from three locations within Missouri: Columbia, Neosho and Blanchard for three tall tower stations. The results obtained from the monofractal analysis showed minor variations in the Hurst exponents for the three stations and heights for all the months in 2009. These values ranged from 0.7 to 0.9 and its corresponding fractal dimension was ranged between 1.3 and 1.1. The results for the MF-DFA showed that the wind speed time series were multifractal in nature as the Hurst exponents were functions of the scaling parameters. Also, the plots of the Renyi Exponent were nonlinear for the stations and the various channels; this is representative of multifractal signals. The fractal dimensions of the time series using multifractal analysis were determined to be greater than these values determined using monofractal analysis. However, there were no indications of consistent increases in the complexity of the systems' multifractality with increasing heights for the various stations' tall towers.

Keywords: Hurst Exponents, Fractal Dimensions, Rescale Range Analysis, Multifractal Detrended Fluctuation Analysis, Wind speed time series

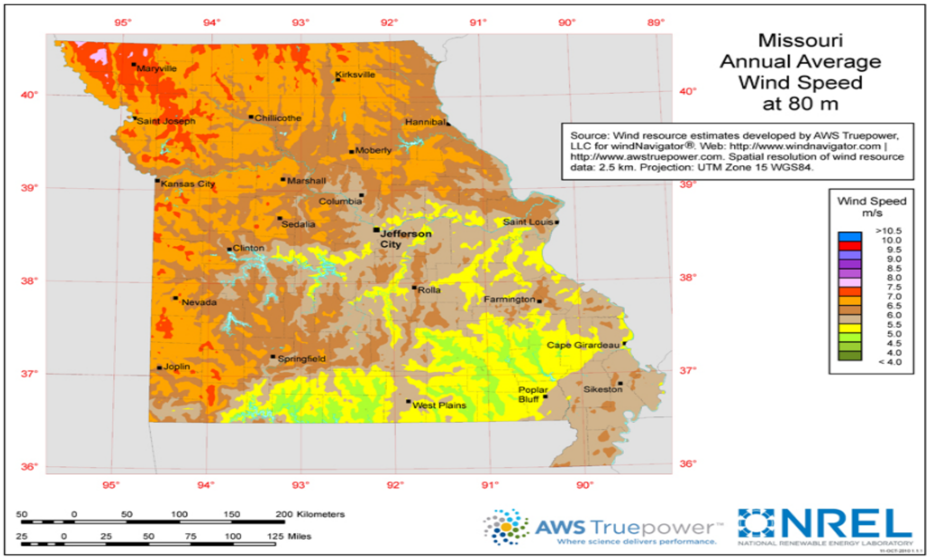
2.2 Introduction

The aim of this study is to determine the internal dynamics of the wind speed time series for three different height levels for three towers in northern, central and southern Missouri. The fractal characteristics of these records provides information on the stochastic processes which generate temporal variations in the series. This information is used in the development of predictive models which ultimately improves the efficacy of wind power as an alternative form of energy.

The subsequent subsections will be an introduction to wind speeds in Missouri, fractals and the relationship between the two. Thereafter, the paper gives a description of the data used in this study. Section 3 seeks to explain the monofractal and multifractal methodologies used and section 4 delves into the analysis of the obtained results for each of these procedures. The final section is the conclusion of the major findings.

2.2.1 Wind Speeds in Missouri

Missouri's average wind speed is approximately 4.5 m/s (Division of Energy, 2019) which is above the 3.5 m/s cut-in wind speed required for small turbines to be operational. The wind speed value for Missouri is higher than some states associated with the wind industry (Division of Energy, 2019). In 2018, six percent of Missouri's electric generation came from renewable energy. Approximately two-thirds of this renewable generation came from wind energy. The wind power generation capacity of 1000 megawatts was derived from 500 turbines (U.S Energy Information Administration, 2019). There is most wind energy potential in the North West regions of Missouri, as seen in Figure 2.1 which is an 80 m average annual wind speed map (Wind Energy Technologies Office, 2019).



Source: Wind resource estimates developed by AWS Truepower, LLC for windNavigator®. Web: http://www.windnavigator.com | http://www.awstruepower.com. Spatial resolution of wind resource data: 2.5 km. Projection: UTM Zone 15 WGS84.

Figure 2.1: AWS True Power and NREL's Wind Resource Map of Missouri.

2.2.2 Fractals

Fractals are associated with objects that are self-similar, that is, they have the same patterns which occur at different scales and sizes. Mandelbrot (Mandelbrot, 1989) stated that is a form of symmetry which is invariant under translations and dilations. These have many details which occur at arbitrary small scales which are too complex to be represented in Euclidean space. Classical geometry and calculus is not suitable for studying fractals and fractal geometry (Falconer, 2004). According to Mandelbrot, when the fractal or Hausdorff dimension is strictly greater than the topological or Euclidean dimension, the set is considered to be fractal and have fractal geometry (Falconer, 2004); the assigned fractal dimension measures the roughness of the surface (Turcotte, 1997). In particular, fractal dimensions can be non-integers which reflects the fact that fractals inhabit space in qualitatively and quantitatively different ways than smooth geometric objects. For example, a smooth curve in the plane is well-approximated by a straight tangent line at each point and hence one dimensional. A fractal, by contrast, does not admit a linear approximation at each point and can have a Hausdorff dimension between one and two. Since the fractal dimension measures the irregularities of a set at various scales, a shape which has a higher fractal dimension is more complex and rough than one that has a lower fractal dimension (Ribeiro and Miguelote, 1998) (Breslin and Belward, 1999).

Fractals can be observed in nature, geometry and algebra as well as mathematical physics. In nature fractals can be seen from small scales such as the scale of two to three atomic diameters in metallic glass alloys (Technology, 2015) to large scales of one hundred thousand light years in a spiral galaxy. Coastlines were characterized as fractal in nature by Mandelbrot; the fractal dimension of a Norwegian coastline was determined to be 1.52 and for a British coastline it was given as 1.31 (Ribeiro and Miguelote, 1998) (Feder, 2013) whilst the fractal dimension of the space distribution of galaxies less than fifty million light years is 1.23 ± 0.04 (Peebles, 1989). Fractals can also be seen in the nonlinear and chaotic behaviour of river and drainage networks as well as hurricanes which is scale invariant (Turcotte, 1997).

In geometry, fractals are observed in for example, the triadic Koch Curve and the

Sierpinski Triangle; these are intermediate shapes of Euclidean Geometry. The Koch Curve is generated from a less detailed starting shape or initiator in which a similar task is added on smaller scales thus making the curve more detailed (Mandelbrot, 1982) (Feder, 2013). That is, each segment of the generator shape is replaced by a smaller copy of the generator itself. Its fractal dimension is 1.26 which is indicative of its infinite length and its area being 0 (Falconer, 2004). The Sierpinski Triangle is generated by the iterative removal of the middle triangle from the previous reconstruction. The fractal dimension of the Sierpinski Triangle is larger than the Koch Curve, its value is 1.58.

We also see fractals in algebra. They are seen in the beginning of modern Mathematics with the middle third Cantor Sets. These sets display properties of self-similarity and have fine structures in which there are details in arbitrary small scales(Mandelbrot, 1982). This uncountably infinite set is formulated from removing in an iterative manner, the middle third of each interval (Falconer, 2004) until the limit of an infinite set of clustered points known as Cantor "dust" is reached (Turcotte, 1997). Since from this process, there are 2^n subsets for n iterations having a magnification factor of 3^n, the fractal dimension given by $D = \log(2^n)/\log(3^n)$ is 0.631 (Turcotte, 1997). The Mandelbrot Set, which led to the development of complex dynamics, is also fractal. This set is defined as all the complex numbers, c for which the function $f_c(z) = z^2 + c$ stays bounded (Mandelbrot, 2013). The image of the Mandelbrot Set shows all the values of c for which the sequence is bounded and all the values of c outside this set for which $f_c(z)$ goes to infinity. It also shows the rate of which the function tends to infinity as seen in the depiction below, Figure 2.2 (c).

There are also fractal connections between non-linear differential equations such as the Navier-Stokes equation (Mandelbrot, 1989). The linear methods of autocorrelation function analysis and spectral analysis are unreliable in the determination of the complex behaviours of non-stationary time series (Philippopoulos et al., 2019). In fluid motion, turbulence is given as effects of singularities of the Navier–Stokes Equation (Mandelbrot, 1989). To study this equation, fractal and multifractal models were developed in which the Hausdorff dimension was determined.

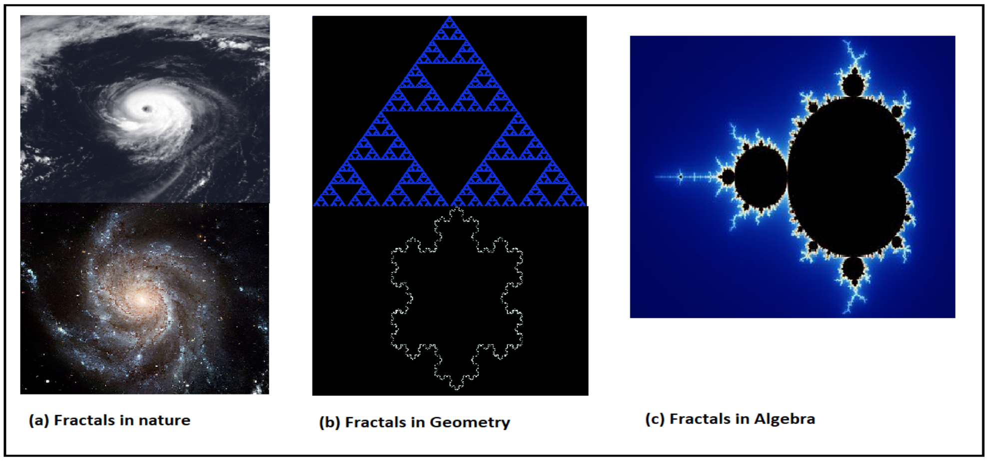

(a) Fractals in nature
(b) Fractals in Geometry
(c) Fractals in Algebra

Figure 2.2: Fractals

2.2.3 Fractals and Wind Speed

To evaluate the wind power and wind potential energy, the analysis of the mean wind speeds and frequency distribution need to be done. This is done to mitigate the problems associated with the intermittency of the wind speeds records, in terms of its the spatial and temporal variations, when trying to integrate wind power into electrical grids (de Figueirêdo et al., 2014). The internal dynamics of the wind speed time series, that is, its monofractal and multifractal characteristics are used to give information on the stochastic processes which are the generators of these temporal variations. This information is useful in the development of predictive models both theoretical and computational in nature (de Figueirêdo et al., 2014). These wind power forecasting tools increase the efficiency of wind power as an alternative renewable source of energy by reducing the unexpected variations in the wind energy conversions systems (WECS) power generation, thus, reducing operational costs in the electricity generation by reducing the requirements of larger primary reserve capacity (W.-Y. Chang, 2014).

2.3 Data

In this study, 10 minutes daily wind speed time series data measured in m/s was used in Missouri, USA for the year of 2009 (Fox, 2011). Three stations were used in this investigation; Columbia, Blanchard and Neosho. Their locations are 038°53.270′N latitude and 092°15.820′W longitude, 040°33.570′N latitude and 095°13.470′W longitude, 036°52.730′N latitude and 094°25.570′W longitude respectively with corresponding site elevations being 255, 328 and 373 m. These are located in North, Central and South Missouri as seen in Figure 2.3. The anemometers were placed on various heights and orientations on the towers. For Columbia, Blanchard and Neosho, the anemometer orientations were 120°and 300°for each of the various sites' tall tower heights of 68, 98, 147 m and 61, 97, 137 m and 50, 70 , 90 m respectively. Channels 1, 3 and 5 are the wind speed times series of the three consecutive heights at an orientation of 120°and Channels 2, 4 and 6 are wind speed values obtained when anemometers were oriented at 300°. The larger of the wind speed values at each time step for all the heights were taken for all of the stations. These were labelled Columbia68,Columbia98 and Columbia147, Blanchard61, Blanchard97 and Blanchard137, Neosho50, Neosho70 and Neosho90. These time series for January to December of 2009 were used in the evaluation of the fractal characteristics of wind speeds within Missouri.

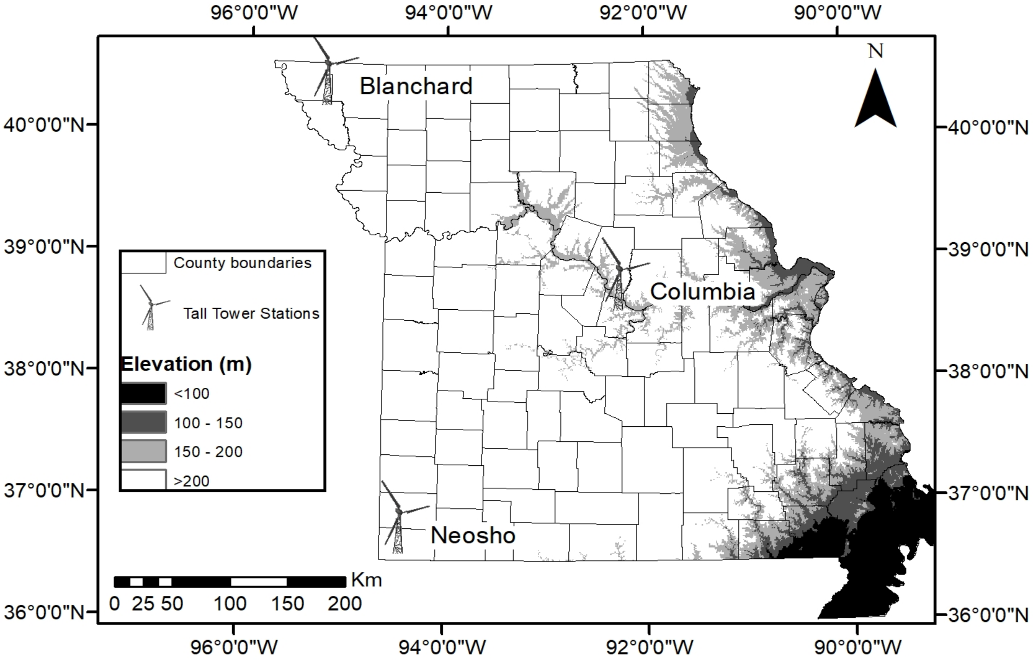


Figure 2.3: Study Locations within Missouri.

2.4 Methodology

2.4.1 Monofractal Analysis: Rescale Range Analysis (R/S Analysis)

There are multiple methods of determining the fractal dimensions of data sets which include the box-counting method, variation method and the Hurst R/S method (Breslin and Belward, 1999). The R/S method gives the scale free irregularity and the long term memory or correlation of the series (Breslin and Belward, 1999). This method was used by Hurst to compare observed ranges of natural phenomena including river discharges, mud sediments and tree rings (Feder, 2013). The scale properties of geophysical variables such as precipitation, temperature, sea level and sunspots using R/S analysis were investigated by Lovejov and Mandelbrot in 1985 and Rangarajan and Sant in 2004 among others (Salomao, Campanha, and Gupta, 2009).

This paper uses the R/S method. To explain the general idea, suppose there is a time series x_i, $i = 1, 2, 3, \ldots, N$. The range R_n is defined to be the difference between the maximum and the minimum accumulative departure from the mean of some $n < N$ points. The dimensionless ration $(R/s)_n$ is given by (2.1).

$$(R/S)_n = \frac{1}{S}\left[\max_n \sum_{i=1}^{n}(x_i - \langle x \rangle) - \min_n \sum_{i=1}^{n}(x_i - \langle x \rangle)\right] \tag{2.1}$$

where

$$\langle x \rangle = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \text{and} \qquad S = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \langle x \rangle)^2} \tag{2.2}$$

From (2.1), as $n \longrightarrow \infty$, $(R/S)_n \longrightarrow Cn^H$ where C is a constant and H is the Hurst Exponent. Thus, from this power law relationship,

$$\ln(R/S)_n = \ln(C) + H\ln(n). \tag{2.3}$$

Given (2.3), a slope of the simple regression line of $\ln(R/S)_n$ against $\ln(n)$ will give the Hurst Exponent H or the degree of correlation. Various values of H corresponds to the following characteristics wind speeds (Cadenas et al., 2019).

1. If $H = 0.5$, then wind speed is random or uncorrelated where future data is not determined by current data. This series is called a Brownian time series or a random walk. This series which display no memory is considered to have 'white noise'.

2. If $0.0 < H < 0.5$, then anti-persistence or mean reverting series: the wind speeds have long term negative auto-correlation in adjacent pairs. That is a long term switching between high and lows among adjacent pairs in the series for a long time into the future. A high will be followed by a low and then a high etc. Thus we will have a more rugged or less smooth time series. This occurs because the future values have a tendency to return to the long-term mean.

The time series is considered to have 'pink noise' which is related to turbulence.

3. If $0.5 < H < 1.0$, then persistence: the wind speeds have long term positive auto-correlation in adjacent pairs. That is a high value in the series will be followed by another high value for a long time into the future.
4. If $H \approx 1$ or $H = 1$, then there is strong predictability wind speeds or the wind speeds are predictable.

From (Mandelbrot, 1985)'s box argument, it is given that the local fractal dimension for self-affine records, D, is $D = 2 - H$.

2.4.2 Multifractal Analysis: Multifractal Detrended Fluctuation Analysis (MF-DFA)

The MF-DFA method was used to study turbulent signals. This procedure was applied to resistor network model, DNA sequences, satellite and microscopic images, financial time series including stock price fluctuation, traffic time series, quantum dynamical theory, weather records, cloud structure, geology and music among others (Salat, Murcio, and Arcaute, 2017) (Kantelhardt et al., 2002) (Shang, Lu, and Kamae, 2008) (Yuan, Zhuang, and Jin, 2009). The four principle methodologies relating fractal theory to measures are the moment method, the histogram method, the multifractal detrended fluctuation analysis method and wavelet transform modulus maxima method (Salat, Murcio, and Arcaute, 2017). These analyses are done when the fractal dimension changes with scale and when the time series is non-stationary. There may be multiple scaling exponents which represents different fractal subsets of the series (Kantelhardt et al., 2002). Unlike the R/S Analysis method, the MF-DFA method can detect non-spurious long-range correlations of a time series when there is non-stationary trends superimposed on it (Li et al., 2015)(Yuan, Zhuang, and Jin, 2009) (Movahed et al., 2006). The scaling of these intrinsic fluctuation of the time series can be determined despite knowing the origin and the shape of the trends present (Movahed et al., 2006). This is especially important for this study as the time resolution of the wind speed data sets is 10 minutes and the analysis is done for a time window of at most one year. Thus the annual trend cannot be estimated and removed from these datasets and as such the trend removing capabilities of MF-DFA is essential (Ludescher et al., 2011). Also, when compared to other multifractal methodologies, the MF-DFA method is less sensitive to the length of the time series and it gives more reliable results using a sample of over 4000 data points (Baranowski et al., 2015).

In this paper the MF-DFA is done. To explain the general idea, consider a non-stationary time series of length N, $x(i), i = 1, 2, 3, \ldots, N$ with compact support (i.e. x(i) = 0 for an insignificant fraction of the series) (Kantelhardt et al., 2002). The trajectory or profile preserves the variability of the time series whilst simultaneously reducing the noise by removing the non-stationary effects (Feng et al., 2009). This profile is given by $Y(i)$, (2.4) (Kantelhardt et al., 2002).

$$Y(i) = \sum_{k=1}^{i} \Big[x(k) - \bar{x} \Big] \tag{2.4}$$

This trajectory is partitioned into N_s non-overlapping intervals of equal length, s, that is, $N_s = \lceil N/s \rceil$. However, N need not be divisible by s thus part of the series may be unaccounted for as the possibility exist that $\lceil N/s \rceil < (N/s)$. To rectify this, a subdivision is done on the right hand side of the sample. This gives a total of $2N$ partitions or intervals (Kantelhardt et al., 2002). The local trend is determined by using a polynomial of degree m to fit the trajectories in each of its partitions . The variance is calculated from (2.5) for the two sets of partitions (Kantelhardt et al., 2002).

$$F^2(s,v) = \begin{cases} \frac{1}{s}\sum_{i=1}^{s}\Big[Y[(v-1)s+i]-y_v(i)\Big]^2 & \text{for } v=1,2,3,\ldots,N_s \\ \frac{1}{s}\sum_{i=1}^{s}\Big[Y[(N-(v-N_s)s+i)-y_v(i)\Big]^2 & \text{for } v=N_{s+1},\ldots,2N_s \end{cases} \tag{2.5}$$

where $y_v(i)$ is the fitting polynomial for that partition. Finally, the q^{th} order fluctuation, $F_q(s)$, is calculated from the average of all the partitions (Kantelhardt et al., 2002). Please see (2.6).

$$F_q(s) = \Big[\frac{1}{2N_s}\sum_{v=1}^{2N_s}\Big[F^2(s,v)\Big]^{\frac{q}{2}}\Big]^{\frac{1}{q}} \tag{2.6}$$

where $q \neq 0$ and $s \geq m+2$, m is the degree of the fitting polynomial. The scale was chosen to be 10: 100 and m was chosen as 1. Thus the inequality for which $F_q(s)$ was defined, holds.

The multifractality of the time series is caused by different long term correlations in the sample. MF-DFA can be used to determine multiple scaling exponents and spectrum parameters to classify the complexity and dynamics of the time series unlike monofractal analysis which characterizes the scaling property by one exponent for the entire data set. The four multi-fractal analyses done in this paper are as follows:

1. $\log(F_q(s))$ against $\log(s)$ where s is the scale and F_q is the q^{th} order fluctuation average. If q is negative then small fluctuations are enhanced and if q is positive then it enhances large fluctuations. To determine if long term correlation exist in the signal there should be a power law variation where F_q increases as a power of s. Thus, the generalized Hurst Exponent is the slope of this log-log plot as $F_q \approx s^{h_q}$ (Kantelhardt et al., 2002) (Fortuna, Nunnari, and Guariso, 2014) (Kavasseri and Nagarajan, 2005) and there is a linear relation in log plot for the various q values.

2. h_q against q or the dependence of the general Hurst Exponent on q. For monofractal time series h_q is independent of q (Fortuna, Nunnari, and Guariso, 2014). The local trend of each segment is calculated from the least square fit of the series and the variance of each segment. Since the scaling does not change, the trend over each segment is the same. For multifractal time series h_q is dependent on q. This dependence of h on q is caused by the fluctuations of scales both large and small. For large positive q values, there will be larger deviations from the least square fit thus larger variances $F^2(s, v)$ (Fortuna, Nunnari, and Guariso, 2014). These large variances are also reflected in the q^{th} order fluctuation and as such there is a relation between the large fluctuations and the Hurst Exponent, h_q. Large fluctuations for multifractal time series implies smaller h_q values (Kantelhardt et al., 2002). Similarly, for negative values of q, there are smaller variances and small fluctuations are characterized by larger scaling exponents h_q (Kantelhardt et al., 2002). Thus we have for a monofractal data set there will be one exponent for all scales where as for a multifractal time series, h_q monotonically decreases with increasing q.

3. τ_q against q or the q^{th} order mass exponent. τ_q is called Rényi Exponent. If τ_q varies linearly with q, then the time series is monofractal whilst the signal is multifractal if it has non-linear variations with q (Kavasseri and Nagarajan, 2005) (de Figueirêdo et al., 2014) (Kantelhardt et al., 2002). The relationship between this exponent and the Hurst Exponent is $\tau_q = qh_q - 1$. This relationship between the two multifractal scaling exponents was proved in (Kantelhardt et al., 2002) by considering a stationary positive and normalized sequence, substituting its simplified version of the variance, standard fluctuation analysis into (2.6) and comparing it with the box probability for the standard multifractal formalism for the normalised series.

4. $f(\alpha)$ against α or the multifractal spectrum. If the signal is a single scale fractal series then $f(\alpha)$ is a constant. A bell-like shape is given if the signal displays multifractal tendencies (Fortuna, Nunnari, and Guariso, 2014). This function is related to Rényi Exponent by the relation $f(\alpha) = q\alpha - \tau_q$ where α is Hölder Exponent and $\alpha = \frac{d\tau_q}{dq}$ (Kantelhardt et al., 2002). Since, $\tau_q = qh_q - 1$, $\alpha = h_q + q\frac{dh}{dq}$ and $f(\alpha) = q[\alpha - h_q] + 1$. Some multifractal spectrum parameters include position of max α_0, width of spectrum W and skew parameter r. The width of the spectrum is given by $W = \alpha_{max} - \alpha_{min}$ and the skew parameter is $r = \frac{\alpha_{max} - \alpha_0}{\alpha_0 - \alpha_{min}}$ (de Figueirêdo et al., 2014). The width of the spectrum determines the degree of the multifractality of the signal where a larger spectrum width coincides with greater dynamics of the data set and stronger multifractality (de Figueirêdo et al., 2014) (Laib et al., 2018). The skewness parameter is classified as $r = 0$ for symmetry, $r > 1$ for a right skewed spectrum and $r < 1$ for a left skewed spectrum. The dominant fractal exponent describing the scaling of small or large fluctuations is also determined by r. For a right skew spectrum, the fractal exponent describes the scaling of small fluctuations whilst large fluctuations are described by a left skew spectrum (de Figueirêdo et al., 2014). The more complex and multifractal signals are signals where α_0 and W are large values as well as $r > 1$ or is right skewed (de Figueirêdo et al., 2014).

2.5 Analysis of Results

2.5.1 Raw Data

The monthly mean wind speeds for the various channels in Columbia, Blanchard and Neosho were plotted in Figure 2.4. Average max wind speeds were recorded and determined for January to December of 2009 in Columbia and for January to August and January to October in Blanchard and Neosho respectively. From the plot, we see a similarity in terms of the wind speed patterns for all three stations. We see that the months of January to April and October to December are peak months whilst there is a decrease in average wind speeds during the period of May to September. From the average wind speeds in Columbia and Neosho, it is evident that the maximum to minimum wind speeds for each month coincided with the highest to lowest height levels, Columbia147 to Columbia68 and Neosho90 to Neosho50. For Blanchard, this holds true with the exception of intermediate height time series, Blanchard97, which had the lowest monthly averages of all the stations.

It was observed that the maximum wind speeds of all the stations for all the months came from the greatest tall tower heights of Blanchard137 and Columbia147. Also, with the exception of Blanchard97, the lowest average wind speeds came from the Columbia and Neosho stations at the lowest heights of 68 m and 50 m respectively.

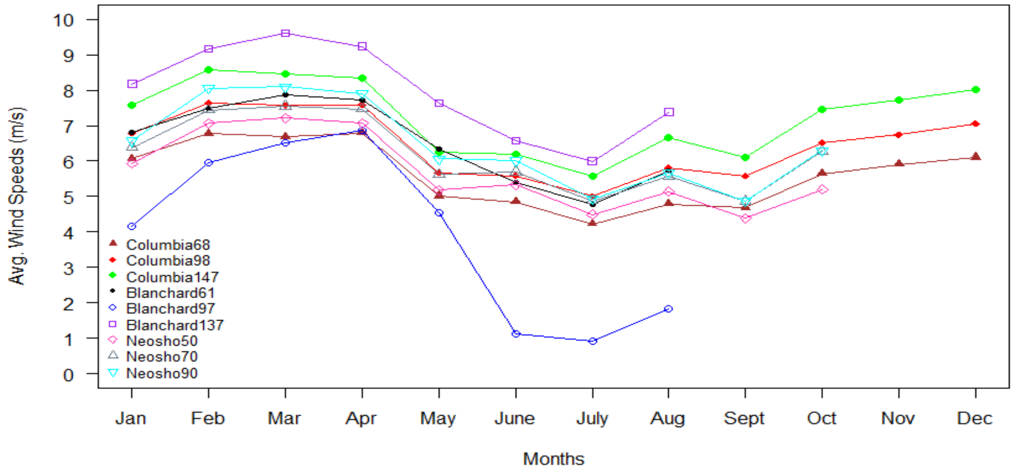


Figure 2.4: Average Max Wind Speeds in Columbia, Blanchard and Neosho in 2009

2.5.2 Monofractal Analysis

Figures 2.5, 2.6 and 2.7 show the monofractal Hurst Exponents for Columbia, Blanchard and Neosho respectively in 2009. From the results obtained there is no clear distinction in the Hurst Exponent values from the various series for all the stations thus indicating that the fractal dimensions of the wind speeds did not alter significantly with increasing heights. The fractal dimensions were consistently in the range of 1.1 to 1.3 for all the stations and months. This may have been as a result of similar variations of wind speeds with height. As such we expect the Hurst exponents and the fractal dimensions to be similar.

In Figure 2.7, it is observed that Neosho had the least monthly variations in the fractal dimensions for all of the heights when compared to the other two stations; its fractal dimensionality was determined to be 1.2 (to one decimal place). However of all the tall towers, this station gives the wind speeds taken over the smallest range of heights. It was determined that the numerically small variations in fractal dimensions of the other two stations, given by Figures 2.5 and 2.6 were not similarly changing with height and months when compared to Figure 2.4. For Columbia, the greatest fractal dimensions occurred in February and December at the lowest height of 68 m and in August at heights of 98 m and 147 m whilst the least fractal dimensions occurred in January for all height levels. Similarly, for Blanchard the fractal dimension of approximately 1.3 was observed for all heights in February. This was also noted in July and August with the exception of Blanchard61 and Blanchard97 respectively.

The Hurst Exponents were determined to be in the range of 0.7 to 0.9. R/S Analysis was used to show that the wind speeds investigated in this study does not follow a random Gaussian process but rather a long term autocorrelation. Since $0.5 < H < 1.0$, this implies that the wind speed had a long term positive autocorrelation in adjacent pairs where a high value will be followed by another high value for a long time into the future. That is, its fluctuations are interconnected because there exist a statistical order in the dynamics of the system (Salomao, Campanha, and Gupta, 2009). There will be fewer peaks than a random series and it will be less rugged than an anti-persistent system (Cadenas et al., 2019). This is consistent with a study done by Fortuna and Guariso (Fortuna, Nunnari, and Guariso, 2014) in which daily and monthly wind speed time series were analyzed from regions within the USA and Italy using two methods, Box Counting Method, D and the Hurst Exponent R/S Range Analysis Method,H. The wind speeds for these regions were determined to be fractal also because the average D values were 1.19 and 1.41 for daily and hourly mean wind speeds respectively. More complexity was discovered for hourly wind speeds than the daily wind speeds as indicated from its higher fractal dimensions. This is indicative of greater details and finer structures which the greater temporal resolution provides. This numerical value is in agreement with our study even though different locations and time scales were used.

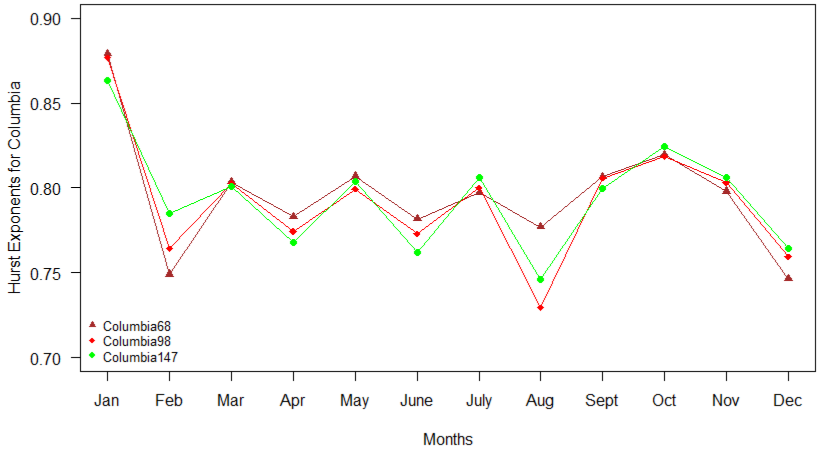


Figure 2.5: Hurst Exponents for Columbia in 2009 (dark red - Columbia68, red- Columbia98, green- Columbia147)

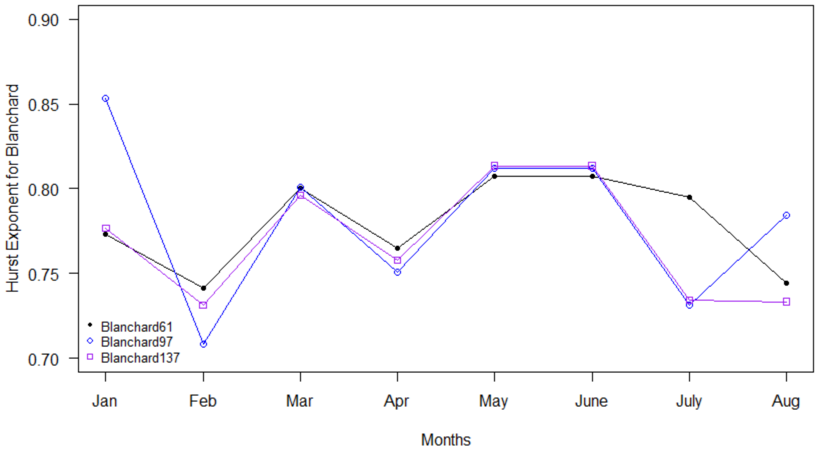


Figure 2.6: Hurst Exponents for Blanchard in 2009 (black- Blanchard61, blue- Blanchard97, purple- Blanchard137)

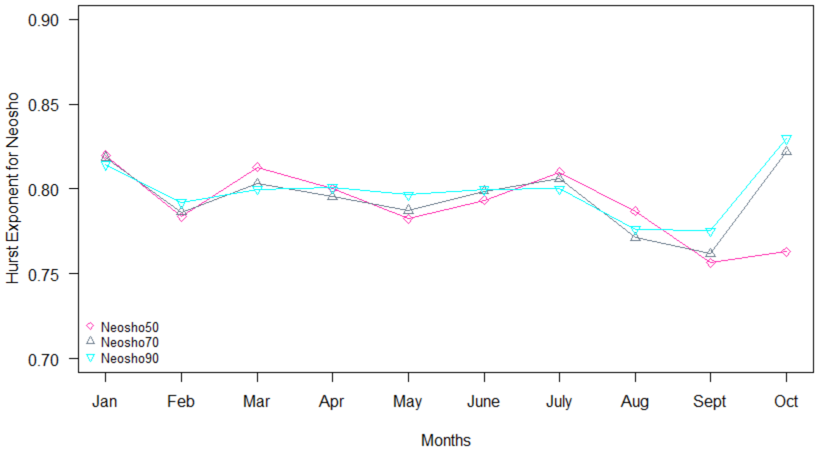


Figure 2.7: Hurst Exponents for Neosho in 2009 (pink-Neosho50, grey- Neosho70, cyan-Neosho90)

2.5.3 Multifractal Analysis

As in Figures 2.8, 2.9 and 2.10, it is evident that for all of the heights, there were increases in scaling function order, F_q as q values were increased from -5 to 0 to 5 for all of the three tall tower stations. We see that $\ln(F_q)$ varies linearly with $\ln(s)$ for a scale of 10 to 100 days with the generalized Hurst Exponent being the slope; this indicates a scale dependence which is characteristic of multifractality. Also, it is observed that as s increases, the distances among the different q values decreases. This occurs because for small segments (small s values), localized periods of small fluctuations given by negative q values, can be differentiated from periods of large fluctuations given by positive q values. This is unlike large segments (large s values) which includes both small and large fluctuations where the tendency for the magnitude differences to cancel occur (Philippopoulos et al., 2019). The hypothesis of the multifractal nature of wind speeds were also supported by the study of Fortuna and Guariso (Fortuna, Nunnari, and Guariso, 2014) for daily mean wind speeds recorded at Aberdeen from 2000 to 2012 in which the regression lines varied for differing q orders. Thus, the Hurst Exponents given by the slope of the plots were changing also for these sets. Similar results were also observed in another study in Northeastern Brazil, Petrolina for both hourly wind speed and max wind speed (de Figueirêdo et al., 2014). Thus we see that for temporal variations of data ranging from the ten minute to daily time series, all showed multi-fractal characteristics.

As seen in Figures 2.11, 2.12 and 2.13, dependence of the Generalize Hurst Exponent, it is seen that q increases as h_q decreases monotonically for all height levels. This is also noted in the slopes of Figures 2.8 to 2.10. Larger fluctuations corresponded with smaller h_q values and similarly, smaller fluctuations corresponded with larger scaling exponents. It is observed that for Columbia68, Columbia98 and Columbia147 in the month of September, when q varied from -5 to 5, h_q decreased from 1.7196 to 1.1730, 1.7171 to 1.2429 and 1.8055 to 1.3216 respectively. Since there is a range of values for the various scales for all height levels then these are indicative of multifractal series. This is also in agreement with a study done by Kavasseri and Nagarajan (Kavasseri and Nagarajan, 2005) for four sites with significant wind potentials in North Dakota where hourly data was taken from a cup anemometer at a height of

20 m. They determined that for one of the sites, when q increased from -6 to 6, the slope decreased from 0.88 to 0.6989.

The Generalized Hurst h_q is related to the Hurst Exponent, H, by $h(2) = H$ for stationary time series where $0 < h(q = 2) < 1$ (Movahed et al., 2006). For non-stationary time series, the scaling exponent of $F_q(s)$ is characterized by $h(q = 2) > 1$ and the relationship between H and h_q is given by $H = h(q = 2) - 1$. This is proved in (Movahed et al., 2006). For September, $h(2)$ values for Columbia68, Columbia98 and Columbia147 were determined to be 1.3536, 1.4290 and 1.4779 respectively. This indicates a non-stationary process with long range correlation behaviour (Shang, Lu, and Kamae, 2008). The corresponding Hurst exponents as well as fractal dimensions, D, for this month and stations at the three heights are 0.3536, 0.4290, 0.4779 and 1.6464, 1.571, 1.5221. These results gives higher fractal dimensions than the monofractal analysis for the time series data. It also showed that the wind speed time series are displaying long-term anti-persistence correlations as in a study done by (Zhang, Zeng, and Meng, 2018) in which the multifractality of multivariate wind speed for both indoor and outdoor records were examined.

From Figures 2.14, 2.15 and 2.16, it is noted that Rényi Exponents τ_q have non-linear variations with q. This is also characteristic of the wind speeds taken at the three heights of each tall tower, being multifractal signals. This is also noted from the hourly non-stationary time series MF-DFA by de Figueirêdo et al. (de Figueirêdo et al., 2014) between the years 2008 and 2011.

The last of the analyses is the multifractal spectrum, Figures 2.17, 2.18 and 2.19. The spectra of $f(\alpha)$ against α are not constant thus indicating that the series are not single scale fractal signals for all the months and height levels in Columbia, Blanchard and Neosho. The results obtained, the signals displayed multifractal tendencies by producing spectra of single-hump like features or bell-shapes with the exception of June C147, July C147, Aug C98 and Dec C68, C98, C147, Jan B97 and B137, Feb B97, Mar B97, Apr B61 and B97, July B137, Jan N50, N70 and N90, Mar N90, Sept N70 and N90, Oct N90. This may have been as a result of artifacts being contained in the observational data which makes the determination of the long-term correlations and multifractality of the records difficult. These artifacts may include additive random noise and short term correlations. Additive random noise can be derived from the limitations in the accuracy of the measuring instruments and short term correlations can be given from the short time scale of our study. The latter induces a strong persistence which is superimposed on the long-range correlations (Ludescher et al., 2011). These artifacts have been proven in (Ludescher et al., 2011) to cause various degrees of underestimation of h_q for small and negative moments which are most affected by noise. Ludescher et al. (Ludescher et al., 2011) also proved that that the multifractality of the positive moments may be corrupted. These graphical anomalies of h_q were noted in our results corresponding with the spectra which did not depict a bell-like shape. This is due to the fact that $f(\alpha)$ is obtained from Legendre transform which utilizes information on the moments (Ludescher et al., 2011).

The MF-DFA parameters of W and r were determined from the singularity spectrum as given by Table 2.1. The width of the spectrum is a measure of multifractality of the time series where a large width characterizes a finer signal structure which is more multifractal in nature. A width which tends to zero, however, is representative of a series that has one scaling exponent or one that is monofractal. From the results of this study, there was no indication of a consistent trend showing that the multifractality increases with increasing height from C69 to C147, B61 to B137 and N50 to N90. From the asymmetry parameter,r, for C68 to C147, some of the spectra are left skewed whilst others are right, also indicating that there is no trend of a dominant fractal exponent as the heights are increased. For Blanchard, predominantly, the dominant scaling is of large fluctuations as described by left skew parameter r. This is indicative of a the prevalence of a fractal exponent describing a structure that is less fine.

It was seen in Kavasseri and Nagarajan (Kavasseri and Nagarajan, 2005) that the spectrum widths for their data taken at height of 20 m were 0.4475 - 0.4862. In de Figueirêdo et al. (de Figueirêdo et al., 2014) the spectrum widths were 0.24 and 0.51 for average and maximum wind speed data taken from a meteorological station of altitude 370.46 m. From this study, the spectral widths for Columbia's single humped multifractal spectra, at tower heights of 68, 98 and 147 m and site elevation of 255 m were given by $0.33 \leq \mathrm{W} \leq 1.05$. This range is similar to the spectral width parameter values obtained by Laib et al. (Laib et al., 2018) for 119 stations in Switzerland using 10 minute time series data; W ranged between 0.206 and 1.15. For Blanchard and Neosho, the single hump widths ranged between 0.53 to 1.09 and 0.56 to 0.99 with the exception of Mar B137 and June N50 whose width values was 1.89 and 2.09 respectively. These differences in the widths from the three stations do not show as much variations as the study done by (Feng et al., 2009) in China using daily wind speed data. However, they represent the non-universal multifractal characteristic of wind speeds due to varying space and time dynamics. The parameters changes with location and heights levels and is as a result of different atmospheric circulation patterns. This is especially valid for wind speeds as seen from a climatic study of 31 years done by (Baranowski et al., 2015), using meteorological variables of precipitation, global radiation, wind speed, relative air humidity and air temperature, the greatest differences in the widths of the spectra were observed for wind speeds between Polish sites. The irregular fluctuations and complexity of the wind speeds is dependent on numerous factors which includes temperature, pressure gradient, turbulence and topography of the various sites (Zhang, Zeng, and Meng, 2018) .

Table 2.1: The Multifractal Spectrum Parameters for Columbia, Blanchard and Neosho for the three height levels and months in 2009

Description		Columbia, C					Blanchard, B		Neosho, N	
Month	Heights	α_0	α_{max}	α_{min}	w	r	w	r	w	r
Jan	C68 B61 N50	0.83	1.46	0.57	0.89	2.42	0.70	1.06	1.29	4.49
Jan	C98 B97 N70	0.85	1.48	0.53	0.95	1.97	1.37	0.18	1.31	4.54
Jan	C147 B137 N90	0.80	1.48	0.43	1.05	1.84	1.61	0.31	1.26	4.07
Feb	C68 B61 N50	1.36	1.57	0.93	0.64	0.49	0.61	0.56	0.57	0.43
Feb	C98 B97 N70	1.39	1.63	1.00	0.63	0.62	11.86	13.83	0.60	0.50
Feb	C147 B137 N90	1.40	1.58	1.06	0.52	0.53	0.64	0.45	0.74	0.72
Mar	C68 B61 N50	1.29	1.47	0.93	0.54	0.50	0.53	0.43	0.03	0.98
Mar	C98 B97 N70	1.31	1.48	0.92	0.56	0.44	14.78	0.05	0.56	0.40
Mar	C147 B137 N90	1.40	1.63	0.93	0.70	0.49	1.89	0.80	0.62	0.59
Apr	C68 B61 N50	1.37	1.57	1.17	0.4	1.00	2.85	3.45	0.7	0.75
Apr	C98 B97 N70	1.38	1.55	1.22	0.33	1.06	10.93	0.07	0.73	1.03
Apr	C147 B137 N90	1.43	1.61	1.22	0.39	0.86	0.62	0.72	0.81	1.31
May	C68 B61 N50	1.29	1.44	0.94	0.50	0.43	0.62	0.72	0.81	0.72
May	C98 B97 N70	1.38	1.80	0.95	0.85	0.98	0.72	0.60	0.99	0.90
May	C147 B137 N90	1.45	1.82	1.01	0.81	0.84	0.73	0.62	0.85	0.55
June	C68 B61 N50	1.29	1.68	0.76	0.92	0.74	1.04	0.65	0.96	0.68
June	C98 B97 N70	1.35	1.69	0.83	0.86	0.65	1.04	0.65	0.99	0.77
June	C147 B137 N90	7.79	8.69	0.86	7.83	0.13	1.09	0.68	2.09	1.38
July	C68 B61 N50	1.34	1.68	1.06	0.62	1.21	0.85	0.77	0.8	0.74
July	C98 B97 N70	1.40	1.75	1.10	0.65	1.17	0.85	0.77	0.66	0.61
July	C147 B137 N90	1.57	2.96	1.11	1.85	3.02	1.73	0.35	0.65	0.59
Aug	C68 B61 N50	1.30	1.46	0.97	0.49	0.48	0.75	0.56	0.81	0.56
Aug	C98 B97 N70	1.62	2.89	1.03	1.86	2.15	0.75	0.56	0.74	0.45
Aug	C147 B137 N90	1.54	1.89	1.09	0.8	0.78	0.76	0.46	0.8	0.74
Sept	C68 B61 N50	1.45	1.94	0.97	0.97	1.02			8.46	0.11
Sept	C98 B97 N70	1.53	1.92	1.04	0.88	0.80			7.34	0.12
Sept	C147 B137 N90	1.58	2.01	1.11	0.9	0.91			0.89	1.07
Oct	C68 B61 N50	1.29	1.70	0.92	0.78	1.11			6.08	0.13
Oct	C98 B97 N70	1.32	1.53	0.94	0.59	0.55			0.70	0.94
Oct	C147 B137 N90	1.34	1.52	0.91	0.61	0.42			0.69	0.77
Nov	C68 B61 N50	1.37	1.78	1.05	0.73	1.28				
Nov	C98 B97 N70	1.43	1.88	1.07	0.81	1.25				
Nov	C147 B137 N90	1.46	1.58	1.18	0.92	1.81				
Dec	C68 B61 N50	2.70	12.19	0.94	11.25	5.39				
Dec	C98 B97 N70	1.75	4.06	1.06	3.00	3.35				
Dec	C147 B137 N90	2.92	13.45	1.05	12.4	5.63				

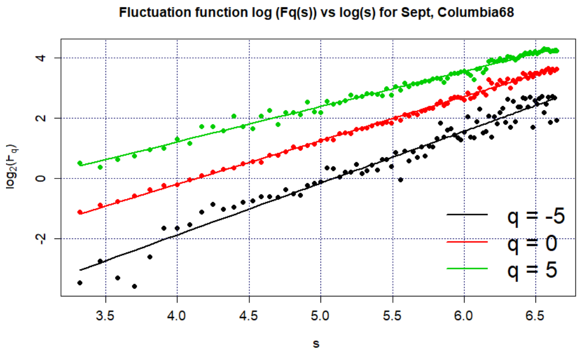


Figure 2.8: MF-DFA performed on 10 minute wind speed data in Columbia for tower height 68 m - Scaling function order F_q. Plot of $Log(F_q)$ against $log(s)$.

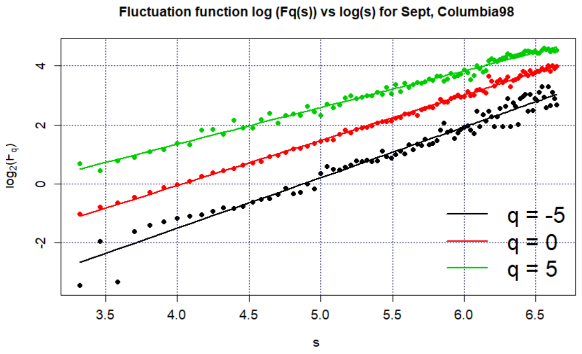


Figure 2.9: MF-DFA performed on 10 minute wind speed data in Columbia for tower height 98 m - Scaling function order F_q. Plot of $Log(F_q)$ against $log(s)$.

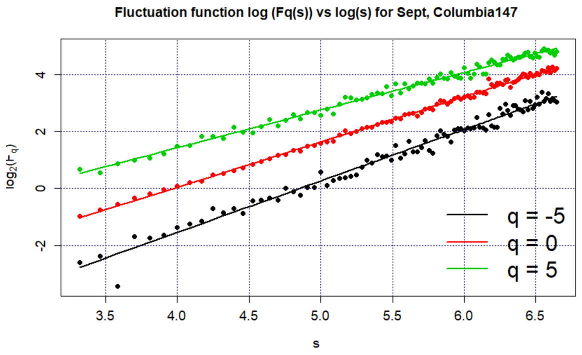


Figure 2.10: MF-DFA performed on 10 minute wind speed data in Columbia for tower height 147 m - Scaling function order F_q. Plot of $Log(F_q)$ against $log(s)$.

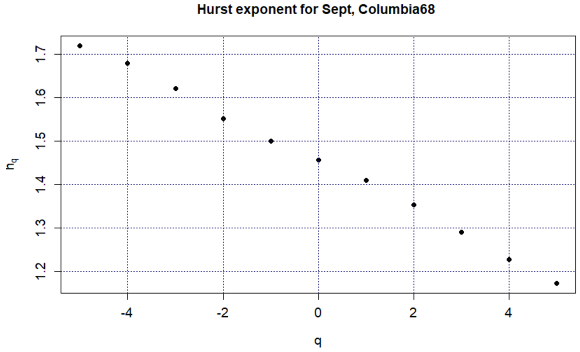


Figure 2.11: MF-DFA performed on 10 minute wind speed data in Columbia for tower height 68 m - Dependence of Gen Hurst Exp on q. Plot of h_q against q.

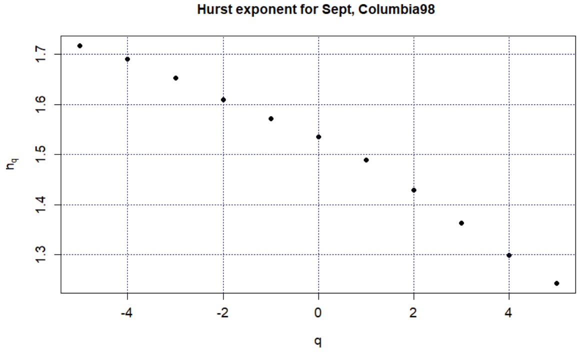


Figure 2.12: MF-DFA performed on 10 minute wind speed data in Columbia for tower height 98 m - Dependence of Gen Hurst Exp on q. Plot of h_q against q.

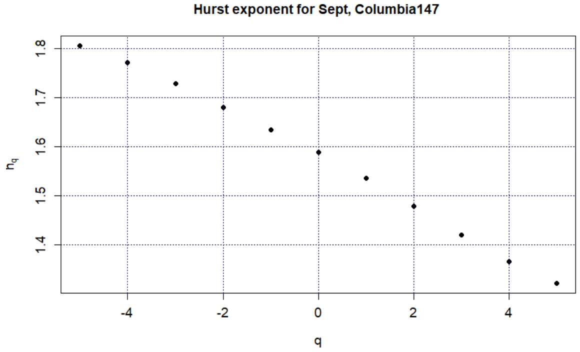


Figure 2.13: MF-DFA performed on 10 minute wind speed data in Columbia for tower height 147 m - Dependence of Gen Hurst Exp on q. Plot of h_q against q.

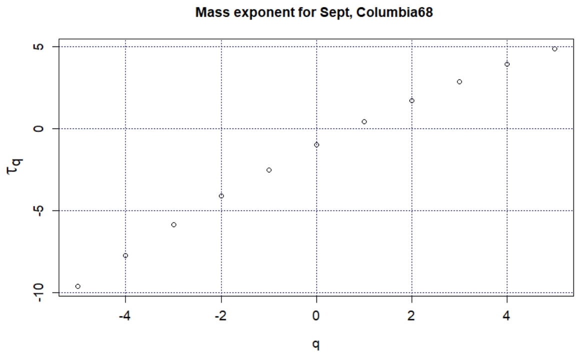


Figure 2.14: MF-DFA performed on 10 minute scaled wind speeds in Columbia for tower height 68 m - q-order Mass exponent. Plot of τ_q against q.

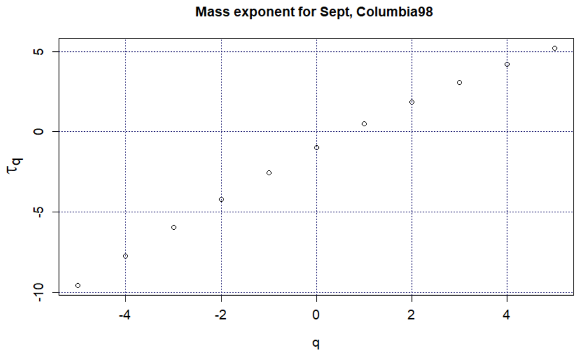


Figure 2.15: MF-DFA performed on 10 minute scaled wind speeds in Columbia for tower height 98 m - q-order Mass exponent. Plot of τ_q against q.

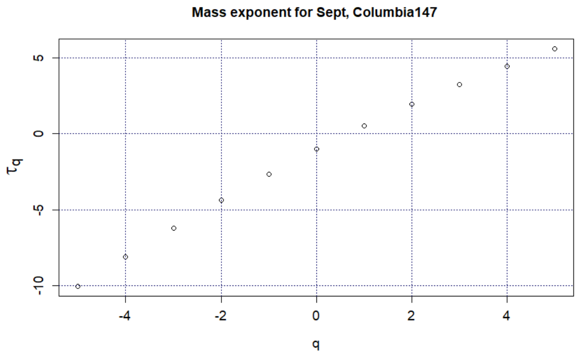


Figure 2.16: MF-DFA performed on 10 minute scaled wind speeds in Columbia for tower height 147 m - q-order Mass exponent. Plot of τ_q against q.

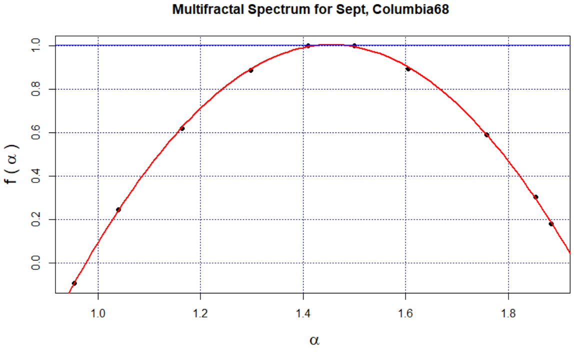


Figure 2.17: MF-DFA performed on 10 minute scaled wind speeds in Columbia for tower height 68 m - Multifractal Spectrum. Plot of $f(\alpha)$ against α

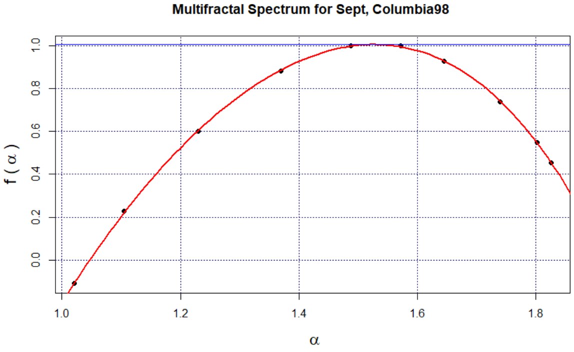


Figure 2.18: MF-DFA performed on 10 minute scaled wind speeds in Columbia for tower height 98 m - Multifractal Spectrum. Plot of $f(\alpha)$ against α

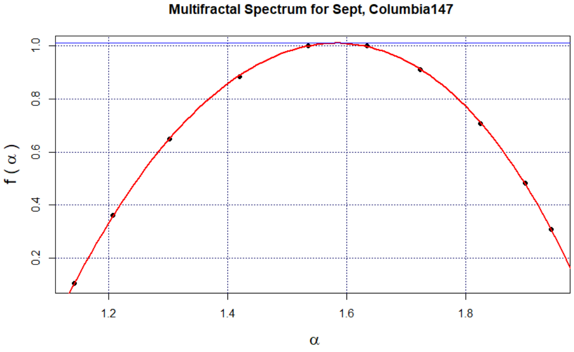


Figure 2.19: MF-DFA performed on 10 minute scaled wind speeds in Columbia for tower height 147 m - Multifractal Spectrum. Plot of $f(\alpha)$ against α

2.6 Conclusion

Winds speeds within Missouri, using monofractal analysis, were determined to be persistent as the Hurst exponent was greater than 0.5 for the three stations at the various height levels for all the months in 2009 using 10 minute data. There were no consistent increases in the fractal dimensions as the height levels were increased nor were they changes in the fractal dimension with months which corresponded with the average max wind speeds for the three stations.

From the MF-DFA, the wind speeds in Columbia, Blanchard and Neosho were determined to be multifractal in nature as there were changes in the fractal dimensions with scales. The fractal dimensions of the time series using multifractal analysis were determined to be greater than these values determined using monofractal analysis. However, the multifractality of the data sets, determined from the widths of the inverse parabolic shaped spectra, did not show any consistently increasing trend with heights which would have been indicative of greater complexity and finer structures of the wind speed records. The range of the widths which changed from various tower locations, reflects the non-universal multifractal characteristics of wind speeds and the different atmospheric circulation patters. From the asymmetry parameter there was no evidence of a dominating fractal exponent as the heights were increased.

Future work entails forecasting using Empirical Dynamical Modelling (EDM), having already established in this paper that the natural system of wind speeds from tall towers within Missouri are complex, dynamical and chaotic. In EDM, the wind time series will be used to construct the attractor from a mathematical theory developed by Takens. The dynamics of the system can be determined by a single time series as shown from Takens' embedding theorem. From the embedding theorem, each variable contains information about the other variables and thus, the wind speed time series and its time lagged co-ordinates will be utilized to study the system (C.-W. Chang, Ushio, and Hsieh, 2017).

References

Baranowski, Piotr et al. (2015). “Multifractal analysis of meteorological time series to assess climate impacts”. In: Climate Research 65, pp. 39–52.

Breslin, MC and JA Belward (1999). “Fractal dimensions for rainfall time series”. In: Mathematics and computers in simulation 48.4-6, pp. 437–446.

Cadenas, Erasmo et al. (2019). “Wind speed variability study based on the Hurst coefficient and fractal dimensional analysis”. In: Energy Science & Engineering 7.2, pp. 361–378.

Chang, Chun-Wei, Masayuki Ushio, and Chih-hao Hsieh (2017). “Empirical dynamic modeling for beginners”. In: Ecological Research 32.6, pp. 785–796.

Chang, Wen-Yeau (2014). “A literature review of wind forecasting methods”. In: Journal of Power and Energy Engineering 2.4.

de Figueirêdo, Bárbara Camboim Lopes et al. (2014). “Multifractal analysis of hourly wind speed records in Petrolina, Northeast Brazil”. In: Revista Brasileira de Biometria 32.4, pp. 599–608.

Division of Energy (2019). Wind Energy. URL: `https://energy.mo.gov/clean-energy/wind`.

Falconer, Kenneth (2004). Fractal geometry: mathematical foundations and applications. John Wiley & Sons.

Feder, Jens (2013). Fractals. Springer Science & Business Media.

Feng, Tao et al. (2009). “A brief description to different multi-fractal behaviors of daily wind speed records over China”. In: Physics Letters A 373.45, pp. 4134–4141.

Fortuna, Luigi, Silvia Nunnari, and Giorgio Guariso (2014). “Fractal order evidences in wind speed time series”. In:

ICFDA'14 International Conference on Fractional Differentiation and Its Applications 2014. IEEE, pp. 1–6.

Fox (2011). "A tall tower study of Missouri winds". In: Renewable Energy 36.1, pp. 330–337.

Kantelhardt, Jan W et al. (2002). "Multifractal detrended fluctuation analysis of nonstationary time series". In: Physica A: Statistical Mechanics and its Applications 316.1-4, pp. 87–114.

Kavasseri, Rajesh G and Radhakrishnan Nagarajan (2005). "A multifractal description of wind speed records". In: Chaos, Solitons & Fractals 24.1, pp. 165–173.

Laib, Mohamed et al. (2018). "Multifractal analysis of the time series of daily means of wind speed in complex regions". In: Chaos, Solitons & Fractals 109, pp. 118–127.

Li, Erhui et al. (2015). "Multifractal detrended fluctuation analysis of streamflow in the Yellow River Basin, China". In: Water 7.4, pp. 1670–1686.

Ludescher, Josef et al. (2011). "On spurious and corrupted multifractality: The effects of additive noise, short-term memory and periodic trends". In: Physica A: Statistical Mechanics and its Applications 390.13, pp. 2480–2490.

Mandelbrot (1982). "The fractal geometry of nature WH Freeman and Company New York". In: NY zbMATH.

— (1985). "Self-affine fractals and fractal dimension". In: Physica scripta 32.4, p. 257.

— (1989). "Fractal geometry: what is it, and what does it do?" In: Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences 423.1864, pp. 3–16.

— (2013). Fractals and chaos: the Mandelbrot set and beyond. Springer Science & Business Media.

Movahed, M Sadegh et al. (2006). "Multifractal detrended fluctuation analysis of sunspot time series". In: Journal of Statistical Mechanics: Theory and Experiment 2006.02, P02003.

Peebles, PJE (1989). "The fractal galaxy distribution". In: Physica D: Nonlinear Phenomena 38.1-3, pp. 273–278.

Philippopoulos, Kostas et al. (2019). "Multifractal Detrended Fluctuation Analysis of Temperature Reanalysis Data over Greece". In: Atmosphere 10.6, p. 336.

Ribeiro, Marcelo B and Alexandre Y Miguelote (1998). "Fractals and the Distribution of Galaxies". In: Brazilian journal of physics 28.2, pp. 132–160.

Salat, Hadrien, Roberto Murcio, and Elsa Arcaute (2017). "Multifractal methodology". In: Physica A: Statistical Mechanics and its Applications 473, pp. 467–487.

Salomao, LR, JR Campanha, and HM Gupta (2009). "Rescaled range analysis of pluviometric records in Sao Paulo State, Brazil". In: Theoretical and Applied Climatology 95.1-2, pp. 83–89.

Shang, Pengjian, Yongbo Lu, and Santi Kamae (2008). "Detecting long-range correlations of traffic time series with multifractal detrended fluctuation analysis". In: Chaos, Solitons & Fractals 36.1, pp. 82–90.

Technology, California Institute of (2015). Atomic fractals in metallic glasses. URL: `https://www.sciencedaily.com/releases/2015/09/150922120015.htm`.

Turcotte, Donald L (1997). Fractals and chaos in geology and geophysics. Cambridge university press.

U.S Energy Information Administration (2019). Missouri State Profile and Energy Estimates. URL: `https://www.eia.gov/state/analysis.php?sid=MO`.

Wind Energy Technologies Office (2019). Wind Energy in Missouri. URL: `https://windexchange.energy.gov/states/mo`.

Yuan, Ying, Xin-tian Zhuang, and Xiu Jin (2009). “Measuring multifractality of stock price fluctuation using multifractal detrended fluctuation analysis”. In: Physica A: Statistical Mechanics and its Applications 388.11, pp. 2189–2197.

Zhang, Xiaonei, Ming Zeng, and Qinghao Meng (2018). “Multivariate multifractal detrended fluctuation analysis of 3D wind field signals”. In: Physica A: Statistical Mechanics and its Applications 490, pp. 513–523.

Chapter 3

Determining chaotic characteristics and forecasting tall tower wind speeds in Missouri using Empirical Dynamical Modeling (EDM)

3.1 Abstract

The chaotic characteristics of the tall tower wind speed data within Missouri was investigated using both quantitative and qualitative methodologies. The phase space diagrams were constructed using the method of time delay. The two parameters needed in the construction of the attractor are the embedding dimension and the time delay. The former was determined using the Cao Algorithm and the latter by Average Mutual Information (AMI). Qualitatively, the phase portraits display chaos for all the wind speed time series for the various stations and height levels. They did not illustrate periodicity nor were they random motions, rather, they depicted a single attractor representative of chaos. Quantitatively the Largest Lyapunov Exponent (LLE) was evaluated. It was determined that for the Columbia station the wind

speeds display chaotic characteristics representative of the positive LLEs. However, the increasing level of chaos characteristics did not coincide with the increasing height levels of the tall tower. Thereafter, a simple non-linear prediction algorithm was used to forecast wind speeds using a moving window. The attractor was constructed using the first 56 days and the subsequent 6 hours or 36 (10 minute) time steps were predicted. The preceding forecast was done when the attractor was reconstructed using the training data of 56 days starting from a 6-hour delay from the previous run. The RMSE, MAE and Correlation were investigated for the model with the errors evaluated cumulatively for all of the 1^{st} through 36^{th} predictions. It was determined that the errors increase as the forecasting steps increased for all stations and height levels. The RMSE plateaus at higher wind speeds for increasing height levels with the exception of the station, Neosho, where it plateaued at all height levels at approximately $3.0\ \mathrm{m\,s^{-1}}$. For Columbia at all height levels, after the 20^{th} time step or 3.33 hours, the model's normalized errors exceeds 1 or 100%. However, using a 50% normalized error cap, it was noted that these values occurred for Columbia's height levels after the 1^{st}, 2^{nd} and 3^{rd} time steps respectively. For Blanchard, this value was given by the 2^{nd} time step for both heights whilst for Neosho, at all heights this percentage occurred after at most, 2 time steps. From the Lyapunov exponent, the prediction horizons or the time limits to obtain accurate predictions from the chaotic system were determined to be 6 time steps for all the height levels in the Columbia station using a 95% confidence band. When a range of confidence bands was used, it was shown that for the 90% confidence, this value was decreased to 4 time steps. This model was compared to the benchmark model of persistence where it was determined that the EDM is comparable to persistence and it beats it in the very short-term range of one time step for Columbia and Blanchard. Seasonality and diurnal cycle analyses were also accomplished. Seasonality was investigated by slicing the results every 6 hours or extracting every 36^{th} forecast error. It was shown that four of the

eight stations' height levels had the season of summer incurring the lowest magnitude of average errors and standard deviations. The diurnal cycle was examined by extracting every four of the 6 time slices done previously. The time of day was analysed by lagging these slices by 6, 12 and 18 hours. It was determined that there was no evident trend where a particular time of day the model incurred more errors and had greater standard deviations for all stations and heights.

Keywords: Phase Space, Largest Lyapunov Exponent, Empirical Dynamical Modeling

3.2 Introduction

The prediction of wind speeds can be categorised in either physical, statistical or hybrid methods. The physical methodologies are mathematical models that utilize large amounts of data from numerical weather prediction (NWP). The statistical methods, however, can be classified as either time series, spatial correlations or artificial intelligence procedures whilst hybrid methods combine two or more approaches (De Freitas, Silva, and Sakamoto, 2018). Alternative methodologies to conventional predictive methods include the nearest neighbour method of chaos theory, artificial intelligence's neural nets as well as wavelets (Szpiro, 1997; Lisi and Villi, 2001).

Chaos theory methodologies are employed in the short-term prediction of forecasting meteorological variables such as wind speed. Short-term wind speed prediction is one of the four temporal ranges of forecast. It is where forecasts are made from one hour to two days ahead and they are used for economic load dispatch planning (De Freitas, Silva, and Sakamoto, 2018). This short term prediction is especially important as the adjustments of power generation to a consistently changing load is required and wind is a variable power source. This will ultimately contribute to a steady power supply (Kawauchi, Sugihara, and Sasaki, 2004). These methodologies are sorted after when initial data, needed to build models from first principle is lacking (Farmer and Sidorowich, 1987). Thus the time series must be established as chaotic. That is, as defined in (De Domenico and Ghorbani, 2010) and (Lisi and Villi, 2001), it is a simple non-linear deterministic system, which is sensitively dependent on its initial conditions. Such systems display random and complex behaviours. Neural nets use historical data and may use a back propagation technique to adjust its weights. A forecast function is then used to predict future variables using past time series inputs (Szpiro, 1997). Finally, wavelets decompose the time series into its various components and then forecast using a weighted sum of these wavelets (Szpiro, 1997). These methods have also been combined in complex dynamics forecasting (Szpiro, 1997).

Natural systems, however, are complex and dynamical, often involving many variables that cannot be measured with sufficient accuracy. If statistical time series models are utilized, these difficulties necessitate the use of non-linear approaches (Chang, Ushio, and Hsieh, 2017) as classical regression analysis cannot fully represent the underlying complex dynamics of especially a chaotic series (Szpiro, 1997). A property of non-linear systems is state dependency, which can be defined as the changing relationships among interacting variables with different states associated with the dynamical system (Chang, Ushio, and Hsieh, 2017). The non-linear statistical methods originates from state space reconstruction which is the lagged co-ordinate embedding of the time series. These methods recover the dynamics of the time series instead of using a set of governing equations. These methodologies are called empirical dynamical modeling (Chang, Ushio, and Hsieh, 2017).

In a study by (Yu et al., 2019), many chaos identification methods are used in the determination of a chaotic series. This is done because wind speed has complex characteristics of deterministic and random signals. There are chaos identification methods that are based on phase space reconstruction, which include the phase diagram, correlation dimension and the largest Lyapunov exponent (Yu et al., 2019). The phase space method is a qualitative direct chaos identifier method and the correlation dimension and the Lyapunov exponent methodologies are quantitative and direct procedures (Zeng et al., 2012). If an attractor is non-periodic (the motion of the system never repeats), finite dimensional and generated by deterministic dynamics, then it is a strange or chaotic attractor. Two points on this attractor at a time will be arbitrarily apart from each other at a later time, that is, nearby points in phase space separate at an exponential rate. This is given by the positive largest Lyapunov exponent (Brandstäter et al., 1983). In this paper, some of these methods will be employed to identify chaos before a non-linear prediction method, referred to as "Lorenz's method of analogues" (Kantz and Schreiber, 2004), is conducted.

The subsequent section, section 2, describes the data utilized in this study. Section 3 outlines the various methodologies used, which are the generation of the wind speed duration curve, the reconstruction of phase space and the determination of the parameters of the embedding dimension and the time delay, the determination of the largest Lyapunov exponent, the forecasting of wind speeds using a non-linear prediction algorithm and the examination of the seasonality and diurnal cycle of the model runs. Section 4 provides the results of the various methods conducted before concluding in section 5.

3.3 Data

This study uses ten-minute interval daily wind speed time series data recorded in 2009, in $\mathrm{m\,s^{-1}}$, for Columbia, Blanchard and Neosho, stations in Missouri (Fox, 2011). Columbia is located in $038°53.270'$N latitude and $092°15.820'$W longitude and has a site elevation of 255 m. Blanchard is located $040°33.570'$N latitude and $095°13.470'$W longitude and has a site elevation of 328 m. Neosho is located 036°52.730′N latitude and 094°25.570′W longitude with a site elevation of 373 m. Please refer to Figure 3.1. The anemometers were placed on various heights and orientations on the towers. For Columbia, Blanchard and Neosho, the anemometer orientations were 120°and 300°for each of the various sites' tall tower heights of 68, 98, 147 m and 61, 97, 137 m and 50, 70 , 90 m respectively. Channels 1, 3 and 5 are the wind speed times series of the three consecutive heights at an orientation of 120°and Channels 2, 4 and 6 are wind speed values obtained when anemometers were oriented at 300°. The larger of the wind speed value at each time step for each height level were taken and labelled as Columbia68,Columbia98 and Columbia147, Blanchard61, Blanchard97 and Blanchard137, Neosho50, Neosho70 and Neosho90.

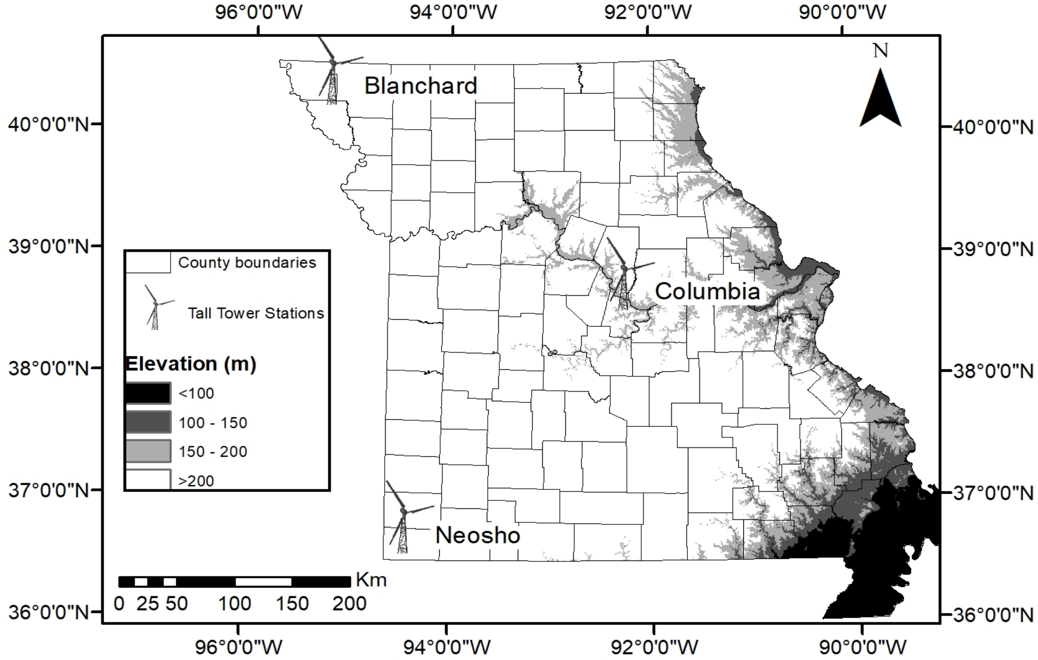


Figure 3.1: Study Locations within Missouri.

3.4 Methods

3.4.1 Wind Speed Duration Curves

Wind speed duration curves (WSDC) provide a simple way to visualize the distribution of wind speeds recorded at the observation sites (Masseran et al., 2012). They are essentially the graph of the wind speed's cumulative distribution function over the time interval. By convention, the independent variable is along the vertical axis and the horizontal axis is scaled to be a percentage.

More precisely, suppose that $\{x_i\}_{i=1}^n$ is a time series of wind speeds and let $\{z_i\}_{i=1}^n$ be the rearrangement of these values so that they are in ascending order: $z_1 < z_2 < \cdots < z_n$. Here it is assumed for simplicity that the readings are distinct and the observation times are evenly spaced. Then the percentage of time in which the wind speed is greater than or equal to z_i can be approximated by Equation 3.1.

$$F(z_i) = 100\left(\frac{n+1-i}{n+1}\right) \tag{3.1}$$

The corresponding WSDC is constructed by plotting the points $(F(z_i), z_i)$ for $i = 1, \ldots, n$.

3.4.2 Takens' Theorem and Reconstruction of the Phase Space

Assume that the weather patterns at the observation sites are governed by a deterministic dynamical system set on a smooth D-dimensional compact manifold M (the phase space). Let $\varphi_t : M \to M$ denote the corresponding flow map, so that, if the system is initially in state $m \in M$, then $\varphi_t(m)$ is its state after time t.

Weather forecasting is thus equivalent to determining φ_t. However, it is typically difficult to do this directly as the dynamical system is hard to formulate explicitly and it is likely to be non-linear, high dimensional, and chaotic. Accurately measuring the many variables involved may also prove impractical.

In the seminal paper (Takens, 1981), Takens gives a number of mathematical results addressing this issue. Suppose that $X : M \to \mathbb{R}$ is a scalar-valued observable quantity; for the present study, this will be the wind speed. Under mild assumptions on the regularity of X, Takens shows that it is possible to reconstruct φ_t from the time evolution of the observable $X(\varphi_t)$. Specifically, by (Takens, 1981, Theorem 2), for a generic time delay $\tau \geq 0$ and time t, the mapping

$$\Phi_t : m \mapsto (X(\varphi_t),\, X(\varphi_{t+\tau}),\, \ldots,\, X(\varphi_{t+2D\tau}))\,|_m \tag{3.2}$$

constitutes a smooth embedding of M into $\mathbb{R}^{2D+1}$. Moreover, in (Takens, 1981, Corollary 5) it is established that, for generic $m \in M$, the ω-limit sets of $\Phi_t(m)$ and $\varphi_t(m)$ are diffeomorphic. Stated more plainly: if the full system has an attractor, the topological and differential structure of that attractor can be inferred from the longtime behavior of the time-delay mapping Φ_t.

Suppose now that $\{x_i\}_{i=1}^n$ is a time series representing readings of the observable $X(\varphi_t)$ at equally spaced times $t_1 < t_2 < \cdots < t_n$. For candidate embedding dimension d and time delay τ, in analogy to (3.2), one forms the Takens reconstruction vector

$$y_i = \big(x_i,\, x_{i+\tau},\, \ldots,\, x_{i+(d-1)\tau}\big) \tag{3.3}$$

with index $i = 1, 2, \ldots, N$ and $N = N(\tau, d) = n - (d-1)\tau$. Note that in what follows the dependence of N on τ and d is often suppressed for the sake of simplifying the notation. Provided that $d \geq 2D + 1$, Takens' embedding theory can then be used to predict the dynamics of the original system from the delayed time series $\{y_i\}_{i=1}^N$.

This leads to the statistical problem of choosing the parameters τ and d, which has been the subject of considerable research. Even though Takens' theorem holds for a generic time delay, it is well known that some care is required in selecting τ when working with experimental data. If the time delay window is too narrow, then the attractor will be projected into a dimension that is too small. Were this to occur, then x_{i+1} will not contain substantially new information relative to x_i, making predictions based on the time series unreliable (Stark, 1994). Conversely, if the window is too large, the components of y_i become noisy (Camplani and Cannas, 2009; Kennel, Brown, and Abarbanel, 1992) making accurate forecasting difficult (Stark, 1994). The next two subsections outline the approach to these questions taken in this work.

Estimate of the Embedding Dimension via Cao's Algorithm

Common technique for selecting the embedding dimension include the method of False Nearest Neighbours (FNN) (Camplani and Cannas, 2009; Kliková and Raidl, 2011) and Cao's Algorithm (Cao, 1997). FNN, which is used in (Kennel, Brown, and Abarbanel, 1992), for example, involves tracking the number of points in the reconstructed time series $\{y_i\}$ that appear to be nearby only because the embedding space is too small. The embedding dimension is taken to be the minimal d so that the number of these "false neighbours" is zero.

In this paper, Cao's Algorithm was used to determine the embedding dimension. Compared to FNN, it has the advantage of being less sensitive to the number of points in the time series, and it does not introduce any subjective parameters (Cao, 1997). Moreover, it can differentiate between deterministic (no uncertainty with respect to time) and stochastic (random signals that cannot be written in mathematical equation) time series (Cao, 1997).

The method can be described as follows. Given a time series $\{x_i\}_{i=1}^{n}$, let $\{y_i^d\}_{i=1}^{N(\tau,d)}$ and $\{y_i^{d+1}\}_{i=1}^{N(\tau,d+1)}$ be the reconstructed time series given by equation (3.3) for the candidate embedding dimensions d and $d+1$, respectively. For each $1 \leq i \leq N-d\tau$, let $y_{n(i,d)}^d$ be the nearest neighbor to y_i^d (that is distinct from y_i^d) with distance measured in the max norm on $\mathbb{R}^d$, and define $n(i, d+1)$ accordingly. As in (Cao, 1997), we set

$$a(i,d) = \frac{\|y_i^{d+1} - y_{n(i,d)}^{d+1}\|}{\|y_i^d - y_{n(i,d)}^d\|} \tag{3.4}$$

where $\|\cdot\|$ is the max norm, and consider the average of this quantity over i:

$$E(d) = \frac{1}{N - d\tau} \sum_{i=1}^{N-d\tau} a(i,d). \tag{3.5}$$

Two points that are close in d-dimensional reconstructed phase space and also close in $(d+1)$-dimensional reconstructed phase space are called true neighbours (Cao, 1997). This is measured by means of the function

$$E1(d) = \frac{E(d+1)}{E(d)}. \tag{3.6}$$

Observe that $E1(d)$ will be constant when $d \geq 2D + 1$, and hence it can be used to identify the minimum embedding dimension (Cao, 1997).

A second consideration that is important for applications is whether the data is stochastic or deterministic. For this, Cao introduces the function

$$E2(d) = \frac{E^*(d+1)}{E^*(d)} \tag{3.7}$$

where

$$E^*(d) = \frac{1}{N - d\tau} \sum_{i=1}^{N-d\tau} |x_{i+d\tau} - x_{n(i,d)+d\tau}|.$$

If the data are deterministic then $E2(d)$ is dependent on d and so one will observe that $E2(d) \neq 1$ for some choice of d. On the other hand, if the data is random, then it must hold that $E2(d) = 1$ for all d. Computing both $E1(d)$ and $E2(d)$ allows one to both find a reasonable embedding dimension and also offers evidence that the underlying process is indeed deterministic (Cao, 1997).

Time Delay Estimation via Average Mutual Information

An appropriate time delay can be determined through the use of an Autocorrelation Function (ACF) or considering the Auto Mutual Information (AMI) (Camplani and Cannas, 2009). This paper uses the latter technique, which was first introduced by Fraser and Swinney (Fraser and Swinney, 1986). This method was used because of its ability to measure the general rather than linear dependence of two variables (Fraser and Swinney, 1986). It is also a common methodology used in the construction of the phase space of the attractor (Zounemat-Kermani and Kisi, 2015).

Let $\mathcal{A}$ and $\mathcal{B}$ be discrete sets. For a random vector (A, B) with state space $\mathcal{A} \times \mathcal{B}$ and discrete joint probability distribution P_{AB}, the mutual information of the random variables A and B is defined by

$$I_{AB} = \sum_{a \in \mathcal{A}} \sum_{b \in \mathcal{B}} P_{AB}(a, b) \log_2 \left(\frac{P_{AB}(a, b)}{P_A(a) P_B(b)} \right), \tag{3.8}$$

where P_A and P_B are the marginal probabilities for A and B, respectively. Intuitively, I_{AB} measures on average how accurately one can determine B given knowledge of A (or vice versa). The idea of AMI is to treat $X(\varphi_t)$ and $X(\varphi_{t+\tau})$ as random variables, then adjust τ so as to maximize their mutual information.

With that in mind, suppose $\{x_i\}_{i=1}^n$ is a (portion of a) time series obtained from experimental data. For a candidate time delay τ, we form the shifted series $\{x_{i+\tau}\}_{i=1}^n$ and introduce the sets $\mathcal{A} = \{x_i : i = 1, \ldots, n\}$ and $\mathcal{B} = \{x_{i+\tau} : i = 1, \ldots, n\}$. Let (A, B) be the random vector taking values on $\mathcal{A} \times \mathcal{B}$ whose probability distribution P_{AB} is generated by a histogram of $\{(x_i, x_{i+\tau})\}_{i=1}^n$ with an appropriate bin size. The mutual information of the components A and B as a function of τ is thus

$$I(\tau) = \sum_{i,j=1}^{n} P_{AB}(x_i, x_{j+\tau}) \log_2 \left(\frac{P_{AB}(x_i, x_{j+\tau})}{P_A(x_i) P_B(x_{j+\tau})} \right). \tag{3.9}$$

We choose the value of τ to be the either the first local minimum of I or when the AMI is monotonically decreasing to the ratio of $\frac{I(\tau)}{I(0)} = 0.2$ or $\frac{1}{e}$ (Kliková and Raidl, 2011).

3.4.3 Largest Lyapunov Exponent (LLE)

The characterizing feature of chaotic systems is that future states are effectively unpredictable despite the underlying dynamics being deterministic. This phenomenon stems from the flow map φ_t exhibiting sensitive dependence on initial conditions (SDOIC) (Kantz and Schreiber, 2004). Lyapunov exponents provide one way to quantify this SDOIC (Rosenstein, Collins, and De Luca, 1993) and thereby gauge how chaotic a system is (Kantz and Schreiber, 2004). Trajectories that are initially close but lie on a chaotic attractor will diverge from one another exponentially fast; the Lyapnuov exponent λ gives the average rate of this divergence (Garcia and Sawitzki, 2015; Rosenstein, Collins, and De Luca, 1993). Note that it is essential here to take an average since the rate will potentially be different in different directions (Kantz and Schreiber, 2004).

As Lyapunov exponents are insensitive to the choice of metric, they can be computed through consideration of the Takens' flow. For a point $m \in M$ and time t, we define the Lyapunov exponents to be the eigenvalues of the linearized flow map there; the eigenvalue with largest real part is called the largest Lyapunov exponent (LLE) and denoted λ. Note that the dynamics of the attractor — should one exist — is dissipative and as such the sum of the Lyapunov exponents is negative there. Conversely, in the applied literature, a positive Lyapunov exponent is often taken as sufficient evidence that the system is chaotic (Rosenstein, Collins, and De Luca, 1993; De Domenico and Ghorbani, 2010; Bryant, Brown, and Abarbanel, 1990). By contrast, if $\lambda < 0$, then there exists (asymptotically) stable fixed points to which one will find trajectories converging exponentially. If one has trajectories that approach or separate from each other slower than an exponential rate, there must exists a limit cycle that is marginally stable; in this case, the LLE is equal to zero (Kantz and Schreiber, 2004). Finally, for random noise, the LLE can be thought of as infinite (Kantz and Schreiber, 2004).

Let $\Delta(t)$ be the distance between two points in phase space for the embedded map at time t. Then

$$S(t) = \ln \frac{\Delta(t)}{\Delta(0)} \tag{3.10}$$

provides an approximate upper bound on λt. To determine $S(t)$ from the time series, we construct a Takens vector, say, y_{N_0} and construct a neighbourhood about this vector of ϵ distance, $B_\epsilon(y_{N_0})$. For each of the nearby Takens vectors, y_i, $i = 1, \ldots, N$, the average error between a forward time step, δT, of both a nearby Taken vector $(x_{i+\delta T})$ and the fixed Taken vector, y_{N_0} $(x_{N_0+\delta T})$ is determined. The logarithm of all the average distances are taken to give a measure of the exponential rate of expansion. The average of all N_0 values are used to smooth out any noise (Kantz and Schreiber, 2004). From Kantz and Schreiber (Kantz and Schreiber, 2004), $S(t)$ is given by

$$S(t) = \frac{1}{N} \sum_{N_0=1}^{N} \ln \left(\frac{1}{|B_\epsilon(y_{N_0})|} \sum_{\substack{i=1,\ldots N: \\ y_i \in B_\epsilon(y_{N_0})}} |x_{N_0+\delta T} - x_{i+\delta T}| \right). \tag{3.11}$$

The largest Lyapunov exponent is estimated by the the slope of the linear region of the curve of $S(t)$ against t (Garcia and Sawitzki, 2015). After the linear region, the curve saturates for large t values since the system is bounded in phase space Rosenstein, Collins, and De Luca, 1993.

This method has been proven by (Rosenstein, Collins, and De Luca, 1993) to work well with smaller time series as well as time series with white noise superimposed with noise-free data. This is especially important, as shown in (Eckmann and Ruelle, 1992), where the length of a series and quality are factors in the accuracy of the extraction of dynamical information.

If the system is chaotic and $\lambda > 0$, then the series can be predicted with a prediction horizon given by

$$t^* = \frac{1}{\lambda} \ln 1.96 \tag{3.12}$$

where t^* is the maximum amount of samples with sampling time that can be predicted with uncertainty; see, for example, (De Domenico and Ghorbani, 2010). The uncertainty used is 1.96ϵ which represents a 95% confidence band. The confidence bands of 90% and 99% were also evaluated in this study. The forecasting error exponentially increases with the forecasting time at a rate given by the LLE (De Domenico and Ghorbani, 2010). Thus we see that even though chaos places a limit on long-term prediction, it affords the ability of short term prediction (Farmer and Sidorowich, 1987).

3.4.4 Forecasting using simple non-linear prediction algorithm

The future state of the system is a function of the present state at some time say, t. That is, there exists a deterministic forecasting function (Kantz and Schreiber, 2004). Since there is no certainty of our present state, inaccuracies grow exponentially over time in chaotic systems. However, the uncertainties increased over a finite rate, even for chaotic systems, and as such, short term forecasts can be made (Kantz and Schreiber, 2004).

Consider a scalar time series given by $\{x_i\}_{i=1}^{n}$. Fixing a time delay τ and embedding dimension d, we form the reconstruction vectors $\{y_i\}_{i=1}^{N}$ according to (3.3). To predict some time δt ahead, we consider $B_\epsilon(y_N)$, the d-dimension ball centered at y_N with radius ϵ. In this analysis, ϵ was chosen to be half the smallest possible reading of the anemometer, 0.05. For all the points y_i lying in this ball, we find their corresponding individual prediction values δt ahead, then average the result. The prediction is given by

$$x_{N+\delta t} = \frac{1}{|B_\epsilon(y_N)|} \sum_{\substack{i=1,\ldots N: \\ y_i \in B_\epsilon(y_N)}} x_{i+\delta t}, \tag{3.13}$$

which represents the average of these values (Kantz and Schreiber, 2004).

Thus we see that the feature of the attractor being a compact object in phase space, and thus, having neighbours is utilized in the prediction of the time evolution of new points on or nearby the attractor (Kennel, Brown, and Abarbanel, 1992).

3.4.5 Errors

In order to test the accuracy of the forecast, the root mean square error (RMSE) and mean absolute error (MAE) were computed. These are given by the following equations.

$$\text{RMSE} = \sqrt{\langle (p_n - x_n)^2 \rangle}, \qquad \text{MAE} = \langle |p_n - x_n| \rangle,$$

where p_n is the predicted observation whilst x_n is the actual observation at the time step n.

The normalized errors were also computed via

$$E = \frac{\sqrt{\langle (p_n - x_n)^2 \rangle}}{\sqrt{\langle (x_n - \bar{x})^2 \rangle}} = \frac{\text{RMSE}}{\text{SD}}. \tag{3.14}$$

Here $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ is the mean of the time series. If $E = 0$ or 0%, then the prediction is perfect. At the other extreme, if $E = 1$ or 100%, then the prediction is no better than the mean (Farmer and Sidorowich, 1987).

3.4.6 Analysis of seasonality and the diurnal cycle

The seasonality effects on the forecast were investigated. For Columbia, Blanchard and Neosho sites the results including the forecast error, found by subtracting the predicted value from the actual value, were sliced for every 36^{th} value. This represents every 6^{th} hour of data. Mathematically, this was given by $36n + 6$. For Columbia the sequence $n = 0, 1, 2, 3, \ldots, 1237$ accounted for its time series of length 44568. Thus, we had a slice of wind speed values, say x_i whose subscript values i range from $i = 6, 42, 78, \ldots, 44538$. For Blanchard, $n = 0, 1, 2, 3, \ldots, 738$ which ran through the series with 26604 data points. Also Neosho has values $n = 0, 1, 2, 3, \ldots, 991$ and a series length of 35712.

The diurnal effects of the forecast were also examined in this study. This was done by further slicing of every 4^{th} 6 hour time slice given above. This is represented as $4m + 1$. It can also be established by slicing of the original data set, $144m + 6$ where $m = 0, 1, 2, 3, \ldots, 309$ for Columbia, $m = 0, 1, 2, 3, \ldots, 184$ for Blanchard and $m = 0, 1, 2, 3, \ldots, 247$ for Neosho. Thus, for Columbia, the indices considered from the wind speed error were 6, 150, 294, 438, …, 44502 (the last value can be given by $144 \times 309 + 6$). For Blanchard and Neosho, the subscript sequence is the same with the exception of the upper limit being 26502 for Blanchard and 35574 for Neosho.

This $144m + 6$ slice of the data was then lagged 6, 12 and 18 hours using the following algorithms $36+[144m+6]$,$72+[144m+6]$ and $108[144m+6]$. This was done to investigate the relationship between the time of day and the errors in the forecast for all height levels and stations within the forecast run of 2009. For Columbia station, for the 6 hour (36 10-minute), 12 hour (72 10-minute) and 18 hour (108 10-minute) shifts, the sequences are given by 42, 186, 330, 474, …,44538 and 78, 222, 366, 510, …, 44430 and 114, 258, 402, 546, …, 44466 respectively. The same sequences were done for Blanchard and Neosho, however their upper limits were 26538, 26574, 26466 and 35610, 35646, 35682 for 6, 12 and 18 hour shifts correspondingly.

3.5 Results

3.5.1 Wind Speed Duration Curve (WSDC)

The WSDC is a graphical analysis tool for the persistence of energy production for a particular station (Masseran et al., 2012). Figures 3.2 to 3.4 depict the WSDC for the various stations, Columbia, Blanchard and Neosho and their individual tower heights for the year 2009. From the graphs we observe that there is more of a wind speed variability among the height levels for Columbia when compared to Neosho. This is expected as Columbia station's tower spans greater height levels than Neosho. It is also observed that Blanchard97 experiences a constant low value of wind speed well below 5 $\mathrm{m\,s^{-1}}$ for approximately 50% of the time. For the various stations and tall towers, if the WSDC is flatter, then the wind regime is more regular, if not it is more irregular. From the results obtained, it can be noted that the wind regime is more irregular as the heights are increased for both Columbia and Blanchard with the exception of Blanchard97. From Table 3.1, the percentage of time the wind speeds are greater than or equal to 4 $\mathrm{m\,s^{-1}}$, which is typically less than the turbine cut-in wind speed (Smith et al., 2002), were higher than 70% excluding Blanchard97. This indicates relatively significant energy persistence for the stations studied. The energy persistence increased with height for stations Columbia and Blanchard neglecting the intermediate height level of Blanchard, Blanchard97. This was not the case for Neosho.

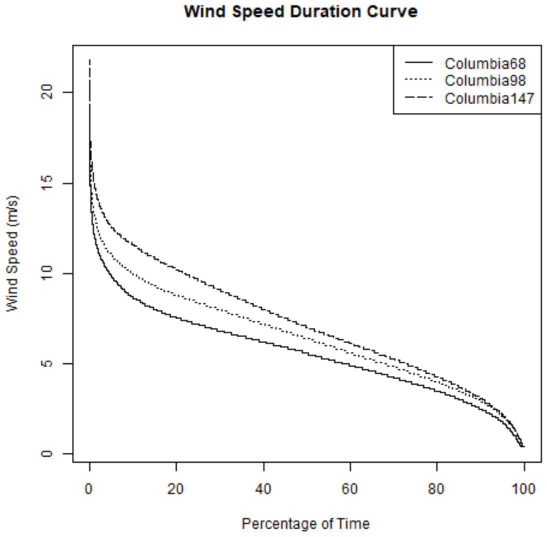


Figure 3.2: WSDC for Columbia Station

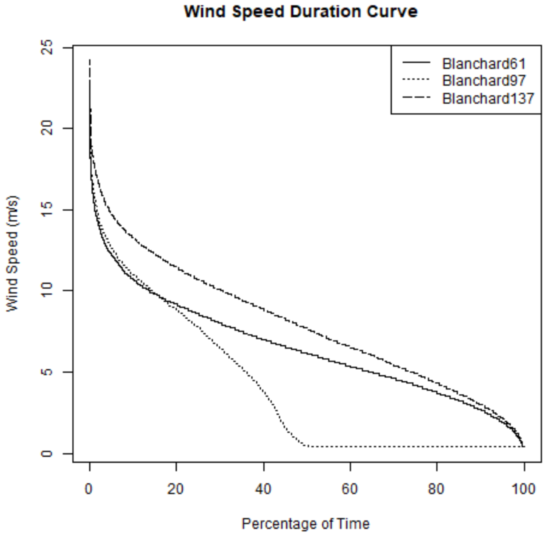


Figure 3.3: WSDC for Blanchard Station

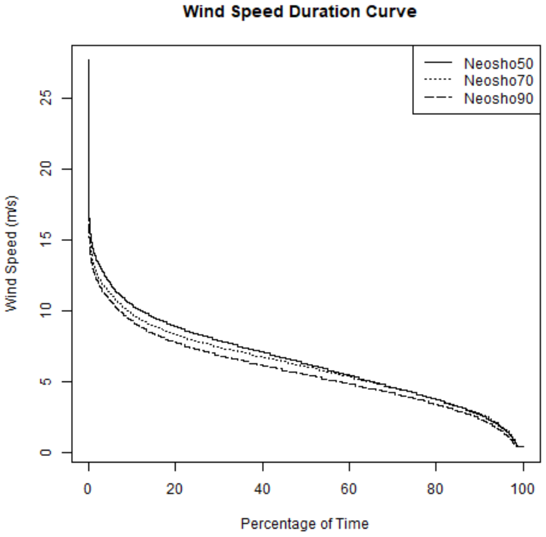


Figure 3.4: WSDC for Neosho Station

Table 3.1: Wind Speed Values greater than or equal to 4 $\mathrm{m\,s^{-1}}$ from WSDC

$y(m/s)$	Station	$x(\%)$
4	Columbia68	73.4
	Columbia98	79.9
	Columbia147	82.7
4	Blanchard61	77.1
	Blanchard97	39.5
	Blanchard137	82.5
4	Neosho50	77.2
	Neosho70	77.1
	Neosho90	72.5

3.5.2 Reconstruction of Phase Space

The two parameters, τ and d, for the various months and height levels for the year of 2009 at Columbia are given by Table 3.2. However, for illustrative purposes we show the results of August consistently throughout the paper. This, as well as the other results, depicts both qualitatively and quantitatively, the chaotic nature of the series. From Figures 3.5, 3.6 and 3.7, we see that at Columbia for the month of August at height levels 68, 98 and 147 m, the first local minima (τ) were observed at 5, 5 and 6 respectively. Similarly, for the corresponding embedding dimensions in August, given by Figures 3.8, 3.9 and 3.10 are 9, 9 and 8. It is observed that $E1(d)$ attains saturation at these values. Also, $E2(d)$ is related to $E1(d)$ and there exist some values of d for which $E2(d) \neq 1$. This implies that the data are deterministic rather than random. From the results obtained we note that there are τ and d values that repeat for this station. Such findings are indicative of required dimensional construction of phase space to capture the underlying dynamics of the system (Zounemat-Kermani and Kisi, 2015). Given that the values for d is for most cases 9, it indicates that the ease of predictability will be similar for this station for all months as dynamics of lower dimensions are easier to predict (Lisi and Villi, 2001).

The phase portraits or diagrams as seen in Figures 3.11, 3.12 and 3.13 depict the non-linear variation of the state of the system with time for Columbia at three different heights. It gives the attractor's spatial structure (Yu et al., 2019). The axes represent the first three time delayed co-ordinates used in the construction of the attractor which can be shown diagrammatically. The number of time delayed, τ, co-ordinates is determined by the embedding dimension, d. Since the trajectories from the system phase space did not show periodicity nor were they random motion but rather illustrated a single attractor, the time series displays chaotic characteristics. These are deterministically chaotic systems. This is shown in also in a study by Zend et al. (Zeng et al., 2012) in which the phase diagram of near surface winds were not closed curves indicative of aperiodic trajectories. In another study done by Yu et al. (Yu et al., 2019), the trajectories of the attractor were not as well-defined as that of Lorenz's because the interference of noise signals from the environment. For all of the months in 2009 and height levels of the tall tower in Columbia, we observed the chaotic characteristics of the wind speed from the phase portraits.

Table 3.2: Values of the Parameters, time delays and embedding dimensions for Columbia station

Month	τ	d	Month	τ	d	Month	τ	d
Jan C68	10	8	Feb C68	6	8	Mar C68	5	9
Jan C98	4	11	Feb C98	6	9	Mar C98	5	9
Jan C147	1	9	Feb C147	6	8	Mar C147	6	9
Apr C68	5	9	May C68	5	9	June C68	4	9
Apr C98	5	9	May C98	5	9	June C98	5	9
Apr C147	5	9	May C147	6	9	June C147	5	8
July C68	4	9	Aug C68	5	9	Sept C68	7	8
July C98	4	9	Aug C98	5	9	Sept C98	7	8
July C147	5	9	Aug C147	6	8	Sept C147	8	7
Oct C68	5	9	Nov C68	5	9	Dec C68	6	9
Oct C98	6	9	Nov C98	5	9	Dec C98	6	8
Oct C147	7	9	Nov C147	6	8	Dec C147	7	8

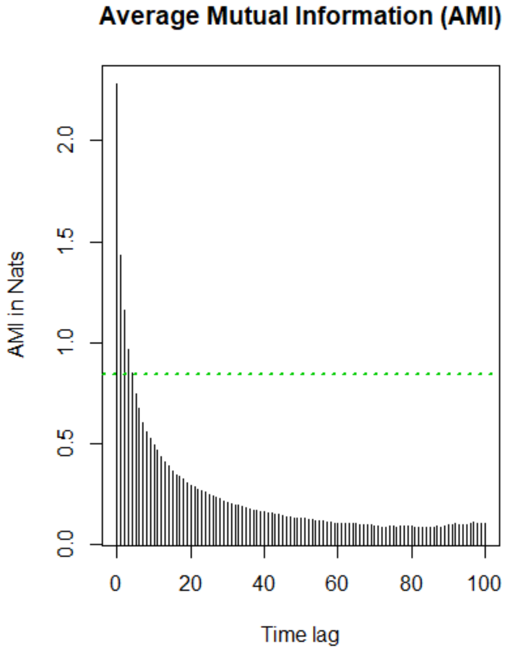


Figure 3.5: Time delay given by the method of AMI, Mutual Information against Time Lag for Aug Columbia68

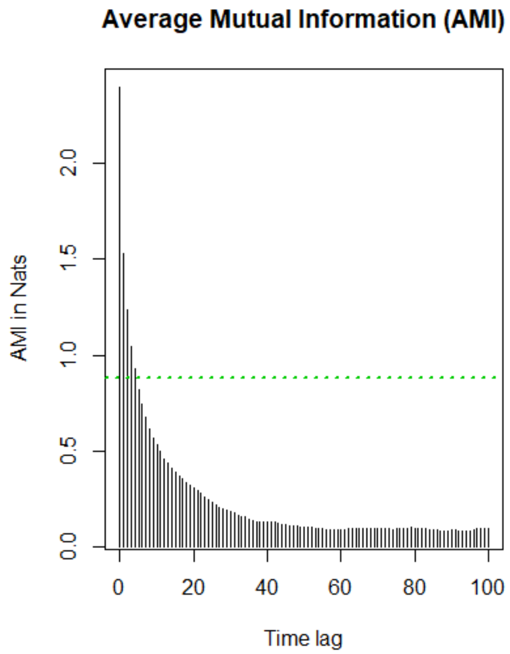


Figure 3.6: Time delay given by the method of AMI, Mutual Information against Time Lag for Aug Columbia98

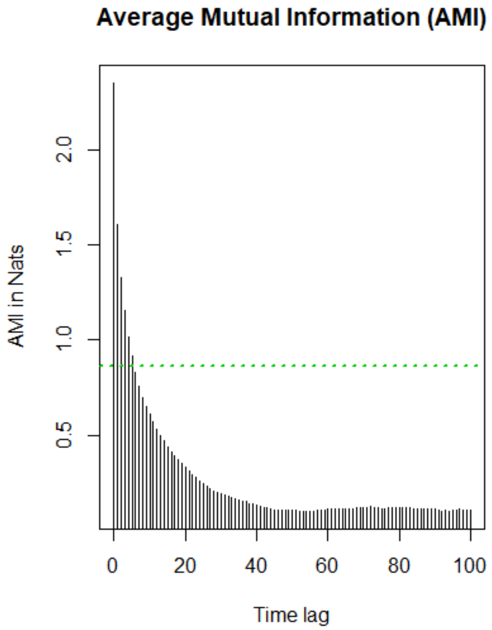


Figure 3.7: Time delay given by the method of AMI, Mutual Information against Time Lag for Aug Columbia147

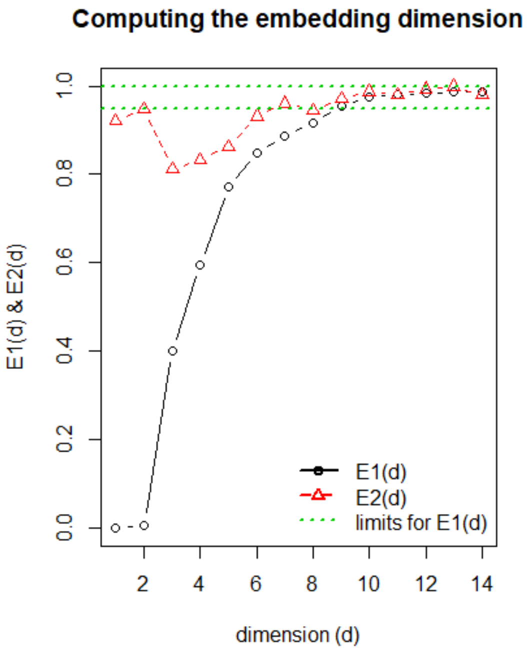


Figure 3.8: Embedding dimension given by Cao's Algorithm, $E1(d)$ and $E2(d)$ against d for Aug Columbia68

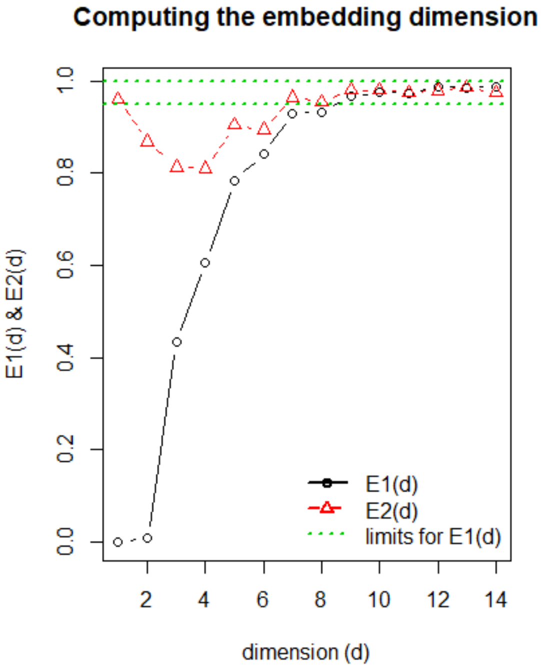


Figure 3.9: Embedding dimension given by Cao's Algorithm, $E1(d)$ and $E2(d)$ against d for Aug Columbia98

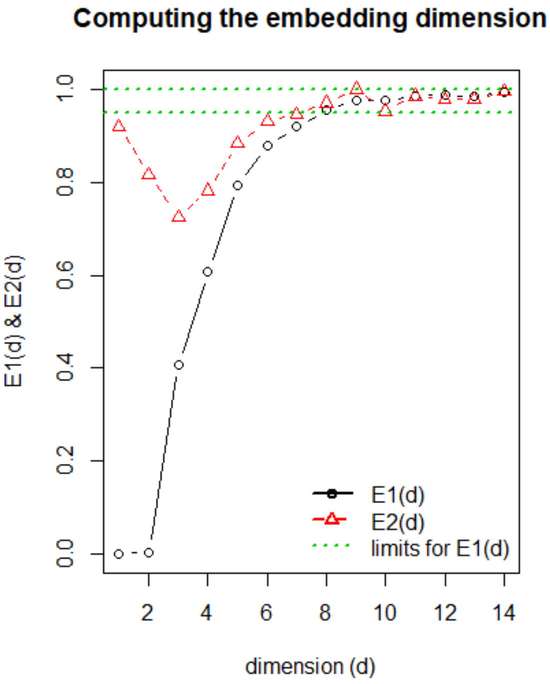


Figure 3.10: Embedding dimension given by Cao's Algorithm, $E1(d)$ and $E2(d)$ against d for Aug Columbia147

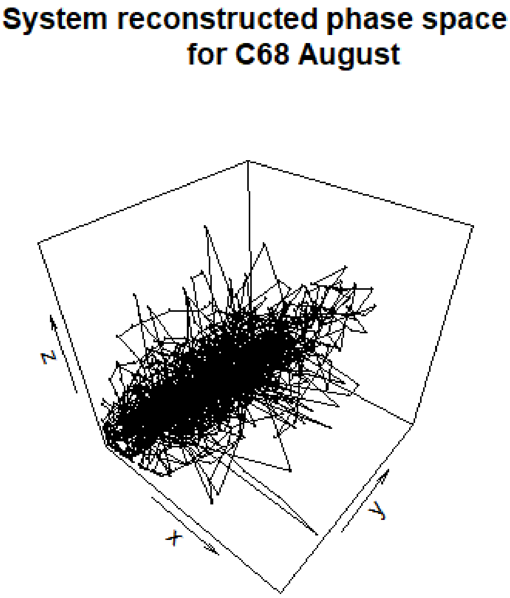


Figure 3.11: Phase space reconstruction for Aug Columbia68 showing the first three time delayed co-ordinates

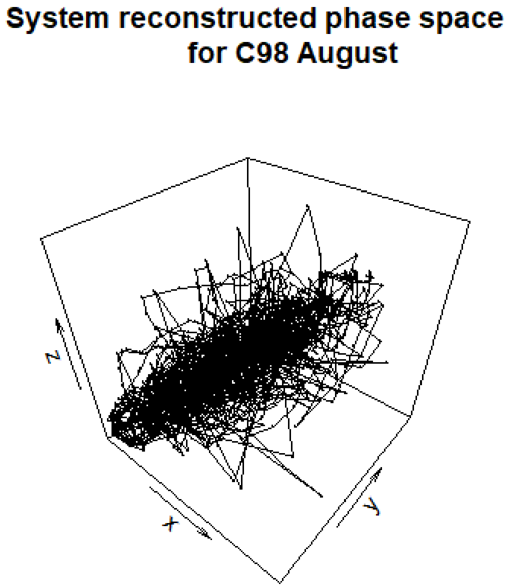


Figure 3.12: Phase space reconstruction for Aug Columbia98 showing the first three time delayed co-ordinates

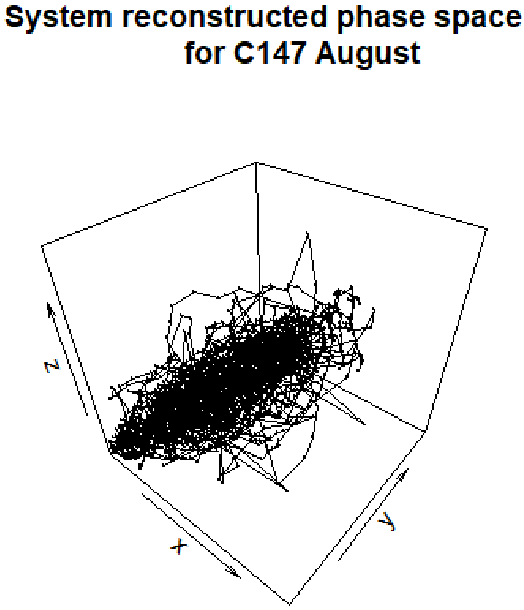


Figure 3.13: Phase space reconstruction for Aug Columbia147 showing the first three time delayed co-ordinates

3.5.3 Largest Lyapunov Exponent (LLE)

The wind speed at all heights of the tall tower display chaotic characteristics as the LLE were positive (Kantz and Schreiber, 2004). This implies that the trajectories diverge exponentially fast and there is an increase in the average exponent of the trajectory divergence characteristic of a non-periodic system as also seen in the phase diagrams of this study. From the graphs in Figures 3.14, 3.15 and 3.16, it can be noted that even though it is not prominent, there is a linear region before the curve saturates regardless of the embedding dimensions. It is also seen in a study done by Yu et al. (Yu et al., 2019) that wind speeds display positive chaotic characteristics. In this study however, the increasing level of chaos characteristics did not coincide with the increasing height levels of the tall tower. This is noted quantitatively from Table 3.3.

The average LLE of the various tall tower heights for the 2009 were 0.1133, 0.11305 and 0.11259 for Columbia68, Columbia98 and Columbia147 respectively. From equation (3.12), we obtain the prediction horizon to be approximately 6 times steps or one hour for all height levels. However, changing the confidence band to 90% and 99%, decreased and increased the prediction horizon to approximately 4 and 8 time steps respectively. Thus, we note an incremental increase in the prediction horizon by 2 time steps when the confidence bands increased from 90 to 95 and to 99%. We expect this increase as increasing the confidence causes an increase in the margin of error thus resulting in a larger interval. If we consider the minimum of all the LLE of each of the increasing height levels, we get the corresponding values of 0.07613, 0.06329 and 0.06452. The prediction horizon, using the 95% confidence band, is increased to approximately 9, 11 and 10 time steps or 1.5, 1.8 and 1.6 hours respectively. Similarly using these minimum LLE, it was determined that for the two largest height levels of this station, for a confidence band of 99%, the prediction horizon increased to an estimated 2.5 hours. Whilst using the lowest confidence band of 90% , the said value was decreased to approximately 1.3 hours. For the lowest height level, confidence bands of 90 and 99% implied roughly 1.2 and 2 hours for the prediction horizon respectively.

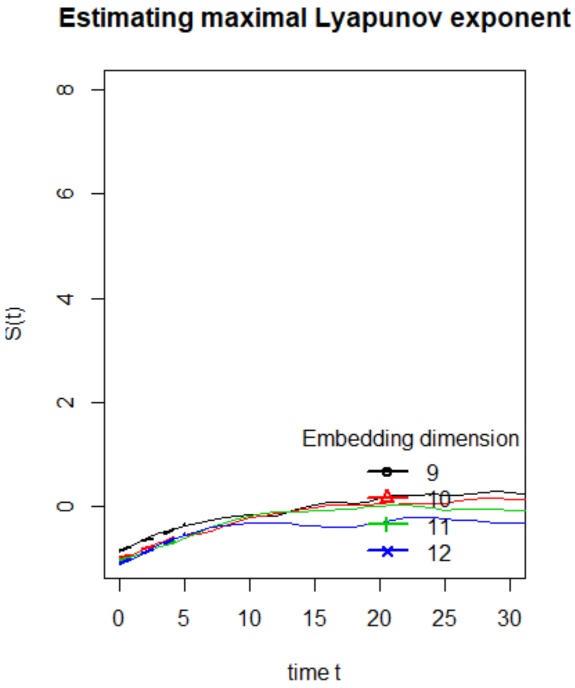


Figure 3.14: Lyapunov Exponents for Aug Columbia68, $S(t)$ against t

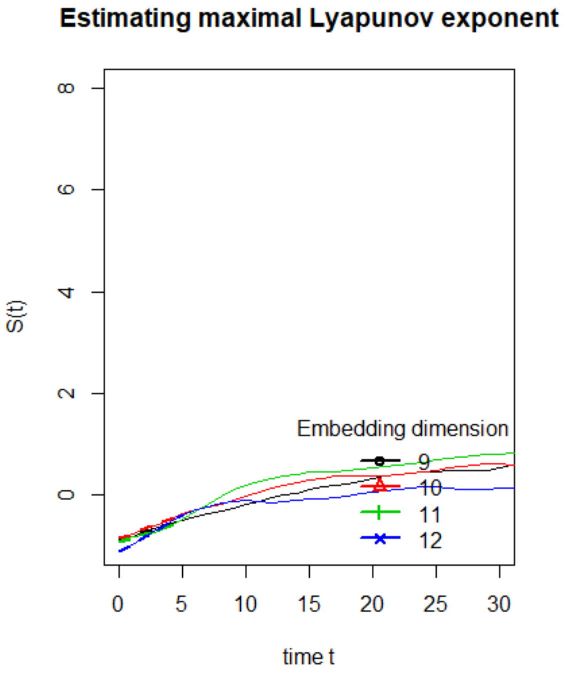


Figure 3.15: Lyapunov Exponents for Aug Columbia98, $S(t)$ against t

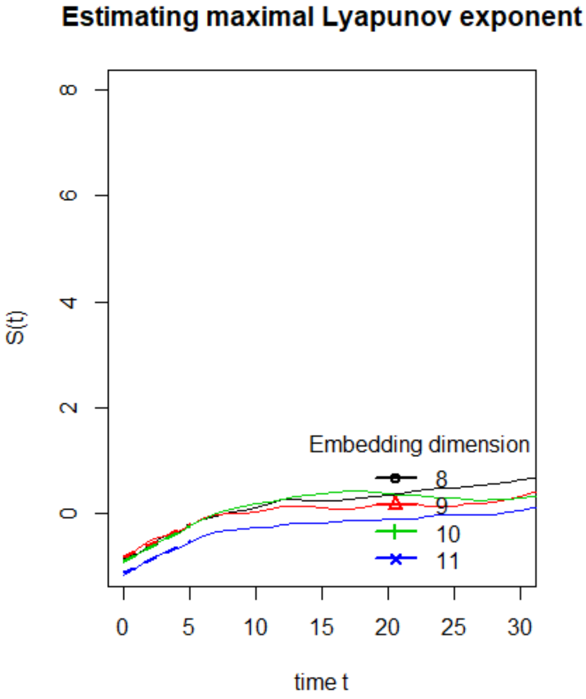


Figure 3.16: Lyapunov Exponents for Aug Columbia147, $S(t)$ against t

Table 3.3: Values of the Parameter, Lyapunov Exponent, for Columbia station

Month	LLE	Month	LLE	Month	LLE
Jan C68	0.21850120	Feb C68	0.11124360	Mar C68	0.12129040
Jan C98	0.23026360	Feb C98	0.06329283	Mar C98	0.10560540
Jan C147	0.28907590	Feb C147	0.09144223	Mar C147	0.08357779
Apr C68	0.11611450	May C68	0.10797860	June C68	0.09807250
Apr C98	0.09545526	May C98	0.07794641	June C98	0.09980393
Apr C147	0.09965277	May C147	0.06451961	June C147	0.11250640
July C68	0.14193490	Aug C68	0.09708272	Sept C68	0.08012870
July C98	0.13593250	Aug C98	0.09956517	Sept C98	0.13722600
July C147	0.06720401	Aug C147	0.12664330	Sept C147	0.10288420
Oct C68	0.11155650	Nov C68	0.07613265	Dec C68	0.07952253
Oct C98	0.10451390	Nov C98	0.07777989	Dec C98	0.12927020
Oct C147	0.10715060	Nov C147	0.11311340	Dec C147	0.09325161

3.5.4 Forecasting using non-linear algorithm

For this forecasting analysis, all the height levels of each station were employed. The attractor was constructed using the first 56 days of the year having determined the embedding dimension and the time delay using that data as the training of the model. Using the forecasting algorithm described in section 3.4, the next 6 hours or 36 time steps (in 10 minute intervals) were predicted. The preceding forecast was done for the next consecutive 6 hours but it was done using a moving window in which the attractor was reconstructed for this run. The training data were assigned to be 56 days starting from a six hour delay of the previous run. This process was iterated until the entire series in this moving window is accounted for. The forecasted data for each run and the actual data were compared and its RMSE, MAE and Correlations were calculated. These errors were determined cumulatively for all of the first predictions or time step through the last prediction or 36^{th} time step.

From the results given in Figures 3.17, 3.18 and 3.19 for all of the height levels in Columbia, it is evident that both the RMSE and the MAE increases as the forecasting time step is increased as expected. It should be noted that the correlation between the actual and the predicted wind speed values is greatest (almost 1) at the first time step and it decreases to under 0.5 for the last time step. It is observed that the RMSE plateaus to approximately 2.5 $\mathrm{m\,s^{-1}}$, 3.0 $\mathrm{m\,s^{-1}}$ and 3.5 $\mathrm{m\,s^{-1}}$for Columbia68, Columbia98 and Columbia147 respectively. The errors are expected to increase as the height level increases because the wind speeds are increasing with height.

Similar analyses were done in Figures 3.20 and 3.21. However, no results are provided for Blanchard97 because of its inoperability for approximately 50% of the time. The max RMSE for Blanchard61 and Blanchard137 were greater than Columbia; their values for the 36th forecasts were approximately 3.5 and 4.5 $\mathrm{m\,s^{-1}}$. For the Neosho tall tower, as depicted in Figure 3.21, the RMSEs for the last prediction step were 3.17, 2.96 and 2.87 $\mathrm{m\,s^{-1}}$ for Neosho50, Neosho70 and Neosho90 respectively. This is approximately 3.0 $\mathrm{m\,s^{-1}}$ for all heights. This as well as the similar numerical trends from the other time steps, indicate that of all the stations, there is least variability between the actual and predicted values for this station. This coincides with Figure 3.4, where Neosho had the least wind speed variability among all the height levels.

Figures 3.22, 3.23 and 3.24 show the normalized errors for the various stations. All of the stations and at all heights follow the same increasing trend, which begins to plateau. From the results of Columbia, for all of the heights, after the 20th time step or after 3.33 hours of forecasting, the model normalized errors exceed 1 or 100%. As such the forecast is no better than the mean of the data after this run. However using a range of normalized errors, we note that for normalized errors exceeding 50 and 75% occurred after the 1st, 2nd, 3rd and the 7th, 7th, 9th time steps for Columbia68, Columbia98 and Columbia147 respectively. From the Lyapunov analysis, the prediction horizon was determined to be 4, 6 and 8 time steps for the various confidence bands, which all gave normalized errors of less than 65, 75 and 80% correspondingly. Similarly, for Blanchard61 and Blanchard137, the normalized error was below this 100% normalized error threshold at the 22nd and 19th time step respectively. The 8th and 2nd time steps corresponded to normalized errors of under 75 and 50% respectively. Also for Neosho, this model is not better than the average projection after the 23rd, 24th and 27th 10-minute forecast for Neosho50, Neosho70 and Neosho90 correspondingly. However the normalized errors of under 50 and 75%, for these respective height levels, occurred at most for the time steps of 2 and 9.

This model was also compared to the benchmark model of persistence. Wind speed persistence for a given site, as defined in, (Koçak, 2002),is a measure in which the duration of the average wind speed persists. The average for every multiple of the 6th hour was recorded as the value of persistence for the next consecutive 6 hours or 36 time steps. This was done for the entire series of 2009 for all height levels for each station. The results are shown also in Figures 3.17 to 3.21. It was determined that for Columbia and Blanchard at all height levels, our model out performed persistence for the first time step however for Neosho, persistence beat our model at all heights of the tall tower. This is expected as Nesoho, of the three stations, had the least annual mean friction velocity of 0.597 (Fox, 2011). Our model, as expected, does better in more turbulent flow with greater roughness. It should be noted though, that the difference in the average RMSE for the EDM and the persistence model at all height levels of Columbia and Neosho did not exceed $0.5\mathrm{m\,s^{-1}}$. This value was approximately $0.6\mathrm{m\,s^{-1}}$ for Blanchard. This, thus demonstrates that the EDM is comparable to persistence. The model displays relatively good accuracy as for this short-term range, persistence, autoregressive, moving average, autoregressive moving average and autoregressive integrated moving average models all perform better than the NWP models (De Freitas, Silva, and Sakamoto, 2018). However, the model displays better accuracy within the very short-term scale, between a few minutes to one hour. This is expected as persistence does well in the short term range. Another reason is due to the chaotic system's SDOIC (Lisi and Villi, 2001). The shorter term forecasts will be more accurate for these deterministic non-linear systems. The degree of accuracy was also determined by the amount of noise in the data. Since it is generally given that high wind speed implies high persistence (Koçak, 2002), it would be expected that the model would not beat persistence for more time steps with increasing heights, and this is seen. We also got the expected result of the average difference between the two models for the 36 time steps to increase with increasing tower height. We obtained this result for two of three stations, Columbia and Blanchard. This small range in forecast where persistence can be beaten is also captured from the prediction horizon given by the LLE. From the results obtained, this is closest in agreement with the prediction horizon using the 90% confidence band and using a normalized error cap of 50%.

For this very short term forecast interval, there are not many research papers available (Soman et al., 2010). However, another model used in a case study in Tasmania, Australia beat persistence in this scale, 2.5 minutes ahead. This is a hybrid model, Adaptive Neuro-Fuzzy Interface System (ANFIS) (Soman et al., 2010). This very short term scale is utilized for efficient trading and optimal use of transmission lines (Potter and Negnevitsky, 2006). There is no widespread acceptance in the industry for a particular very short-term forecasting system as there is no reliable forecasting technique for this scale where wind speeds have the most variations. Persistence is often deemed sufficient for this time scale as historically predicting for this scale was viewed as unnecessary (Potter and Negnevitsky, 2006).

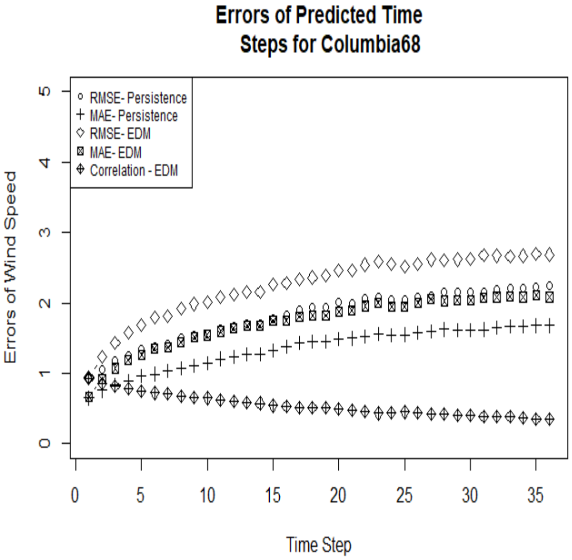


Figure 3.17: Errors and Correlations for Columbia68, RMSE ($\mathrm{m\,s^{-1}}$), MAE ($\mathrm{m\,s^{-1}}$) and Correlations against Time Step for EDM and Persistence

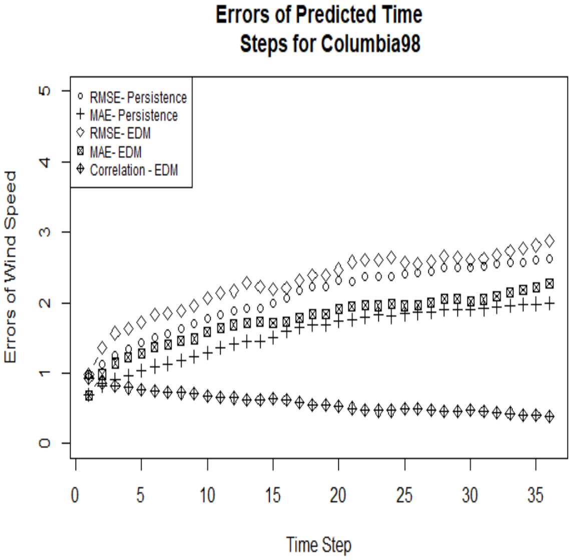


Figure 3.18: Errors and Correlations for Columbia98, RMSE ($\mathrm{m\,s^{-1}}$), MAE ($\mathrm{m\,s^{-1}}$) and Correlations against Time Step for EDM and Persistence

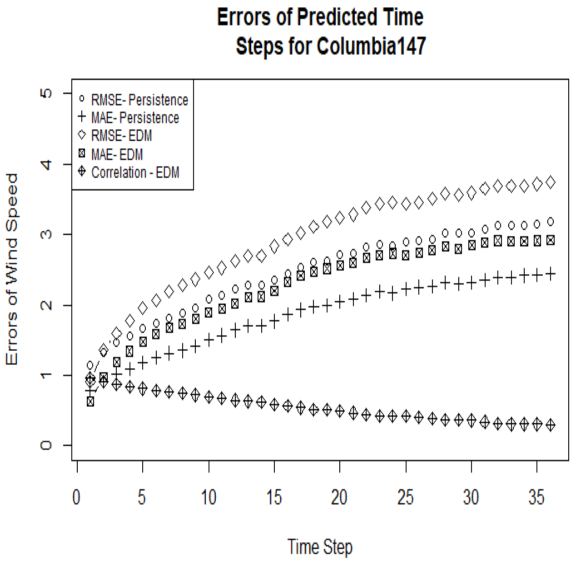


Figure 3.19: Errors and Correlations for Columbia147, RMSE ($\mathrm{m\,s^{-1}}$), MAE ($\mathrm{m\,s^{-1}}$) and Correlations against Time Step for EDM and Persistence

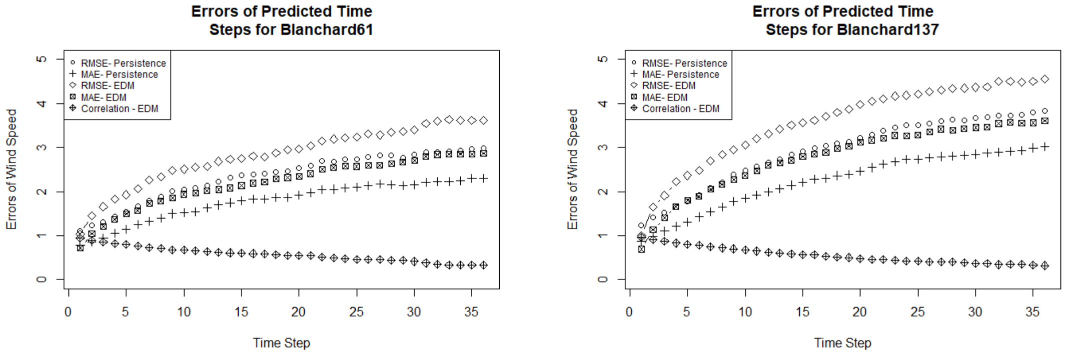


Figure 3.20: Errors and Correlations for Blanchard Tall Tower, RMSE ($\mathrm{m\,s^{-1}}$), MAE ($\mathrm{m\,s^{-1}}$) and Correlations against Time Step for EDM and Persistence

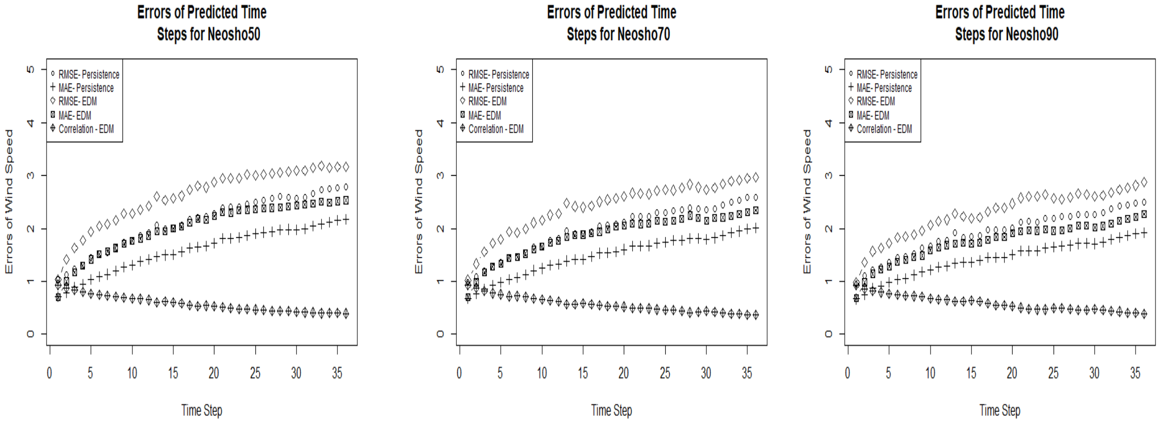


Figure 3.21: Errors and Correlations for Neosho Tall Tower, RMSE ($\mathrm{m\,s^{-1}}$), MAE ($\mathrm{m\,s^{-1}}$) and Correlations against Time Step for EDM and Persistence

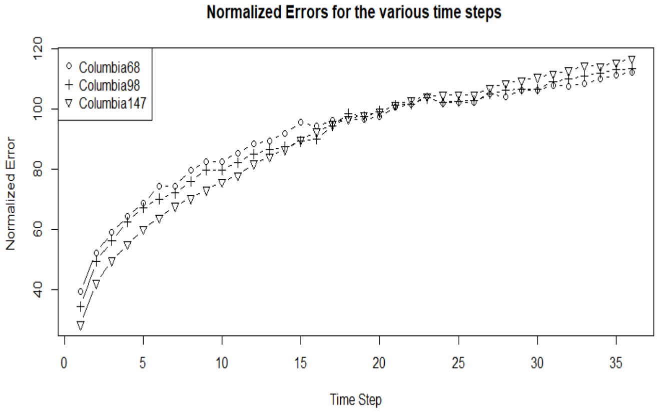


Figure 3.22: Normalized Errors for height levels of Columbia Tall Tower, Normalized Errors (%) against Time Step

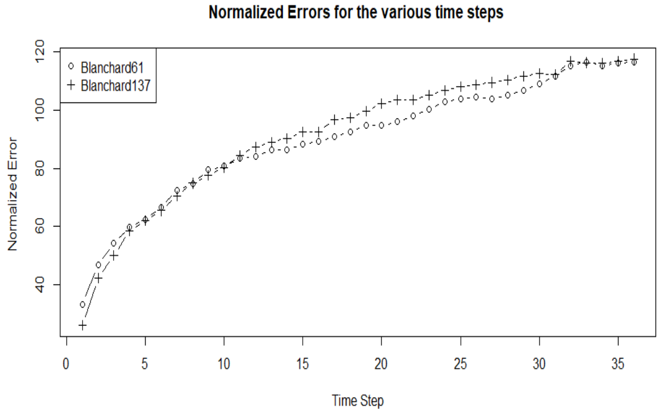


Figure 3.23: Normalized Errors for height levels of Blanchard Tall Tower, Normalized Errors (%) against Time Step

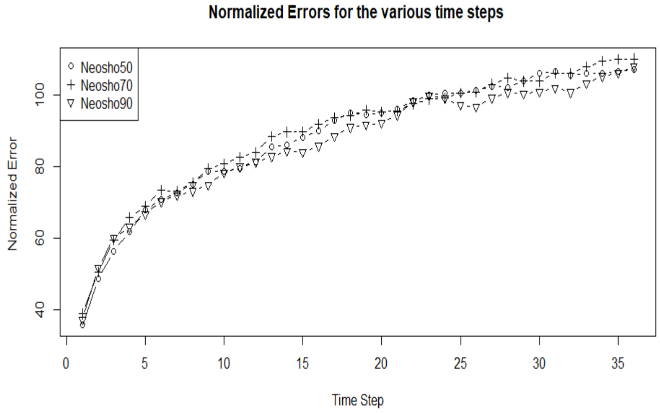


Figure 3.24: Normalized Errors for height levels of Neosho Tall Tower, Normalized Errors (%) against Time Step

3.5.5 Analysis of seasonality and the diurnal cycle

From Table 3.4, it is clear that for the slices and shifts of the data, all stations and height levels had a mean error of approximately 0 $\mathrm{m\,s^{-1}}$. This is due to a roughly even distribution of negative and positive errors. They deviated from their respective means, for most cases, by an estimated value of 2 $\mathrm{m\,s^{-1}}$. The exception was Blanchard137 after the first 6 hour shift from $144m + 6$ slicing of the data, its deviation from the mean of approximately 0 $\mathrm{m\,s^{-1}}$ was an estimated wind speed value of 3 $\mathrm{m\,s^{-1}}$.

The seasonality analysis was done for all three stations with particular concentration on Columbia as this station had all of the monthly data for 2009. In the Northern Hemisphere for the year of 2009, spring began on Friday 20th March. Summer started Saturday June 20th while for fall and winter they commenced on September 22nd and December 21st respectively. Spring corresponded to the 7272nd window. This coincided with a numeric value of forecast 3276. The start of summer 2009 corresponded with window 20520 which is numeric value 16524 whilst for fall and winter the windows values were 34056 and 47016 which coincided with numerical values of forecast 30060 and 43020 respectively.

Figures 3.25, 3.26 and 3.27 show the various errors for the four seasons in the forecast period in 2009 for Columbia68, Columbia98 and Columbia147 respectively. The seasons of spring, summer and fall will be of focus as winter of 2009 was represented as a few days in this analysis as data of 2010 were not included in this study. Also, since the data spanned one year, the seasonality aspect of the model could not be fully investigated. The average errors and standard deviations of these seasons were, however, determined and compared to that of the entire model run for 2009. It was given that for Columbia68, summer had the lowest magnitude of errors between model run and actual values; its value was -0.03809 with standard deviation of 1.511631. This is the only season whose average error was under the average error for the entire model run for 2009. Fall average errors were a bit higher than for the summer; their values were 0.064728 and 1.770366 respectively. Spring has the largest average errors and standard deviations; its values were -0.19993 and 1.954639. This is expected as from the average wind speeds for the various months, shown in (Balkissoon, Fox and Lupo, 2020), summer had the least average wind speeds as well as incurred the least monthly variations. For Columbia98, once again summer had the least average errors in the simulation. The mean error was approximately the same as Columbia68 with

Table 3.4: Means, Standard Deviations, Min Values and Max Values of errors, in $\mathrm{m\,s^{-1}}$, for various analyses

Analyses	Station	Mean	Standard Deviation	Min Value	Max Value
36n+6	Columbia68	-0.04976043	1.791448	-8.6	9.6
	Columbia98	0.09388809	1.914046	-8.6	15
	Columbia147	0.08977431	2.064804	-9.4	10
144m+6	Columbia68	-0.191192	1.679333	-5.05	6
	Columbia98	0.2311889	1.670056	-5.4	6.9
	Columbia147	0.1320323	2.066888	-8	6.7
36+144m+6	Columbia68	0.07817742	1.819659	-5.9	9.6
	Columbia98	0.06601613	2.088458	-6.1	15
	Columbia147	-0.04369969	1.976215	-5.6	8.25
72+144m+6	Columbia68	-0.1004072	1.800744	-7.9	5.3
	Columbia98	0.08013061	1.946351	-6.9	7.4
	Columbia147	0.1764266	2.11462	-9.4	10
108+144m+6	Columbia68	0.01442377	1.857963	6.2	-8.6
	Columbia98	-0.002137463	1.929332	7.8	-8.6
	Columbia147	0.09463323	2.102374	9.7	-7.8
36n+6	Blanchard61	-0.1574112	2.063242	-6.6	10.9
	Blanchard137	-0.07311908	2.480952	-9.6	8.8
144m+6	Blanchard61	-0.1845045	2.071191	-6.6	4.8
	Blanchard137	0.004801802	2.261244	-9.6	6.3
36+144m+6	Blanchard61	-0.1805598	2.129216	-5.1	10.9
	Blanchard137	0.02518018	2.847701	-8.6	8.8
72+144m+6	Blanchard61	-0.2122522	1.995355	-5	8.1
	Blanchard137	-0.5742613	2.379635	-9.05	7.1
108+144m+6	Blanchard61	-0.05175725	2.068112	6.6	-6.6
	Blanchard137	0.2535688	2.335961	7.4	-6.05
36n+6	Neosho50	-0.007078053	2.050773	-9.8	12.5
	Neosho70	0.006690188	1.937222	-7.8	10.5
	Neosho90	-0.02743856	1.831548	-7.7	8.5
144m+6	Neosho50	0.2418952	2.111877	-7.1	8.1
	Neosho70	0.01000672	2.003076	-7.8	7.7
	Neosho90	-0.1898185	1.839123	-7.1	5.6
36+144m+6	Neosho50	-0.03209293	2.105925	-9.8	5.3
	Neosho70	0.1394792	1.853771	-7.2	8.8
	Neosho90	-0.07311972	1.724716	-7.7	5.25
72+144m+6	Neosho50	-0.2335916	1.950025	-6.65	7.2
	Neosho70	-0.229328	1.865122	-6	6.9
	Neoaho90	-0.0705506	1.879079	-5.9	6.3
108+144m+6	Neosho50	-0.004522849	2.015396	-6.7	12.5
	Neosho70	0.1066028	2.01141	-4.8	10.5
	Neosho90	0.2237346	1.864748	-5.23	8.5

the standard deviation increasing by an estimated 0.21. Fall, for this height level, had the highest error value of approximately 0.10 while spring had the lower value of 0.06. The standard deviation of the errors in the model, for spring were higher than for fall with an estimated value of 2.17 opposed to 1.82. Both summer and spring have lower average errors than the year model run, which from Table 3.4 is around 0.09. For the largest of the three height levels, Columbia147, all seasons considered were under the entire 2009 model run. Spring and summer both had the same average error of 0.05 while fall had a slightly smaller mean error value of 0.04. The standard deviation values were largest for spring (2.162856) and smallest for summer (1.896873). Thus, it was observed that for the station at Columbia, with the exception of the last height level in which all height levels incurred the same average deviations from the actual wind speed values, the summer season was the most accurate from this model. Also, for all heights on the tall tower, the standard deviations were largest for spring and smallest for summer.

For Blanchard61 and Blanchard137, all of the spring season was in the model run; however, for summer, 98 of the 376 readings of the slices representative of fall 2009 were missing. Similar statistics were calculated and it was observed that for the lowest height level there was larger average error and a lower standard deviation value for summer than spring, a magnitude difference of 0.2276 and 0.21228 respectively. For Blanchard137, summer had lower average errors and deviations when compared to spring. Considering station Nesoho at height level 50 m, once again, summer incurred the least average error of 0.017019 with standard deviation 1.936593 when compared to the other two seasons of spring and fall whose magnitude of average error were estimated 0.03 and 0.05 respectively. However, it must be noted that the fall season has 205 data slices missing from the 360 slices representative of fall 2009. For Neosho70, spring, however, had the least magnitude in its average error whose value is 0.00351 whilst the fall had the least standard deviation of 1.559813. For the

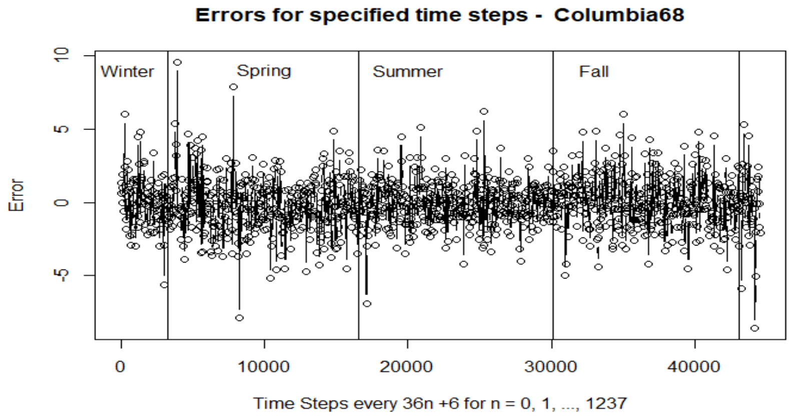


Figure 3.25: Seasonality Analysis for Columbia68, Error ($\mathrm{m\,s^{-1}}$) against Time Step every $36n + 6$

final height level of Neosho, it was determined that the modulus of the average error for all the seasons considered were less than that value for the model run. This value as well as the standard deviation were the smallest for the summer season. Thus, four of the eight stations' height levels had the season of summer incurring the lowest magnitude of average errors and standard deviations. This is expected as the weakest winds occur during this season whilst the strongest winds are observed from October through spring. This is due to the passage of cyclones during these cooler periods. There is also during the summer, more heating which causes more instability and turbulence with less wind shear and near surface winds. The increased vegetation cover during this season are responsible as well for the increased turbulence due to its surface roughness and lower albedo and thus deeper boundary layer (BL) (Fox, 2011).

The diurnal analysis methodology, explained in Section 3.6, was investigated. The model took 56 days to initialize and it began to forecast on February 26th for all stations and height levels. The $144m + 6$ slicing of the forecast represents the extraction of the data every day at 6:50 am. The subsequent extracting of the data 6, 12 and 18 hours later represented times of 12:50 pm, 6:50 pm and 12:50 am respectively. The results obtained for this slicing for the station of Columbia at height level 68 m are

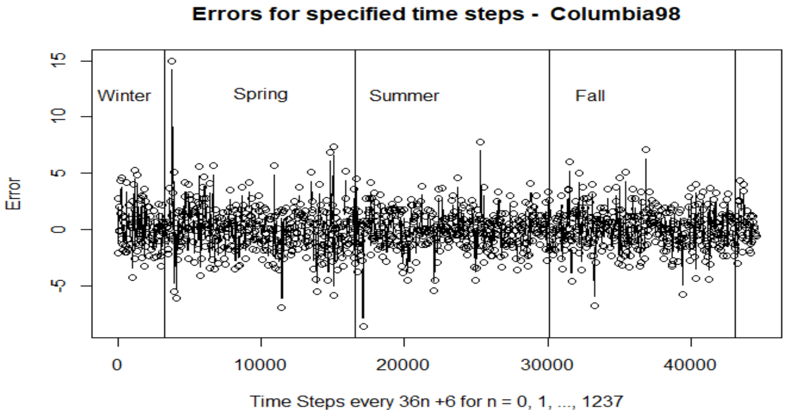


Figure 3.26: Seasonality Analysis for Columbia98, Error $(\mathrm{m\,s^{-1}})$ against Time Step every $36n + 6$

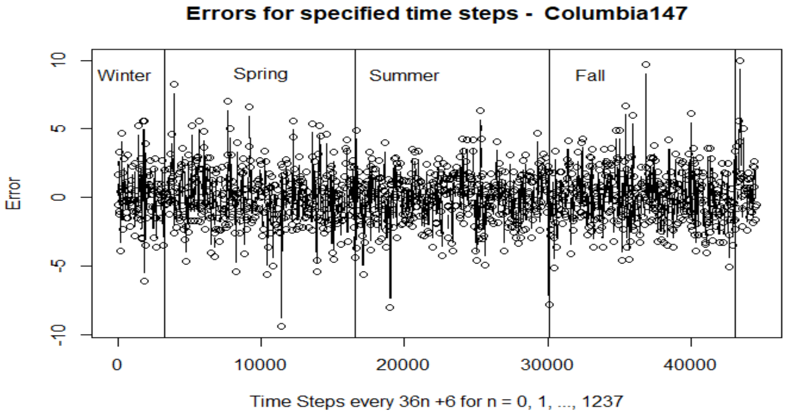


Figure 3.27: Seasonality Analysis for Columbia147, Error $(\mathrm{m\,s^{-1}})$ against Time Step every $36n + 6$

shown in Figures 3.28 to 3.31. Statistical parameters for the entire forecast period for all slices and stations can also be viewed in Table 3.4.

Two cases occurred where the hours of 6:50 am and 12:50 pm had the highest standard deviation values when compared to the early morning and evening hours of 12:50 am and 6:50 pm. This occurred for the lowest heights of Neosho and Blanchard. While for the highest levels of Columbia and Neosho we see the reverse occurrence where 6:50 pm had the highest standard deviation followed by 12:50 am, 6:50 am and 12:50 pm. This might be as a result of the nocturnal low level jet and strong overnight wind, which is common in the Midwest. This is due to the reduced convection and turbulence during the cool night hours which causes more stability with larger wind shear as a result of the reduced transfer of energy and momentum between the layers. The boundary layer is shallower and as such the geostrophic winds are closer to the surface (Fox, 2011). However, there is no evident trend where a particular time of the day the model incurred more errors and greater standard deviations for all stations and heights. This may imply that the lack of sensitivity to the diurnal cycle by the stations height levels means that the model handles any transitions well.

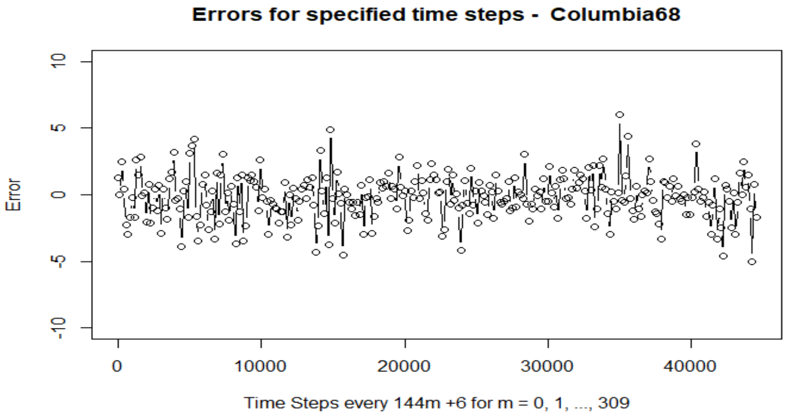


Figure 3.28: Diurnal Analysis for Columbia68, Error ($\mathrm{m\,s^{-1}}$) against Time Step every $(144m + 6)$

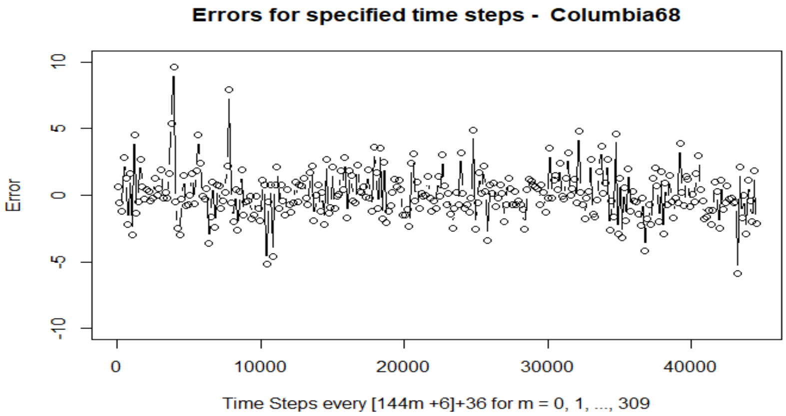


Figure 3.29: Diurnal Analysis for Columbia68, Error ($\mathrm{m\,s^{-1}}$) against Time Step every $[(144m + 6) + 36]$

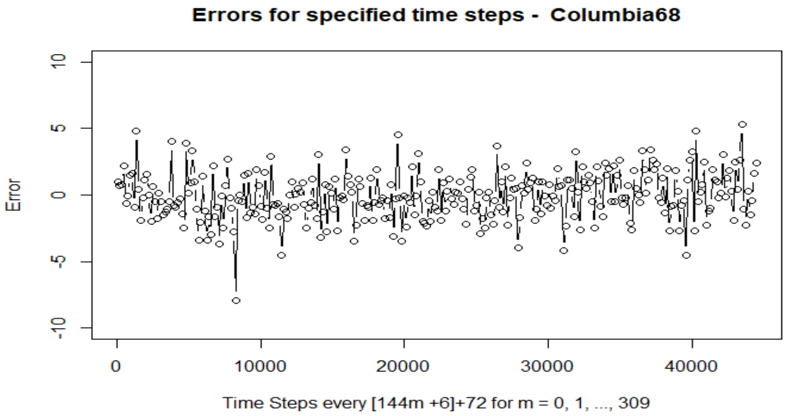


Figure 3.30: Diurnal Analysis for Columbia68, Error ($\mathrm{m\,s^{-1}}$) against Time Step every $[(144m + 6) + 72]$

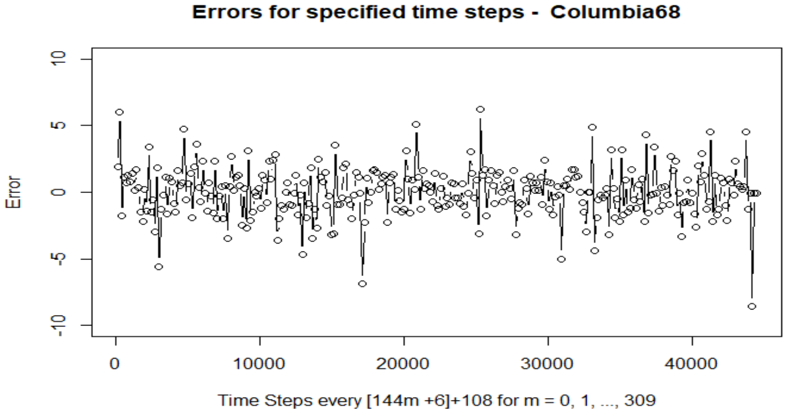


Figure 3.31: Diurnal Analysis for Columbia68, Error ($\mathrm{m\,s^{-1}}$) against Time Step every $[(144m + 6) + 108]$

3.6 Conclusions

Chaotic characteristics were established for wind speeds within Missouri for various height levels using both a qualitative method, phase space reconstruction, and a quantitative method, determination of the Lyapunov Exponent. Having established that the time series were chaotic in nature, a non-linear prediction algorithm was applied. Empirical Dynamical Modeling employed the dynamics of the time series instead of a set of governing equations. It was determined that this non-linear statistical method, which utilizes phase space reconstruction is relatively accurate as it is comparable to persistence. It however, beats this benchmark model for a prediction horizon of 1 time steps (10 minutes). It thus works best in the very short term as expected and predicted by prediction horizon given by the LLE. The normalized errors using a 50% cap also corroborate this time scale. This prediction model showed, as expected, the errors are increasing as the prediction time step increases; the errors eventually plateau. Seasonality and diurnal effects of the model were also investigated. It was shown that for the summer, four of the eight height levels had the lowest magnitude of average errors and standard deviations. This is expected, as there is less synoptic forcing in Missouri for the summer. It was also determined that the model handles the transitions of the diurnal cycle well as there is no time of day considered for which the model incurred more errors and had greater standard deviations for all stations and height levels.

The limitations of this study are due to the chaotic system suitability context that the data should be noise free and deterministic to fully capture the dynamics of multi-dimensional system by one observation. Thus forecasting using this method is not a simple task due to the non-linearity of the dynamical system and because the noise hides this dynamic (Lisi and Villi, 2001).

It will be interesting to test this model with exogenous variables. Future work entails comparing the results obtained from this model with the findings from Artificial Intelligence (AI) methods such as Artificial Neural Network (ANN).

References

Balkissoon, Fox and Lupo (2020). “Fractal characteristics of tall tower wind speeds in Missouri”. In: Renewable Energy.

Brandstäter, A et al. (1983). “Low-dimensional chaos in a hydrodynamic system”. In: Physical Review Letters 51.16, p. 1442.

Bryant, Paul, Reggie Brown, and Henry DI Abarbanel (1990). “Lyapunov exponents from observed time series”. In: Physical Review Letters 65.13, p. 1523.

Camplani, Massimo and Barbara Cannas (2009). “The role of the embedding dimension and time delay in time series forecasting”. In: IFAC Proceedings Volumes 42.7, pp. 316–320.

Cao, Liangyue (1997). “Practical method for determining the minimum embedding dimension of a scalar time series”. In: Physica D: Nonlinear Phenomena 110.1-2, pp. 43–50.

Chang, Chun-Wei, Masayuki Ushio, and Chih-hao Hsieh (2017). “Empirical dynamic modeling for beginners”. In: Ecological Research 32.6, pp. 785–796.

De Domenico, M and M Ali Ghorbani (2010). “Chaos and scaling in daily river flow”. In: arXiv preprint arXiv:1002.0076.

De Freitas, Nicksson CA, Marcelino P dos S Silva, and Meiry S Sakamoto (2018). “Wind Speed Forecasting: A Review”. In: Int. J. Eng. Res. Appl 8, pp. 4–9.

Eckmann, J-P and David Ruelle (1992). “Fundamental limitations for estimating dimensions and Lyapunov exponents in dynamical systems”. In: Physica D: Nonlinear Phenomena 56.2-3, pp. 185–187.

Farmer, J Doyne and John J Sidorowich (1987). “Predicting chaotic time series”. In: Physical review letters 59.8, p. 845.

Fox (2011). “A tall tower study of Missouri winds”. In: Renewable Energy 36.1, pp. 330–337.

Fraser, Andrew M and Harry L Swinney (1986). “Independent coordinates for strange attractors from mutual information”. In: Physical review A 33.2, p. 1134.

Garcia, CA and G Sawitzki (2015). “nonlinearTseries: Nonlinear Time Series Analysis”. In: R Package Version 0.2 3.

Kantz, Holger and Thomas Schreiber (2004). Nonlinear time series analysis. Vol. 7. Cambridge university press.

Kawauchi, Seiji, Hiroaki Sugihara, and Hiroshi Sasaki (2004). “Development of very-short-term load forecasting based on chaos theory”. In: Electrical Engineering in Japan 148.2, pp. 55–63.

Kennel, Matthew B, Reggie Brown, and Henry DI Abarbanel (1992). “Determining embedding dimension for phase-space reconstruction using a geometrical construction”. In: Physical review A 45.6, p. 3403.

Kliková, B and Aleš Raidl (2011). “Reconstruction of phase space of dynamical systems using method of time delay”. In: Proceedings of WDS. Vol. 11, pp. 83–87.

Koçak, Kasım (2002). “A method for determination of wind speed persistence and its application”. In: Energy 27.10, pp. 967–973.

Lisi, Francesco and Vigilio Villi (2001). “Chaotic forecasting of discharge time series: A case study 1”. In: JAWRA Journal of the American Water Resources Association 37.2, pp. 271–279.

Masseran, Nurulkamal et al. (2012). “Evaluating the wind speed persistence for several wind stations in Peninsular Malaysia”. In: Energy 37.1, pp. 649–656.

Potter, Cameron W and Michael Negnevitsky (2006). “Very short-term wind forecasting for Tasmanian power generation”. In: IEEE Transactions on power systems 21.2, pp. 965–972.

Rosenstein, Michael T, James J Collins, and Carlo J De Luca (1993). “A practical method for calculating largest Lyapunov exponents from small data sets”. In: Physica D: Nonlinear Phenomena 65.1-2, pp. 117–134.

Smith, Kevin et al. (2002). Evaluation of wind shear patterns at midwest wind energy facilities. Tech. rep. National Renewable Energy Lab., Golden, CO (US).

Soman, Saurabh S et al. (2010). “A review of wind power and wind speed forecasting methods with different time horizons”. In: North American Power Symposium 2010. IEEE, pp. 1–8.

Stark, Jaroslav (1994). “Analysis of time series”. In: Centre for Nonlinear Dynamics and its Applications, University Collage London.

Szpiro, George G (1997). “Forecasting chaotic time series with genetic algorithms”. In: Physical Review E 55.3, p. 2557.

Takens, Floris (1981). “Detecting strange attractors in turbulence”. In: Dynamical systems and turbulence, Warwick 1980. Springer, pp. 366–381.

Yu, Kaijun et al. (2019). “Comparative study of chaos identification methods for wind speed time series under different environmental measurement”. In: Ekoloji Dergisi 107.

Zeng, Ming et al. (2012). “Nonlinear analysis of the near-surface wind speed time series”. In: 2012 5th International Congress on Image and Signal Processing. IEEE, pp. 1893–1897.

Zounemat-Kermani, Mohammad and Ozgur Kisi (2015). “Time series analysis on marine wind-wave characteristics using chaos theory”. In: Ocean Engineering 100, pp. 46–53.

Chapter 4

Classification of tall tower meteorological variables and forecasting wind speeds in Columbia, Missouri

4.1 Abstract

Tall tower meteorological variables in Columbia, Missouri are clustered using Self-Organizing Maps after the optimal number of clusters was determined using the Elbow and Silhouette methods among others. The optimal number of clusters, k was given as 4 for all methods. The data were then grouped into three Intervals which consisted of approximately 50 percent and over of vectors or rows from the data frame. These intervals were then used as training and testing for the forecast models of Long Short-Term Memory Networks with pressure and wind speeds as inputs as well as lagged wind speeds as inputs. Other models using these intervals in our analyses include Moving Autoregressive Integrated Moving Average (ARIMA) and persistence. From the results obtained from the ARIMA, the metric of the root mean square error (RMSE) ranged from approximately 0.6 to 1.0 $\mathrm{m\,s^{-1}}$ for forecast horizon 2 to 12 in increments of 2. Interval2 had the upper and lower values and thus showed most variability in errors because it encompassed most of spring, all of summer and the beginning of fall. The moving ARIMA showed lower errors than the LSTM with pressure and wind speeds inputs for all the intervals. This may be attributed to the difficulty in representing the system's non-linearity and high dimensionality by using just the wind speeds and pressure as inputs. The lagged co-ordinates of the wind speed was then examined and used as inputs for the LSTM. The metric used for the evaluation of prediction of the forecast horizons of 6, 12, 18, 24, 30 and 36 or 1, 2, 3, 4, 5 and 6 hours ahead is the Normalized Root Mean Square Error (NRMSE). These models were compared to the benchmark model of persistence. It was determined that all of the models beat persistence and the LSTM with the lag series outperforms the LSTM with pressure and wind speed as inputs. The Moving ARIMA is now beaten by the lagged series LSTM in all intervals for at least 2 time forecast horizons of 6 and 12 or 1 and 2 hours.

Keywords: Self-Organizing Maps (SOMs), Autoregressive Integrated Moving Average (ARIMA), Long Short-Term Memory (LSTM) Networks

4.2 Introduction

4.2.1 Wind speeds

Wind speeds closer to the ground are subjected to resistance and friction. Even though these winds are highly positively correlated with each other, the correlations grow weaker with height as noted in both this study and Cao et al.'s (Cao, Ewing, and Thompson, 2012). Due to local surface characteristics and large scale forcing mechanisms such as pressure and temperature differences, wind is one of the most difficult meteorological variables to forecast (Sfetsos, 2000). Also, since the atmosphere is highly nonlinear and high dimensional, it is especially difficult to forecast this variable in the much needed higher resolutions and longer time horizons (Hu et al., 2019). The higher resolution shows more details of the faster variations in wind speeds caused by turbulence and other factors. The importance of such forecasts stems from the ability to aid in the scheduling, dispatching and adjusting electricity reservations (Hu et al., 2019). In our work we are looking at short term forecasting at high resolution (10 minute wind speeds at hub height).

4.2.2 Forecasting of wind speeds

Due to the stochastic nature of wind speeds, forecasting this variable is important for its optimal integration into the power grids (Sandhu, Nair, et al., 2019). These short term forecasts, can be used by plant managers to adjust turbine components to achieve more efficiency. Another advantage of short term forecasts is the ability to make turbines operable closer to extreme weather events before shutting down. Daily short term forecasts are also important as they relate to the operability of the turbines in terms of their cut-in and cut-off wind speeds as they aid in the reduction of structural damage to infrastructure (Cao, Ewing, and Thompson, 2012).

There are numerous methods that have been used to forecast wind speed values, some of which are illustrated in Figure 4.1. Please see the acronyms [1] associated with this chart. The methods incorporated in this paper are Artificial Intelligence (AI) methods which are compared to statistical models as well as the benchmark of persistence. The significant difference between these two methods is that statistical multiple linear regression is written in terms of a set of linear operators whilst Ar-

---

[1] AI-Artificial Intelligence, ANN-Artificial Neural Networks, ANFIS-Adaptive Neuro-Fuzzy Interface System, SVM-Support Vector Machine, AR-Auto-Regressive, MA-Moving Average, ARMA-Auto-Regressive Moving Average, ARIMA-Auto-Regressive Integrated Moving Average

tificial Neural Networks(ANNs) are representative of a linear combination of simple nonlinear functions (Basheer and Hajmeer, 2000). A mapping is done from random input vectors to output vectors without the assumption that there is a fixed relationship between the two (Li and Shi, 2010). ANNs have the ability to learn from past data by recognizing patterns among the observations and using these to forecast into the future (Li and Shi, 2010). Research has indicated the superiority in prediction accuracy of ANNs to statistical regression especially as the non-linearity of the problem increases (Basheer and Hajmeer, 2000). Previous studies, namely (Balkissoon, Fox and Lupo, 2020) and (Balkissoon et al., 2021), done in Missouri, found the wind speeds to be chaotic in nature, hence motivating the choice of this method to address the complexity and non-linearity of the data.

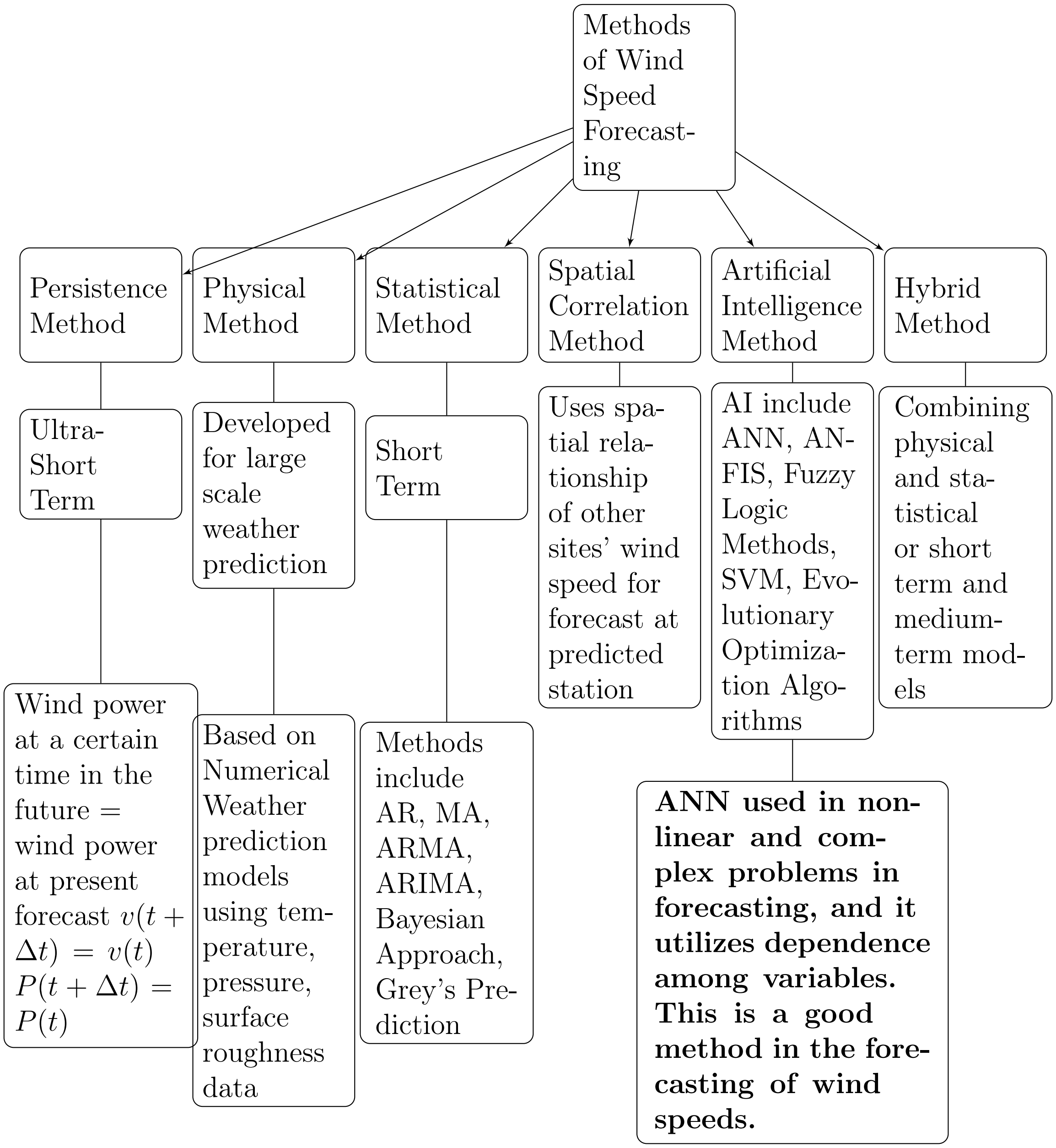
4.2.3 Wind Power

From the relationship $P = \frac{1}{2}\rho A V^3$ where P is the available power at the turbine,ρ is the density of air, A is the area swept by the turbine and V is the wind speed, the two meteorological variables which determines the available turbine power are ρ and V. The latter variable has the greater influence as the power varies as the cube of V. The air density is dependent on pressure and temperature as seen from the following equation (Mabel and Fernandez, 2008)

$$\rho = D\left(\frac{273.15}{T}\right)\left[\frac{B - 0.3783e}{760}\right] \tag{4.1}$$

where D is $1.168kg/m^3$ - the density of dry air at standard atmospheric temperature $(25°C)$ and pressure $(100kPa)$ and B is the barometric pressure in torr, e is the moist air vapour pressure in torr. Hence, as seen in subsequent sections of the methods, these two meteorological parameters, will be considered, together with wind speeds, when determining the inputs to the Neural Network.

Figure 4.1: Methods for wind speed forecasting



4.3 Data

Columbia, Missouri is located in 038°53.270′N latitude and 092°15.820′W longitude and has a site elevation of 255m as seen in Figure 4.2. Ten-minute tall tower wind speed, wind direction and temperature data in 2009 from this region were used in our study (Fox, 2011). The respective units are $\mathrm{m\,s^{-1}}$, degrees and degrees Celsius, respectively. The anemometer orientations were 120°and 300°for the tall tower height of 68 m. Channels 1 and 2 represent the respective wind speed times series. The larger of the wind speed values at each time step were taken and labelled as Max1. The wind direction time series at this height level was given from channel 7 and sine of these angles was labelled as Direction1. The temperature time series from 2 m logger height were also utilized in our analyses, taken from channel 11, it is referred to as Temp. Hourly maximum pressure data was taken from University of Missouri, Extension's Missouri Historical Agricultural Weather Database. Each hourly pressure value was repeated five times. This time series, labelled as Pressure, along with Max1, Direction1 and Temp were combined and used in all of the analyses for Columbia68 as detailed in the Methods section below.

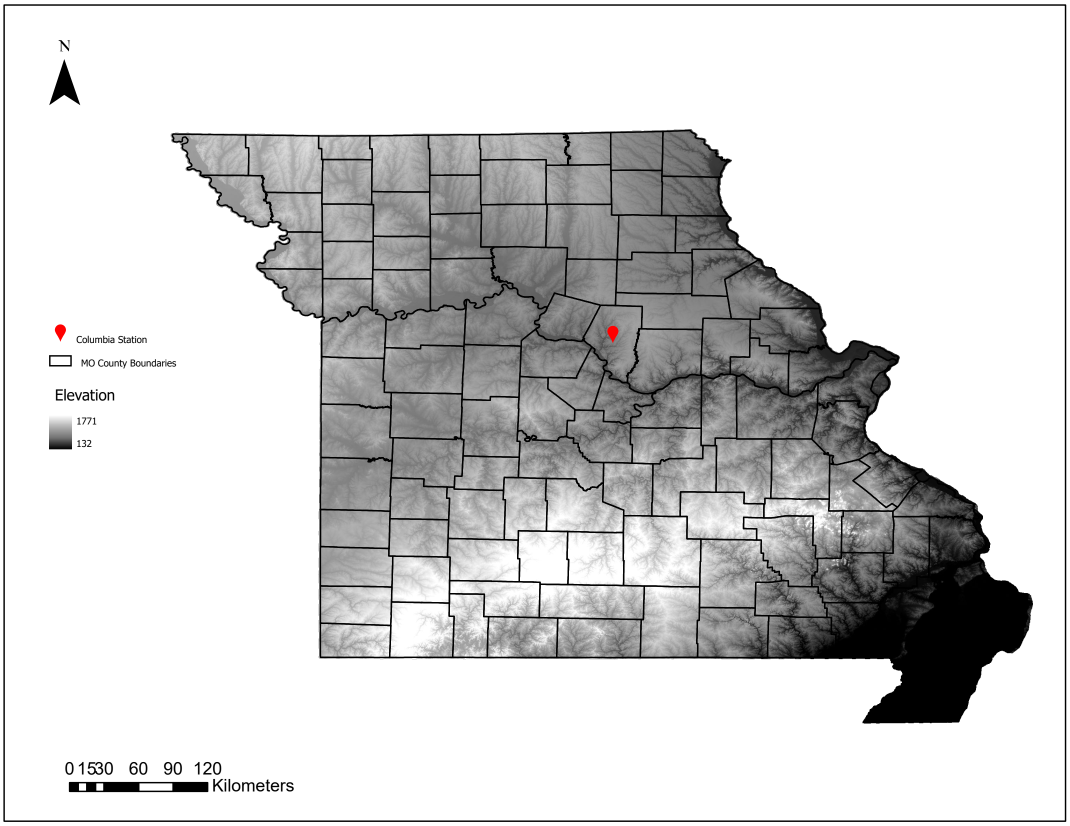


Figure 4.2: Tall tower location

4.4 Methods

Artificial Neural Networks (ANN) have to date been applied to a multitude of fields in solving complex problems. These data-driven models are utilized especially as physically-based mathematical models are difficult to construct given the high non-linearity of natural systems (Lin, Wang, and Chen, 2016). As defined by Basheer and Hajmeer (Basheer and Hajmeer, 2000) and Ramasamy et al. (Ramasamy, Chandel, and Yadav, 2015) , ANNs can be considered as a system of densely interconnected processing elements, also called artificial neurons or nodes, which have the ability to conduct parallel computations of input data .Complex relationships are derived from the input and the output. The input variables are multiplied by weights and biases are added to these products. These are then passed through transfer functions for the generation of the outputs (Ramasamy, Chandel, and Yadav, 2015).

ANNs, being abstractions of biological systems, have the advantages of processing data that are highly nonlinear. These robust systems have the ability to learn and generalize imprecise and fuzzy data (Basheer and Hajmeer, 2000). Data are allowed to be processed faster and have a better fit amidst inaccuracies from measurement errors. The system in its learning is self updating and has the ability to unlearn data as well (Basheer and Hajmeer, 2000).

There are many applications of ANNs, which include modelling, classification, pattern recognition and multivariate data analysis problems (Basheer and Hajmeer, 2000) (Ramasamy, Chandel, and Yadav, 2015). Here, we will focus our attention on the clustering of the data into various clusters and then on subsequent modelling and forecasting. Clustering as described by (Basheer and Hajmeer, 2000), is formed by investigating the similarities and differences of the inputs based on their inter-correlations. Kohonen networks or Self-Organizing Feature Maps (SOMs), the unsupervised learning ANN where the actual values are not required for the training set, are used in this study. Forecasting is also done by training the ANN using a training set of historical data. A Recurrent Neural Network (RNN) is utilized especially for its dynamic memory capabilities where the outputs of neurons are fed as inputs to the same neurons or other neurons in the preceding layers (Basheer and Hajmeer, 2000). Details of these methods are given in the following sub-sections.

4.4.1 Methods determining the number of clusters

Clustering analyses as mentioned in (Nourani et al., 2012), are statistical methods which are used to partition multivariate data into subsets. There are numerous methods that can be used to determine the optimal number of clusters to classify the data into relatively homogeneous groups, such as the Elbow Method, Silhouette Analysis and Gap Statistic. These methods, which are used in this study, are outlined below. We have considered in the analyses this multivariate data set (of length n) of variables wind speed, direction, temperature and pressure. We denote these points as x_i for $i = 1, \ldots, n$.

The Elbow Method is determined by plotting the within-cluster sum of squares (WCSS) against the number of clusters, say k. The $WCSS(k)$ gives the sum of the squared distances between each data point, say x_i, in all clusters and their associated centroids denoted as $\bar{x_j}$ (which is the geometric center or the arithmetic mean position of all the points in the plane figure). This can be written as follows.

$$WCSS(k) = \sum_{j=1}^{k} \sum_{x_i \varepsilon cluster j} \|x_i - \bar{x_j}\|^2 \tag{4.2}$$

The changes in $WCSS$ with a range of k determines the optimal number of clusters in accordance with the Elbow method. The value in which k elbows or the point where the rate of decrease in $WCSS$ is relatively minimal when compared to its previous k values.

The Silhouette Clustering method was also used in our analysis. This method examines the within cluster-consistency by comparing how similar objects from a cluster are to another. Its value, $S(i)$ range from -1 to 1 where the lower end of the interval indicates that the configuration has too much or few clusters. The closer this value is to 1 however, is indicative of an object that is well matched to its cluster or poorly matched to the other clusters.

The mean similarity of point i and all other points in the same cluster, C_i, is given by equation 4.3 where $|C_i|$ denotes the number of elements in C_i and $d(i,j)$ give the distances between data points i and j in cluster C_i. In this average the distances,

$d(i,i)$ are not considered hence the consideration of $|C_i|-1$ in the formulation below.

$$a_i = \frac{1}{|C_i|-1} \sum_{j \varepsilon C_i, i \neq j} d(i,j) \tag{4.3}$$

The smallest mean dissimilarity of point i and all the other points of another cluster, C_k, is given by b_i in equation 4.4.

$$b_i = min_{k \neq i} \frac{1}{|C_k|} \sum_{j \varepsilon C_k} d(i,j) \tag{4.4}$$

This is the second best fit cluster for point i based on the distance metric. The Silhouette value for point i, we define as S_i, is given in terms of a_i and b_i as seen in equation 4.5.

$$S_i = \begin{cases} \frac{(b_i)-(a_i)}{max\{a_i,b_i\}} & \text{if } |C_i| > 1 \\ 0 & \text{if } |C_i| = 1 \end{cases} \tag{4.5}$$

This can be further simplified as seen below, depending on the inequality relations between the mean similarity and dissimilarity.

$$S_i = \begin{cases} 1 - \frac{a_i}{b_i} & \text{if } a_i < b_i \\ 0 & \text{if } a_i = b_i \\ \frac{b_i}{a_i} - 1 & \text{if } a_i > b_i \end{cases} \tag{4.6}$$

The Gap Statistic, another consideration used in this paper, is outlined as follows. As previously denoted, let C_i be the i^{th} Cluster and $|C_i|$ be representative of the number of elements in this cluster. Let the pairwise distances between elements say i and j in C_i, d_i, be given by equation 4.7.

$$d_i = \sum_{i,j \varepsilon C_i} d(i,j) \tag{4.7}$$

For a given number of clusters k, the within cluster distance for that particular partitioning P_k, is given by equation 4.8. A better classification is indicative of a smaller W_k value.

$$W_k = \sum_{i=1}^{k} \frac{1}{2\left|C_i\right|} d_i \tag{4.8}$$

Considering the data in which the 'true' number of clusters is given by G, W_k should drop as k increases until it reaches G where it will decrease at a much slower rate. Thus, there will be an 'elbow' point in W_k; this k value corresponds to the optimal number of clusters. The Gap Method is used to compare the original data with the expected curve, $E_n^* \left\{log(W_k)\right\}$ where E_n^* gives the expectation of sample n from the reference distribution. The Gap Statistic is the value of k which maximizes $Gap_n(k)$ or from equation 4.9, the cluster value where W_k is at the furthest distance from the expected curve (Yan and Ye, 2007).

$$Gap_n(k) = E_n^* \left\{log(W_k)\right\} - log(W_k) \tag{4.9}$$

4.4.2 Self Organizing Maps (SOM)

A Self Organizing Map (SOM) is an unsupervised clustering method as there is no additional information being supplied to the model by a 'supervisor' (Kalinić et al., 2015). In this model, high-dimensional data sets are reduced to the two-dimensional map in which nodes with most similarity are nearest to each other and vice-versa (Pearce et al., 2014). It does this dimensionality reduction via the usage of cluster centers which can then be interpreted as an 'abstract representation' of any given vector from that particular cluster (Kalinić et al., 2015). It preserves topology where vectors that are near in input space are also mapped to nearby neurons in the SOM (Nourani et al., 2012)(Pearce et al., 2014). This resulting map is a projection of a multidimensional space rather than a geographical space (Pearce et al., 2014). There are two modes of operation, training which builds the map using input examples through a method called vector quantization and mapping which classifies new input vectors (Burguillo, 2014).

This Kohonen Neural Network is used in many applications (Wehrens, Buydens, et al., 2007) such as Pearce et al.'s air quality classifications (Pearce et al., 2014) and in geoscience for the extraction of climate and atmospheric circulation patterns (Kalinić et al., 2015). Previous studies using SOMs also include Berkovic's (Berkovic, 2017) determination of the wind regimes, choosing from various map sizes, the number of nodes in the rows and columns. However in our study, since we utilized SOMs for the purposes for clustering our data to be later incorporated in our forecasting algorithm, we defined our map size based on the formulation written in (Tian, Azarian, and Pecht, 2014). The number of neurons, M of the map is determined from the number of observations, N. It is given by the following expression (Burguillo, 2014).

$$M \approx 5\sqrt{N} \tag{4.10}$$

According to (Browell, Drew, and Philippopoulos, 2018), the methodology of the SOM can be achieved via the processes of competition- where the Best Match Unit (BMU) is identified, cooperation -where the topological neighbourhood of the 'excited' neurons are identified and finally adaptation -where BMU and excited neurons are updated in accordance to the input vector.

In more detail the methodology of the SOM is as follows (Tian, Azarian, and Pecht, 2014).

1. The weight vector of each of the neurons in the map is initialized randomly.
2. The training observed data, say x_t, is 'passed' to the map as an input vector and Euclidean distance between the all the neurons and this vector is calculated.

The neuron with the smallest distance is termed the Best Matching Unit or (BMU). For each input observation, the BMU is identified. We denoted this unit as c henceforth.

3. A neighbourhood of c is selected and using a neighbourhood function given by h_{ci}, the weighted vectors of the neighbouring neurons, i are updated.

   $$h_{ci}(t) = a(t)e^{-\frac{\|r_c - r_i\|^2}{2R^2(t)}} \tag{4.11}$$

   $$W_i(t+1) = W_i(t) + h_{ci}(t)\left[x_t - W_i(t)\right] \tag{4.12}$$

   Where, from equation 4.11, h_{ci} is the neighbouring function and t is an index of iteration, $a(t)$ is the learning rate, r_c is vector of c, r_i is the vector of the neuron i and R is the radius around c. This function is a monotonically decreasing function of t as the learning rate decreases with the iterations during the training process and the radii around c decreases with t. This process ensures that neurons i closest to c are being adjusted the most.

   The neurons are updated in accordance to equation 4.12 where $W_i(t+1)$ and $W_i(t)$ represent the weighted vector of neuron i at the $t+1$ and t indices of iterations respectively, h_{ci} is the neighbourhood function above and x_t is the observed input vector.

4. This process is repeated in the iterative training until the clusters are identified based on their distances.

The data described in Section 4.2, were normalized between 0 and 1 by subtracting from each element in that particular column of the data frame, its mean. These values are then divided by the standard deviation of the column to give the z or standard scores. This standardizing of the variables was done using the scale command. These analyses were done in R studio (Wehrens, Buydens, et al., 2007). The SOM grid was then created using the relation of 4.10 where $N = 52,560$ data observations for each variable. The grid size used was 41 by 28 of hexagonal nodes corresponding to the factor pair of 1,148. This value was used instead of the calculated numeric of 1,146 because it had more factor pairs.

The following is a list of the metrics to be plotted and their description will be shown in the results section.

1. Node Count- This map gives the number of samples that are mapped to each of the nodes of the map. This value should be relatively uniform throughout the SOM. Large values in some areas of the map is indicative for the need of a larger map whilst empty nodes indicates that a smaller map may be more appropriate. Generally, it is used to determine high density areas in the map where ideally there should be a homogeneous distribution (Rakotomalala, 2005).

2. Neighbourhood Distance or U-Matrix- This map gives the distance between each node and its neigbouring neurons. It represents the Euclidean distance amongst the codebook vectors of the respective neighbourhoods (Rakotomalala, 2005). Larger distances indicates dissimilarities and thus cluster boundaries as nodes from the same cluster have the tendency to be closer.

3. Heat Maps - These maps separately give the distribution of each of the parameters throughout the map. These are done for the four variables, both scaled and unscaled.

4. Clustering of the codebook vectors - This map consists of the codebook vectors which is the data structure that carries the neuron's weight vector in a 2D grid. The number of clusters or groups is input as well as the specification to add the cluster boundaries.

After the clusters are identified, the cluster associated with each of the $x(t)$ vectors was determined. Continuous intervals of the clustered 2009 Columbia, MO data set, representing approximately 50% and over of data points in that particular cluster, were established. There were three intervals in which the majority of the vectors or rows from the data frame belonged to two of the identified four clusters, denoted Cluster1 to Cluster4 (we will explain more of this in the subsequent results section). For example, Interval1 ranged from 1 to 16,000 rows in which Cluster3 consisted of 50.2% of the vectors. Interval2 which started at 16,001 and ended 40,500 inclusively, comprised 78.92% of rows from Cluster2. Interval3 included vectors from 40,501 to 52,500 in which Cluster3 represented approximately 48% of this interval. It should be noted that there were predominately two clusters which we will also show in subsequent results of the clustering of the codebook vectors. Another note to mention is that the entire time series of length was not used. Instead, 52,500 rows were utilized in our analyses. There were 16,000, 24,500, and 12,000 points in Cluster1, Cluster2 and Cluster3 respectively.

These intervals are then separately trained and tested in time series forecasting using the Recurrent Neural Network explained in the subsequent subsection.

4.4.3 Recurrent Neural Networks (RNN), Long Short- Term Memory Networks (LSTM)

Recurrent Neural Networks (RNN) allow information to persist via one or more hidden states and loops that pass information from one step to another of the network. However, for this, there exists the vanishing gradient problem as the gradients asymptotically reduce to 0 from the repeated multiplication of weights for various time steps. Long Short-Term Memory networks (LSTMs) are a special type of RNN that can learn these long-term dependencies. The LSTM has memory blocks called cells where information is stored in the cell state, c_t and the hidden state, h_t. A diagrammatic representation of the architecture of such memory blocks or cells is seen in Figure 4.3. Information is regulated by gates by optionally allowing certain data through using sigmoid and tanh activation functions. The output of the sigmoid function is a number between 0 and 1, where 0 and 1 mean no and all information goes to the cell state, respectively. Generally notated, the inputs to the gates are the output hidden state from the previous step , h_{t-1}, and the output cell state from the previous step, c_{t-1} and current input, x_t, which are pointwise multiplied by weight matrices, W, and then added to a bias, b.

There are three major gates: the forget, the input, and the output gates.

1. The forget gate: As seen in Figure 4.3, the input of this gate is x_t and h_{t-1} for that time step. These inputs are multiplied by weight matricies and added to a bias. This value is then inputted to the sigmoid function and a vector is outputted which corresponds to each value in the cell state, c_{t-1}. Please refer to equation 4.13. This vector output is multiplied to the cell state. If a 0 is output from the sigmoid function for a particular value, the forget gate wants the cell state to disregard that information whilst if 1 is the sigmoid output, the forget gate wants the cell state to remember this data.

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \tag{4.13}$$

2. The input gate: The gate determines the information being stored in the cell state. The sigmoid layer decides the data to be updated and the tanh layer, whose output values ranges from -1 to 1, creates a vector of possible values that could be added to the cell state. Please refer to equations 4.14 and 4.15.

$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] + b_i\right) \tag{4.14}$$

$$g_t = \tanh\left(W_g \cdot [h_{t-1}, x_t] + b_g\right) \tag{4.15}$$

The old cell state, c_{t-1} is then used to update the new cell state c_t. This is done representatively by equation 4.16. The old state is multiplied by f_t to forget the information decided upon earlier and then it is added to the product of i_t and g_t which is indicative of the new possible values scaled to the update amount decided upon for each value. Note that $*$ is representative of the Hadamard or entrywise product.

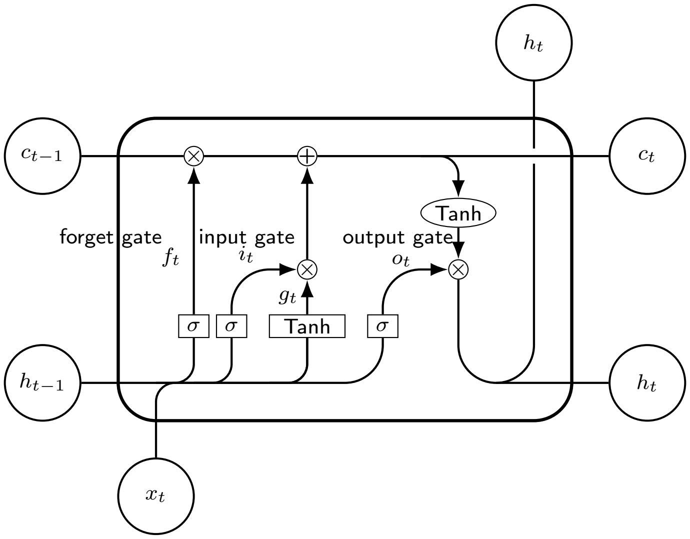
$$c_t = (f_t * c_{t-1}) + (i_t * g_t) \tag{4.16}$$

3. The output gate: A vector is created from scaling the values in the cell state using a tanh function. The sigmoid function is once again used as a filter to regulate what is to be outputted from the vector mentioned previously. This can be represented by equation 4.17. This is sent as the output and as the hidden state of the next cell.

$$o_t = \sigma\left(W_o \cdot [h_{t-1}, x_t] + b_o\right) \tag{4.17}$$

$$h_t = o_t * \tanh\left(c_t\right) \tag{4.18}$$

Figure 4.3: LSTM Architecture



4.4.4 Moving AutoRegressive Integrated Moving Average Method (ARIMA)

A moving AutoRegressive Integrated Moving Average Method (ARIMA) is used as another model in our analysis. This is a statistical method which uses the relationship within the time series data in its construction. Data cannot be white noise, that is, purely random with mean = 0 and standard deviation being a constant as forecasting into the future would not be possible. If this condition is met, AutoRegressive, AR(p), Moving Average, MA(q) and AutoRegressive Moving Average, ARMA(p,q) methods can be utilized. If the data are not stationary (that is not constant mean and variances), differencing needs to be performed. An AutoRegressive Integrated Moving Average, ARIMA (p,d,q) can be used where the Integrating part represents the *d* or the differencing factor.

The AR method, a time series model, is regressed from its previous values up to an order determined by the *p* parameter. This can be seen mathematically from equation 4.19. The Partial Autocorrelation function (PACF) determines how many lags are to be incorporated in the AR method; large PACF values gives the order of the model. For lag p, the relationship between x_t and x_{t-p} is determined, filtering all the intermediate linear influence from $x_{t-1}, x_{t-2}, \ldots, x_{t-(p-1)}$.

$$x_t = \beta_0 + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \ldots + \beta_p x_{t-p} + \varepsilon_t \tag{4.19}$$

Where x_t and x_{t-1}, $\ldots$,x_{t-p} are the current and previous values respectively and β_0 is a constant term and β_1,$\ldots$,β_p are the coefficient representing what part of x_{t-1},$\ldots$, x_{t-p} are relevant in explaining the current value etc.

The MA model is written in terms of a linear combination of past error. It gives the extent the series is related to its past errors. Generally it can written as equation 4.20. The Autocorrelation function determines the number of lags for the MA model. It is given by the lag value which is statistically different from 0 and above the error band, followed by consecutive insignificant ACF values for subsequent lags.

$$x_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q} \tag{4.20}$$

Where x_t is the current value, ε_t and $\varepsilon_{t-1}, \ldots, \varepsilon_{t-q}$ are errors from the current and previous predictions respectively and $\theta_1, \ldots, \theta_q$ represent the corresponding part which is relevant in explaining the current value.

The ARMA method is the linear combination of the linear models, AR and MA as such they too are linear models. This method thus, takes into account past values and errors in its formulation. Generally it can be written as equation 4.21.

$$x_t = \beta_0 + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \ldots + \beta_p x_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q} \tag{4.21}$$

The differencing parameter,d is introduced in the ARIMA models to remove trends and seasonality. The first order difference is given by $\Delta_1 x_t = x_t - x_{t-1}$. This and higher orders can be written in terms of B, the backward shift operator where $Bx_t = x_{t-1}$ and $B\left(Bx_t\right) = B\left(x_{t-1}\right) = x_{t-2}$. Generally for shifting an observation some m periods, $B^m x_t = x_{t-m}$. Thus, the first and second differences in terms of operator B are $\Delta_1 x_t = x_t - x_{t-1} = x_t - Bx_t = \left(1 - B\right) x_t$ and $\Delta_2 x_t = \Delta_1 x_t - \Delta_1 x_{t-1} = \left(1 - B\right)^2 x_t$ respectively. The second difference can be shown to be via expansion, $x_{t-2} - 2x_{t-1} + x_t$.

To determine the number of differencing to use we examine the autocorrelations. If the series has positive autocorrelations out to a large number of lags then the series may need differencing. If for lag 1, the autocorrelation is zero or negative then the series does not need higher order differencing. However, if for lag 1 the autocorrelation is less than or equal to -0.5, then the series may be over-differenced. A model with no differencing implies that the series is stationary whilst the assumptions are made that for the first and second differencing of the series, the original series has a constant average trend and has time varying trends respectively. An ARIMA(1,1,0), ARIMA(0,1,1) and ARIMA(1,1,1) can be written mathematically as equations 4.22a and 4.22b, 4.23a and 4.23b, 4.24a and 4.24b respectively.

$$\Delta_1 x_t = \beta_0 + \beta_1 \Delta_1 x_{t-1} \tag{4.22a}$$

$$\Rightarrow x_t = \beta_0 + x_{t-1} + \beta_1 \left(x_{t-1} - x_{t-2}\right) \tag{4.22b}$$

$$\Delta x_t = c + \Theta_1 \varepsilon_{t-1} \tag{4.23a}$$

$$\Rightarrow x_t = c + x_{t-1} + \Theta_1 \varepsilon_{t-1} \tag{4.23b}$$

$$\Delta x_t = \beta_0 + \beta_1 \Delta_1 x_{t-1} + \Theta_1 \varepsilon_{t-1} \tag{4.24a}$$

$$\Rightarrow x_t = \beta_0 + x_{t-1} + \beta_1 \left(x_{t-1} - x_{t-2}\right) + \Theta_1 \varepsilon_{t-1} \tag{4.24b}$$

4.4.5 Model Configuration

Model: LSTM (pressure and wind speeds as inputs) The Pytorch structure of the codes for this model was motivated/developed by (Kent, n.d.).

- The data were loaded, preprocessed (by taking the larger wind speed of the orientations at each time step) and plotted.

- The target variable was specified as wind speed along with the forecast lead (how much we are forecasting ahead, h). The target was specified as the lag/shift of the wind speed by the forecast lead. The features were given as wind speed and pressure. The data were then split into the training (75%) and testing (25%) sets from the observations. The train and test data were then standardized where the values are not restricted to a particular bounding range like normalization.

- A sequence of observations from the train and test set were constructed. This sequence was given as a block of data from some i^{th} row - sequence length through row i. For i less than the sequence length, the 1^{st} row was padded by repeating it as many times deemed necessary. Thus, the outputs have the number of rows in the block equal to the sequence length.

- These sequences from data set was set in Pytorch's dataloader to select mini-batches. However, in our model the batch sizes selected were the entire respective train and test data sets for the Intervals. Thus we had two features (columns), fifty sequence length (rows) and one batch the length of the train and test sets.

- A shallow regression LSTM model was then utilized with one hidden layer of 100 hidden units. The loss function is used to calculate the error or the difference between the predicted and the actual values. The loss function chosen was Mean Square Error (MSE). The optimizer is used to make changes to the weights; it does this to try to lower the model loss function. The optimizer chosen was the Adaptive Moment Estimation (Adam) algorithm with a learning rate of 0.01. An epoch is the number of times the algorithm traverses the training data. The model was trained using 20 epochs and was then evaluated.

4.5 Results

From Figure 4.4, we can see that the elbow occurs at 4, indicative that this is the optimal k. In figure 4.5, 4 has the largest $S(i)$ value indicating that for $k = 4$, the objects are well matched to their respective clusters. Similarly, from Figure 4.6, the value which maximizes $Gap_n(k)$ is $k = 4$. From the analysis of multiple methods, the bar chart in Figure 4.7 indicates that most of the methods result in an optimal k of 4. This is an important consideration, as mentioned in Pearce et al. (Pearce et al., 2014), because a grid with too few classes losses important information via generalizations whilst too many classes will result in loss of statistical power as there will exist smaller within class sample sizes.

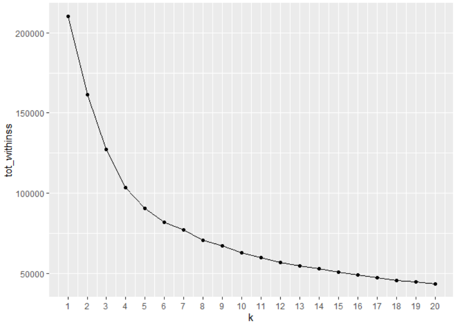


Figure 4.4: Elbow Method

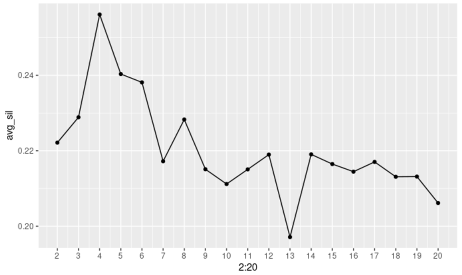


Figure 4.5: Silhouette Clustering Method

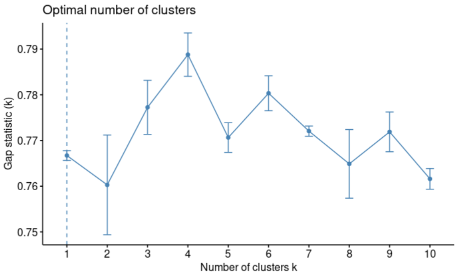


Figure 4.6: Gap Statistic

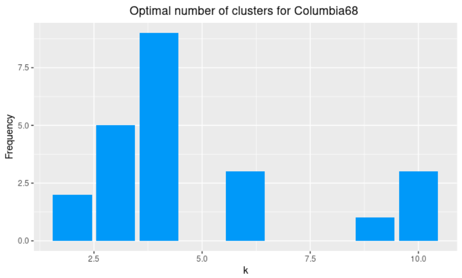


Figure 4.7: Methods determining optimal k

The grid was a hexagonal structure consisting of 1148 nodes. This structure consisted of no x and y axes but rather nodal positions, which were numbered as bottom left having the least value, whose node numbering increases from left to right (Lakshminarayanan, 2020). As mentioned in (Pearce et al., 2014), limitations of SOMs include its grid having a finite structure, which imposes restrictions on the map in the provision of precise information on clustering dissimilarity. Another restriction is using set of numbers to define the grid that in turn generalize its shape, be it a rectangle or a square (Pearce et al., 2014).

From the results of the SOMs, the node count plot can be seen in Figure 4.8. Since the distribution of the counts is relatively uniform throughout the domain of the SOM, the map size is appropriate. Figure 4.9 shows the neighbourhood distance in which cluster boundaries can be identified via large nodal distances. From this map, it is evident that there exist areas where there are greater distances representative of the upper end of the scale and the lighter colours. This is seen for example in the north eastern portion of the map. From the clustering of the codebook vectors in Figure 4.14, we do note that this is separated as part of a cluster. This is contained in a smaller cluster whilst there are two major clusters where the adjacent nodes are grouped in the same cluster. This grid also shows, for each node, all the variables (as colour coded) in various sector representations. The radii of the sectors varies with respect to its variable (Lakshminarayanan, 2020). The unscaled heat map for all of the variables used in this study are seen from Figures 4.10 to 4.13.

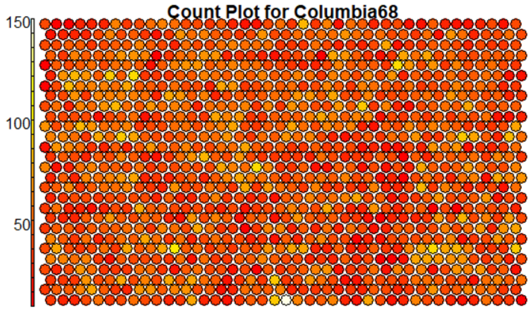


Figure 4.8: Node Count Plot

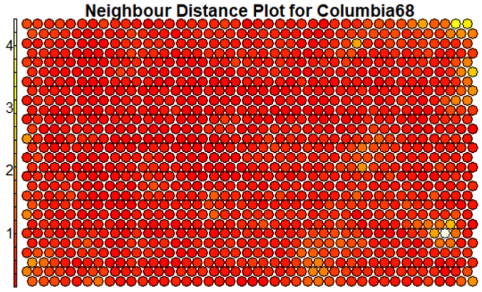


Figure 4.9: Neighbourhood Distance or U-Matrix

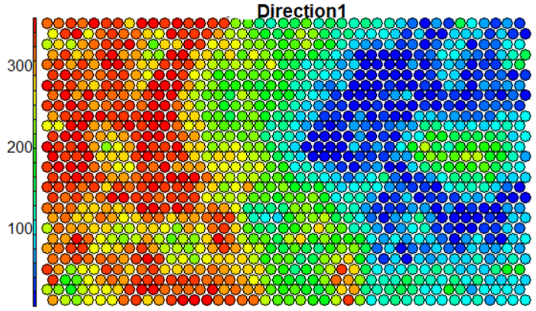


Figure 4.10: Heat Map- wind direction

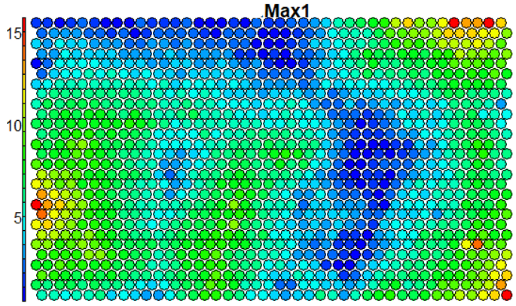


Figure 4.11: Heat Map- wind speed

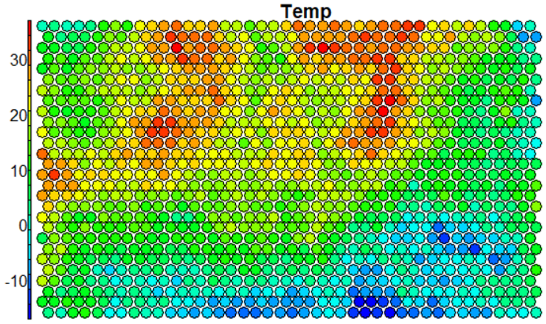


Figure 4.12: Heat Map- Temperature

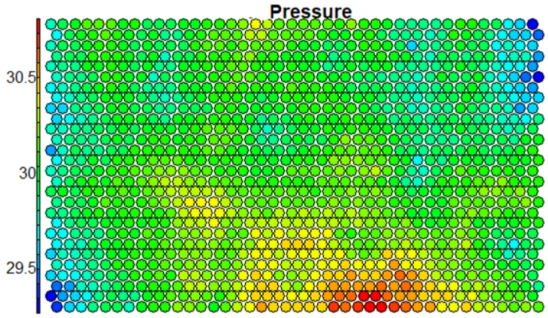


Figure 4.13: Heat Map- Pressure

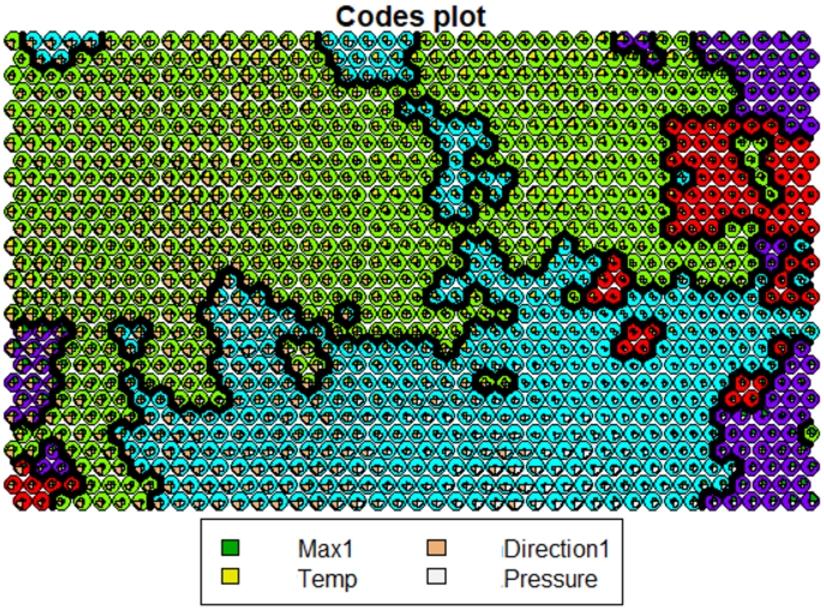


Figure 4.14: Clustering of codebook vectors

From the clusters of the SOMs, continuous intervals belonging to a particular cluster were identified. These intervals are representative of approximately 50% and more of the rows from the data frame belonging to a particular cluster where Interval1, Interval2 and Interval3 belonging to Cluster3, Cluster2, and Cluster3 ranged from 1 to 16,000, 16,001 to 40,500 and 40,501 to 52,500 rows respectively. The three intervals identified by our clustering are graphed in Figures 4.15, 4.16 and 4.17 where both the test and training sets are visualized. These clusters from the SOM were utilized to optimize model performance in forecasting as done in Browell et al.'s article (Browell, Drew, and Philippopoulos, 2018). The forecast horizon is from 20 minutes to 2 hours. As mentioned in (Browell, Drew, and Philippopoulos, 2018), for these time scales which are used to balance the power systems by operators, statistical methodologies inclusive of ARIMA are superior to that of results obtained from Numerical Weather Predictions (NWP). This can be attributed to its low computational cost and ease of including of new data (Browell, Drew, and Philippopoulos, 2018).

From table 4.1, the RMSE and the MAE for these intervals and various time steps, h, using the moving ARIMA model, can be seen. These values ranged from approximately 0.6 to 1.0 $\mathrm{m\,s^{-1}}$. These results are somewhat comparable to that of (Browell, Drew, and Philippopoulos, 2018). Browell et al. (Browell, Drew, and Philippopoulos, 2018) used vector autoregression in the spatial consideration of multiple locations and for this model they obtained RMSEs of 0.96, 1.55, 2.00 $\mathrm{m\,s^{-1}}$ for one, three and six hours ahead. Another study by (Sandhu, Nair, et al., 2019) using both hourly and 10 minute data in which 39 and 173 points were forecasted respectively for each data set, have RMSEs of 1.27 $\mathrm{m\,s^{-1}}$ for the hourly dataset and 0.96 $\mathrm{m\,s^{-1}}$ for the 10 minute dataset. For our time step or forecast horizon of h equal to 6 (one hour ahead), for Interval1 and Interval3 this value was approximately 0.8 $\mathrm{m\,s^{-1}}$ whilst for Interval2, it was an estimated 0.1 $\mathrm{m\,s^{-1}}$ less that the other two intervals.

In our analyses, the upper and lower values from this range resulted from the run of Interval2. This is expected as this interval encompassed most of the spring, all of the summer and the beginning of the fall. As such it is expected that the model shows the most variability in errors for this interval. It is expected as well that this interval has the lowest errors as it has highest learning ability of the neurons due to its largest training set (Lin, Wang, and Chen, 2016). This can be seen graphically in Figure 4.18. From Table 4.2, we see that the results were comparable to that of the intervals defined by the SOMs. We note also that spring has the largest RMSE from the moving ARIMA as expected due to the prevalence of convective storms.

The moving ARIMA was also trained using three quarters of the entire data set, despite having this advantage of more information variability in training/learning phase, these results did not deviate significantly from the interval and the seasonal analyses.

ANN are powerful and are frequently used in time series forecasting due to their high parallelism, among other characteristics (Ramesh Babu and Arulmozhivarman, 2012). However, the ARIMA model is widely used and has given more accurate results for very short term forecasts (Ramesh Babu and Arulmozhivarman, 2012).

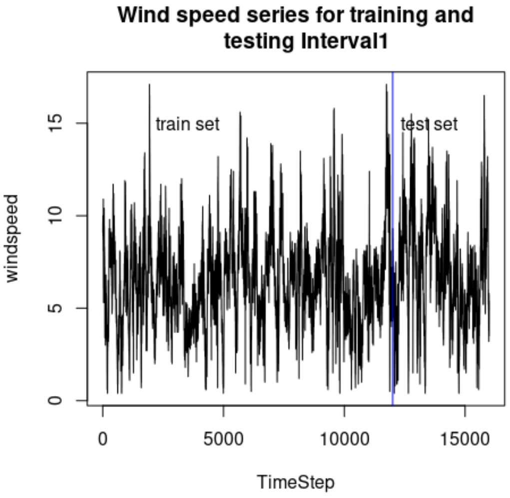


Figure 4.15: Interval 1

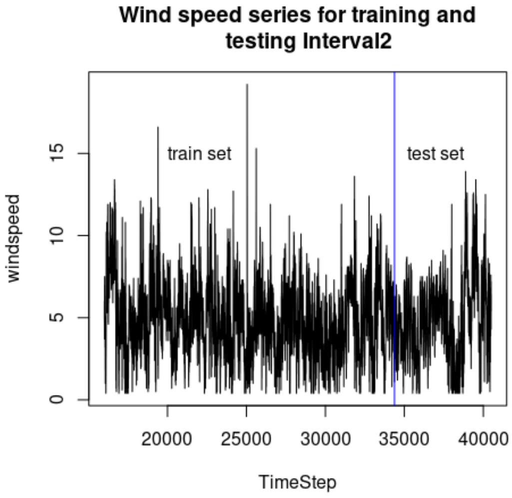


Figure 4.16: Interval 2

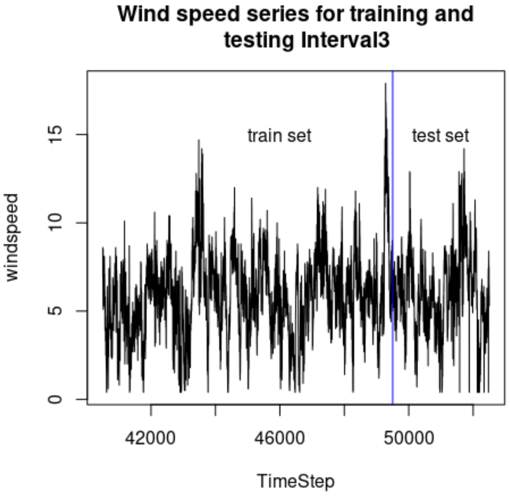


Figure 4.17: Interval 3

Table 4.1: Moving ARIMA Results for the Intervals

h	**Interval1**	**Interval1**	**Interval2**	**Interval2**	**Interval3**	**Interval3**
(10-mins)	$RMSE$	MAE	$RMSE$	MAE	$RMSE$	MAE
2	0.7624309	0.5587051	0.6152723	0.4480265	0.738208	0.4980114
4	0.7817789	0.5783413	0.665906	0.5017876	0.756581	0.5186229
6	0.8008848	0.5957457	0.7360105	0.5666321	0.7666935	0.5284694
8	0.8154731	0.6082994	0.8106469	0.6321702	0.7712227	0.5328341
10	0.8257323	0.6171721	0.8831231	0.6938833	0.7727915	0.5341433
12	0.832471	0.622839	0.9509406	0.7507482	0.7735839	0.5348155

Table 4.2: Moving ARIMA Results for 2009 data set and the seasons

h	**2009 data set- RMSE**	**Spring- RMSE**	**Summer- RMSE**	**Fall- RMSE**
2	0.6943994	0.763676	0.5139139	0.6625789
4	0.711852	0.7787825	0.5504232	0.68182
6	0.730347	0.7998983	0.5724217	0.7019841
8	0.7445687	0.8185765	0.5821634	0.7278857
10	0.7543516	0.8330311	0.5854124	0.7558021
12	0.7608936	0.8438839	0.5858682	0.7844518

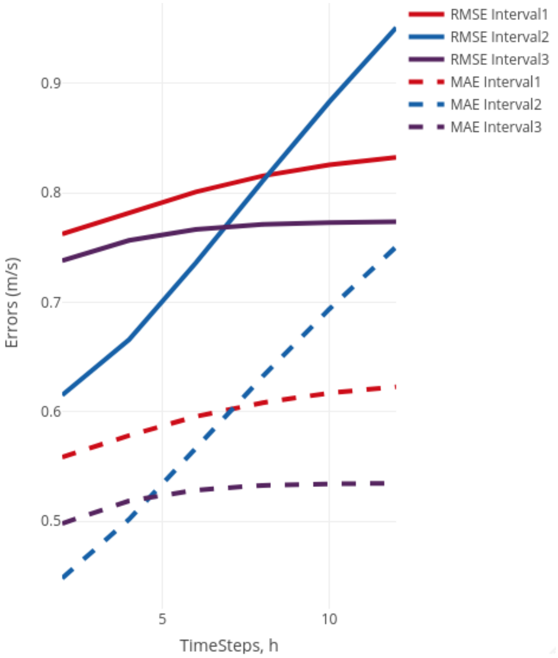


Figure 4.18: ARIMA Errors for the Intervals

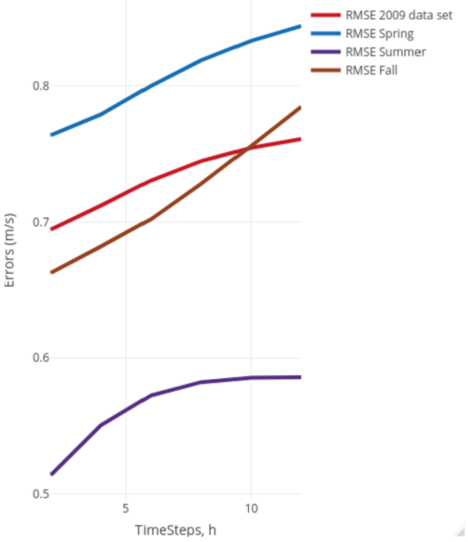


Figure 4.19: ARIMA Errors for 2009 data set and the seasons

The LSTM methodology was applied for the wind speed and pressure time series. Pressure was chosen because it had the greatest magnitude correlation with wind speed when compared with the other meteorological variables of wind direction and temperature. The RMSE results can be seen from Table 4.3. The test forecasted series for the various intervals, together with the actual series, can be seen in Figures 4.20 to 4.22. From the results obtained, ARIMA incurs smaller RMSE than the LSTM model for all intervals. Though there have been studies for which ARIMA outperforms ANN and SVM as mentioned in (Sandhu, Nair, et al., 2019), there have been RNN methods used in wind speed forecasting which performs better that ARIMA. In (Sandhu, Nair, et al., 2019), the errors are approximately 11 to 14 percent less in the RNN model compared to their ARIMA method. Another study, (Cao, Ewing, and Thompson, 2012), univariate ARIMA saw higher errors than univariate RNN. The same result was observed when comparing multivariate ARIMA with a multivariate RNN.

The reason for such results can be attributed to the difficulty of representing the high dimensional and non-linear system using the one-dimensional wind speed time series (Hu et al., 2019). As such the series is lagged using the time delay τ and the embedding dimension d for each of the intervals and these lagged co-ordinates were input to the LSTM model. The τ value was determined to be 3 using Auto Mutual Information (AMI) with the exception of Interval2 whose value was given by 2. The d value was determined to be 6 using Cao Algorithm for all intervals. The τ value was taken at the first local minimum for the AMI and the d value, as when $E1(d)$ attains saturation. Please refer to (Balkissoon et al., 2021) for more information on the methodologies of these parameters as well as Figures 4.23 and 4.24. Another study that uses the lags of the series in the training of the ANN as input variables was (Cadenas and Rivera, 2009). It was determined in their study that the best model was the simplest consisting of two layers and two input and one output neurons (Cadenas and Rivera, 2009).

The results obtained can be seen in Figures 4.25 to 4.27 for Interval1 to Interval3 respectively. The persistence model for each interval was constructed by calculating the average for every multiple of the 6th hour and recording these as the values of persistence for the next consecutive 6 hours or 36 time steps. The time forecast horizon, h looked at for this analysis are 6, 12, 18, 24, 30 and 36. The models under comparison are the LSTM with lagged wind speeds as inputs, the Moving ARIMA, the LSTM with pressure and wind speeds as inputs and persistence. The Normalized Root Mean Squared Error (NRMSE) metric for model evaluation was determined for all of the models. For values of this metric exceeding 1 or 100% implies that the forecast is no better than the mean of the data after this run.

It can be noted that all models performed better than the persistence model which stayed consistently between 0.8 and approximately 1.0 for the three intervals. The h value of 36 for Interval1 and Interval2 have values which are over 1.0 or representative of a forecast no better than the mean. The LSTM with the lagged wind speeds as inputs, denoted as Lagseries, outperformed the LSTM with the pressure and wind speeds as inputs, denoted as Pressureandwind, for all of the intervals. The Moving ARIMA method is now beaten by the lagged LSTMs for up to the 18th time step in Interval2 and up to the 12th time forecast horizon in Interval1 and Interval3. The second interval as mentioned previously has most of spring which have convective storm events, so it is expected that if any interval is to do best in the non-linear model of the LSTM when compared to the linear model of the ARIMA model, it would have been Interval2. The NRMSE of the LSTM Pressureandwind tends to one faster than the LSTM Lagseries for all of the intervals though up to the h value of 36, they do not exceed 1. As expected when the entire test set was forecasted for the models ($h=$ test set), the NRMSE for most of the intervals exceeded 1; for the other cases, they were 0.97 and 0.98.

The tabulated results of the RMSE values for each of these models can be seen in

Table 4.4. The forecasted and the actual series for $h = 36$ for LagSeries1 to LagSeries3 can be seen in Figures 4.28 to 4.30 respectively. Similarly, these plots for the ARIMA1 to ARIMA3 test set can be seen in subsequent figures whilst correspondingly the error defined as the difference between the actual test data and the predicted test data can be viewed in Figures 4.34 to 4.36. It can be noted, especially for the Moving ARIMA results, there was a significant match between the predicted and the actual series. The differences in the actual test data and the predicted test data were varying about the zero marker thus indicating that the trends were well captured by the model.

Table 4.3: LSTM RMSE

h	**Interval1- RMSE**	**Interval2- RMSE**	**Interval3- RMSE**
2	0.940142796	0.975498677	0.834158971
4	1.155461471	1.219467954	0.927660403
6	1.296958276	1.257914418	0.992396502
8	1.444971761	1.35340903	1.026433109
10	1.451004208	1.432323828	1.137951695
12	2.767638696	1.502783097	1.14651633

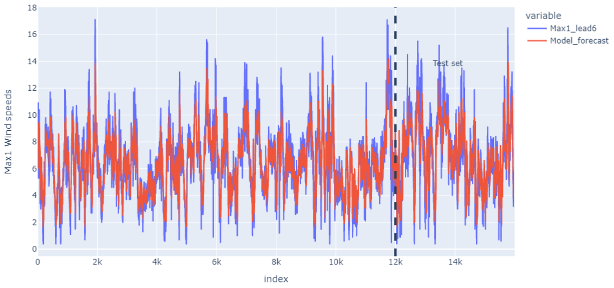


Figure 4.20: LSTM Interval 1

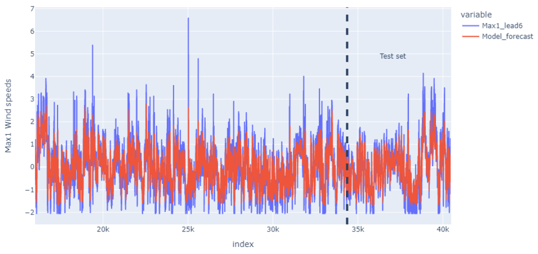


Figure 4.21: LSTM Interval 2

Figure 4.22: LSTM Interval 3

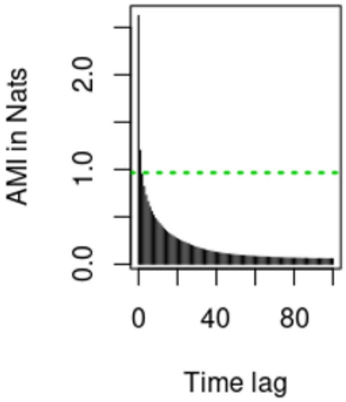


Figure 4.23: Tau for Interval2

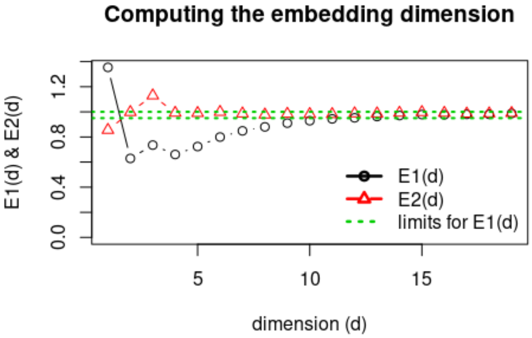


Figure 4.24: Embedding dimension for Interval2

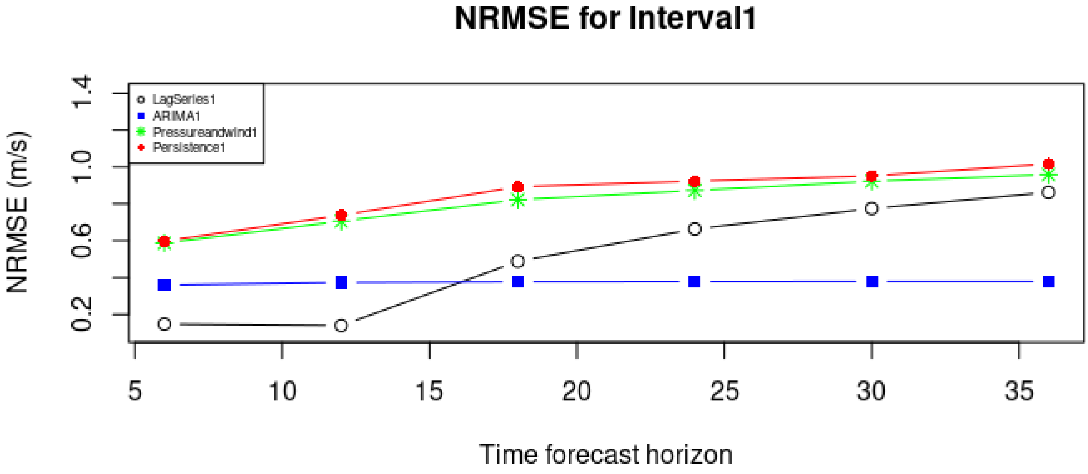


Figure 4.25: NRMSE for Interval1

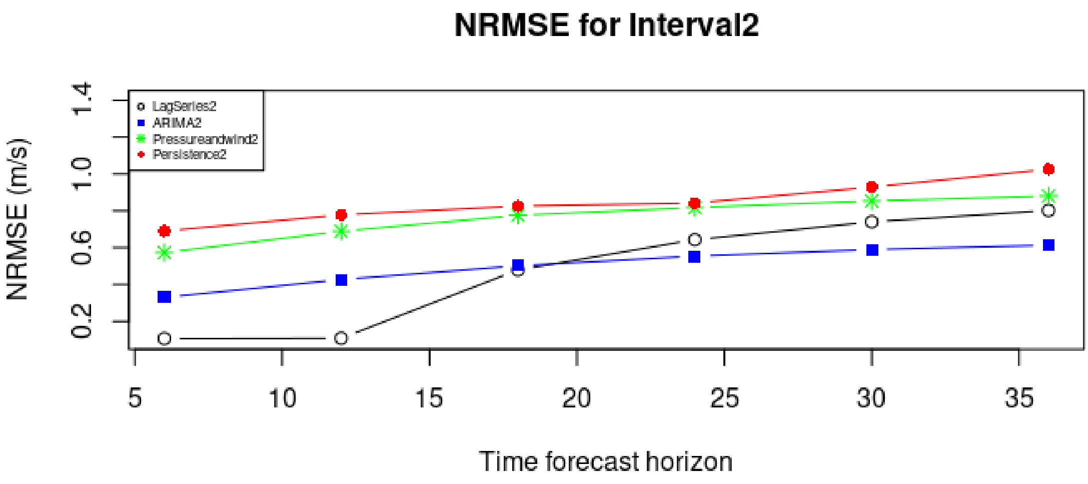


Figure 4.26: NRMSE for Interval2

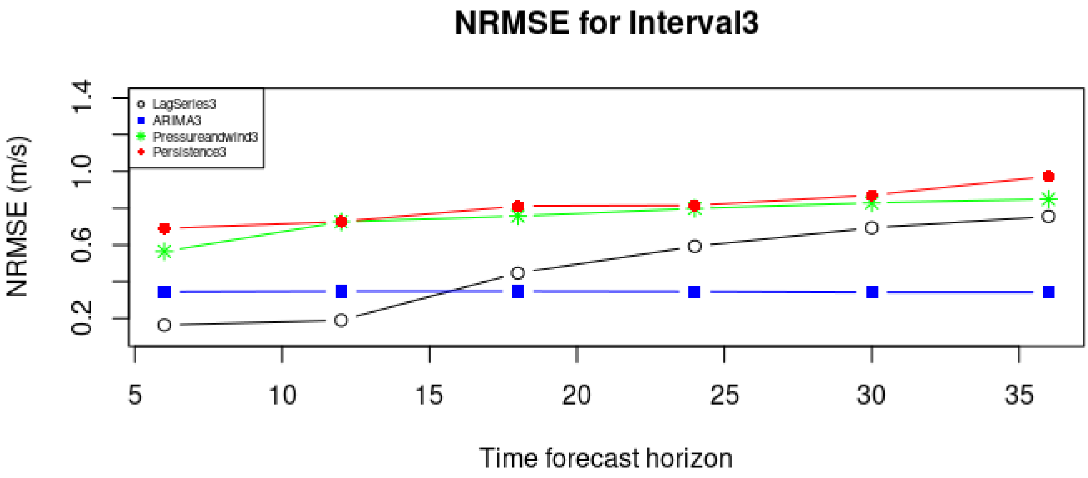


Figure 4.27: NRMSE for Interval3

Table 4.4: Models' RMSE Results for the Intervals

Interval	**h**	**LagSeries**	**ARIMA**	**Pressureandwind**	**Persistence**
Interval1	6	0.372907494	0.8008848	1.307214932	1.331339773
	12	0.352250278	0.832471	1.572144369	1.643609466
	18	1.087379921	0.8412903	1.831567231	1.987401039
	24	1.477212704	0.8436806	1.943241466	2.055514112
	30	1.724707138	0.8443778	2.05444378	2.119308164
	36	1.920543992	0.8444589	2.13412727	2.263135526
Interval2	6	0.238592146	0.7360105	1.276806808	1.538227218
	12	0.242534049	0.9509406	1.53582464	1.734079072
	18	1.067046011	1.117465	1.72837262	1.83946146
	24	1.432849673	1.233804	1.822299134	1.875548504
	30	1.649253577	1.312496	1.9014607	2.069048477
	36	1.786658699	1.365862	1.961462493	2.284102311
Interval3	6	0.376359712	0.7666935	1.257796263	1.536880065
	12	0.436734176	0.7735839	1.617069345	1.619311738
	18	1.028754688	0.7734066	1.685952464	1.80858766
	24	1.36522671	0.7698904	1.778578867	1.814853169
	30	1.596357379	0.7602228	1.84599341	1.939363159
	36	1.739204251	0.7602228	1.891493873	2.168188039

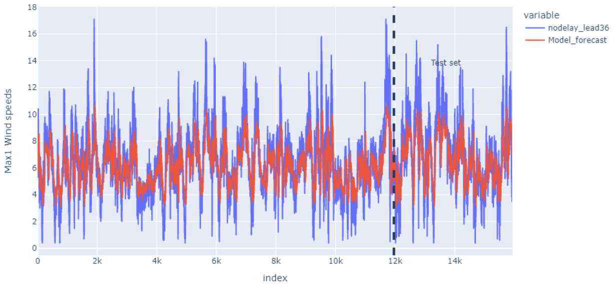


Figure 4.28: LagSeries LSTM1 for $h = 36$

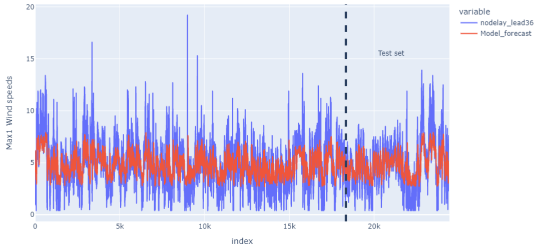


Figure 4.29: Lagseries LSTM2 for $h = 36$

Figure 4.30: LagSeries3 LSTM for $h = 36$

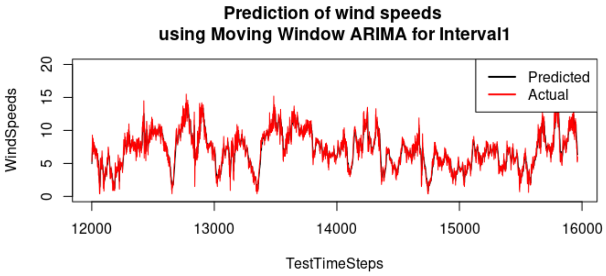


Figure 4.31: Forecasts for Moving ARIMA1 for $h = 36$

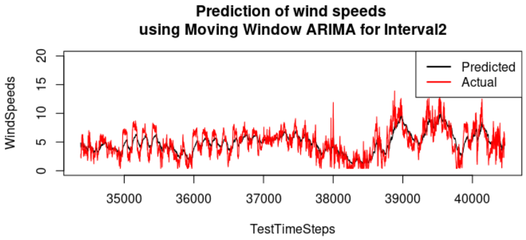


Figure 4.32: Forecasts for Moving ARIMA2 for $h = 36$

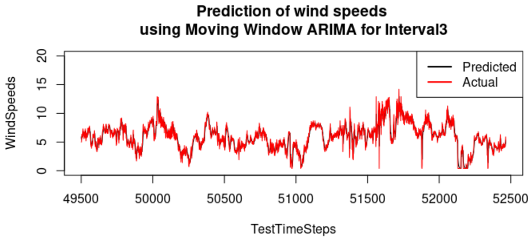


Figure 4.33: Forecasts for Moving ARIMA3 for $h = 36$

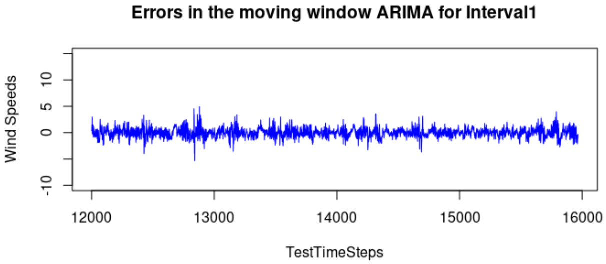


Figure 4.34: Errors for Moving ARIMA1 for $h = 36$

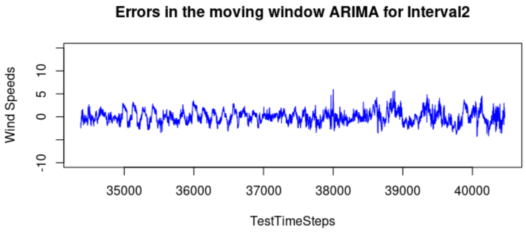


Figure 4.35: Errors for Moving ARIMA2 for $h = 36$

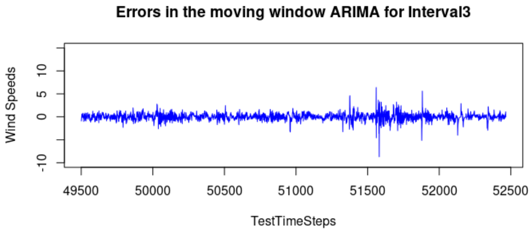


Figure 4.36: Errors for Moving ARIMA3 for $h = 36$

4.6 Future Work and Additional Analyses

In the model runs, the forecast variable - in our case wind speed - can be further processed to determine if there are any patterns in the wind speed forecast values (in terms of its accuracy) when its actual values is less than or greater than some x value or the difference between consecutive actual values rates are higher than some y value.

Yearly analysis can be done to see if there are the same number of clusters and accuracy in forecasting (seasonal analysis - using yearly data) is similar.

4.7 Conclusion

The optimal number of clusters was determined to be 4 using the Elbow and Silhouette methods among others. SOMs were then used to cluster the data after which three continuous intervals belonging to a particular cluster, which represented approximately 50% and over of the input vectors or rows from the data frame were identified. These intervals were then inputs for the LSTMs with inputs pressure and wind speeds, the lagged series LSTMs with embedding dimension d and time delay τ, the Moving Window ARIMA and persistence models. It was determined that the Moving ARIMA model is outperformed by the lagged LSTM for at most 180 minutes from the runs of the defined intervals. The lagged series improved upon the LSTM with the wind speed and pressure series. All of these models however, performed better than the benchmark of persistence for all time steps.

References

Balkissoon, Sarah et al. (2021). “Determining chaotic characteristics and forecasting tall tower wind speeds in Missouri using empirical dynamical modeling (EDM)”. In: Renewable Energy 170, pp. 1292–1307.

Balkissoon, Fox and Lupo (2020). “Fractal characteristics of tall tower wind speeds in Missouri”. In: Renewable Energy.

Basheer, Imad A and Maha Hajmeer (2000). “Artificial neural networks: fundamentals, computing, design, and application”. In: Journal of microbiological methods 43.1, pp. 3–31.

Berkovic, Sigalit (2017). “Winter wind regimes over Israel using self-organizing maps”. In: Journal of Applied Meteorology and Climatology 56.10, pp. 2671–2691.

Browell, Jethro, Daniel R Drew, and Kostas Philippopoulos (2018). “Improved very short-term spatio-temporal wind forecasting using atmospheric regimes”. In: Wind Energy 21.11, pp. 968–979.

Burguillo, Juan C (2014). “Using self-organizing maps with complex network topologies and coalitions for time series prediction”. In: Soft Computing 18.4, pp. 695–705.

Cadenas, Erasmo and Wilfrido Rivera (2009). “Short term wind speed forecasting in La Venta, Oaxaca, México, using artificial neural networks”. In: Renewable Energy 34.1, pp. 274–278.

Cao, Qing, Bradley T Ewing, and Mark A Thompson (2012). “Forecasting wind speed with recurrent neural networks”. In: European Journal of Operational Research 221.1, pp. 148–154.

Fox (2011). “A tall tower study of Missouri winds”. In: Renewable Energy 36.1, pp. 330–337.

Hu, Rui et al. (2019). “High resolution wind speed forecasting based on wavelet decomposed phase space reconstruction and self-organizing map”. In: Renewable Energy 140, pp. 17–31.

Kalinić, Hrvoje et al. (2015). “Comparison of two meteorological models using self-organizing maps”. In: OCEANS 2015-Genova. IEEE, pp. 1–6.

Kent, Brian (n.d.). “How to use PyTorch LSTMs for time series regression”. https://www.crosstab.io/articles/time-series-pytorch-lstm.

Lakshminarayanan, Sivakkumaran (2020). “Application of self-organizing maps on time series data for identifying interpretable driving manoeuvres”. In: European transport research review 12.1, pp. 1–11.

Li, Gong and Jing Shi (2010). “On comparing three artificial neural networks for wind speed forecasting”. In: Applied Energy 87.7, pp. 2313–2320.

Lin, Gwo-Fong, Tsung-Chun Wang, and Lu-Hsien Chen (2016). “A forecasting approach combining self-organizing map with support vector regression for reservoir inflow during typhoon periods”. In: Advances in Meteorology 2016.

Mabel, M Carolin and Eugene Fernandez (2008). “Analysis of wind power generation and prediction using ANN: A case study”. In: Renewable energy 33.5, pp. 986–992.

Nourani, V et al. (2012). “Classification of groundwater level data using SOM to develop ANN-based forecasting model”. In: Int J Soft Comput Eng 2.1, pp. 2231–07.

Pearce, John L et al. (2014). “Using self-organizing maps to develop ambient air quality classifications: a time series example”. In: Environmental Health 13.1, pp. 1–14.

Rakotomalala, Ricco (2005). “Tanagra Data Mining”. In: Version 1, p. 39.

Ramasamy, P, SS Chandel, and Amit Kumar Yadav (2015). “Wind speed prediction in the mountainous region of India using an artificial neural network model”. In: Renewable Energy 80, pp. 338–347.

Ramesh Babu, N and P Arulmozhivarman (2012). “Forecasting of wind speed using artificial neural networks”. In: Int. Rev. Mod. Sim 5.5.

Sandhu, KS, Anil Ramachandran Nair, et al. (2019). “A comparative study of ARIMA and RNN for short term wind speed forecasting”. In: International Conference on Computing, Communication and Networking Technologies. IEEE, pp. 1–7.

Sfetsos, Athanasios (2000). “A comparison of various forecasting techniques applied to mean hourly wind speed time series”. In: Renewable energy 21.1, pp. 23–35.

Tian, Jing, Michael H Azarian, and Michael Pecht (2014). “Anomaly detection using self-organizing maps-based k-nearest neighbor algorithm”. In: PHM Society European Conference. Vol. 2. 1.

Wehrens, Ron, Lutgarde MC Buydens, et al. (2007). “Self-and super-organizing maps in R: the Kohonen package”. In: Journal of Statistical Software 21.5, pp. 1–19.

Yan, Mingjin and Keying Ye (2007). “Determining the number of clusters using the weighted gap statistic”. In: Biometrics 63.4, pp. 1031–1037.

Chapter 5

Summary/Conclusion

The investigation of the fractal characteristics of wind speeds in Columbia, Neosho and Blanchard for various height levels was conducted. The wind speeds showed fractal characteristics, but there were no significant trends that showed that the fractality of the wind speeds was increasing with height levels.

Chaotic characteristics were further investigated both quantitatively and qualitatively. The phase space diagrams all depicted evidence of a chaotic system. This was also corroborated quantitatively by the Largest Lyapunov Exponent (LLE). The LLE also showed that there was no increasing level of chaos which coincided with increasing height levels. A simple non-linear prediction algorithm, Empirical Dynamical Modeling, was then used to forecast every 6 hours ahead. This showed, as expected, that the errors increase as the forecast horizon increased for each station and height level. For Columbia, after the 20th time step, the model normalized error exceeds 1. From the Lyapunov Exponent the prediction horizon was determined to be 6 time steps using a 95% confidence band. But, when using a 90% confidence band, this value decreased to 4 time steps. The model, however, beats persistence in Columbia for the very short term range of one time step.

Another forecasting algorithm was employed, Neural Networks. Firstly Self-

Organizing Maps were used to cluster the data. The optimal number of clusters were determined by the Elbow and Silhouette methods. This value, k was given as 4. The three intervals defined from these clusters were inputs into the Long Short-Term Memory Network (LSTM). This model was compared to a Moving ARIMA and persistence. The Moving ARIMA incurred less errors than the LSTM for all intervals. However, the lagged wind speed LSTM, considering the forecasting times of 60, 120, 180, 240, 300 and 360 minutes, not only outperforms persistence and the LSTM with wind speed and pressure inputs, but it beats the Moving ARIMA at most 180 minutes. The lagged LSTM may have outperformed the LSTM with pressure and wind speeds inputs because the latter may encounter difficulty in representing the system's non-linearity and high dimensionality.

VITA

Sarah is a Fulbright scholar from Trinidad and Tobago. Sarah comes from a highly family oriented background where she is the last of three children. Her undergraduate studies were in Mathematics and Physics. After which she pursued a Masters in Mathematics at the University of the West Indies, St. Augustine Campus. This decision came after her undergraduate studies when she was tracking Hurricane Irene in the summer of 2011. She wanted to study more about these complex dynamical systems. This was a motivating factor of her pursuit of a PhD. in Atmospheric Sciences in the University of Missouri under the advisership of Dr. Neil Fox and Dr. Anthony Lupo.

Sarah, throughout her PhD tenure, has been involved in the Preparing tomorrow's leaders in Science (PTLS) as a participant and a mentor. She was also a guest speaker for the University of Wisconsin students who were enrolled in this program. She was also a speaker for high school students as part of the Science on Wheels program. She has presented her research in various conferences during her studies.

One quote which resonates with Sarah, from her favorite scientist, Nikola Tesla is

" If you want to find the secrets of the universe, think in terms of energy, frequency and vibration."