SHORT-TERM WIND POWER PREDICTION

by

Fatemeh Marzbani

A Thesis Presented to the Faculty of the

American University of Sharjah

College of Engineering

in Partial Fulfillment of

the Requirements

for the Degree

Master of Science in

Electrical Engineering

Sharjah, United Arab Emirates

June 2014

Approval Signatures

We, the undersigned, approve the Master's Thesis of Fatemeh Marzbani.

Thesis Title: Short-term Wind Power Prediction

Signature **Date of Signature**

___________________________ _______________

Dr. Ahmed Osman
Associate Professor, Department of Electrical Engineering
Thesis Advisor

___________________________ _______________

Dr. Mohamed Hassan
Associate Professor, Department of Electrical Engineering
Thesis Co-Advisor

___________________________ _______________

Dr. Taha Landolsi
Associate Professor, Department of Computer Science and Engineering
Thesis Co-Advisor

___________________________ _______________

Dr. Khaled Assaleh
Professor, Department of Electrical Engineering
Thesis Committee Member

___________________________ _______________

Dr. Mohammad Jaradat
Visiting Associate Professor, Department of Mechanical Engineering
Thesis Committee Member

___________________________ _______________

Dr. Mohamed El Tarhuni
Department Head, Department of Electrical Engineering

___________________________ _______________

Dr. Hany El Kadi

Associated Dean, College of Engineering

___________________________ _______________

Dr. Leland Blank
Dean, College of Engineering

___________________________ _______________

Dr. Khaled Assaleh
Director of Graduate Studies

Acknowledgements

First and foremost I would like to express my deepest gratitudes to my advisors, Dr. Ahmed Osman, Dr. Mohamed Hassan, and Dr. Taha Landolsi for their excellent guidance, remarks, advice, and patience. I would like to thank you for your supports through my research and your painstaking efforts for proof reading my drafts. Thank you for making time for our weekly meetings. It has been an honor to work with you.

I would also like to acknowledge the financial and academic supports of the Electrical Engineering Department of the American University of Sharjah. I would like to thank the Electrical Engineering Department Head Dr. Mohamed El-Tarhuni for his excellent guidelines and advices during my study. I am most grateful to Dr. Ayman El-Hag, Dr. Khaled Assaleh, Dr. Verica Gajic, Dr. Rached Dhaouadi, Dr. Mamoun F. Abdel-Hafez, and my advisors for teaching me my Master's courses. I offer my sincerest appreciation to Dr. Ahmed Osman, Dr. Aly Elrefaie, Mr. Kashif Bangash, Mr. Ibrahim Abu Seif, Mr. Wasel El Tahir, and Mr. Narayanan Madathumpadical. I enjoyed the teaching experience while working with you.

My special thanks goes to all my friends at AUS for their kind friendship and support. In addition, I would like to thank my supportive friend Sanaa Hamid for her helps and warm encouragements. Sanaa, thank you for always being there for me. My greatest appreciation also goes to my friends in Iran, Canada, the USA, and Germany. You have been always a source of love and energy to me.

I am sincerely grateful to my family: my father, my mother and my lovely sister for their unconditional support, understanding, and endless love. Dad, thanks for allowing me to be as ambitious as I wanted. Words cannot describe the opportunities you have given me through my life. Thank you for teaching me how to be independent and confident. Without you I would have never been the person that I am today. Mom,

you are always a source of support, care, and love from heaven. As dad always says we are so blessed to have you. I am thankful for all the things you have sacrificed for me. Regardless of what may happen next, I promise to make you proud. Sahba, you are the joy and happiness of our family. Thank you for all your supports and cares. I cannot wait to see your university days. You are so smart and a brilliant future is waiting for you. I love you so much.

To my parents and sister…

Abstract

Environmental considerations in addition to energy crises have forced many countries to consider alternative energy sources; renewable energies are known as the best alternatives. Among renewable energies, wind power is the most promising energy source. The chaotic nature of the wind is a major challenge against the integration of wind power into grids. Integration of wind power results in several problems due to the fluctuations inherent in wind power, such as power quality, stability, and dispatch issues. The prediction accuracy of wind power affects its integration into power systems. Several wind power forecasting techniques have been proposed and developed. However, not all of them are able to provide sufficient accuracy. The main contribution of this thesis is to provide accurate short-term wind power prediction. A simple, yet effective adaptive-parameter regression model is developed. Specifically, the proposed approach uses a window of previous observations to obtain the model parameters that minimizes the prediction error. Regression-based models are affected by measurement errors. Thus, other models with the capability of moderating the impact of measurement errors are needed. In order to cope with such errors, two hybrid grey-based short-term wind power prediction techniques are proposed: GM(1,1)-ARMA and GM(1,1)-NARnet. These techniques are combined with ARMA models and Nonlinear Auto Regressive Neural Network (NARnet) models, respectively. GM(1,1)-ARMA and GM(1,1)-NARnet are applied to wind power data and the obtained results are compared with those obtained from ARMA, the traditional grey model, as well as the persistent model. The efficiency of both of the proposed techniques is confirmed. In contrast to the GM(1,1)-ARMA model, the GM(1,1)-NARnet model utilizes the nonlinear components of wind power during the forecasting procedure which results in more accurate prediction.

Search Terms—Wind Power Forecast; Time-Series Analysis; ARMA Models; Grey Theory; Nonlinear Time Series Analysis.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AAFTA	-	Amplitude Adjusted Fourier Transform
ACF	-	Auto Correlation Function
AGO	-	Accumulated Generating Operation
AESO	-	Alberta Electrical System Operator
AIC	-	Akaike Information Criteria
ANFIS	-	Adaptive Neuro Fuzzy Inference System
ANN	-	Artificial Neural Network
AR	-	Auto Regressive
ARMA	-	Auto Regressive Moving Average Process
ARIMA	-	Auto Regressive Integrated Moving Average Process
DEM	-	Digital Elevation
HILARM	-	High-Resolution Limited Area Model
IAAFT	-	Iterative Amplitude Adjusted Fourier Transforml
IAGO	-	Inverse Accumulated Generating Operation
MA	-	Moving Average
MLP	-	Multi Layer Perceptrone
NN	-	Neural Network
NWP	-	Numerical Weather Prediction
RMSE	-	Root Mean Square Error
RPE	-	Relative percentage Error
SCADA	-	Supervisory Control And Data Acquisition
WAsP	-	Wind Atlas Application and Analysis Program
WPF	-	Wind Power Facilities

Chapter 1: Introduction

1.1 Wind Power

In recent years, environmental considerations and energy crises are the main factors which have forced the world to accelerate from fossil fuels to alternative sources of energy. Renewable energies are known as the best alternatives. Among the renewable energy sources, such as wind, biomass, geothermal, solar, etc. Wind power is the most promising energy source. However, its intermittent nature introduces challenges to the operators when integrated to power grids. In order to cope with the chaotic nature of wind, various forecasting techniques can be utilized. Accurate wind power prediction provides sufficient information to the grid operators which can be used in the electricity market, economic dispatch, unit commitment problems, regulation actions, and more.

1.2 Wind Power Prediction Time Horizons

Over the past few decades, several prediction approaches have been developed. Time horizon, the time period for which the wind power is going to be predicted, is an important feature for any prediction tool [1]. The prediction techniques can be classified based on the time horizons into four groups. First, there are long-term forecasting methods which refer to one day to one week ahead prediction. Such predictors are utilized for commitment and reserve requirement decisions as well as maintenance scheduling. Second, there are medium-term forecasting methods which provide six hour to one day ahead predictions. Generator on-line/off-line decisions and operational security in the one day-ahead electricity market are examples of medium-term forecasting applications. Third, there are short-term forecasting methods which account for thirty minute to six hour ahead forecasts [2]. Such approaches can be developed for economic load dispatch planning and load increment/decrement decisions. The last group is very short-term forecasting approaches taking care of a few seconds to thirty minutes ahead predictions. Such predictors are used in the electricity market in addition to clearing and regulation actions. Among different prediction time horizons, short-term prediction techniques are the most frequently used [2].

1.3 Fundamental Wind Power Prediction Techniques

There are five classes of basic wind power forecasting techniques: the persistent method, physical approaches, hybrid structures, statistical approaches, and some new techniques. The persistent method is a benchmark approach for evaluating the performance of other predictors. This method performs very well for very short and short-term predictions. Physical approaches, such as Numerical Weather Predictors (NWP) [3, 4], are very accurate techniques when utilized for long-term applications. Artificial Neural Networks (ANN) and time series models are two subclasses of statistical approaches. Such methodologies are mainly used for short-term forecasts [2]. Adaptive Neuro Fuzzy Inference System (ANFIS) [5] and combinations of correlation and Neural Networks (NN) [6] are examples of hybrid structures. Such approaches are efficient when used for very short-term, medium-term, and long-term forecasts.

In this section, a description of the five groups of fundamental forecasting methods is presented. This is followed by reviewing some examples of each group.

1.3.1 Persistent Method.

The persistence method, also known as Nave Predictor, is the simplest forecasting approach. It assumes that the next measurement of the wind power is the same as the measurement at the current time. For a given time series $\{y_n\}$, the persistent forecast is defined by $y_{(n+1)} = y_{(n)}$. The persistent method works well for a few steps ahead predictions when data patterns change slightly. However, it breaks down for highly varied data. It is usually used as the classical benchmark to examine the efficiency of other prediction methodologies [4].

1.3.2 Physical Models.

Physical models are capable of predicting the wind speed or power based on a number of physical and climatology inputs, such as temperature, obstacles, terrain, and pressure. The NWP models, a subclass of physical models, are well established techniques originally developed by meteorologists for prediction of planetary scales. Such methodologies benefit from mathematical fluid mechanic models. The NWP models are very accurate in long-term forecasts. Nevertheless, they lack sufficient accuracy in short-term predictions [2].

McCarthy [7] is thought to have developed the earliest physical model for the Central California Wind Resource Area. The model was built around the climatological study of the site to predict daily average wind speed. The obtained results show that the proposed model outperform the persistence.

In subsequent years, there have been developments in physical methods. In what follows, some examples of such developments are expressed.

In order to enhance the prediction results, NWP models solve conservative equations such as heat, water, mass, and momentum numerically on four dimensional grids: latitude, longitude, elevation and time at any given site. Since the wind patterns may become highly significant in small sites, the NWP models must be non-hydrostatic for short-term applications and represent the wind travel topography by incorporating the Digital Elevation Model (DEM). The accuracy of NWPs is significantly affected by highly varied data. Moreover, the mathematical complexity of such models is another constraint; usually they run on super computers and it takes a few hours to obtain a result [8].

The High-Resolution Limited Area Model (HIRLAM), an example of an NWP models, is introduced in [9]. This technique is an automatic on-line prediction system for wind farm production output providing up to thirty six hour ahead prediction. This research is a large scale forecast using the Wind Atlas Application and Analysis Program (WAsP) program to supply intricate data when zoomed into one site. Finally, the Model Output Statistics (MOS) module is utilized to reduce the remaining error. The obtained results confirm the out-performance of the proposed methodology compared to the persistence.

A research study on NWP models is carried out in [10]. It shows that NWP models are suitable for several hour ahead predictions or more. Additionally, it is confirmed that NWP models (using DEMs and MOS corrections) fail in very short-term scale predictions.

Landberg and Watso studied different models including NWP, HIRLAM, and NN in [11]. The obtained results are compared with the persistence model. It is shown that all of the investigated techniques are more efficient than the persistence. It is also asserted that the HIRLAM/WAsP model performed the best for the site, as it had a large value of wind speed mean.

1.3.3 Other Prediction Techniques.

There are several prediction approaches which can be classified as examples of the other prediction techniques subclass: spatial correlations, fuzzy logic, and wavelet transformation. In what follows, such techniques are expressed. Some research papers developing these methods are briefed.

(a) Spatial Correlation Models

In spatial correlation models, the spatial relationships between the wind power time series (or speed) at different adjacent sites are considered. The wind power (or speed) at the given site is predicted based on the predicted points of its neighbor sites. The complexity of this method stems from the essential requirements of the measurements at many spatial correlated sites and the need for timely transmission. The most popular spatial correlation method utilizes the cross correlation curves of the wind power time series at two given sites (for example A and B). The predicted value of the wind power (or speed) at time $(t + \Delta t)$ at site B is calculated as follows [12]:

$$p\left(B, t + \Delta t\right) = m_B + \left[\rho(\Delta t)\sigma_B/\sigma_A\right]\left[p(A, t) - m_A\right], \tag{1.1}$$

where $\rho(\Delta t)$ is the peak correlation coefficient (usually taken equal or greater than 0.8), σ_{A} and σ_{B} are the standard deviations of wind power (or speed) time series at sites A and B, respectively. The mean of the wind power (or speed) time series at sites A and B are represented by m_A and m_B, respectively. The current wind power at the given sit A is depicted by $p\left(A,t\right)$.

In what follows, some studies of spatial correlation models for wind power (or speed) prediction are described.

A novel spatial correlation predictor is proposed in [13]. In this technique, the predicted value of the wind speed at each site is provided using the measurements at the other sites. Two given points which belongs to site A and B are considered in this methodology: A' and B', respectively. The wind speed at B' is found using the following equation:

$$\begin{aligned} &v_F\left(t\right) = v\left(A,t\right) = a_0 + a_1 v\left(A',t\right), \\ &v\left(B',t+\Delta t\right) = b_0 + b_1 v\left(A',t\right), \end{aligned} \tag{1.2}$$

where the relation between the local and upper wind speed at site A is considered through coefficients a_0 and a_1. The wind speed at A' is adjusted to the wind characteristic at B' using b_0 and b_1. This approach is applied to seven years data collected at six different South and Central Aegean Sea islands in Greece. The obtained results confirm the efficiency of the proposed technique.

The spatial correlation model is used in [14] to reduce the prediction error. A comparison is provided between the prediction errors of a single site and an ensemble of wind farms. It is demonstrated that the prediction error of the spatial correlation model is a smaller value compared to that of a single site due to the spatial smoothing effect. During the error reduction process, different numbers of sites are considered. The obtained results show that it is sufficient to include only a few sites.

(b) **Fuzzy Logic**

Jan Lukasiewicz introduced multi value logic, also known as fuzzy logic, for the first time in the 1930s [15]. The classical operators deal with two values: true (1) or false (0). However, according to fuzzy logic, the truth values lay within the range of 0 (completely false) and 1 (completely true). In other words, fuzzy logic deals with degrees of truth and membership, accepting the concept of partly true and partly false. An approximate reasoning technique, called possibility theory, stems from this research. In 1965, the possibility theory was expressed as a formal mathematical logic system [16]. Fuzzy logic refers to this new logic which manipulates the fuzzy terms.

The theory of fuzzy logic has been used in a broad range of applications such as information retrieval [17], pattern recognition [18], electricity price estimation [19, 20], and wind speed and power prediction [21].

A probabilistic fuzzy predictor for short-term wind speed is proposed in [22]. This methodology is capable of taking into account both the deterministic and nondeterministic components of the wind. In order to evaluate the efficiency of this method, it is applied to different wind speed data sets. It is found that the proposed algorithm shows acceptable performance in a complex stochastic environment.

(c) **Wavelet Transformation Techniques**

Wavelet transformation is a technique for approximating real world signals. This algorithm, decomposes a signal into a series of wavelets. Such decomposition stores the signals more efficiently compared to the Fourier transformation. The objective of Wavelet transformation is to provide future information of the available signal [23].

This algorithm is used in various applications, such as medical image processing [24], compressing images [25], detecting self similarity [26], earthquake magnitude predictions [27], and wind power forecasts [28].

A novel wavelet-based prediction algorithm is developed in [29] to provide one day ahead wind speed forecasts. In this research, the wavelet theory is utilized to decompose nonlinear time series into several approximate stationary time series. This technique is applied to real wind speed data collected at a weather station. The obtained results demonstrate that the wavelet theory is successful in wind speed forecasting.

1.3.4 Hybrid Techniques.

Hybrid techniques refer to the combination of one or more prediction approaches. The objective of such methodologies is to benefit from each individual model. Hence, it provides more accurate prediction than any of the individual methods. Hybrid predictors have been widely used for wind power (or speed) prediction. In this subsection, some of such research papers are reviewed.

A hybrid wind speed prediction algorithm is introduced in [30]. This predictor is comprised of an NN and wavelet transformation. First, the wavelet transformation is used to decompose the wind speed signal. Then, the NN is employed to predict future values of the wind speed.

In [31], a new prediction technique comprised of fuzzy logic and the NN is presented for wind power prediction. In order to examine the performance of the proposed algorithm, it is applied to the actual wind power data collected from a wind farm in China. The calculated Root Mean Square Error (RMSE) for the predicted values is smaller than 20 %. This technique can be used for an on-line forecasting system based on the adaptability feature of neuro-fuzzy networks.

.

1.3.5 Statistical Approaches. Compared with other predictors, the statistical methods are quite cost effective in terms of algorithm complexity. Such algorithms, provide the future values of the wind power by analyzing the historical time series. The objective of such techniques is to formulate the patterns of the measurements. Once the pattern is established, the future values can be obtained by extrapolating the determined pattern. In other words, the problem is mathematically formulated using the historical data. The popularity of the statistical models stem from their simplicity, moderate cost, and the ability to provide timely predictions.

These techniques include Auto Regressive (AR), Moving Average (MA), Auto Regressive Moving Average (ARMA), and Auto Regressive Integrated Moving Average (ARIMA) models [32]. The AR models were originally introduced by the mathematician Yule in 1927. A few years later in 1931, MA and ARMA models were presented by another mathematician, Walker. In the following paragraphs, some studies which employ the time series models for wind power prediction are described.

A time series based model is presented in [33]. This predictor utilizes the non-Gaussian distribution and non-stationarity feature of the wind speed to provide future values of the wind speed. The proposed ARMA process considers the positive correlations between the consecutive wind speed measurements. The ARMA model is expressed by:

$$X_t = \sum_{i=1}^{p} \phi_i X_{t-i} - \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}, \tag{1.3}$$

where ϕ_i and θ_j are non-zero coefficients and ε_t is a normal white noise process. This algorithm is applied to a data set of two years of the wind speed time series. The obtained results demonstrate that the proposed predictor can be efficient for short-term prediction with a confidence interval of 95%.

In [34], the ARMA model is developed to predict the hourly average wind speed up to ten hours ahead. This model is applied to wind speed data measured during nine years at five different sites with different topographic characteristics in Spain. Due to the non-stationary nature of the wind speed, it is essential to pre-analyze the available data by employing the standardization and transformation process. The proposed model beats the persistent method in terms of prediction accuracy.

The wind speed is modeled and predicted using a time-variant AR process in [35]. First, the parameters of such predictors are modeled using the integrated random walk process. Then, the model parameters are estimated by employing a Kalman filter. The proposed methodology is capable of predicting the wind speed one to a few hours ahead successfully.

1.4 Short-term Wind Power Prediction Techniques

Wind power prediction is defined as the expected power produced by the wind turbines at look-ahead periods. In order to collect the required data for wind power prediction, Supervisory Control And Data Acquisition (SCADA) systems are installed at each wind turbine of a wind farm. Such data sets are expressed in kW or MW based on the nominal capacity of the wind farm [36].

Wind power prediction depends on various parameters such as temperature, wind direction, obstacles, height, and the geographical area. Therefore, a unique forecasting technique can not be developed for all wind farms. In what follows, the recently developed short-term wind power prediction techniques are described.

Time series based models and learning-based predictors are well known for their accuracy in short-term prediction. Time series based models include ARMA, ARIMA, and AR modesl with exogenous input (ARX), grey predictors, Linear predictors, and exponential smoothing [2].

Among the most popular wind power short-term prediction techniques are the ARMA models. Such predictors are linear methods which provide the future values of a time series as a linear combination of a number of immediate past values using the time series analysis [37]. A Bayesian approach is utilized to develop an AR(6) model for one hour ahead wind power prediction in [38]. The efficiency of this technique is examined by applying it to two years of wind speed data time series. In comparison with the persistent model, this approach results in more accurate predictions. However, the lower orders of the AR models are not successful.

In [39], six hour ahead wind power prediction is provided by developing an ARMA model. The proposed predictor outperforms the persistent method by 18% and 7% for the sixth and the first hour, respectively. In addition, the obtained results affirm that the performance of the ARMA models differs according to the time periods to which they are applied. The predicted values of the first hour outperform the persistent method. However, ten minute ahead predicted values do not have sufficient accuracy.

The Mexican Electrical Power Research Institute (MEPRI) utilizes an adaptive and recursive statistical predictor for short-term wind power forecasts [40]. The proposed technique is comprised of several forecasting approaches; a weight is given to each model based on its previous efficiency. The number and weights of the utilized models vary over time.

The performance of the ARX and NN based models are investigated in [41]. The obtained results for these two techniques are compared with those obtained from the persistence model for one to thirteen hour ahead predictions. It is shown that the ARX model is inefficient compared to the persistent method for prediction time scales less than thirteen hours ahead. However, the NNs beat the persistence method.

The traditional grey prediction model GM(1,1) is developed in [42] for wind power forecasting. Based on the fact that the wind power prediction procedure is highly influenced by the terrain and obstacles, the grey model is a suitable algorithm to cope with such uncertain characteristics of wind power. The efficiency of the GM(1,1) is approved by applying it to a real set of wind power data recorded in Taiwan.

In [43], the wind power is predicted indirectly using the wind speed data. The proposed method is a short-term predictor using a support vector machine. First, the wind speed is predicted. Then, the power-wind speed characteristics of the wind turbines are used to predict the wind power. The obtained results affirm the accuracy of the proposed techniques compared to the persistent model and the Radial Basis Function (RBF) NN.

The one hour ahead values of the upper and lower bounds as well as the average of the wind speed are forecasted using a Gaussian process along with the Kernel machine approach and Bayesian estimation [44]. The efficiency of the proposed algorithm is affirmed by applying it to a historical wind speed data set. The obtained results are compared against the Multi Layer Perceptrone (MLP) and NNs; an improvement of

27% of the mean absolute error is confirmed.

Three prediction error criteria are used in [45] to compare the accuracy of three neural networks: Back Propagation (BP), linear element, and Radial Basis Function (RBF). Different tests varying in input size and learning rates are performed. The obtained results affirm the importance of the chosen input size, learning rates, and network type. It is also shown that a robust method along with universal criteria results in consistency.

A robust prediction approach, called Mycielski, is proposed in [46]. According to this technique, the future value of the process is equal to the data chain which has the longest repeating pattern among all other previous data. The accuracy of this predictor is evaluated by applying it to three case studies; the maximum obtained prediction RMSE is 1.5%.

In order to speed up the learning process and reduce the NN size, the original data set is pre-processed [47]. The RMSE and coefficient of determination are used as evaluation criteria. The proposed predictor results in smaller values of RMSE and bigger values of determination coefficient as well as a reduction in computational time.

1.5 Thesis Organization

The rest of this thesis is organized as follows. In Chapter 2, the wind power time series analysis, the ARMA prediction models, the grey system theory, and the proposed grey forecasting models are explained. Chapter 3 presents the wind power time series pre-analysis and the prediction results. The presented thesis is concluded in Chapter 4.

Chapter 2: Research Methodology

In this chapter, a brief introduction to time series is given. The structural concepts of time series analysis, time series models, and model identification techniques are illustrated. Then the grey system theory and the traditional grey model GM(1,1) are described. The chapter concludes with introducing new hybrid grey-based models, such as GM(1,1)-ARMA and GM(1,1)-NARnet.

2.1 Time Series

Time series refers to a set of observations of a random variable. Therefore, a time series is a stochastic process. This observation occurs typically at equally-spaced time intervals. Time series are used in various applications, such as finance (daily rates of exchange), medicine (heart and brain activities), economics (annual unemployment data), and engineering (daily electricity demand).

Time series analysis is comprised of different techniques. Such techniques are utilized to analyze the measurements time series in order to elicit the statistical characteristics of the data. The time series analysis is concerned with describing the data, fitting proper models, and providing future values of the data.

Time series modeling is a statistical problem which assumes that the measurements vary according to some probability distributions. Based on the obtained model, time series prediction uses the previously measured values to provide future data [48].

2.2 Structural Concepts

In this section, some structural time series concepts are introduced. Such concepts play an essential role in understanding different time series characteristics as well as stochastic time series models. The following subsection provides a simple description of the stochastic processes. This is followed by an introduction to the Auto Correlation Function (ACF) of time series and the white noise process. The most frequently used time series models, such as AR, MA, ARMA, and ARIMA are discussed toward the end of this section.

2.2.1 Stochastic Processes.

Stochastic processes (also known as random processes) refer to a family of random variables $\{z(t)|t \in T\}$. Such variables are indexed over a time set T. For a given probability sample space, a random variable assigns a number to each outcome of the sample space, while a stochastic process assigns a sample function $z(t,s)$ to each outcome.

Stochastic processes can be categorized based on the values they take or the time instants at which they occur. A stochastic process is said to be continuous in time if it occurs at continuous time instants. However, those happening at discreet time intervals are known as discrete time stochastic processes. When the outcome of a process is continuous, it is called a continuous value process. Nevertheless, discrete value processes take discrete values [49].

2.2.2 Auto Correlation Function.

The correlations between different values of a stochastic process can be expressed using the ACF of that process. For a random process Y, the k^{th} component of its ACF describes the correlation between y_t and y_{t+k}. The ACF of a time series is a graph showing the sample autocorrelations at lags $k = 1, 2, ...n$. The mathematical expression for the k^{th} ACF component r_k is given by [50]:

$$r_k = \frac{\sum_{t=1}^{n-k} (y_t - \bar{y})(y_{t+k} - \bar{y})}{\sum_{t=1}^{n} (y_t - \bar{y})^2}, \tag{2.1}$$

where $\bar{y}$ is the mean of the time series calculated using the following formula.

$$\bar{y} = \frac{1}{n}\sum_{t=1}^{n} y_t. \tag{2.2}$$

In which n is the length of the time series.

In the following subsection, the behavior of the ACF of a time series is studied in order to investigate the stationarity of a historical data set.

2.2.3 Stationary Processes.

In time series analysis, stationarity is the basis of a large number of time series models. Therefore, it is essential to have a good understanding of the stationarity concept and stationary time series.

There are two kinds of stationary stochastic processes: strictly stationary and weak (wide sense) stationary. The time series $\{y_t, t \in Z\}$ is strictly stationary if for all values of k and τ, the joint distributions of $\{y_{t1}, y_{t2}, ..., y_{tk}\}$ are the same as those of $\{y_{t1+\tau}, y_{t2+\tau}, ..., y_{tk+\tau}\}$. For such time series, all moments of all degrees such as expectation, variance, third or higher order moments are the same. In other words, the distributions of $\{y_t, y_n\}$ and $\{y_{t+m}, y_{n+m}\}$ only depend on $n - t$.

For real-world time series, strict stationarity is too strict. Therefore, weak or wide sense stationarity is usually used in time series analysis. In the present thesis, stationary means weak stationary, unless otherwise is specified. Weak stationary time series have three major stochastic properties. First, the mean of such time series is constant. Second, its variation is finite. Third, its second moments (covariance) is only a function of $t - n$ but it is neither the function of t nor that of n. Such properties are mathematically expressed in the following equations, respectively [51]:

$$E\left(y_t\right) = \mu, \quad \forall t \in Z. \tag{2.3}$$

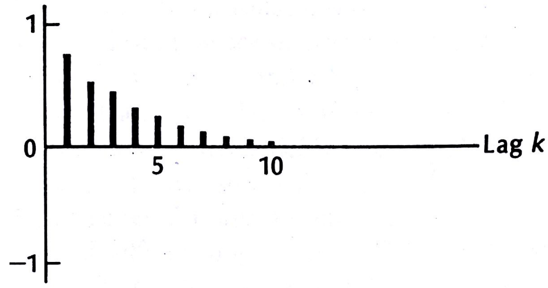
$$E\left(y_t^2\right) < \infty, \quad \forall t \in Z. \tag{2.4}$$

$$Cov\left(y_t, y_n\right) = Cov\left(y_{t+m}, y_{n+m}\right), \quad \forall t, n, m \in Z. \tag{2.5}$$

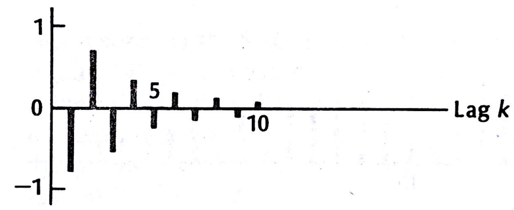
where μ is a constant.

Many stationary time series approaches are based on the assumption of the stationarity of the historical data. Hence, it is crucial to investigate the stationarity of time series before performing any time series analysis.

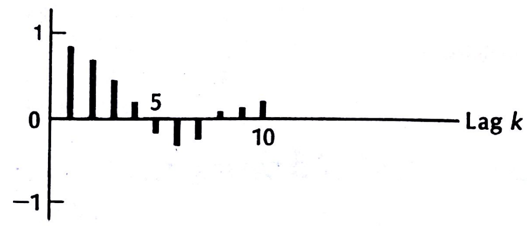
In order to scrutinize the stationarity of a time series, one should investigate the ACF of the data set. A time series is said to be stationary if its ACF cuts off at small lags or dies down quickly. The ACF of a time series can die down in three forms (as illustrated in Fig.1: a damp exponential fashion, damp exponential fashion with oscillation, and damp exponential fashion. When the ACF of a time series dies down extremely slowly, the non-stationarity of the time series is confirmed.



(a) Damped exponential fashion.



(b) Damped exponential fashion with oscillation.



(c) Damped sinusoidal fashion.

Figure 2.1: Different ACF dying down patterns [48].

Another common method to examine the stationarity of a time series is to study the variation of the mean or variance of the data over time. The mean and variance of a non-stationary time series changes over time; however, the mean and variance are time-invariant for stationary data.

2.2.4 Differencing.

When the non-stationarity of a set of data is affirmed, it is essential to employ proper time series transformation in order to remove the non-stationary characteristics of the data [52].

One of the most common transformations used to remove non-stationarity from a time series is differencing. The differencing procedure includes determining the correct smallest differencing order. This selected order typically results in the minimum variance. However, over-differencing must be avoided since it can cause data loss. Usually, for non-stationary time series, first or second order differencing is adequate. The term ”differencing” refers to subtracting the value of an earlier measurement from the value of a later one. The first order differencing ($d = 1$) of a time series can be represented by the following equation.

$$z_t = \nabla y_t = y_t - y_{t-1}, \quad t = 2, ..., n. \tag{2.6}$$

For some time series, second order differencing should be applied to remove the non-stationarity. In other words, the first order differencing operator should be applied to the data twice. The second order differencing ($d = 2$) can be mathematically expressed as

$$\begin{aligned} z_t = \nabla^2 y_t = (y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) = \\ y_t - 2y_{t-1} + y_{t-2}, \quad t = 3, ..., n. \end{aligned} \tag{2.7}$$

When the non-stationary time series is differenced, it is ready to be fitted by ARMA models. There are different tools which can be used to identify the appropriate ARMA model to the data. In the following sections some of them are discussed.

2.2.5 White Noise Process.

A white noise process $\{a_t\}$ is a stochastic process comprised of uncorrelated random variables. Such processes are rarely found in real applications. However, they are a very important building block for constructing time series models [48]. A white noise process is characterized by the following properties: constant mean $E(a_t) = \mu_a$ and variance $Var(a_t) = \sigma_a^2$ as well as uncorrelated random variables $Cov(a_t, a_{t+k}) = 0$.

The ACF of a white noise can be expressed as:

$$r_k = \begin{cases} 1, & k = 0 \\ 0, & k \neq 0 \end{cases} \tag{2.8}$$

2.2.6 Time Series Models.

In time series analysis, a time series can be expressed using four useful representations: AR, MA, ARMA, and ARIMA. Such models are linear time series models applicable to the stationary time series. Therefore, the stationarity of the available data should be investigated before these models are employed [52]. The following paragraphs provide brief explanations of such models and their mathematical formula.

An auto regressive process is denoted as AR(p). The name "autoregressive" refers to self regression and p is the model order. The AR(p) model is illustrated in the following equation.

$$z_t = \varepsilon_t + \varphi_1 z_{t-1} + \varphi_2 z_{t-2} + ... + \varphi_p z_{t-p}, \tag{2.9}$$

where ε_t is a white noise process, φ_i represents the model coefficient at the i^{th} lag, and z_t shows y_t when it is reduced to a zero-mean set of data. The time series z_t can be obtained using the following equation.

$$z_t = y_t - \bar{y}, \quad t = 1, ..., n. \tag{2.10}$$

From another point of view, time series can be modeled as an unevenly weighted moving average of the white noise process ε_t. This time series model can be expressed by a MA(q) model as follows:

$$z_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + ... + \theta_q \varepsilon_{t-q}, \tag{2.11}$$

where the model coefficients are denoted by θ_i. Such coefficients should be estimated using the historical time series.

The ARMA(p, q) is obtained by combing AR(p) and MA(q) models.

$$z_t = \varphi_1 z_{t-1} + \varphi_2 z_{t-2} + ... + \varphi_p z_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + ... + \theta_q \varepsilon_{t-q}. \tag{2.12}$$

Non-stationary time series are very common in real time applications. Auto Regressive Integrated Moving Average ARIMA(p, d, q) models can be utilized to model such data sets. ARIMA models are simply ARMA models when z_t is replaced by its differenced time series of order d. Fitting an ARIMA model to a set of historical data can contribute to understanding the physical dynamic of the process as well as removing the persistence from the data. Moreover, it can be used to predict the behavior of the time series [48].

2.3 Model Identification

The most critical stage in time series analysis is to identify the proper models for the available data. Hence, an appropriate understanding of the stochastic process is required, specifically the behavior of the time series ACF [48].

For real applications, the ACF of a given time series is unknown and should be estimated. The objective of the model identification is to match the patterns in the estimated ACF with the known patterns of the ACF of the ARMA models. For instance, based on the fact that for an MA(2) model the ACF cuts off at the second lag, the existence of a large spike at the second lag of the estimated ACF demonstrates the MA(2) as a possible underlying process.

There are several different model identification techniques. In the following sections, two model identification methods are described: model identification using

ACF and adequacy checking tests (residual analysis test [53] and Akaike Information Criteria (AIC) [48]). Employing such techniques may result in identifying different model orders. The model whose prediction error is smaller compared to the others is chosen as the proper model order for the available data.

(a) Model Identification Using ACF

The ACF of the time series can be investigated in order to identify the proper statistical models for the historical data. Generally, AR(p) is a suitable model for time series with the following properties [48]:

1. The ACF of the time series dies down in a quick exponential fashion.
2. The r_k takes large values for $k \leq p$ while for $k > p$ they are of small values.

The MA(q) model is the best fit for a time series if:

1. The standard error of the ACF values are statistically large for $k \leq q$ and statistically small for $k > q$.
2. The ACF dies down quickly in an exponential way.

For time series whose characteristics are best fitted by the ARMA(p, q) model, their ACF dies down exponentially and slowly. Therefore, the model orders can not be identified using only the ACF investigation. Other techniques should be employed in order to identify the proper models to the time series.

2.3.1 Model Identification Using Adequacy Checking Tests.

There are different adequacy checking tests for model identification in time series analysis. Among the family of such tests, the whiteness of the residuals and the AIC are the most frequently used ones.

(a) **Whiteness of the Residuals**

Residuals are the prediction error, described by

$$a_t = z_t - \varphi_1 z_{t-1} - \varphi_2 z_{t-2} - \ldots - \varphi_p z_{t-p}. \tag{2.13}$$

After developing a proper model, the obtained residuals are free of any sequential contingency. Indeed, the residuals have the same stochastic properties as uncorrelated random variables. Such properties include independent and normal distribution with zero mean and standard deviation of σ_a. In order to identify any possible correlation in the residual time series, their ACF should be investigated. When all ACF components are statistically zero, the whiteness of the residuals is confirmed. In order to affirm that all components are statistically zero, the procedure discussed below can be used.

For 95% and 99% confidence values, the model is accepted if all of the ACF components of the residuals satisfy the following inequalities, respectively [54].

$$\frac{-1.96}{\sqrt{n}} < r_k < \frac{1.96}{\sqrt{n}}, \qquad \text{for} \quad k = 1, 2, \ldots, k_{\max}. \tag{2.14}$$

$$\frac{-2.58}{\sqrt{n}} < r_k < \frac{2.58}{\sqrt{n}}, \qquad \text{for} \quad k = 1, 2, \ldots, k_{\max}. \tag{2.15}$$

where n is the length of the time series and r_k is the k^{th} ACF component. Employing Eq. 2.14 results in models with higher orders, while Eq. 2.15 leads to simpler models with lower orders.

The above inequalities investigate the auto-correlation of all residual components individually. In order to check the adequacy of the overall model, a statistic can be introduced and examined. This statistic is able to conclude whether the first k ACFs of the residuals, considered together, signify the model adequacy. The following two statistical values have been proposed by Box and Pierce (1970) and Ljung and Box (1978), respectively [48]:

$$Q = n' \sum_{l=1}^{k} r_l^2, \tag{2.16}$$

and

$$Q^* = n'(n'+2) \sum_{l=1}^{k} (n'-l)^{-1} r_l^2. \tag{2.17}$$

In which, $n' = n - d$, n is the size of the original time series, and d represents the order of differencing used to transfer the original data to a stationary time series. In addition, the residual ACF at the l^{th} lag is represented by r_l. Both of these statistics can be used to test the adequacy of the models. However, the history confirms the out-performance of Q^*.

A suitable model is supposed to detect the correlations between the available time series data; the residuals should be left uncorrelated. In other words, the auto-correlation values for the residual time series should be small. Hence, Q^* should take small values. The larger the auto-correlation of the residuals are, the more correlated the residuals and the larger the Q^*. Therefore, the inadequacy of a model is asserted when Q^* is of a large value. This test is known as the Portmanteau lack of fit test [55]. According to this test, a model is rejected if the following condition holds: a probability of Type I whose error is equal to α is established. The null-hypothesis is rejected when:

$$Q* > \chi^2_{[\alpha]}(k_{\max} - m), \tag{2.18}$$

where $\chi^2_{[\alpha]}(k_{\max} - m)$ represents a point on the chi-square distribution with $k - m$ degree of freedom such that there is an area of α under the curve of this distribution above this point. Figure 2 shows a chi-square distribution. The number of parameters to be estimated is indicated by m. Usuallyα is chosen equal to 0.05. However, it can be chosen from the range of 0.01 to 0.05; the choice of k is somewhat arbitrary [48].

For a thorough study, this test is employed in four forms. The appropriate model is the one which satisfies inequalities (2.14) and (2.18) with $k_{max} = 10$ or $k_{max} = 20$, respectively and $\alpha = 0.05$ or the model is accepted if inequalities (2.15) and (2.18) with

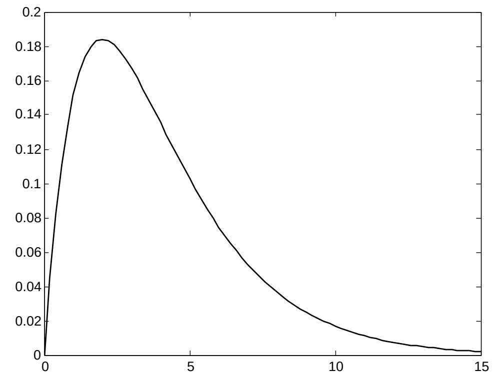


Figure 2.2: Chi-square distribution.

$k_{max} = 10$ or $k_{max} = 20$, respectively, and $\alpha = 0.01$ are satisfied [53].

(b) The Akaike Information Criteria (AIC)

Another useful technique to choose a proper model among a set of candidate models is the AIC [48]. According to this test, the model which minimizes the Kullback-Leibler distance between the original time series and the modeled one is chosen. The AIC test chooses the best fitted model with few parameters.

In this test, an estimator is proposed for the Kullback-Leibler variable $I(h; g)$. This variable is a measure of the asymmetrical distance between the true model of the time series $h(y)$ and the proposed model $g(y|\beta)$ in which the vector of the parameters is represented by β. The proposed estimator is expressed in the following equation [56].

$$AIC_n = -2Ln(L_n) + 2k_n. \tag{2.19}$$

In this estimator, the sample likelihood of the n^{th} model is depicted by L_n. The number of independent parameters estimated for the n^{th} model is represented by k_n. Indeed, over- optimization is avoided by employing the penalty term $2k_n$.

According to this adequacy checking test, the best fitted model is the one whose AIC value is the minimum. Additionally, for the proper model, when the true model coefficients are replaced with those obtained by the maximum likelihood estimate, the smallest information loss occurs [56].

2.4 Grey System Theory

Grey system theory was initially established in 1982. The name ”grey system” refers to the color of the information of the process under investigation. According to this theory, unknown information is represented by black, completely known information by white, and partially known and partially unknown data by grey. Therefore, a black system is one whose information is completely unknown, while a system with completely known information is called a white system. A system with partially known and partially unknown information is called a grey system [57].

The grey system’s objective is to link the separation between science and nature. In the following section, fundamental concepts of grey system theory as well as the mathematical procedure to establish grey models are illustrated.

2.4.1 General Concepts of Grey System.

In various research fields, differential equations are able to deeply explain the aspects of the system variables. The differential equations are useful when the assumption of differentiability is proved. However, for discrete sequences this assumption is not valid. Hence, using differential equations includes some difficulties.

Grey system theory introduced the concept of grey equations. This concept enables researchers to establish models similar to the differential equations for discrete data. In what follows, some definitions are described. These definitions are the building blocks for grey equations [58].

1. **Derivative and Background Value**

   Any first order differential equation is comprised of three components: $\frac{dy}{dt}$ derivative of unknown function y, y the background value of $\frac{dy}{dt}$ and parameters a and b.

$$\frac{dy}{dt} + ay = b. \tag{2.20}$$

2. **Infinite Density Information**

   The information density of $y(t)$ is infinite on a time set T, when

$$y(t + \Delta t) - y(t) \neq 0, \quad \Delta t \to 0. \tag{2.21}$$

   Accordingly, any function satisfying the differential equation has infinite density.

3. **Arithmetic Horizontal Mapping**

   Assume two different sets A and B when $a_1, a_2 \in A$ and $b \in B$. The operator R is said to be a horizontal mapping of a_1 and a_2 when $a_1\ R\ b = a_2\ R\ b$.

   When R is the absolute difference operation $a\ R\ b = |a - b|$, it is called arithmetic horizontal mapping. For a first order differential equation, the derivative and the back ground components satisfy the horizontal mapping relation when

$$\frac{dy}{dt}\ R\ y(t + \Delta t) = \frac{dy}{dt}\ R\ y(t). \tag{2.22}$$

   In which $y(t)$ and $y(t + \Delta t)$ are two components of the background set.

4. **Grey Derivative**

   Consider a set of time sequence Y. The following equation illustrates the grey derivative of a general time sequence.

$$d(k) = y(k) - y(k - 1). \tag{2.23}$$

5. **Grey Differential Equation type**

   The following equation is known as a grey differential type, in which $d(k)$ is the grey derivative.

$$d(k) + ax(k) = b. \tag{2.24}$$

6. **Grey Differential Equation**

   An equation of grey differential type is a grey differential equation if it satisfies the following conditions:

   (a) There exists a horizontal mapping relation between the background value and the grey derivative.

   (b) The information density is infinite.

2.5 Traditional Grey Model GM(1,1)

The traditional grey model GM(1,1) (also known as grey model first order one variable) is the most frequently used model among the family of the grey models. This model has been utilized in a broad range of applications. The GM(1,1) model can be established by developing four phases: Accumulated Generating Operation (AGO), GM(1,1) differential equation, parameters estimation, and calculating the forecasted values [58]. Each phase is described mathematically and also formulated in what follows.

2.5.1 Accumulated Generating Operation (AGO).

Assume the non-negative sequence of original measurements $Y^{(0)}$

$$Y^{(0)} = \left\{ y^{(0)}(1), y^{(0)}(2), ..., y^{(0)}(n) \right\}, \tag{2.25}$$

where $y^{(0)}(k) \geq 0,\ k = 1, 2, ..., n.$

The first stage of establishing the conventional GM(1,1) is to apply the AGO to the measurement time series $Y^{(0)}$. Compared to the original time series, the obtained data $Y^{(1)}$ includes less randomness. The AGO is mathematically expressed as

$$y^{(1)}(k) = \sum_{i=1}^{k} y^{(0)}(i), \;\; k = 1,\; 2,\; \ldots,\; n. \tag{2.26}$$

In which the number of the measurements is represented by n ($n \geq 4$) and the AGO time series of the observation is represented as

$$Y^{(1)} = \left\{ y^{(1)}(1), y^{(1)}(2), ..., y^{(1)}(n) \right\}. \tag{2.27}$$

The grey derivative of the $Y^{(0)}$ can be obtained as follows.

$$d(k) = y^{(1)}(k) - y^{(1)}(k-1) = \\ \sum_{j=1}^{k} y^{(0)}(j) - \sum_{j=1}^{k-1} y^{(0)}(j). \tag{2.28}$$

2.5.2 GM(1,1) Differential Equation.

The relation between the dependent variables and independent ones is expressed by the grey differential equations. Specifically for the traditional GM(1,1), the dynamic of the model is characterized only by one independent variable [58].

The mathematical formulation for the GM(1,1) differential equation can be expressed as follows:

$$y^{(0)}(k) + az^{(1)}(k) = b, \tag{2.29}$$

where the mean generated sequence of the consecutive neighbors of $y^{(1)}$ is represented by $z^{(1)}$.

The whitening equation of Eq.2. can be expressed using the following equation:

$$\frac{dy^{(1)}}{dt} + ay^{(1)} = b, \tag{2.30}$$

where the independent variable is represented by y and the grey model coefficients are depicted by a and b.

2.5.3 Parameter Estimation.

When the grey differential equation is constructed, the least square estimate technique is employed to estimate the grey coefficients. The obtained coefficients are illustrated in the following equations:

$$A = [a, b]^T = \left[B^T B\right]^{-1} B^T Y. \tag{2.31}$$

In which

$$B = \begin{bmatrix} -Z^{(1)}(2) & 1 \\ -Z^{(1)}(3) & 1 \\ \vdots & \vdots \\ -Z^{(1)}(n) & 1 \end{bmatrix}, \tag{2.32}$$

and,

$$Y = \left[y^{(0)}(2), y^{(0)}(3), ..., y^{(0)}(n)\right]^T. \tag{2.33}$$

where $Z^{(1)}$ is the mean generated sequence of the consecutive neighbors.

An equation of grey differential type is said to be a grey differential equation when there exists a horizontal mapping between the grey derivative and the background values. Assume the following grey differential type equation

$$y^{(0)}(k) + ay^{(0)}(k) = b. \tag{2.34}$$

It can be shown that the horizontal mapping condition is not satisfied by the grey derivative $y^{(0)}(k)$ and the background set $\left\{y^{(1)}(k), y^{(1)}(k-1)\right\}$.

$$|y^{(0)}(k) - y^{(1)}(k)| = y^{(1)}(k-1), \tag{2.35}$$

but

$$|y^{(0)}(k) - y^{(1)}(k-1)| = |2y^{(0)}(k) - y^{(1)}(k)|. \tag{2.36}$$

Therefore, the horizontal mapping relation does not hold for such background values. In order to satisfy this condition, it is suggested to replace the background values by the mean of $y^{(1)}$ [58]. The mean generated by the consecutive neighbors is expressed as:

$$z^{(1)}(k) = 0.5(y^{(1)}(k) + y^{(1)}(k-1)). \quad (2.37)$$

It can be simply shown that there exists a horizontal mapping relation between the background value $z^{(1)}(k)$ and the grey derivative components: $y^{(1)}(k)$ and $y^{(1)}(k-1)$.

2.5.4 Calculating the forecasted values.

In order to obtain the predicted values, one should solve the differential equation. Based on Eq.2.30, the AGO solution for the $y^{(1)}(k)$ is as follows [58]:

$$\hat{y}^{(1)}(k+1) = e^{-ak}\left(y^{(0)}(1) - \frac{b}{a}\right) + \frac{b}{a}, \quad k = 1,\ 2,\ \ldots,\ n. \quad (2.38)$$

2.5.5 Inverse Accumulated Generating Operation (IAGO).

The original predicted time series can be obtained utilizing the the inverse of the AGO (IAGO). The IAGO formula is given by

$$\hat{y}^{(0)}(k+1) = e^{-ak}\left(y^{(0)}(1) - \frac{b}{a}\right)(1 - e^{a}),\ k = 1,\ \ldots,\ n. \quad (2.39)$$

The traditional GM(1,1) has a rolling mechanism; the time series of the measurements gets updated when a predicted point becomes an observation. Through this mechanism, the predicted value of $y^{(0)}(k+1)$ is obtained by employing the GM(1,1) to $y^{(0)} = \{y^{(0)}(1), y^{(0)}(2), ..., y^{(0)}(k)\}$ when $k < n$. The original time series gets updated by removing $y^{(0)}(1)$. Hence the updated set of observations can be represented by $y^{(0)} = \{y^{(0)}(2), y^{(0)}(2), ..., y^{(0)}(k+1)\}$. Consequently, the updated set of observations is used to provide the predicted value of $y^{(0)}(k+2)$. In other words, the grey model parameters gets updated when a new observation is added to the training data set.

The most important peculiarity of the grey predictors is that no prior probability distribution of the observation time series is required. Moreover, grey models are capable of using as few as four data pieces as the training data set.

The best prediction accuracy is obtained using the grey models when applied to roughly exponential data. However, for most real world applications, the time series of the measurement does not have such characteristics. Therefore, it makes it difficult to describe the process completely using the grey model individually.

Individual predictors are capable of describing only some aspects of the available data. Thus, employing only one prediction tool results in neglecting some valuable information of the observation time series. Here lies the main concept of hybrid predictors. This can be defined as how to combine the outputs of different predictions with added precision. The combined models are able to describe the data pattern more adequately due to their ability to employ various useful aspects of the data. Therefore, they provide more accurate predictions.

In the following subsections, two hybrid grey-based predictors are introduced and their mathematical procedures are illustrated.

2.6 Hybrid GM(1,1)-ARMA

The hybrid GM(1,1)-ARMA is comprised of the traditional grey model GM(1,1) and ARMA model. This technique takes advantages of both the grey and ARMA models. The following steps describe the general scheme of the GM(1,1)-ARMA prediction approach [59].

1. Assume that Y_t is the original time series.
2. Establish the classical grey prediction model and obtain the corresponding residual time series e_t.
3. Check the ACF of the residual time series in order to affirm its stationarity.

4. Apply the adequacy checking tests, described in Section (2.3.1) to the residual time series and determine the orders of the proper ARMA(p, q) model.

5. Based on the obtained models, forecast the future values of the residual time series.

6. Modify the GM(1,1) forecasted values by utilizing the predicted values of residuals employing ARMA (p, q).

As previously discussed, ARMA models are linear models which employ only the linear deterministic components of a time series to provide the future values. Consequently, the nonlinear predictable components will be left in the residual time series. In order to enhance the prediction accuracy, it is essential to utilize a technique that benefits from the nonlinear components of the measurements.

2.7 Nonlinear Analysis of Time Series

A time series can be deconstructed into notional components using two principles: rates of change and predictability-based decomposition techniques. Decomposition based on predictability refers to the idea of decomposing a time series into predictable and unpredictable parts [60]. Such decomposition is mathematically expressed by:

$$y_t = f(y_{t-1}, y_{t-2}, ..., y_{t-p}, a_{t-1}, a_{t-2}, ..., a_{t-q}) + a_t, \tag{2.40}$$

where y_t is the measurement signal and $f(.)$ represents the deterministic component of the signal that is a function of past values of y and a. The unpredictable components of the time series are depicted as a_t.

The Wold decomposition theory, named after Herman Wold, explicitly proves that any stationary discrete-time stochastic process is comprised of deterministic (predictable) and non-deterministic (unpredictable) components [61]. In more detail, the deterministic part is comprised of linear and nonlinear components.

$$y_t = g(y_{t-1}, y_{t-2}, ..., y_{t-p}, a_{k-1}, a_{k-2}, ..., a_{k-q}) + \\ h(y_{t-1}, y_{t-2}, ..., y_{t-p}, a_{k-1}, a_{k-2}, ..., a_{k-q}) + a_k, \tag{2.41}$$

where the linear and nonlinear deterministic components of the signal are represented by the linear function $g(.)$ and a net nonlinear function $h(.)$ (a function whose Taylor series include no linear components).

As discussed previously, the best fitted ARMA models only consider the linear deterministic components of the time series. Therefore, the nonlinear deterministic components are left in the residuals and are not used during the prediction procedure. Such components can be taken into account using the nonlinear prediction algorithms. However, it is necessary to investigate the significance of nonlinear deterministic components of the measurements before employing any nonlinear technique.

To assess the significance of the nonlinear components of the time series, the surrogate data method can be used. This method is described in the following subsection.

2.7.1 Surrogate Data Approaches.

In order to examine the presence of nonlinearity in time series, the Surrogate data method was initially developed in 1992 [62]. This approach is an indirect technique which employs a hypothesis using the surrogate data to investigate the existence of nonlinear features [63]. The procedure of this technique can be summarized in three steps. First, a null hypothesis which is an assumption for the data is defined. Then, a set of N time series are generated based on the dynamic of the observation time series. Such time series are known as surrogate data. Finally, the same nonlinear statistic Q is calculated for all of the surrogates $Q_1, ...Q_N$ as well as the original data Q_0. The obtained values are ranked from minimum to maximum (Fig.2.3). The rejection or acceptance of the hypothesis is based on the position of the nonlinear statistic value of the original data Q_0 among the ranked values. The hypothesis is rejected if Q_0 falls at either end of the ranked values of the nonlinear measures. The hypothesis is accepted if Q_0 lies within the body of ranked values. In what follows, three major components of this test are described:null hypothesis, a set of surrogate time series, and nonlinear measures.

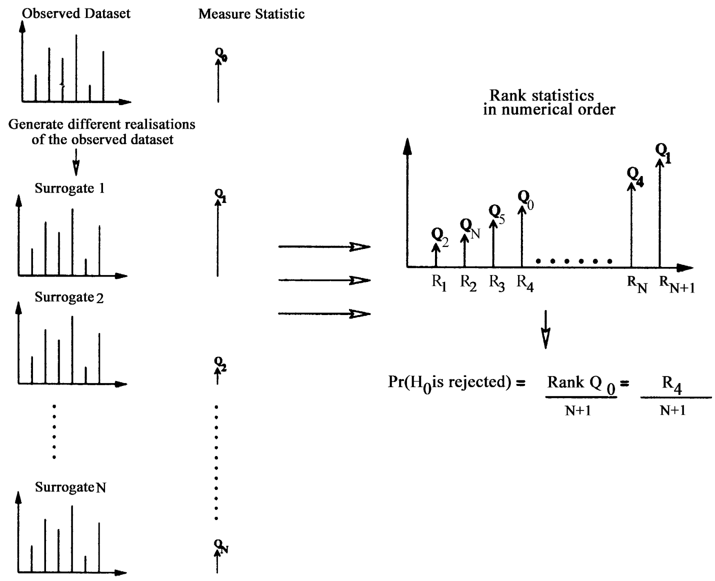


Figure 2.3: Surrogate data approach schematic.

(a) Null Hypothesis

The null hypothesis H_0 is determined based on the under-investigation time series characteristic as well as the objective of the test. An improper null hypothesis results in a meaningless test. Hence, it is essential to properly define the null hypothesis. The objective of this test is to study the presence of nonlinearity in the time series. Therefore, the null hypothesis states that the time series is generated from some linear stochastic process [63]. The null hypothesis rejection and acceptance schematic is shown in Fig.2.4. The null hypothesis can be mathematically formulated based on the utilized nonlinear measures. The nonlinear measures can be categorized into two

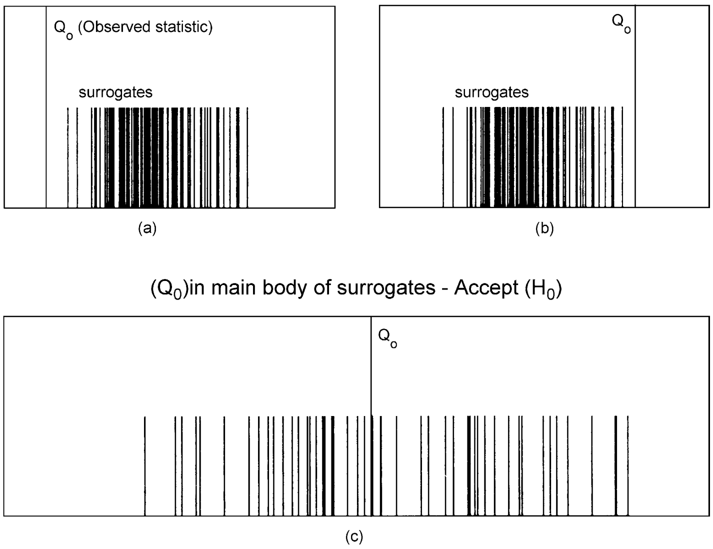


Figure 2.4: Null hypothesis rejection and acceptance schematic.

classes: one sided and two sided measures. For the one sided measures, the null hypothesis is rejected when the inequality (2.42) is satisfied.

$$(M+1)(1-\alpha) < \text{rank}(r_0), \tag{2.42}$$

where M is the total number of surrogate time series and α is the significance level. Assume that the value of a nonlinear measure for the original data y_t^0 and the i^{th} surrogate time series are indicated as $r_0 = r(y_t^0)$ and $r_k = r(y_k^i)$, respectively. When $r_0, r_1, ..., r_M$ are sorted in an increasing manner, the smallest and largest values take rank 1 and $M+1$.

When a two sided nonlinear measure is used, the null hypothesis is rejected if any of inequalities (2.43) and (2.44) are satisfied.

$$(M+1)(1-0.5\alpha) < \textit{rank}(r_0), \tag{2.43}$$

or

$$rank(r_0) \leq 0.5\alpha(M+1). \tag{2.44}$$

(b) Surrogate Time Series

The surrogated time series should satisfy two conditions. First, such time series should be compatible with the null hypothesis. Second, they should be generated with the same linear stochastic properties of the original time series, such as autocorrelation, power spectrum, histogram, etc. In other words, the marginal CDF $F_x(x)$ and autocorrelation r_x of the original and surrogate time series should be of the same values. However, when one condition is fully satisfied, the deviation of the other is possible. For the sake of computational simplicity, all of the proposed surrogate algorithms assume that $F_z(z) = F_x(x)$ and $r_z \cong r_x$.

There are various methodologies to generate surrogate data based on the original time series. The Amplitude Adjusted Fourier Transform (AAFT) is the most well-known surrogate algorithm [62]. The procedure of this algorithm can be described in three steps. First, the marginal cdf is normalized using the ”guassianisation” transformation. Then, any possible nonlinear structure is removed by employing phase randomization so that the linear correlations are not affected. Finally, the original marginal CDF is obtained using the inverse gaussianisation. The AAFT has been used in a broad range of applications. However, it is not capable of matching the linear correlations properly. This drawback stems from the assumption of the monotonic characteristic of the null hypothesis transform [64]. Therefore, the bias in the linear correlations may lead to false rejections when the utilized nonlinear measure is sensitive to linear correlations. In order to approximate the linear correlations more accurately, an improved AAFT (IAAFT) algorithm is proposed in [65]. The proposed approach is comprised of two stages. First, acquire the linear correlation. Second, obtain marginal CDF of the original time series. The obtained surrogates from the IAAFT have similar amplitude spectra to the original time series. Moreover, their apmlitude distribution is identical to the observations. The surrogate using the IAAFT tends to be more powerful than the AAFT.

(c) Nonlinear Measures

Various nonlinear measures can be investigated during the surrogate data approaches. In this thesis, three nonlinear measures are examined: third order auto-covariance, asymmetry due to time reversal, and standard deviation test. In what follows, the mathematic formulation of these measures are expressed.

1. **Third order auto-covariance**

   Third order auto covariance is a higher-order extension of the traditional auto covariance introduced by Schmitz [66]. This measure is a two sided measure whose mathematical description is given by:

   $$T_{C3}(\tau) = \frac{\langle y_k y_{k-\tau} y_{k-2\tau} \rangle}{|\langle y_k y_{k-\tau} \rangle|^{1.5}}, \tag{2.45}$$

   where the time lag is represented by τ and $\langle . \rangle$ depicts the mean function.

2. **Asymmetry due to time Reversal**

   A stochastic process is said to be time reversal when it is invariant with respect to the reversal of the time scale. In other words, for a stationary time reversal process, the joint distribution of $\{X(t_1), ..., X(t_n)\}$ and $\{X(-t_1), ..., X(-t_n)\}$ are of the same values. The following equation describes a possible measure for the asymmetry due to time reversal:

   $$T_{REV}(\tau) = \frac{\left\langle (y_k - y_{k-\tau})^3 \right\rangle}{\left\langle (y_k - y_{k-\tau})^2 \right\rangle^{1.5}}. \tag{2.46}$$

   The value of τ is taken as unity for the sake of convenience in comparison. For the nonlinear time series, the time lag τ has a significant role, while linear time series can be precisely described using $\tau = 1$. Hence, when the null hypothesis is rejected, it can be strongly concluded that the time series is nonlinear.

3. **Standard Deviation**

   This measure divides each time series into several sections with the same length and calculates their corresponding standard deviation values [67].

$$m_j^y = \sqrt{\frac{1}{N^*}\sum_{i=1}^{N^*}(y_i - \bar{y})^2}, \tag{2.47}$$

where m_j^y represents the standard deviation of the i^{th} section, y_i is the i^{th} sample of the j^{th} section, $\bar{y}$ is the mean value of the j^{th} section, and N^* is the length of each section.

Next, the standard deviation of all of the m_j^y is calculated using the following formula.

$$F^y = \sqrt{\frac{1}{n}\sum_{j=1}^{n}\left(m_j^y - \bar{m}^y\right)^2}, \tag{2.48}$$

where n is the number of sections.

When the significance of the nonlinear components of the observation time series is affirmed, it is essential to develop nonlinear predictors to consider such components during the prediction procedure. In the following section, a nonlinear grey- based prediction methodology is proposed.

2.8 Hybrid GM(1,1)-NARnet

In real world applications, time series of the measurements have significant fluctuating characteristics. Hence, neglecting the nonlinear components results in an imprecise forecast. In order to improve the prediction accuracy, nonlinear models should be employed instead of linear ones.

Compared with ARMA models, nonlinear autoregressive predictors provide future values of a time series by the aid of a nonlinear combination of immediate past values of the time series. Therefore, they are used in this study to establish a hybrid grey-based predictor: GM(1,1)-NARnet.

Neural network based models are powerful tools dealing with intricate systems. A neural network is comprised of several different elements. The most basic unit in the neural network construction is neurons, also called the processing elements. Neurons are connected to each other by weighted links. A group of connected neurons makes a layer. By specifying different connection manners for the neurons, different neural network structures will be constructed [68]. The most commonly used neural networks are the MLP networks. Such networks are a subclass of a general class of structures called feed-forward neural networks. In the MLP network architecture, the neurons are grouped into connected layers. Typically, an MLP neural network is comprised of an input layer (the first layer), output layer (the last layer), and one or more hidden layers. The following equation describes the NN procedure mathematically.

$$Y_j = f\left(\sum_i W_{ij} X_{ij}\right), \tag{2.49}$$

where the output of the j^{th} node is represented by Y_j, the connection weight between node j and node i is given by W_{ij}, X_{ij} is the input signal form node i to node j, and $f\,()$ is the nonlinear transfer function between layers.

In this study, in order to enhance the prediction accuracy, a hybrid method is proposed. This technique improves the prediction accuracy by taking advantage of the GM(1,1) and NARnet prediction models. The network architecture includes 3 layers: an input layer, an output layer and one hidden layer. The training algorithm employs the Levenberg-Marquardt back-propagation method constrained on the minimum mean square error. In what follows, the structure of the utilized network is described.

Among different architectures of the NN, the most frequently used structure for time series prediction is the single hidden layer feed forward network. The relation between the out-put of the network and the inputs can be mathematically expressed as:

$$y_t = w_0 + \sum_{j=1}^{q} w_j.g\left(w_{0,j} + \sum_{i=1}^{p} w_{i,j}.y_{t-i}\right) + \varepsilon_t, \tag{2.50}$$

wherew_j and $w_{i,j}$ are the model parameters. The number of inputs and hidden nodes are represented by p and q, respectively.

In this structure there is no activation function assigned to the input layer; the input layer only transfers data to the hidden layer. The activation function used in this study is the sigmoid function described as follows:

$$S(y_t) = \frac{1}{1 - \exp(y_t)}. \tag{2.51}$$

Therefore, the NN provides the future values of a time series using a nonlinear mapping between the previous measurements and the future values. Such nonlinear mapping can be described using the following equation:

$$y_t = f(y_{t-1}, ..., y_{t-p}, W) + \varepsilon_t. \tag{2.52}$$

In which W is the vector of all parameters and $f(.)$ is the function obtained by the network using the model parameters. Therefore, the NN is performing as a nonlinear autoregressive model.

Different training algorithms have been developed for NNs, such as back propagation algorithm, steepest decent method, Gauss Newton algorithm, Levenberg-Marquardt, etc. The Levenberg-Marquardt training algorithm is a numerical approach to minimize a nonlinear function. Compared to other training methods, its convergence is fast. The training process is evaluated using the Sum Square Error (SSE). In this study, the previously described NN is trained using the Levenberg-Marquardt algorithm. During the training procedure, the available data is randomly divided into three kinds of target time steps: training, validating, and testing. The training data sets are used to train and adjust the network according to the corresponding error. The generalization of the network is evaluated using the validation data. The testing data is unseen to the network and is used as a measure of the network's performance after training.

The GM(1,1)-NARnet algorithm can be described as follows:

1. Let $Y^{(0)}$ represent the original measurements and select a proper window size of data as the training set.

2. Establish the GM(1,1) forecasting model and calculate the predicted AGO wind power time series.

3. Obtain the original predicted wind power time series by employing the IAGO. Then calculate the corresponding residuals e_t.

4. Build a proper NARnet predictor by finding the proper number of hidden layers and nodes. Then, train the network and obtain the forecasting values of the grey residuals.

5. Modify the grey predicted time series by utilizing the values of the NARnet model to the grey residuals.

The frame work of the algorithm is summarized in the flowchart shown in Fig.2.5.

2.9 Prediction Evaluation Metrics

In this thesis, the prediction accuracy is evaluated using three different metrics: Relative Percentage Error (RPE), RMSE, and Black-check error. The PRE consists of simple yet powerful criteria to assess the forecast precision. The following formula describes how the RPE is calculated.

$$\text{RPE}(k+1) = \left(\frac{\left| \hat{P}^{(0)}(k+1) - P^{(0)}(k+1) \right|}{P^{(0)}(k+1)} \right) * 100, \tag{2.53}$$

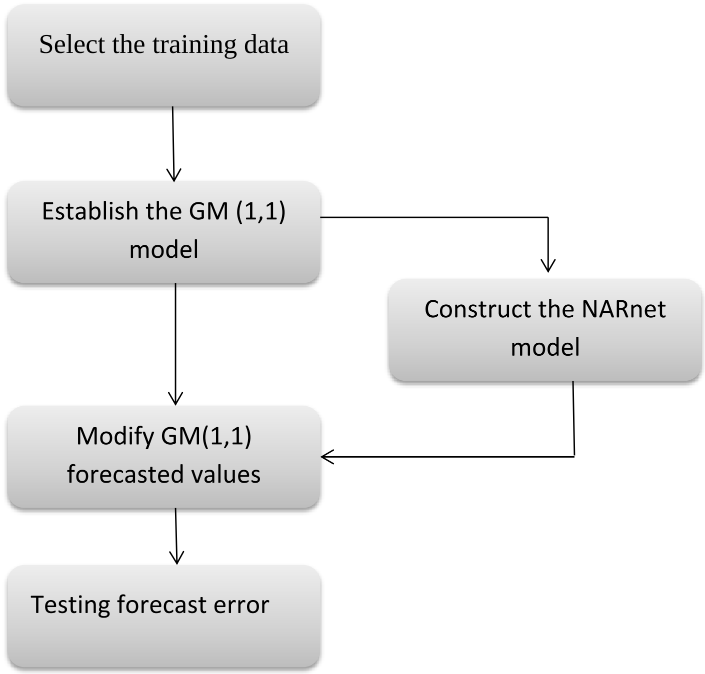


Figure 2.5: GM(1,1)-NARnet frame work.

where the predicted value is depicted by $\hat{P}^{(0)}(k+1)$ and $P^{(0)}(k+1)$ represents the original wind power value.

The RMSE, also known as quadratic mean, is another useful index for checking the prediction accuracy. It is a statistical measure of the magnitude of a varying quantity. Specifically, it is useful for the positive and negative variates.

$$\text{RMSE} = \sqrt{\frac{1}{n} * \sum_{k=1}^{n} \left(\hat{P}^{(0)}(k+1) - P^{(0)}(k+1)\right)^2}. \tag{2.54}$$

The Back-check error is a common index to evaluate the accuracy of the grey models. This index is a ratio of the black-check error C and the small error probability P described by [69]:

$$C = \frac{S_2}{S_1}, \tag{2.55}$$

in which

$$S_1^2 = \frac{1}{n} * \sum_{t=1}^{n} \left(P(t) - \frac{1}{n} \sum_{t=1}^{n} P(t) \right)^2 \tag{2.56}$$

and

$$S_2^2 = \frac{1}{n} * \sum_{t=1}^{n} \left(e(t) - \frac{1}{n} \sum_{t=1}^{n} e(t) \right)^2 . \tag{2.57}$$

The error probability is formulated as

$$P = P \left\{ \left| e(t) - \frac{1}{n} \sum_{t=1}^{n} e(t) \right| < 0.6745 S_1 \right\}, \tag{2.58}$$

where

$$e(t) = P(t) - \hat{P}(t) = (e(1), e(2), ..., e(n)). \tag{2.59}$$

The principles of the grey model perception is illustrated in Table 2.1. According to this table, the model accuracy can be categorized into four grades based on the calculated error probability and the black-check error [69].

Table 2.1: The back-check error principles.

Perception Accuracy	p	C
First Grad	$0.95 \leq p$	$C \leq 0.35$
Second Grad	$0.8 \leq p \leq 0.95$	$0.35 \leq C \leq 0.5$
Third Grad	$0.7 \leq p \leq 0.8$	$0.5 \leq C \leq 0.65$
Fourth Grad (Unqualified)	$p \leq 0.7$	$0.65 \leq C$

Chapter 3: Results

In this chapter, first the wind power data preparation is explained. Then, the wind power time series are pre-analyzed using the ACF investigations and observations of the mean and variance variation over the time. The proper ARMA orders are identified for wind power data using two adequacy checking tests: checking the whiteness of the residuals and the AIC. This is followed by utilizing the obtained models for predicting the wind power. In the next stage, the traditional grey model, the Hybrid GM(1,1)-ARMA, and the GM(1,1)-NARnet are developed.

3.1 Wind Power Data Preparation

The wind power captured by the rotor of a wind turbine is a function of different parameters such as wind speed, wind direction, and air density. The theoretical wind power captured by a wind turbine can be calculated as follows [70]:

$$P_{rotor} = 0.5\rho\pi R^2 C_p(\lambda, \beta)v^3, \tag{3.1}$$

where ρ is the air density, R is the rotor radius, and C_p is a function of the blade pitch angle β and the tip speed ratio λ that is given by:

$$\lambda = \frac{\omega_r R}{v}. \tag{3.2}$$

where ω_r is the rotor rotational speed.

In this thesis, the historical data published by the Alberta Electrical System Operator (AESO) [71] is the data set under investigation. During the prediction process, all variables included in Eq. 3.1 are considered constant except the wind speed. The AESO wind power is obtained from the wind speed time series collected from five wind power farms in South western Alberta. The nameplate capacity for individual

Wind Power Facilities (WPFs) varies between 31 and 75 MW, while the installed total capacity is 254 MW [71].

In order to simulate short-term wind power generation and fluctuation, the AESO has developed a ten minute model of the wind speed data. This technique is comprised of two stages. In the first stage, a weighted moving average algorithm is applied to the wind speed time series. Then, the Multi-turbine power curves are used to convert the wind speed data into wind power. In what follows, these stages are described.

3.1.1 Weighted Moving Average Wind Speed Time Series.

A change in the wind speed typically propagates in the average wind direction by the speed that is comparable to the average speed of the wind. As a result, the measured wind speed within the site during the time period can be used to represent the wind speed time series.

In order to represent the wind speed fluctuations, a weighted moving average is generated. This moving average has a window size that is chosen based on the average of the time series representing the wind speed as well as the spatial distances.

The weighted moving average wind speed time series is generated by specifying a time window. The window size is chosen based on the average of the wind speed time series and the spatial distance. The weighted moving average of the wind speed can be obtained as follows: [71],

$$v_j = \frac{1}{N+1} \sum_{i-\frac{N}{2}}^{i+\frac{N}{2}} w_i v_i, \tag{3.3}$$

where v_i is the i^{th} element of the original wind speed time series, v_j is the j^{th} element in the weighted time series, w_i is the weight associated with the wind speed time series, and N is the number of measurements included in the averaging process.

3.1.2 Spatial Wind Speed Distribution.

It is assumed that at any specific time and for any individual wind turbine the wind speed is normally distributed. The standard deviation of the wind speed is a function of two variables; the spatial dimension and the turbulence intensity of the site. It is proved that for distances smaller than fifty km, the standard deviation can be approximated as a linear function of the distance [72].

3.1.3 Multi-turbine Power Curves.

The wind power output is simulated using a modified multi turbine power curve and a weighted moving average process. The wind speed time series is assumed to be normally distributed. This normal distribution of the wind speed is employed to generate a multi turbine power curve that can be written as:

$$P_j^m = \sum\nolimits_i P_{j+i}^s \times P_i^s, \tag{3.4}$$

where P_j is the j^{th} element of the single turbine power curve and p_i is the spatial distribution probability.

Figure 3.1 represents a comparison between the single turbine and multiple turbines power curve for the AESO wind power facility. For the single turbine power curve, the output power is zero for a wind speed of 4m/s or less. Afterwards, the power curve follows a rapid growth before it gets to a constant value at 16 m/s. The power output is constant above this level until it reaches the cut-off wind speed (25m/s). In comparison with the single turbine power curve, for the multiple wind turbine power curve the minimum wind speed threshold is around 4m/s and the cut-off speed is around 26m/s. Such different threshold levels stem from the fact that in large wind farms, there are some wind turbines whose minimum wind speed level is smaller than that of others. Likewise, there are some wind turbines with larger cut-off values. Hence, the multiple power curve is smoother and has a lower rate of change.

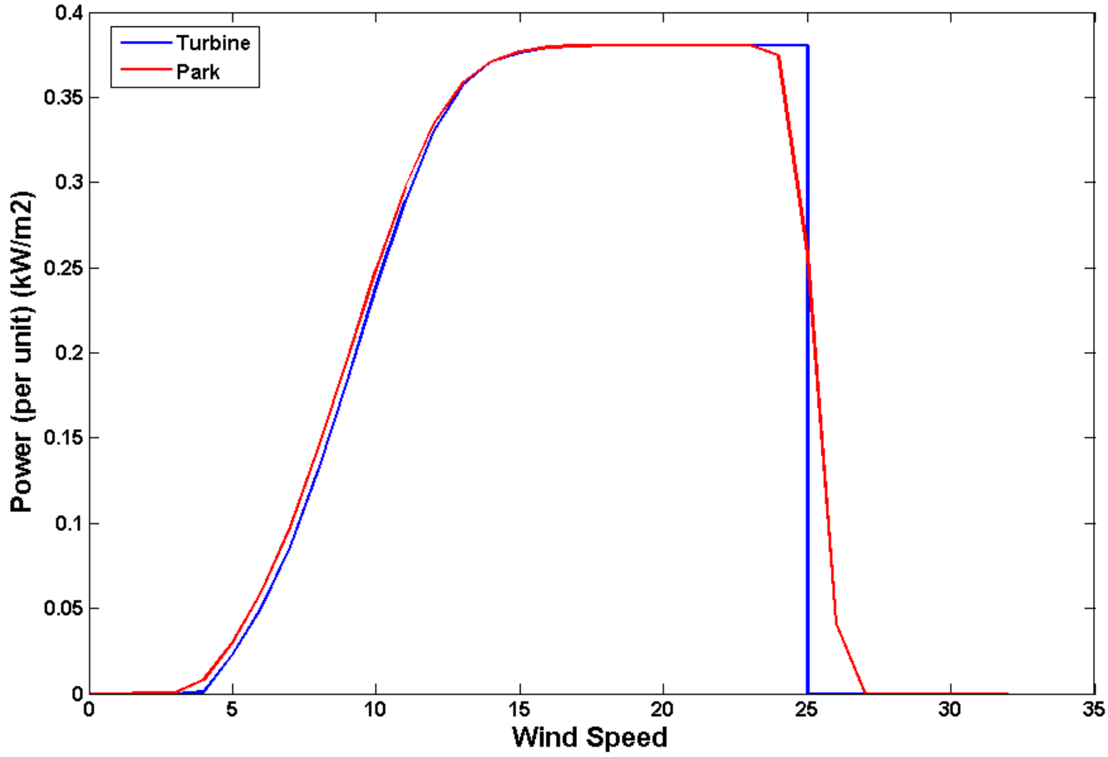
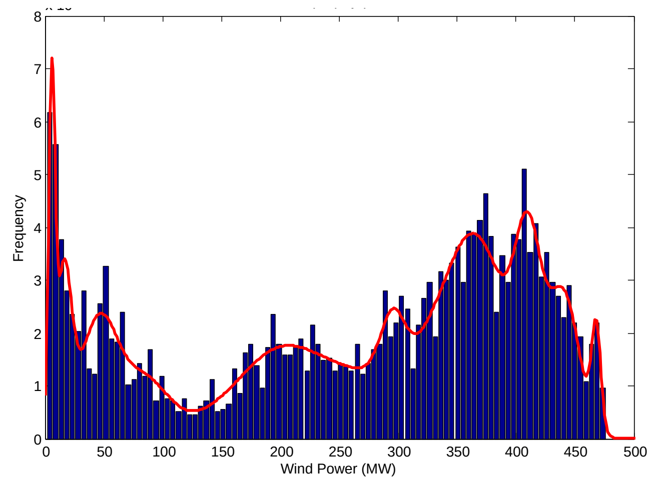


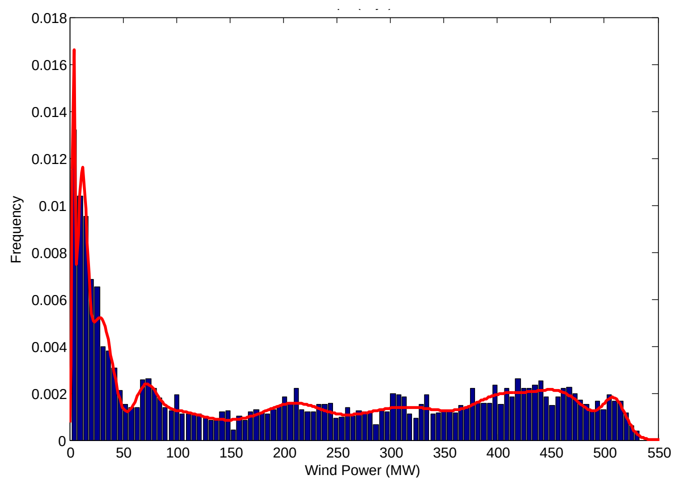
Figure 3.1: Single turbine and multiple turbine power curve of a wind power facility.

3.2 Wind Power Case Study

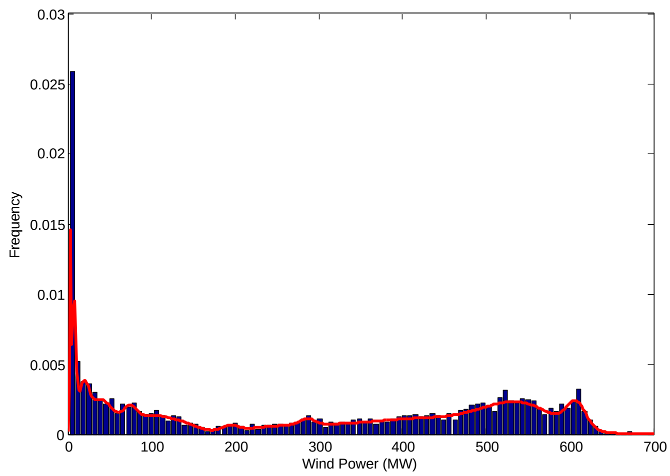
In this study, three years of wind power are investigated: 2009, 2010, and 2011. The wind power data is recorded for every ten minutes. Figures 3.2 and 3.3 show the histogram of the wind power measurements for the three years and the month of January for three years, respectively. In Fig. 3.2 and 3.3, the solid red line shows the fitted GMM model to the time series. The order of the proper GMM model is chosen based on the minimum Log-Likelihood value to ensure accurate model coefficients. Because of the fluctuations seen in Fig. 3.3, a higher order GMM model is recommended for such data.



(a) Jan. 2009.

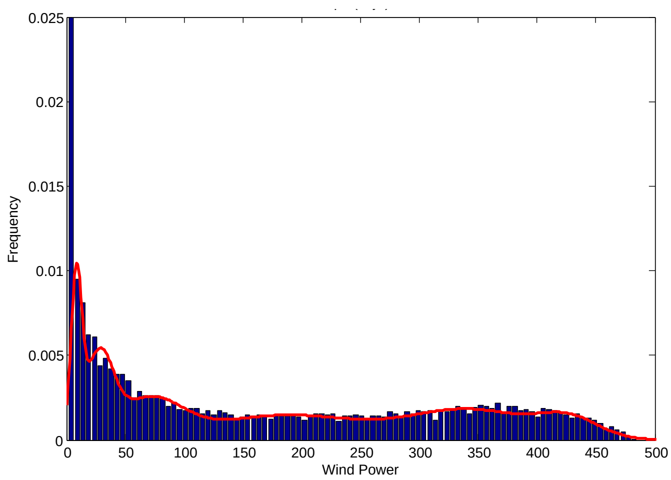


(b) Jan. 2010.

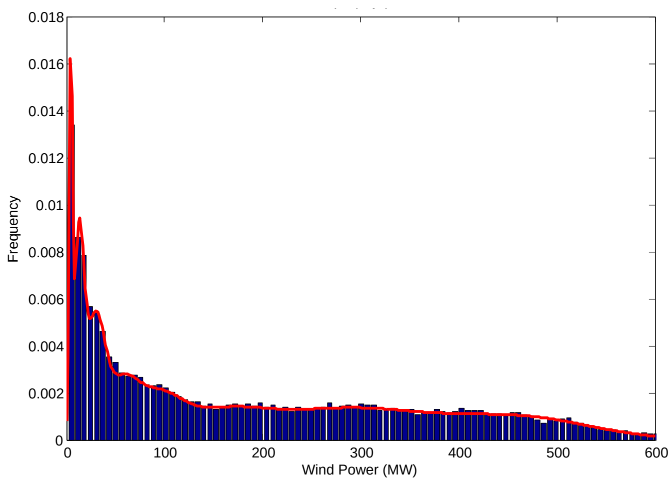


(c) Jan. 2011.

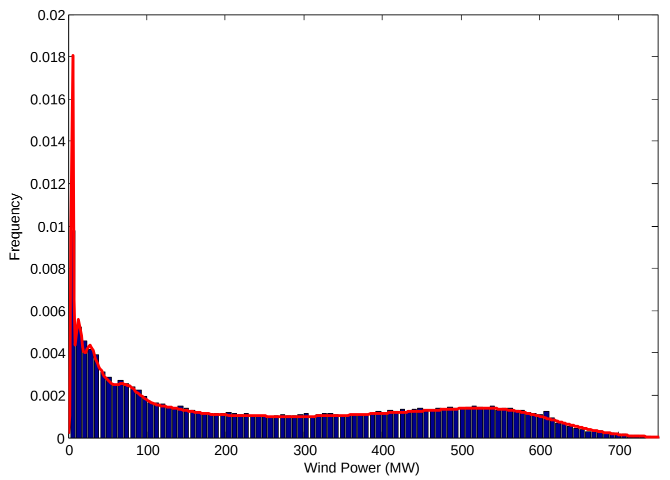
Figure 3.2: Month of January histogram.



(a) Year 2009.



(b) Year 2010.



(c) Year 2011.

Figure 3.3: Yearly histogram.

3.3 Wind Power Time Series Pre-analysis

As discussed previously, stationarity is the basis of many time series models such as ARMA models. Such models are applicable only to stationary data. Thus, it is essential to investigate the stationarity of the measurements' time series. Different methods can be used to study the stationarity of a time series. The most frequently used methods include the ACF investigation and the time series statistical moments observation over time.

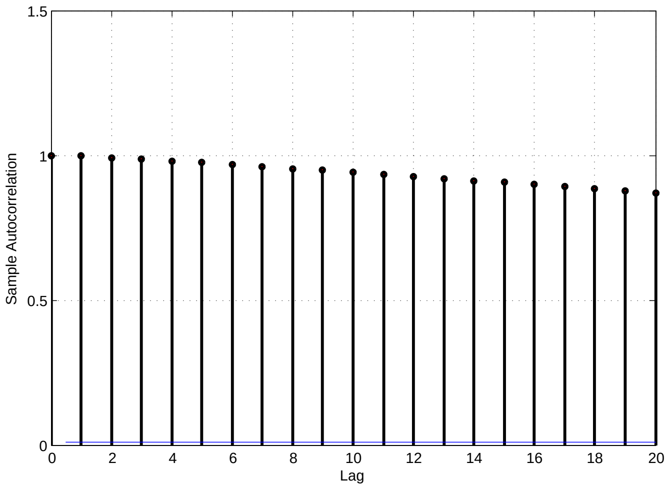
In the ACF analysis technique, the ACF of the time series is plotted. The time series is considered as stationary when its ACF cuts off at small lag, while exponential or sinusoidal decay of the ACF is an indicate of the non-stationarity of the data.

Another technique used to investigate stationarity is the statistical moments observation over time. In this method, different orders of time series statistical moments should be analyzed. When the time series moment of a specific order is time invariant, higher orders should be studied. The non-stationarity of the time series is proven if a moment of a specific order is time variant. In this section, a pre-analysis is done on the wind power time series to examine its stationary characteristic.

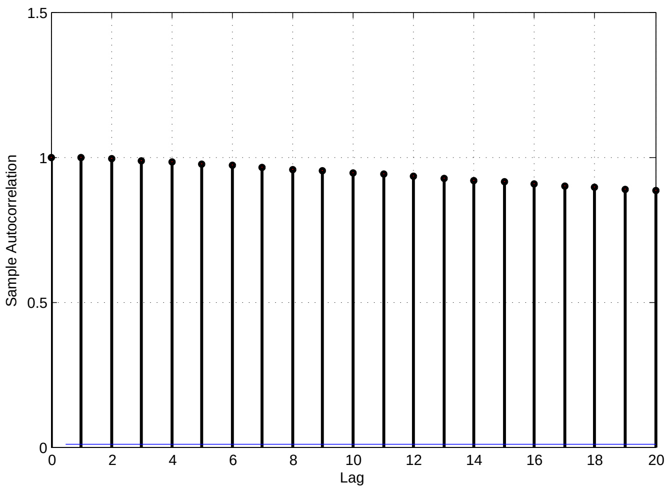
3.3.1 ACF Investigation.

The time series ACF investigation is one of the most popular techniques among researchers for proving the stationarity or non-stationarity of data sets.

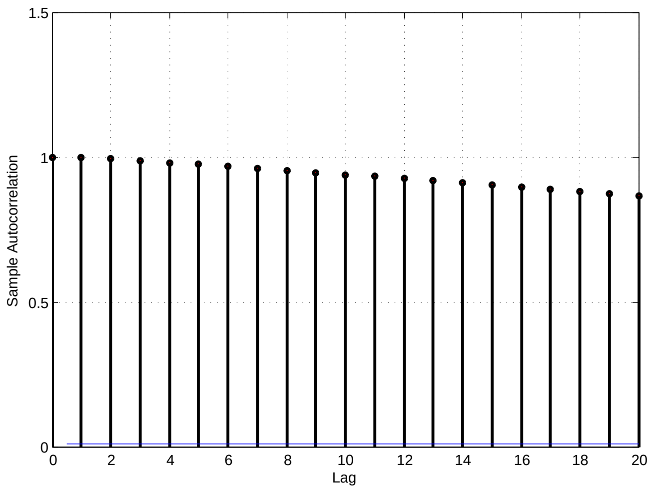
The ACF components represent the correlation between two consecutive time series measurements. A time series is stationary if its ACF components diminish immediately at small lags, while for non-stationary time series they show a linear or sinusoidal decaying trend. Figure 3.4 shows that in terms of the corresponding ACF, the time series behave similarly for the three years. Such similar behavior of the time series could also be seen from the ACF for the same month for the years as shown in Fig. 3.5. The slow decaying behavior of the ACF of both yearly and monthly data indicates the non-stationarity of the time series.



(a) Year 2009.

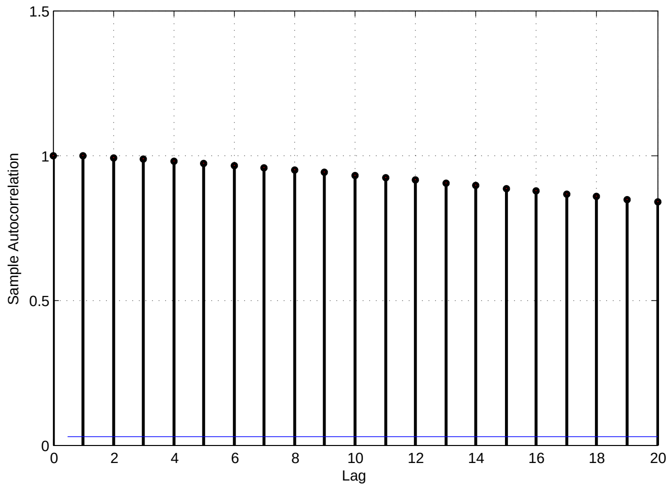


(b) Year 2010.

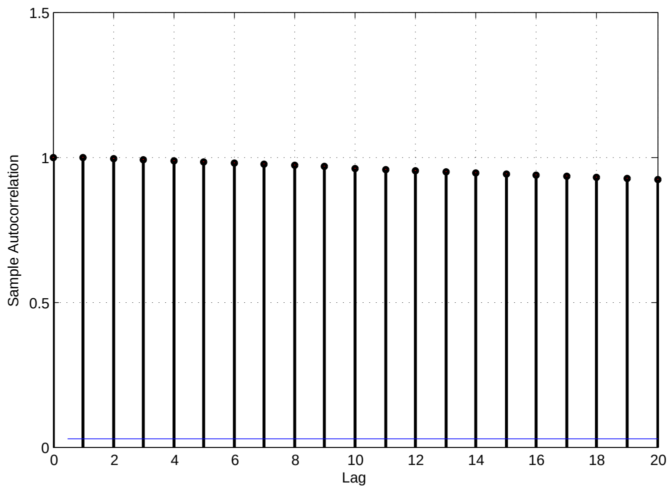


(c) Year 2011.

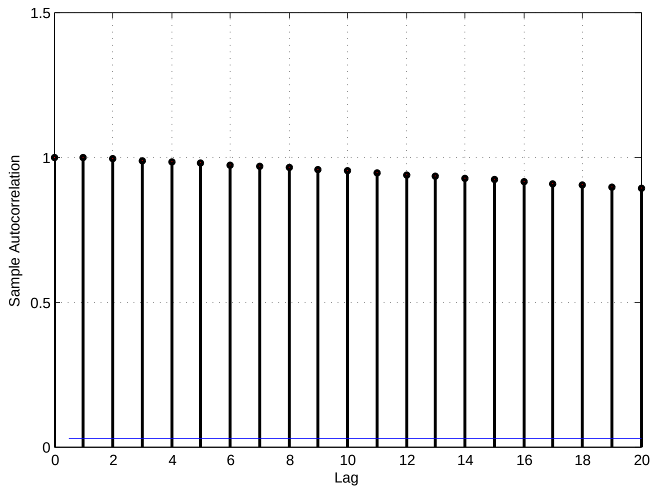
Figure 3.4: Sample ACF for original yearly measurements.



(a) Jan.2009.



(b) Jan.2010.



(c) Jan.2011.

Figure 3.5: Sample ACF for original monthly measurements.

3.3.2 Mean and Variance Observation.

Another method to investigate the stationarity of a time series is to observe the variation of its mean and variance over the time. For stationary time series, the mean and variance are not time-variant.

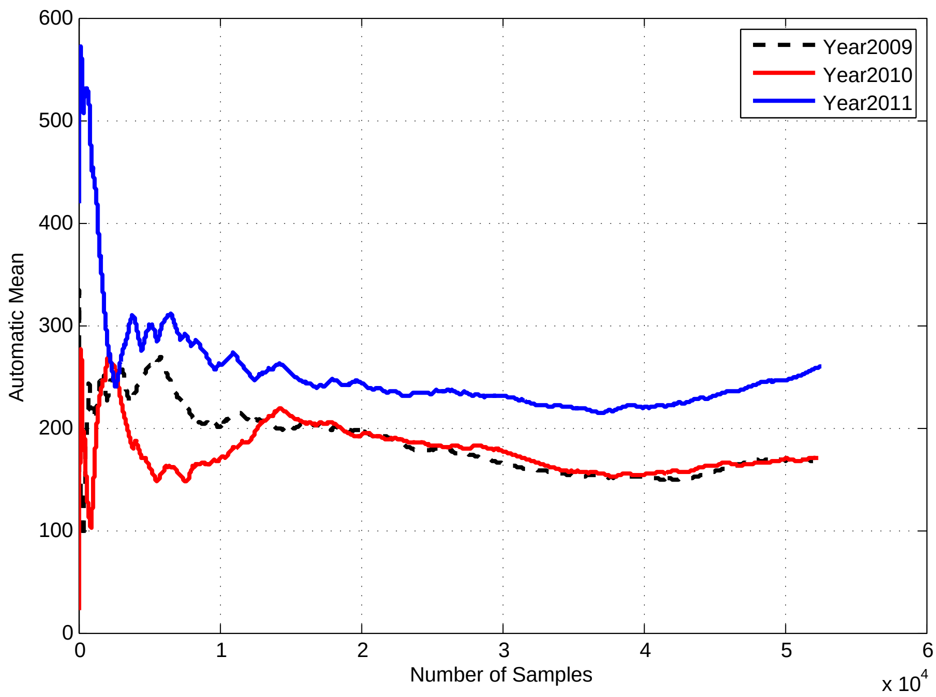
Figure 3.6 shows the mean variation of yearly and monthly wind power data. The mean of the wind power is changing over time for both yearly and monthly wind power time series. In Fig. 3.7 the variation of the automatic variance of wind power time series is illustrated: for all time series the variance is changing over the time. Such variations indicate the non-stationarity of the time series. When the non-stationarity of a time series is confirmed, the transformation techniques should be utilized to remove such characteristics.

3.3.3 Differencing Wind Power Time Series.

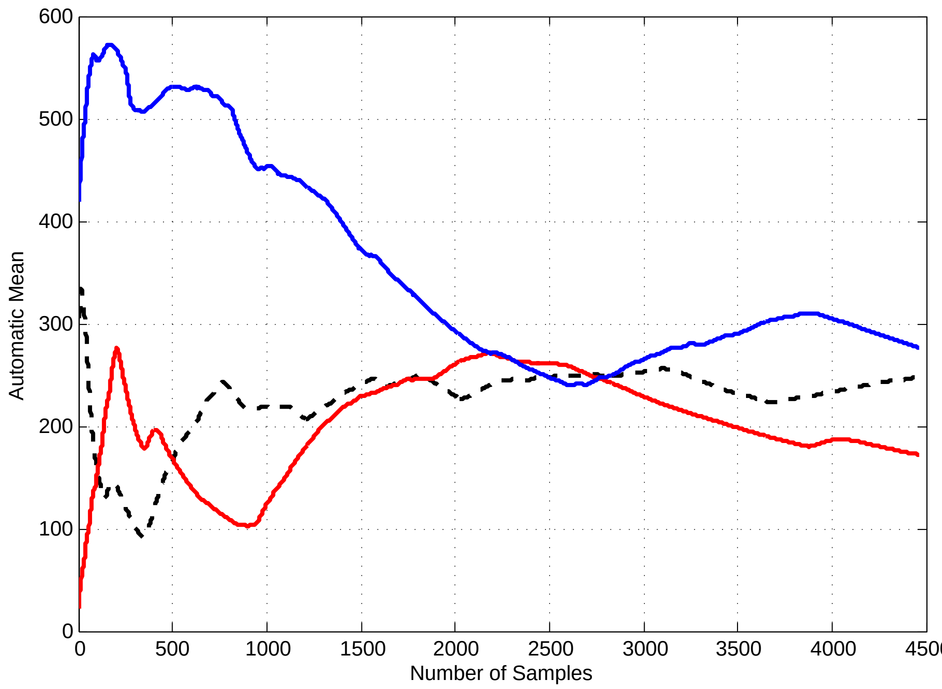
The non-stationarity of the wind power under investigation is proved using the ACF investigation and the observation of the variation of the mean and variance over the time. Therefore, the data set should be transformed to a stationary time series.

Differencing the time series is one of the most frequently used transformations to remove the non-stationary nature of the time series. This transformation utilizes the correct smallest differencing order which typically results in the minimum variance. During the differencing process, over-differencing must be avoided since it can cause data loss.

For the wind power time series used in this thesis, it is found that a second order differencing transformation is sufficient to remove the non-stationarity of the data. Figures 3.8 and 3.9 depict the ACF of the differenced wind power data for yearly and monthly wind power data sets, respectively. The obtained time series are stationary since their ACF diminishes at small lags. When the stationarity of the wind power time series is confirmed, it can be modeled using time series models such as ARMA. In the next section, the proper ARMA model identification techniques are explained.

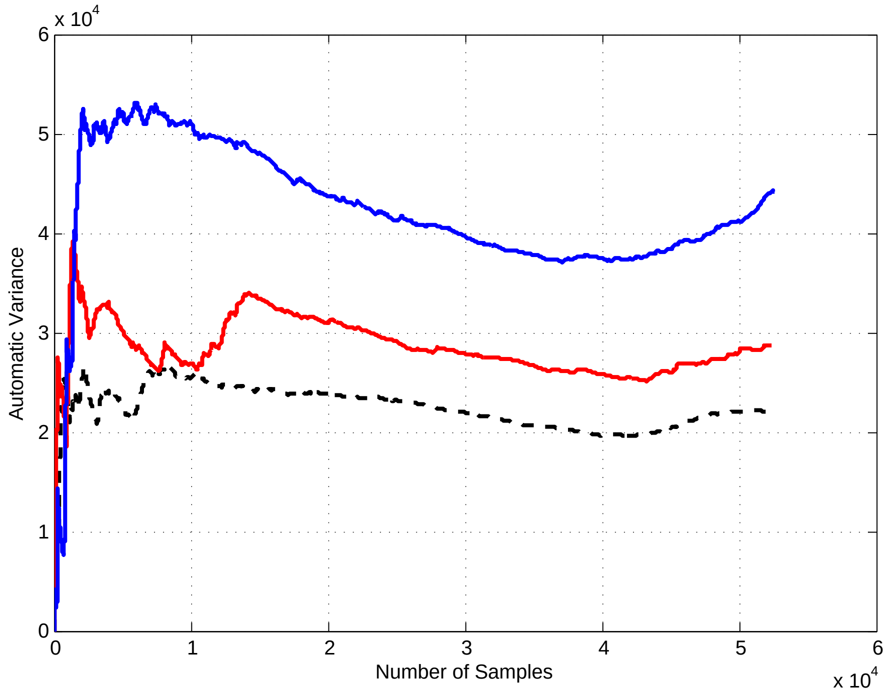


(a) Yearly wind power data.

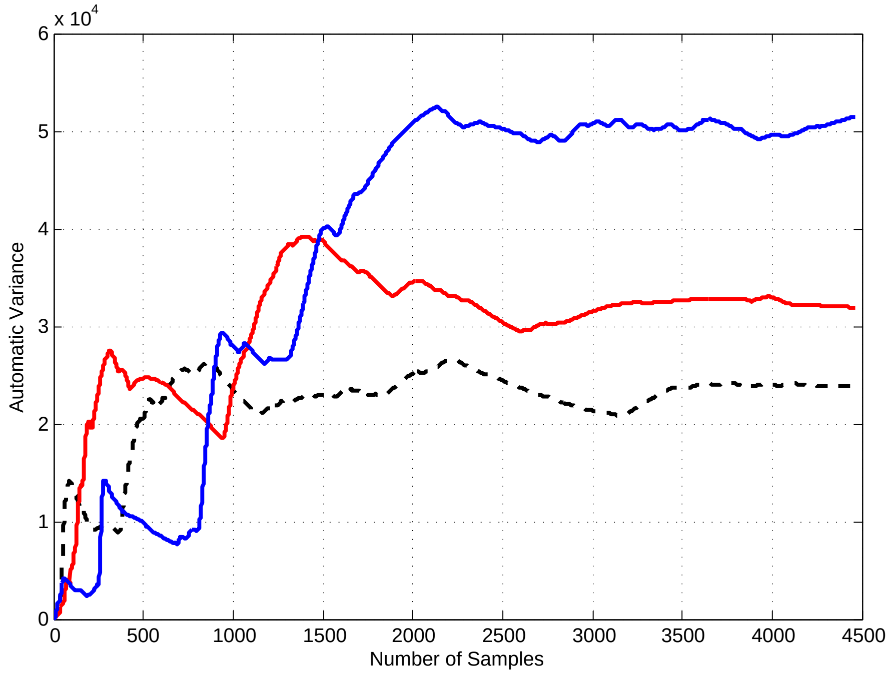


(b) Monthly wind power data.

Figure 3.6: Variations of the automatic mean of yearly and monthly wind power data.

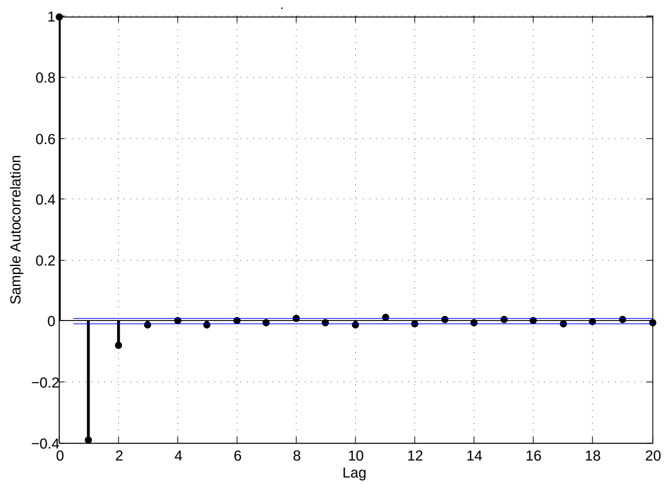


(a) Yearly wind power data.

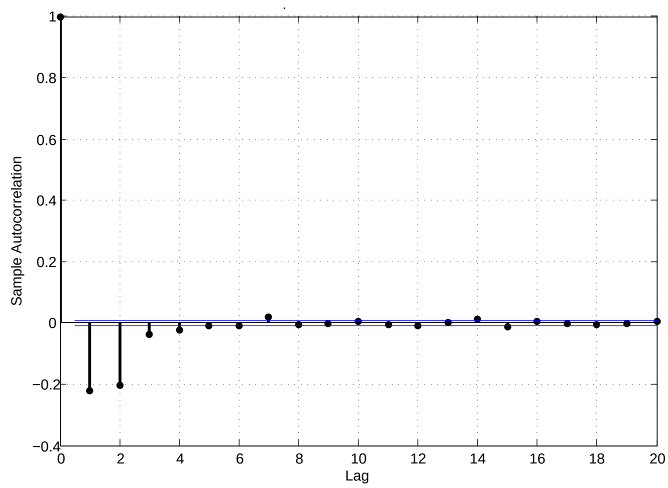


(b) Monthly wind power data.

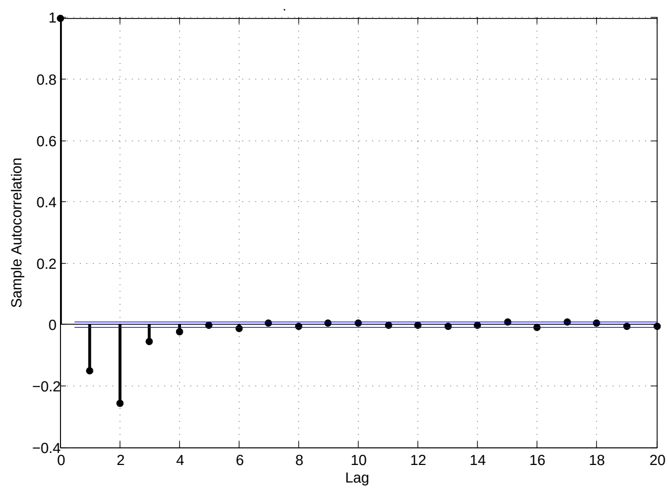
Figure 3.7: Variations of the automatic variance of yearly and monthly wind power data.



(a) Year 2009.

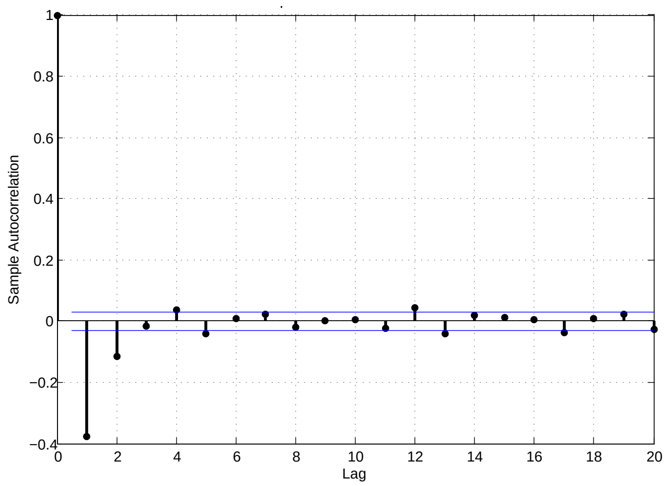


(b) Year 2010.

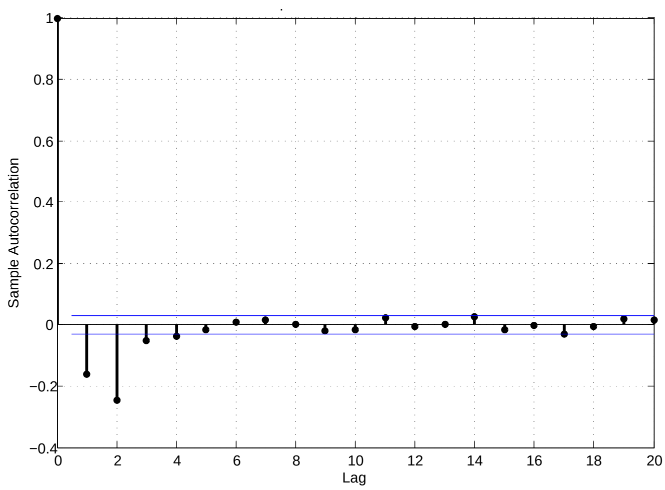


(c) Year 2011.

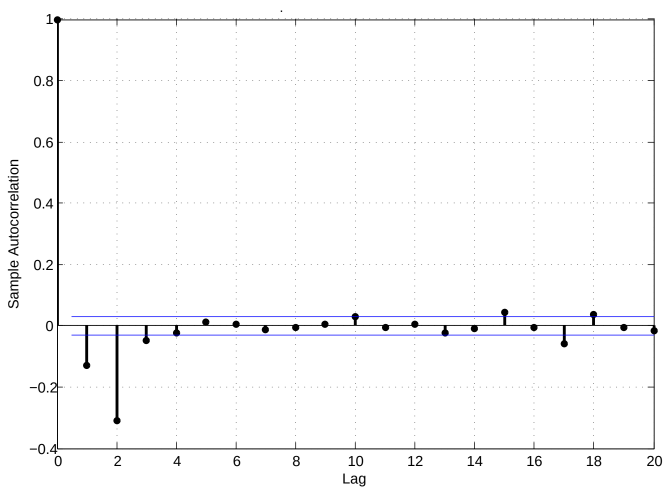
Figure 3.8: Sample ACF for yearly differenced time series.



(a) Jan.2009.



(b) Jan.2010.



(c) Jan.2011.

Figure 3.9: Sample ACF for monthly differenced time series.

3.4 Choosing The Proper ARMA Orders

The ARMA models are linear stationary models which provide the future values of a time series as a linear combination of a number of immediate past values. Such models are characterized by two parameters: p and q. The numbers of the past values required for an efficient prediction are represented by such parameters. The major issue with the ARMA models lies in the proper identification of the model orders: p and q. In the following subsections, two adequacy checking tests are developed to determine the best fitted ARMA models for the wind power time series.

3.4.1 Checking the Whiteness of the Residuals.

The process of determining the proper ARMA model is made up of two phases: the parameter estimation and the validation phase. In the parameter estimation phase, several model orders are tested while their coefficients are estimated [48].

A suitable ARMA model is capable of capturing the correlated useful components. Such components can be used during the prediction procedure. In other words, fitting the proper ARMA models to the time series leaves the residuals (ε_t) with non-correlated components. Hence, as explained in Section 2.3.1, the ACF of the residuals is statistically zero.

In this thesis, for the sake of simplicity one hundred ARMA models are exhaustively investigated for the proper ARMA model identifications. Specifically, the parameters (p, q) for the candidates models are selected from the set of values of $p = 0, 1, 2, 3, ..., 10$ and $q = 0, 1, 2, 3, ..., 10$. For all candidate models, the whiteness of the prediction residuals are investigated to determine the best fitting ARMA models for the wind power time series.

Four scenarios of the whiteness of the residuals adequacy checking test, described in Section 2.3.1, are applied to the monthly and yearly wind power time series. The obtained ARMA (p, q) models are presented in Tables 3.1 and 3.2.

Table 3.1: Proper ARMA models of monthly time series for years 2011, 2010 and 2009 from whiteness of residuals tests 1 to 4

2011											
Jan.	Feb.	Mar.	Apr.	May	Jun.	Jul.	Aug.	Sep.	Oct.	Nov.	Dec.
(7,10)	(7,1)	(8,10)	(9,3)	(3,5)	(6,8)	(9,8)	(4,7)	(2,9)	(8,9)	(10,4)	(3,10)
(8,2)	(6,8)	(8,9)	(9,6)	(3,7)	(6,6)	(10,3)	(4,10)	(3,3)	(9,3)	(9,8)	(4,3)
(7,7)	(6,6)	(8,8)	(8,8)	(2,8)	(6,1)	(8,8)	(3,9)	(2,4)	(8,5)	(9,6)	(3,5)
(7,9)	(6,4)	(8,3)	(9,2)	(1,10)	(5,7)	(9,1)	(3,10)	(2,1)	(8,1)	(9,2)	(3,7)
2010											
Jan.	Feb.	Mar.	Apr.	May	Jun.	Jul.	Aug.	Sep.	Oct.	Nov.	Dec.
(5,3)	(10,6)	(10,9)	(8,10)	(6,2)	(8,6)	(7,9)	(6,1)	(7,7)	(10,3)	(9,1)	(5,10)
(4,9)	(10,9)	(10,10)	(8,4)	(6,5)	(8,4)	(7,6)	(5,6)	(5,10)	(9,9)	(8,6)	(6,6)
(4,4)	(9,7)	(10,4)	(7,9)	(5,6)	(7,9)	(7,3)	(5,1)	(5,8)	(9,4)	(7,8)	(5,2)
(4,7)	(10,1)	(9,8)	(7,7)	(5,2)	(8,1)	(7,5)	(5,4)	(5,9)	(8,8)	(8,4)	(5,4)
2009											
Jan.	Feb.	Mar.	Apr.	May.	Jun.	Jul.	Aug.	Sep.	Oct.	Nov.	Dec.
(7,9)	(8,6)	(10,1)	(7,3)	(6,7)	(9,10)	(9,9)	(5,7)	(6,10)	(5,3)	(9,1)	(9,9)
(8,2)	(8,3)	(10,5)	(6,9)	(6,3)	(10,6)	(10,3)	(7,3)	(7,4)	(5,7)	(8,7)	(10,2)
(7,5)	(6,8)	(9,5)	(6,1)	(5,7)	(9,9)	(8,10)	(6,1)	(6,3)	(4,6)	(8,3)	(9,1)
(7,8)	(7,2)	(9,7)	(6,4)	(5,10)	(9,7)	(8,7)	(6,4)	(5,10)	(4,10)	(8,5)	(9,7)

Table 3.2: Proper ARMA models of yearly time series for years 2011, 2010 and 2009 from whiteness of residuals tests 1 to 4

Test / Data	**Year 2009**	**Year 2010**	**Year 2011**
1	(10,3)	(10,3)	(9,7)
2	(10,7)	(9,10)	(9,10)
3	(9,7)	(9,3)	(8,10)
4	(10,1)	(9,6)	(9,3)

As seen in Table 3.1 for some of the wind power time seres, the obtained ARMA models are of high order, while for others they are of small value. Such differences in the obtained orders stem from the different original measurements values. The wind power data differs from year to year or month to month due to various reasons such as different numbers of installed wind turbines, temperature, climate, etc.

Table 3.2 shows that for all yearly data sets the four developed scenarios result in high order ARMA(p, q) models. Such high orders increase the number of required previous measurements as well as the complexity of the prediction algorithm.

3.4.2 Akaikes Information Criterion (AIC).

Another useful test to select the proper ARMA model is the AIC [48]. The AIC test is also employed on the monthly and yearly wind power time series. The obtained model parameters are summarized in Tables 3.3 and 3.4. Results shown in Table 3.3

demonstrate that for a specific month of different years, the obtained ARMA models are of different values. The climate and topology changes can result in such differences. The obtained ARMA models summarized in Table 3.4 are of high order. This confirms the fluctuating and highly varied nature of wind power.

Table 3.3: Proper ARMA models of monthly time series for years 2009, 2010 and 2011 obtained from AIC test

Year / Month	Jan.	Feb.	Mar.	Apr.	May	Jun.	Jul.	Aug.	Sep.	Oct.	Nov.	Dec.
2009	(7,10)	(7,10)	(9,9)	(6,6)	(6,9)	(10,10)	(9,7)	(6,6)	(6,7)	(4,9)	(8,9)	(9,5)
2010	(4,10)	(10,8)	(10,7)	(8,5)	(5,9)	(8,3)	(7,7)	(5,7)	(6,10)	(10,10)	(8,9)	(5,7)
2011	(7,10)	(6,9)	(9,8)	(6,8)	(2,10)	(6,8)	(10,7)	(4,9)	(2,5)	(8,6)	(9,10)	(3,8)

Table 3.4: Proper ARMA models of yearly time series for years 2009, 2010 and 2011 obtained from AIC test

Year	Proper ARMA Model
2009	(10,10)
2010	(9,8)
2011	(9,9)

3.5 Wind Power Prediction Using the Proper ARMA Models

In this section, the best fitting ARMA models obtained from Subsections 3.3.1 and 3.3.2 are used to provide one hour ahead (six steps ahead) wind power prediction. In order to compare the accuracy of different predictions, the prediction RMSE values are calculated.

The obtained RMSE values for monthly wind power data for the years of 2009, 2010, and 2011 are presented in Tables 3.5, 3.6, and 3.7, respectively. For most cases, the RMSEs of the AIC are smaller than those obtained from the four scenarios of checking the whiteness of the residuals. Table 3.5 demonstrates that the smallest prediction RMSE value for January. (12.8448 MW) is obtained using the AIC. The wind power measurement varies between 0 and 674.4433 MW during January. 2009. Therefore, the prediction error of 12.8448 MW is an indicator of an accurate forecast.

Table 3.5: RMSE of wind power prediction of monthly time series for year 2009 obtained from AIC and whiteness of residuals tests

2009					
Month / Test	**AIC**	**Whiteness of Res-1**	**Whiteness of Res-2**	**Whiteness of Res-3**	**Whiteness of Res-4**
Jan.	12.8448	12.8827	12.9188	12.9074	12.8447
Feb.	7.8114	7.8462	7.8590	7.8432	7.8638
Mar.	13.8916	13.9796	13.9457	13.9431	13.9231
Apr.	11.3544	11.4200	11.4161	11.4355	11.4187
May.	12.5008	12.5319	12.5464	12.5387	12.5358
Jun.	10.5162	10.5523	10.5598	10.5623	10.5658
Jul.	10.9217	10.9158	10.9602	10.9421	10.9446
Aug.	10.7304	10.7392	10.7414	10.7541	10.7482
Sep.	12.1949	12.2580	12.2664	12.2937	12.2836
Oct.	11.7569	11.7649	11.7487	11.7586	11.7571
Nov.	13.9311	14.0092	13.9540	14.0117	14.0121
Dec.	9.3840	9.3991	9.4271	9.4302	9.3971

Table 3.6: RMSE of wind power prediction of monthly time series for year 2010 obtained from AIC and whiteness of residuals tests

2010					
Month / Test	**AIC**	**Whiteness of Res-1**	**Whiteness of Res-2**	**Whiteness of Res-3**	**Whiteness of Res-4**
Jan.	10.4241	10.4450	10.4192	10.4275	10.4367
Feb.	7.6271	7.6329	7.6246	7.6269	7.6425
Mar.	13.0249	13.0927	13.0692	13.1332	13.0927
Apr.	13.1600	13.1193	13.1729	13.1670	13.1687
May.	11.1140	11.1577	11.1332	11.1108	11.1576
Jun.	11.2274	11.2283	11.2273	11.2268	11.3264
Jul.	12.5677	12.5621	12.5768	12.5938	12.5926
Aug.	4.9119	4.9385	4.9243	4.9419	4.9255
Sep.	11.5642	11.5802	11.5771	11.5971	11.5849
Oct.	10.8880	10.9882	10.9161	10.9856	10.9769
Nov.	13.2135	13.3077	13.2282	13.2801	13.2717
Dec.	12.3499	12.3743	12.3453	12.4156	12.3708

Table 3.7: RMSE of wind power prediction of monthly time series for year 2011 obtained from AIC and whiteness of residuals tests

2011					
Month / Test	**AIC**	**Whiteness of Res-1**	**Whiteness of Res-2**	**Whiteness of Res-3**	**Whiteness of Res-4**
Jan.	14.7571	14.7571	14.8909	14.8861	14.7772
Feb.	14.9787	15.0420	14.9774	14.9836	15.0434
Mar.	11.4031	11.4008	11.4181	11.4248	11.4509
Apr.	12.7737	14.1787	14.1672	14.1674	14.1788
May.	11.3632	11.4011	11.4013	11.3991	11.3985
Jun.	12.3134	12.3134	12.3287	12.3343	12.3305
Jul.	14.0835	14.1648	14.2071	14.1685	14.2184
Aug.	10.8481	10.8564	10.8372	10.8536	10.8531
Sep.	11.8220	11.8132	11.8191	11.8218	11.8226
Oct.	13.0905	11.7869	11.8021	11.8114	11.8198
Nov.	17.9881	18.0785	18.0418	18.0579	18.0845
Dec.	17.6512	17.6473	17.6867	17.6569	17.6536

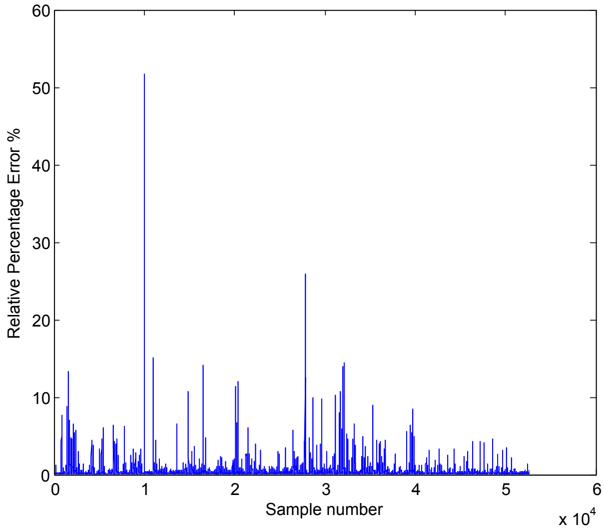
The obtained ARMA models are applied to the yearly wind data sets. Table 3.8 shows the RMSE values corresponding to the ARMA models. For most of the monthly and yearly time series, the RMSE values obtained from the AIC tests are smaller compared to those of the four scenarios of checking the whiteness of residuals.

Table 3.8: RMSE of wind power prediction (MW) of yearly time series for year 2009, 2010 and 2011 obtained from four tests

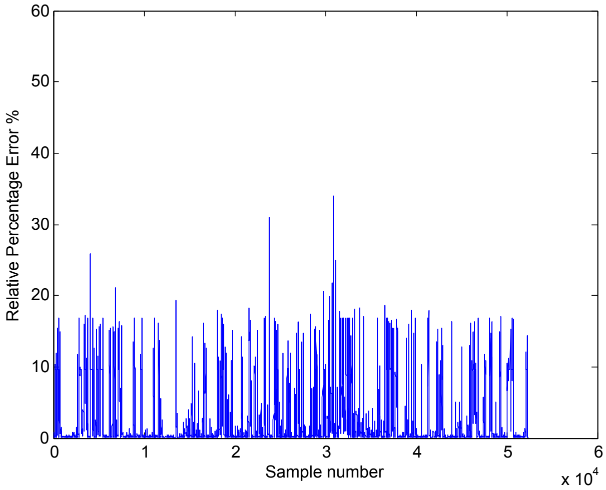
RMSE(MW)	2009	2010	2011
AIC	11.6975	12.1245	13.9217
Whiteness of Res-1	11.7475	11.5879	14.3226
Whiteness of Res-2	11.7041	11.5861	14.4210
Whiteness of Res-3	11.7469	11.6354	14.3318
Whiteness of Res-4	11.7478	11.6349	14.4237
Min RMSE	AIC	Whiteness of Res-2	AIC

The proper ARMA models for yearly data summarized in Table 3.2 are used to provide future values of yearly time series. The obtained ARMA prediction errors are illustrated in Fig. 3.10. In Fig.3.10, the prediction error for yearly data are plotted. Figure 3.11 shows the predicted values of the monthly data versus the measurements. Figure 3.12 demonstrates that the prediction error histograms for April and the year 2011 are normally distributed.

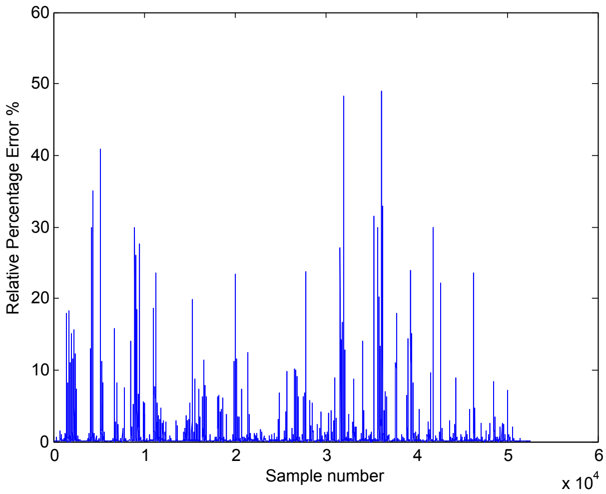
The efficiency of the ARMA models is affected by the accuracy of the measurements. Therefore, any measurement errors affect the accuracy of the predicted values obtained by the ARMA models. Other predictors capable of moderating the effects of such errors on the prediction accuracy should be developed. The grey models deal with systems with partially known and partially unknown information. Hence, they can be good alternatives for ARMA models. In the following section, grey models are utilized to provide short-term wind power predictions.



(a) Year 2009.

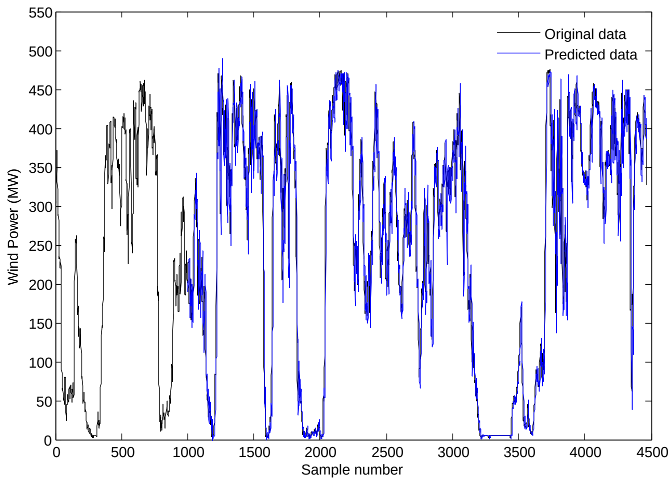


(b) Year 2010.

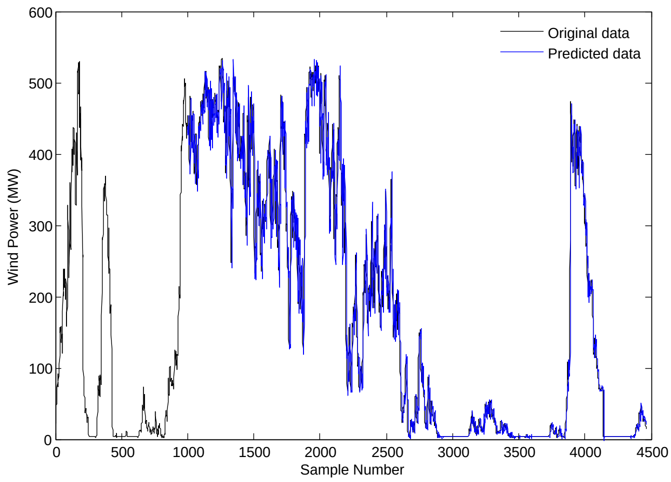


(c) Year 2011.

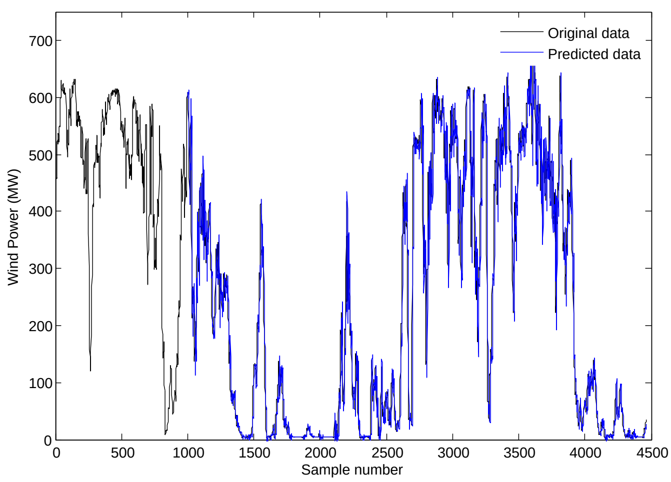
Figure 3.10: Wind power ARMA prediction RPE for yearly measurements.



(a) Jan.2009.

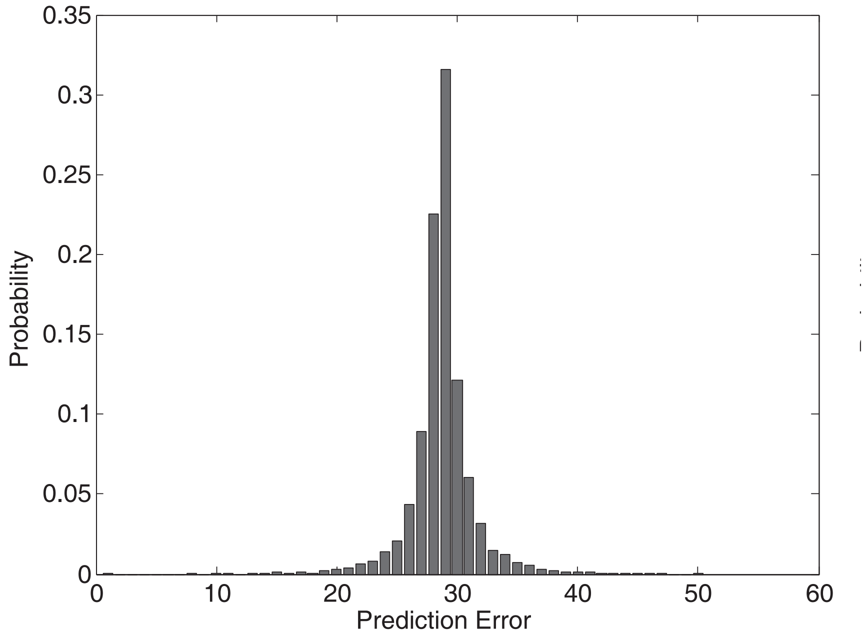


(b) Jan.2010.

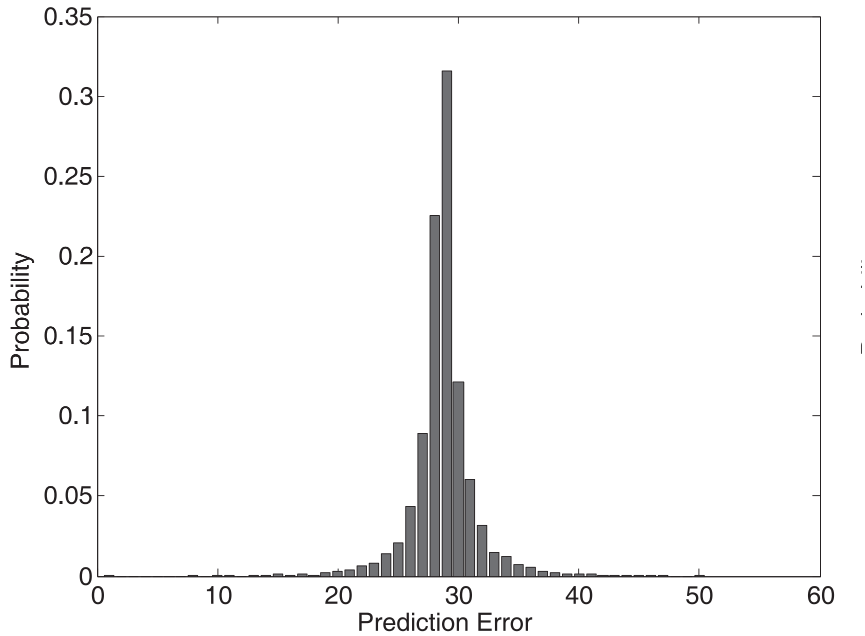


(c) Jan.2011.

Figure 3.11: Wind power prediction error for monthly data versus measurements.



(a) April 2011.



(b) The year 2011.

Figure 3.12: Prediction error histogram.

3.6 Wind Power Prediction Using Grey Models

The AESO wind power time series is collected from the WPFs using the real-time telemetered data points. There exist sample points at which the wind power data includes some errors. Hence, the measurement errors can reduce the efficiency of the ARMA models.

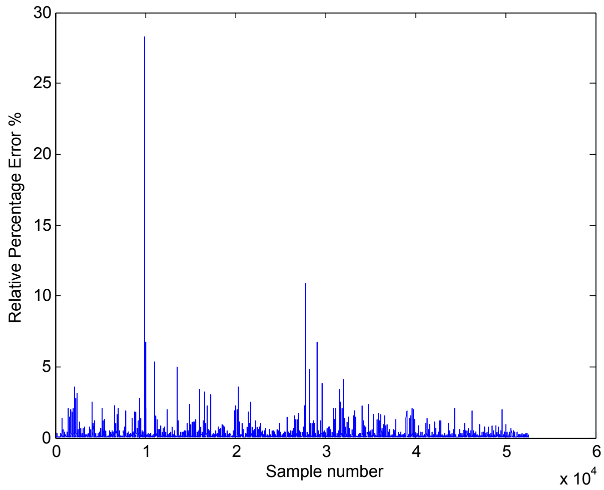
In order to cope with such errors, traditional grey models can be developed and applied to the wind power time series. The nature of wind power is fluctuating and intermittent. Therefore, it is considered as a grey system and the grey model can be used to provide wind power predictions.

3.7 Traditional Grey Model: GM(1,1)

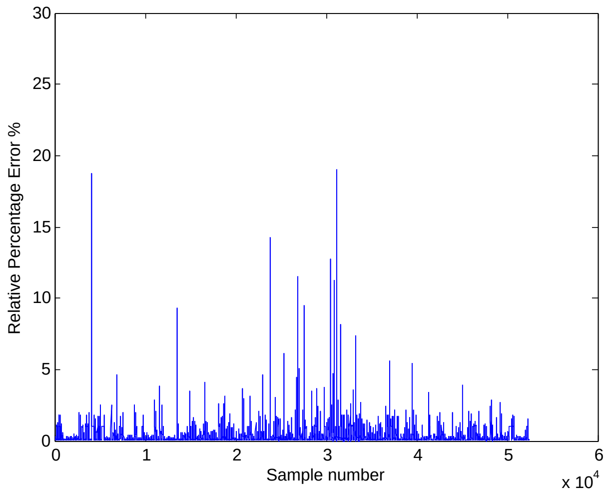
In order to overcome the measurement error effect on the forecast precision, the classical grey forecasting model is applied to the yearly and monthly wind power data. The obtained grey prediction errors are plotted in Fig.3.13 for yearly wind data. In Fig. 3.14, the grey predicted monthly data versus the original measurements are presented.

3.8 Hybrid GM(1,1)-ARMA

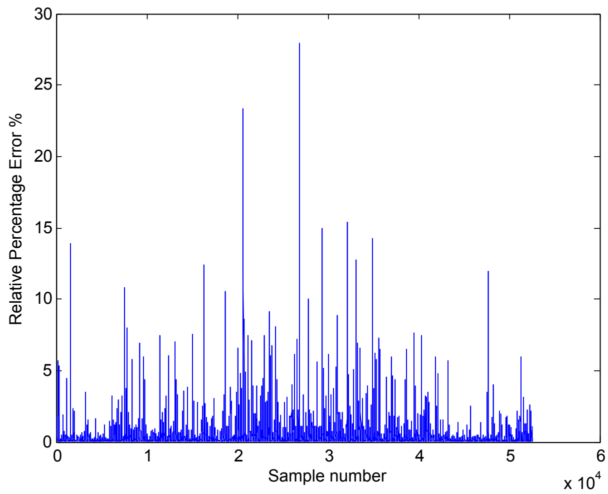
In order to benefit from different aspects of the wind power information, GM(1,1) is combined with the ARMA model. This forecasting algorithm is applied to both sets of wind power time series: yearly and monthly wind power data. The GM(1,1)-ARMA prediction errors for yearly data are illustrated in Fig. 3.15. The obtained predicted values for monthly data versus the original data are shown in Fig. 3.16.



(a) Year 2009.

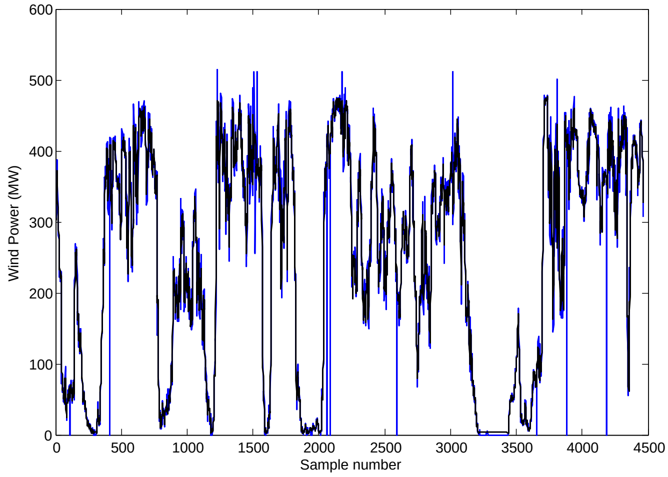


(b) Year 2010.

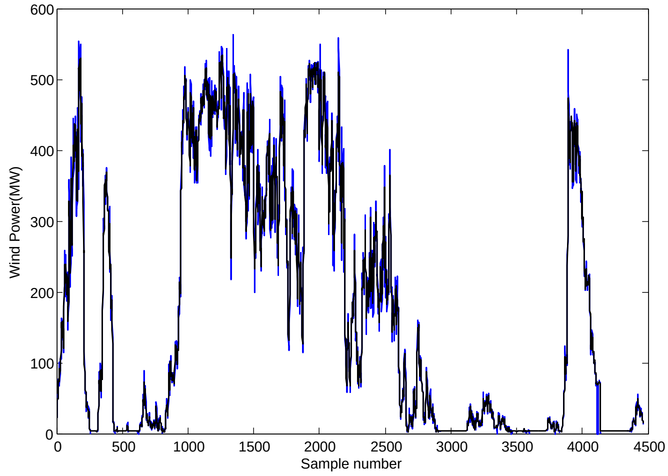


(c) Year 2011.

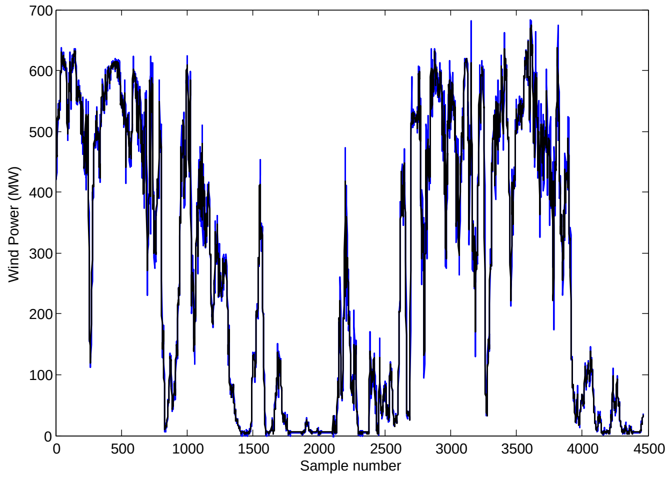
Figure 3.13: Wind power GM(1,1) prediction RPE for yearly measurements.



(a) Jan.2009.

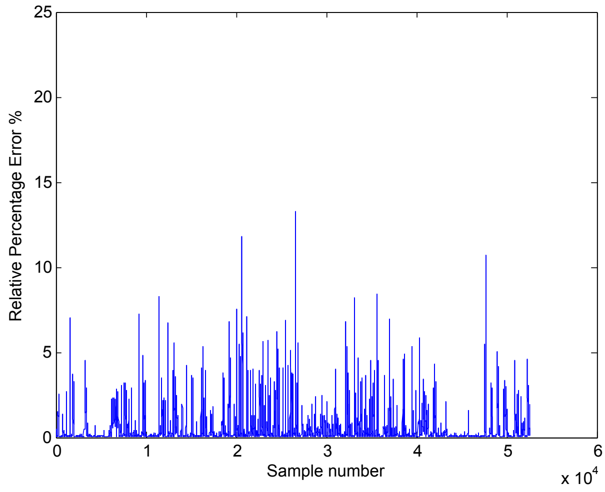


(b) Jan.2010.

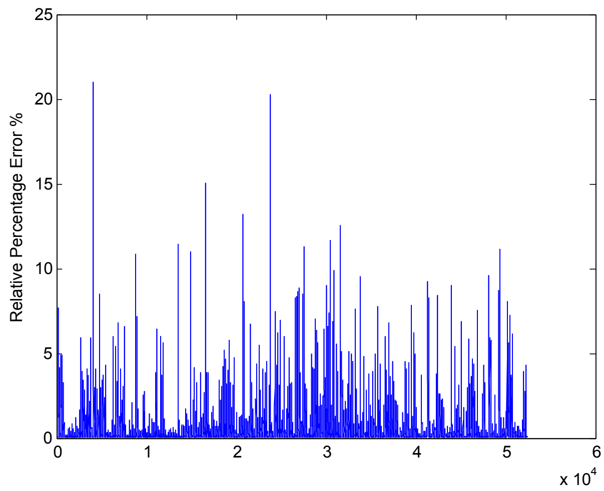


(c) Jan.2011.

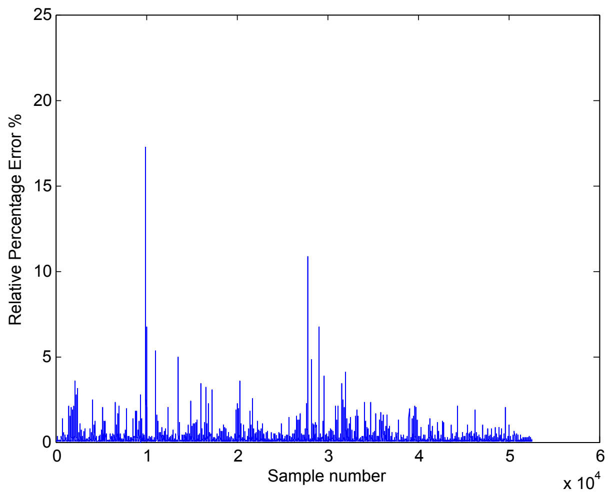
Figure 3.14: Wind power monthly measurement versus GM(1,1) predicted values.



(a) Year 2009.

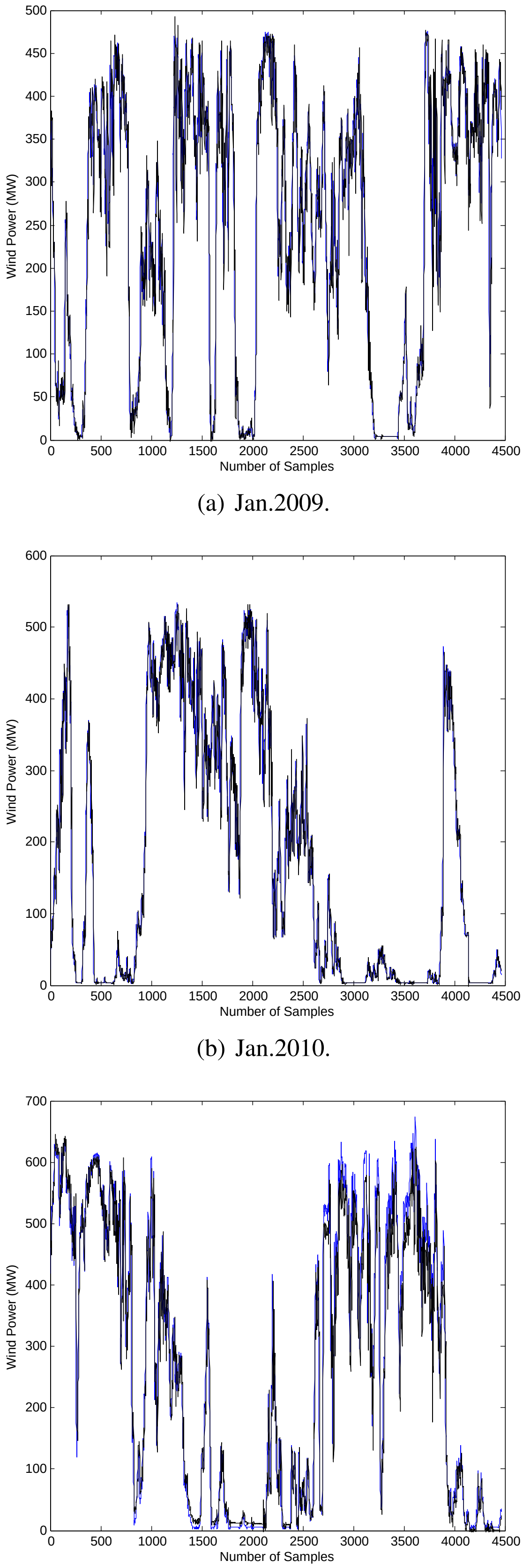


(b) Year 2010.



(c) Year 2011.

Figure 3.15: Wind power GM(1,1)-ARMA prediction RPE for yearly measurements



(a) Jan.2009.

(b) Jan.2010.

(c) Jan.2011.

Figure 3.16: Wind power monthly measurement versus GM(1,1)-ARMA predicted values.

3.9 Nonlinear Analysis of Wind Power Time Series

In this section, in order to assess the significance of the nonlinear components in the wind power time series, the nonlinear analysis is applied to the wind power data using the surrogate data approach.

This analysis includes two stages. First, for each wind power time series, the highest ARMA order obtained in Section 2.3.1 is chosen. Such high orders of ARMA models are capable of capturing the entire linear deterministic components of the time series. Thus, the residuals of the wind power are free of linear components and include only nondeterministic and nonlinear deterministic components.

Next, the surrogate data approach is applied to the wind power residuals. In this study, the values of the significance level α and the number of surrogate time series M are chosen 0.1 and 19, respectively. The IAAFT algorithm is utilized to generate the surrogate data and three nonlinear measures explained in Section 2.7.1 are used to investigate the presence of nonlinear components: third order auto covariance, time reversibility, and standard deviation.

This technique is applied to the monthly wind power time series of 2011, 2010, and 2009. The obtained results are summarized in Tables 3.9, 3.10, and 3.11, respectively. In these tables, the presence of nonlinear and linear components are depicted by N and L, respectively. It is demonstrated that for a specific data set, different nonlinear measures may lead to different results. Therefore, it is essential to consider all three tests in nonlinear or linear decision making. It is shown that almost all of the wind power time series include nonlinear components. Thus, it is essential to employ the nonlinear predictors to benefit from nonlinear components during the prediction procedure and enhance the accuracy of the prediction.

3.10 Hybrid GM(1,1)-NARnet

As discussed previously, the ARMA models are linear models in which only the linear components of the time series are utilized in the prediction procedure. Therefore,

Table 3.9: Nonlinear analysis of monthly wind power data of 2011

Year 2011	Jan.	Feb.	Mar.	Apr.	May	Jun.	Jul.	Aug.	Sep.	Oct.	Nov.	Dec.
C3	N	N	L	L	L	L	L	L	N	N	N	N
REV	N	N	L	L	N	L	N	N	L	L	N	L
SD	L	N	L	L	L	L	L	L	L	N	L	L

Table 3.10: Nonlinear analysis of monthly wind power data of 2010

Year 2010	Jan.	Feb.	Mar.	Apr.	May	Jun.	Jul.	Aug.	Sep.	Oct.	Nov.	Dec.
C3	L	L	N	L	L	L	L	L	L	N	N	N
REV	L	N	N	L	N	L	L	N	N	L	L	N
SD	L	L	N	L	L	L	L	L	L	L	L	N

Table 3.11: Nonlinear analysis of monthly wind power data of 2009

Year 2009	Jan.	Feb.	Mar.	Apr.	May	Jun.	Jul.	Aug.	Sep.	Oct.	Nov.	Dec.
C3	N	L	N	L	L	L	L	L	L	N	N	L
REV	N	L	N	N	N	N	N	N	N	L	L	N
SD	L	L	L	N	L	L	L	N	N	L	L	L

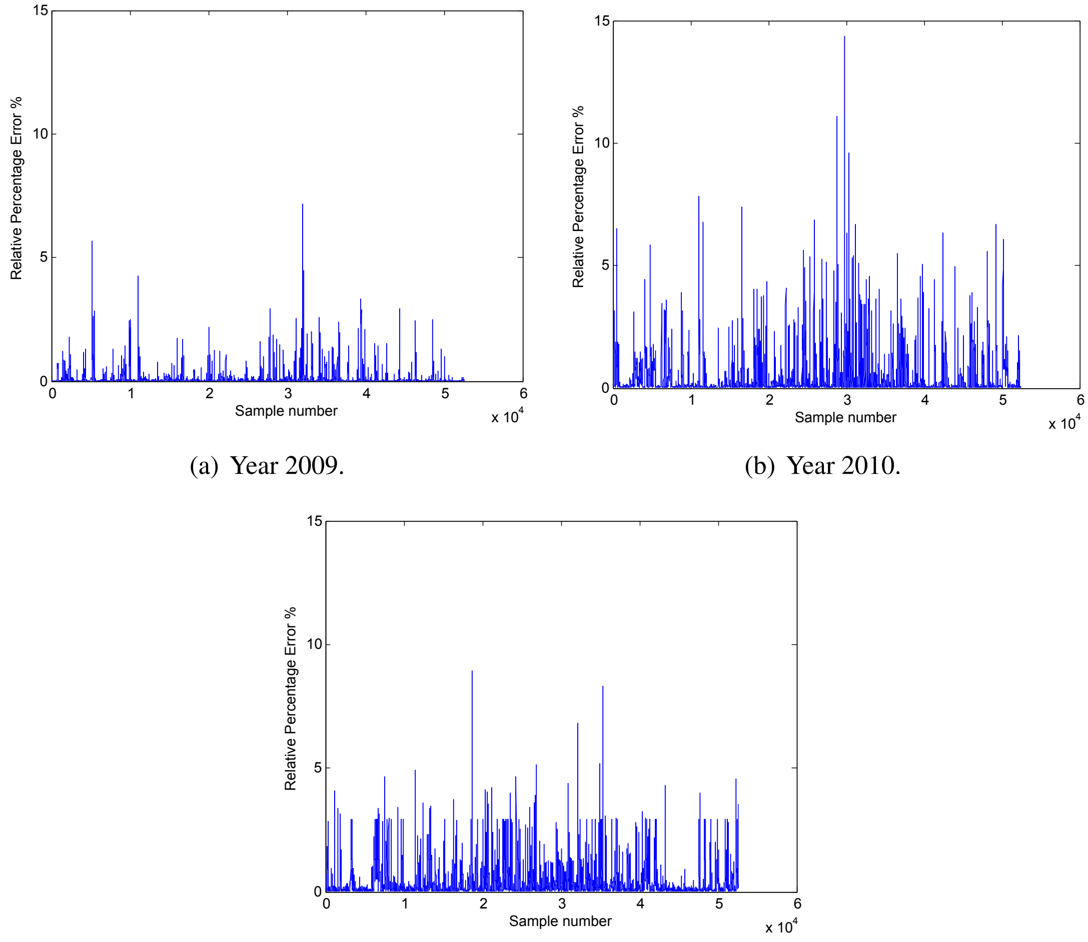
the nonlinear components are neglected and left in the residuals. Since wind power has strong nonlinear components, utilizing nonlinear techniques results in a more accurate forecast. Hence, the GM(1,1)-NARnet algorithm is applied to the processed and unprocessed wind power time series. This algorithm benefits from both the grey and NARnet models. The architecture of the utilized network has four layers: an input layer, an output layer and two hidden layers. The Levenberg-Marquardt back propagation approach is used to train the network. In order to establish the GM(1,1)-NARnet model the following steps are completed:

1. A proper length of the data is chosen as the training data set.
2. The predicted AGO values of the wind power are calculated by establishing the GM(1,1).
3. The IAGO is applied to the AGO predicted values to find the original predicted values.

4. A NARnet predictor including four layers is constructed and trained to find the predicted values of the grey model residuals.

5. The grey model predicted values are modified using the values of the NARnet model.

The GM(1,1)-NARnet obtained prediction error for yearly data is illustrated in Fig. 3.17. These values are much smaller compared with those obtained from the traditional GM(1,1) and GM(1,1)-ARMA (Fig. 3.13 and 3.15). The GM(1,1)-NARnet predicted values for monthly data versus the measurements are represented in Fig. 3.18.

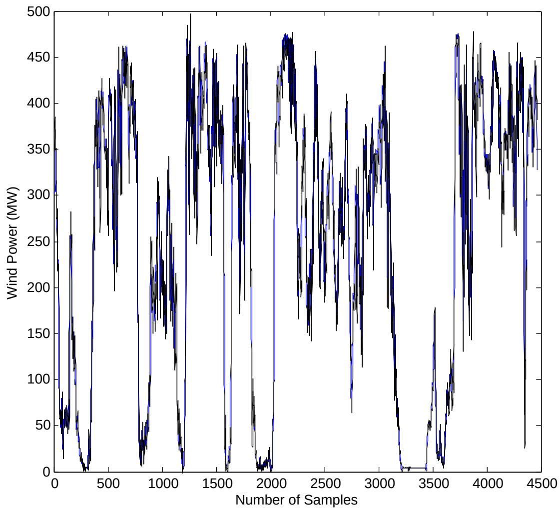


(a) Year 2009.

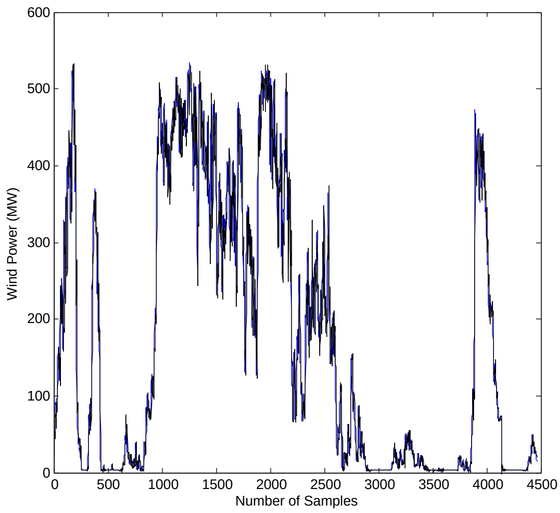
(b) Year 2010.

(c) Year 2011.

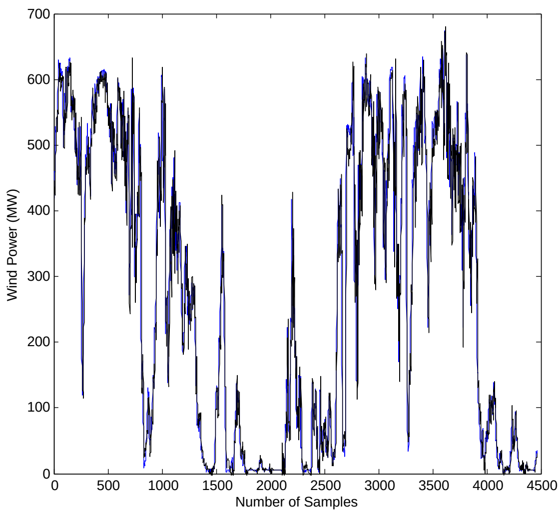
Figure 3.17: Wind power GM(1,1)-NARnet prediction RPE for yearly measurements for yearly measurements.



(a) Jan.2009.



(b) Jan.2010.



(c) Jan.2011.

Figure 3.18: Wind power monthly measurement versus GM(1,1)-NARnet predicted values.

3.11 Comparison of Results

In this section, we aim to provide hourly wind power prediction. The available wind power time series are recorded at 10 minute intervals. The hourly wind power prediction is obtained by applying the following averaging process to the 10 minutes predicted data.

$$P_{ave}(j) = \frac{1}{6}\sum_{i=1}^{6} P_m(i), \quad j = 1, 2, ..., n. \tag{3.5}$$

where $P_{ave}(j)$ denotes the hourly average predicted value of the wind power during the hour j, and $P_m(i)$ represents the ten minute measured data during the hour j.

The prediction algorithms described in Chapter 2 are applied to the monthly data of 2011 to obtain the predicted values of the wind power for five hours ahead. Table 3.12 provides a comparison between the persistence and grey prediction algorithms: GM(1,1), GM(1,1)-ARMA. It is illustrated that both GM(1,1) and GM(1,1)-ARMA outperform the persistence method in terms of prediction accuracy. Moreover, it is seen that the GM(1,1)-ARMA is more accurate compared with the traditional GM(1,1). Such accuracy stems from utilizing both the GM(1,1) and ARMA.

Table 3.12: Hourly wind power prediction RPE for Jan2011.

Time (hr)	1	2	3	4	5
GM(1,1)	0.7539	0.9538	0.1667	0.1354	0.5143
GM(1,1)-ARMA	0.4225	0.0504	0.2865	0.7992	0.1829
Persistence	0	5.2913	9.4806	3.8371	1.1482

In Fig. 3.19 a visual comparison is provided for the predicted values using GM(1,1), GM(1,1)-ARMA, and GM(1,1)-NARnet.

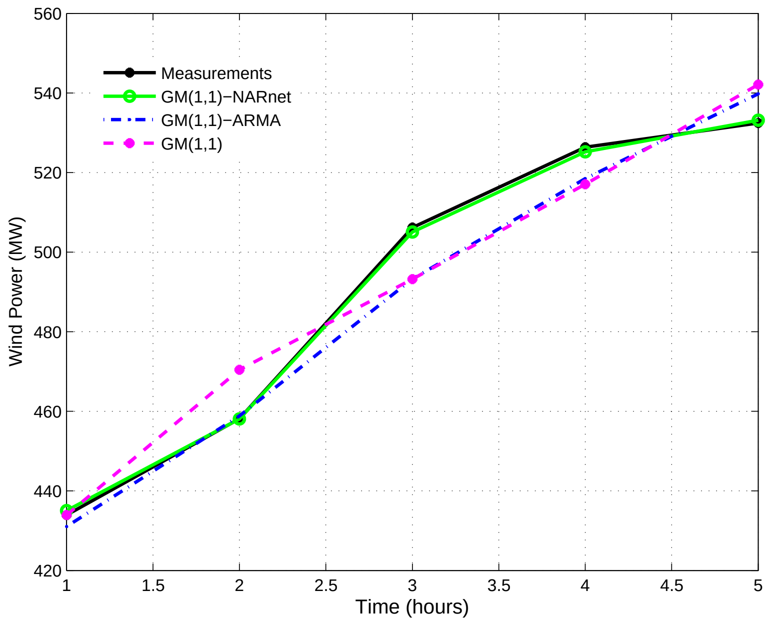


Figure 3.19: One hour ahead wind power prediction using GM(1,1), GM(1,1)-ARMA and GM(1,1)-NARnet.

Chapter 4: Conclusions and Future Work

In summary, the integration of wind power introduces several problems to electrical power systems, such as power system stability and reliability issues, economic dispatch issues, and elevation of of operation costs. Such problems stem from the intermittent nature of wind power. In order to combat the uncertain characteristics of wind, precise wind power prediction is required. An accurate prediction has significant effects on power system operations. It enables power system operators to efficiently schedule the power generation so that it can meet customer demands. In addition, precise wind power forecasts are important for electricity transmission and energy traders.

In the present thesis, we aim to provide accurate short-term wind power prediction for dispatch planning using the wind power time series published by the Alberta Electrical System Operator. In the data pre-analysis stage, the wind power time series is analyzed and transformed to a stationary time series. In the next stage, two adequacy checking tests are utilized to determine the best fitting ARMA models. These tests include checking the whiteness of the residuals and the Akaike Information Criteria (AIC). The first test is comprised of four scenarios that are applied to the wind power time series along with the AIC. The obtained ARMA orders are used to predict the wind power time series. It is demonstrated that the prediction RMSE values obtained from using the checking of the whiteness of the residuals are of smaller values compared with those obtained from the AIC.

The performance of the ARMA models is affected by the measurement errors as well as nonlinear components included in the original wind power time series. Hence, other prediction algorithms capable of dealing with such errors should be used as alternatives. In this thesis, the traditional grey model is utilized for wind power prediction. The wind power time series includes random variations. Therefore, it is difficult for the individual predictors to completely describe the data. In this thesis, a hybrid GM(1,1)-ARMA model is proposed and developed for one hour head wind power prediction.

The obtained results confirm that the GM(1,1)-ARMA beats the persistence method and traditional grey model in terms of prediction accuracy.

A nonlinear analysis is applied to the wind power time series to investigate the persistence of the nonlinear components. This analysis is done using the Surrogate data approach. It is found that almost all of the wind power data sets include nonlinear components. Hence, nonlinear prediction techniques can be utilized to benefit from the nonlinear components of the time series during the prediction procedure.

GM(1,1)-NARnet is proposed to enhance the prediction accuracy. The obtained results confirm the out-performance of GM(1,1)-NARnet over the traditional grey model, GM(1,1)-ARMA, and the persistence model.

Future work can include employing the proposed prediction methodologies for load forecast and using the predicted values of wind power and load forecast in economic dispatch problems. In order to enhance the prediction error, the grey model residuals can be modified using Fourier series of the residuals. Utilizing the grey Verhulst models can result in more accurate prediction due to the fact that such models are capable of uniforming the model parameters between the differential equations and the difference equation.

References

[1] M. Lydia and S. Kumar, "A comprehensive overview on wind power forecasting," in *IPEC, 2010 Conference Proceedings*, pp. 268–273, 2010.

[2] S. Soman, H. Zareipour, O. Malik, and P. Mandal, "A review of wind power and wind speed forecasting methods with different time horizons," in *North American Power Symposium (NAPS), 2010*, pp. 1–8, 2010.

[3] L. Landberg, "A mathematical look at a physical power prediction model," in *Wind Energy*, pp. 23–28, 1998.

[4] C. Monteiro, H. Keko, R. Bessa, V. Miranda, A. Botterud, J. Wang, and G. Conzelmann, "Wind power forecasting: state-of-the-art 2009," in *Decision and Information Sciences Division, Argonne National Laboratory*, pp. 23–28, 2009.

[5] A. Sfetsos, "A comparison of various forecasting techniques applied to mean hourly wind speed time series," *Renewable Energy*, vol. 21, no. 1, pp. 23–35, 2000.

[6] P. Chang and L. Li, "Ocean surface wind speed and direction retrievals from the SSM/I," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 36, no. 6, pp. 1866–1871, 1998.

[7] M. McCarthy, "Wind speed forecasting in the central california wind resource area," *EPRI-DOE-NREL Wind Energy Forecasting Meeting*, 1998.

[8] M. Negnevitsky and C. Potter, "Innovative short-term wind generation prediction techniques," in *Power Engineering Society General Meeting, 2006. IEEE*, 2006.

[9] L. Landberg, "Short-term prediction of the power production from wind farms," *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 80, no. 12, pp. 207–220, 1999.

[10] M. Negnevitsky, P. Johnson, and S. Santoso, "Short term wind power forecasting using hybrid intelligent systems," in *Power Engineering Society General Meeting, 2007. IEEE*, pp. 1–4, 2007.

[11] L. Landberg and S. Watson, "Short-term prediction of local wind conditions," *Boundary-Layer Meteorology*, vol. 70, no. 1-2, pp. 171–195, 1994.

[12] R. Schlueter, G. Park, R. Bouwmeester, L. Shu, M. Lotfalian, P. Rastgoufard, and A. Shayanfar, "Simulation and assessment of wind array power variations based on simultaneous wind speed measurements," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-103, no. 5, pp. 1008–1016, 1984.

[13] M. Alexiadis, P. Dokopoulos, H. Sahsamanoglou, and I. Manousaridis, "Short-term forecasting of wind speed and related electrical power," *Solar Energy*, vol. 63, no. 1, pp. 61–68, 1998.

[14] U. Focken, M. Lange, K. Mnnich, H.-P. Waldl, H. G. Beyer, and A. Luig, "Short-term prediction of the aggregated power output of wind farmsa statistical analysis of the reduction of the prediction error by spatial smoothing effects," *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 90, no. 3, pp. 231–246, 2002.

[15] T. P. T. Chen, G., *Introduction to Fuzzy Systems*. CRC Press, 1 ed., 1991.

[16] "Introduction to fuzzy systems," in *Electronic Technology Directions to the Year 2000, 1995. Proceedings.*, pp. 94–103, May 1995.

[17] Y.-J. Horng, S.-M. Chen, Y.-C. Chang, and C.-H. Lee, "A new method for fuzzy information retrieval based on fuzzy hierarchical clustering and fuzzy inference techniques," *IEEE Transactions on Fuzzy Systems*, vol. 13, no. 2, pp. 216–228, 2005.

[18] Cui-Yan, J.-Y. Tao, and W.-J. Tao, "Pattern recognition based on fuzzy three-valued logic reasoning," in *Proceedings of the International Conference on Information Engineering and Computer Science (ICIECS)*, pp. 1–4, 2009.

[19] M. Sheikh-El-Eslami and H. Seifi, "Short-term electricity price forecasting using a fuzzy stochastic predictor," in *Power Engineering Society General Meeting, 2006. IEEE*, p. 5, 2006.

[20] N. Amjady, "Day-ahead price forecasting of electricity markets by a new fuzzy neural network," *IEEE Transactions on Power Systems*, vol. 21, no. 2, pp. 887–896, 2006.

[21] I. Damousis, M. Alexiadis, J. Theocharis, and P. Dokopoulos, "A fuzzy model for wind speed prediction and power generation in wind parks using spatial correlation," *IEEE Transactions on Energy Conversion*, vol. 19, no. 2, pp. 352–361, 2004.

[22] G. Zhang, H.-X. Li, and M. Gan, "Design a wind speed prediction model using probabilistic fuzzy system," *IEEE Transactions on Industrial Informatics*, vol. 8, no. 4, pp. 819–827, 2012.

[23] A. R. G. H. Sidney Burrus, C., *Introduction to Wavelets and Wavelet Transforms: A Primer*. Prentice-Hall, 1997.

[24] D.-Z. Tian and M.-H. Ha, "Applications of wavelet transform in medical image processing," in *Proceedings of the International Conference on Machine Learning and Cybernetics*, vol. 3, pp. 1816–1821 vol.3, 2004.

[25] S. P. Raja and A. Suruliandi, “Analysis of efficient wavelet based image compression techniques,” in *Proceedings of the International Conference on Computing Communication and Networking Technologies (ICCCNT)*, pp. 1–6, 2010.

[26] C. Akujuobi, N. K. Ampah, and M. Sadiku, “Application of wavelets and self-similarity to enterprise network intrusion detection and prevention systems,” in *Proceedings of IEEE International Symposium on Consumer Electronics (ISCE)*, pp. 1–6, 2007.

[27] G. Hloupis and F. Vallianatos, “Wavelet-based rapid estimation of earthquake magnitude oriented to early warning,” *Geoscience and Remote Sensing Letters, IEEE*, vol. 10, no. 1, pp. 43–47, 2013.

[28] S. Mishra, A. Sharma, and G. Panda, “Wind power forecasting model using complex wavelet theory,” in *Proceedings of the International Conference on Energy, Automation, and Signal (ICEAS)*, pp. 1–4, Dec 2011.

[29] C. Lei and L. Ran, “Short-term wind speed forecasting model for wind farm based on wavelet decomposition,” in *Proceedings of the 3rd International Conference on Electric Utility Deregulation and Restructuring and Power Technologies (DRPT)*, pp. 2525–2529, 2008.

[30] Y. Chuanan and Y. Yongchang, “A hybrid model to forecast wind speed based on wavelet and neural network,” in *Proceedings of the International Conference on Control, Automation and Systems Engineering (CASE)*, pp. 1–4, 2011.

[31] J. Xia, P. Zhao, and Y. Dai, “Neuro-fuzzy networks for short-term wind power forecasting,” in *Proceedings of the International Conference on Power System Technology (POWERCON)*, pp. 1–5, 2010.

[32] J. Wang, J. Liang, J. Che, and D. Sun, “ARMA model identification using particle swarm optimization algorithm,” in *Proceedings of the International Conference on Computer Science and Information Technology (ICCSIT)*, pp. 223–227, 2008.

[33] L. Kamal and Y. Z. Jafri, “Time series models to simulate and forecast hourly averaged wind speed in Quetta, Pakistan,” *Solar Energy*, vol. 61, no. 1, pp. 23–32, 1997.

[34] J. Torres, A. Garca, M. D. Blas, and A. D. Francisco, “Forecast of hourly average wind speed with ARMA models in Navarre (Spain),” *Solar Energy*, vol. 79, no. 1, pp. 65–77, 2005.

[35] Z. Huang and Z. Chalabi, “Use of time-series analysis to model and forecast wind speed,” *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 56, no. 23, pp. 311–322, 1995.

[36] A. Kusiak, H. Zheng, and Z. Song, "Wind farm power prediction: a data-mining approach," *Wind Energy*, vol. 12, no. 3, pp. 275–293, 2009.

[37] S. Rajagopalan and S. Santoso, "Wind power forecasting and error analysis using the autoregressive moving average modeling," in *Power Energy Society General Meeting, 2009. PES '09. IEEE*, pp. 1–6, 2009.

[38] M. Miranda and R. Dunn, "One-hour-ahead wind speed prediction using a bayesian methodology," in *Power Engineering Society General Meeting, 2006. IEEE*, p. 6, 2006.

[39] S. M. Milligan, M. and Y. Wan, "Statistical wind power forecasting models: Results for USA wind farms," in *Windpower*, p. 6, 2003.

[40] A. Rodriguez-Garcia and E. De-la Torre-Vega, "A statistical wind power forecasting system: A Mexican wind farm case study," in *European Wind Energy Conference and Exhibition*, 2009.

[41] E. Panteri and S. Papathanassiou, "A statistical wind power forecasting system: A Mexican wind farm case study," in *the European Wind Energy Conf., National Technical University of Athens (NTUA), School of Electrical and Computer*, 2008.

[42] C.-Y. Huang, Y.-W. Liu, W.-C. Tzeng, and P.-Y. Wang, "Short term wind speed predictions by using the grey prediction model based forecast method," in *Green Technologies Conference (IEEE-Green), 2011 IEEE*, pp. 1–5, 2011.

[43] J. Zeng and W. Qiao, "Support vector machine-based short-term wind power forecasting," in *Power Systems Conference and Exposition (PSCE), 2011 IEEE/PES*, pp. 1–8, 2011.

[44] H. Mori and E. Kurata, "Application of gaussian process to wind speed forecasting for wind power generation," in *Proceedings of IEEE International Conference on Sustainable Energy Technologies (ICSET)*, pp. 956–959, 2008.

[45] G. Li and J. Shi, "On comparing three artificial neural networks for wind speed forecasting," *Applied Energy*, vol. 87, no. 7, pp. 2313–2320, 2010.

[46] F. O. Hocaolu, M. Fidan, and mer N. Gerek, "Mycielski approach for wind speed prediction," *Energy Conversion and Management*, vol. 50, no. 6, pp. 1436–1443, 2009.

[47] M. Monfared, H. Rastegar, and H. M. Kojabadi, "A new strategy for wind speed forecasting using artificial intelligent methods," *Renewable Energy*, vol. 34, no. 3, pp. 845–848, 2009.

[48] B. Bowerman and R. Oconnel, *Forecasting and time series*. Duxbury press, 1993.

[49] F. Beichelt and L. Paul Fatti, *Stochastic Processes and Their Applications*. CRC Press, 2001.

[50] H. Abarbanel, *Analysis of Observed Chaotic Data*. Springer, 1996.

[51] R.-H. Lindgren, G. and M. Sandsten, *Stationary Stochastic Processes for Scientists and Engineers*. CRC press, 2013.

[52] W. W.S.Wei, *Time series analysis*. Pearson, 1990.

[53] H. Samet and M. E. H. Golshan, "Employing stochastic models for prediction of arc furnace reactive power to improve compensator performance," *Generation, Transmission Distribution, IET*, vol. 2, no. 4, pp. 505–515, 2008.

[54] H. Lutkepohl, *Introduction to multiple time series analysis*. Springer-Verlag Pub, 1991.

[55] G. Ljung and G. Box, "On a measure of lack of fit in time series models," *Biometrika*, vol. 65, no. 2, pp. 297–303, 1978.

[56] H. Akaike, "Information theory and an extension of the maximum likelihood principle," *Proceedings of the International Symposium in information Theory*, vol. 2, no. 4, pp. 267–281, 1973.

[57] D. Julong, "Control problems of grey systems," *Systems and Control letters*, vol. 5, pp. 288–294, 1982.

[58] K. Wen, *Grey systems: modeling and prediction*. Yangs Scientific, 2004.

[59] L. Y. R.Wang and Y. Li, "A hybrid forecasting method for day-ahead electricity price based on GM(1,1) and ARMA," in *Proceedings of the IEEE International Conference on GSIS*, pp. 557–581, 2009.

[60] M. Kendall, *Time-Series (Second ed.)*. Charles Griffin, 1976.

[61] H. Wold, *A Study in the Analysis of Stationary Time Series*. Uppsala, 1938.

[62] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, and J. D. Farmer, "Testing for nonlinearity in time series: The method of surrogate data," *Physica*, vol. 58, no. 1-4, pp. 77–94, 1992.

[63] T. Han and D. Goodenough, "Investigation of nonlinearity in hyperspectral imagery using surrogate data methods," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 10, pp. 2840–2847, 2008.

[64] M. palus, "Testing for nonlinearity using redundancies: Quantitative and qualitative aspects," *Physica D*, vol. 80, no. 1, pp. 186–205, 1995.

[65] T. Schreiber and A. Schmitz, "Improved surrogate data for nonlinearity tests," *Phys. Rev. Lett*, vol. 77, no. 4, pp. 635–638, 1995.

[66] T. Schreiber and A. Schmitz, "Discrimination power of measures for nonlinearity in a time series," *Phys. Rev. E*, vol. 55, pp. 5443–5447, 1997.

[67] R. Andrzejak, K. Lehnertz, C. Rieke, F. Mormann, P. David, and C. Elger, "Indications of nonlinear deterministic and finite dimensional structures in time series of brain electrical activity: dependence on recording region and brain state," *Phys. Rev. E,*, vol. 64, pp. 5443–5447, 2001.

[68] "Introduction to artificial neural networks," in *Proceedings of the Electronic Technology Directions to the Year 2000*, pp. 36–62, 1995.

[69] D. Wang, S. Yin, C. Chen, J. Wu, and G. Wang, "Application of GM(1,1) model on predicating the outpatient amount," in *Advances in Multimedia, Software Engineering and Computing Vol.2* (D. Jin and S. Lin, eds.), vol. 129 of *Advances in Intelligent and Soft Computing*, pp. 71–76, Springer Berlin Heidelberg, 2012.

[70] A. Kusiak, H. Zheng, and Z. Song, "Power optimization of wind turbines with data mining and evolutionary computation," *Renewable Energy*, vol. 35, no. 3, pp. 695–702, 2010.

[71] "Alberta Electric System Operator, Alberta,Canada." , http://www.aeso.ca.

[72] B. Ernst, B. Oakleaf, M. Ahlstrom, M. Lange, C. Moehrlen, B. Lange, U. Focken, and K. Rohrig, "Predicting the wind," *Power and Energy Magazine, IEEE*, vol. 5, no. 6, pp. 78–89, 2007.

Vita

Fatemeh Marzbani was born in Iran in 1989. She acquired her high school and pre-university education in Iran and at the National Organization for Development of Exceptional Talents (NODET) where she was ranked among the top 5% pre-university graduated students. In 2007, she participated in the very competitive Iranian national universities entrance exam. She ranked in the top 0.5 % among 400,000 participants. She then enrolled in the school of Electrical Engineering and Computer Science at Shiraz University. In 2011, she graduated with the degree of Bachelor of Science (Honours) in Electrical Engineering minor control. She then joined the Master of Science in Electrical Engineering (MSEE) program at the American University of Sharjah, UAE. She was awarded a graduate teaching and research assistantship during her studies. During her Master's, she published two research papers in two IEEE conferences: ICMSAO 2013 (Tunisia) and EPECS 2013 (Turkey). Power system control and analysis, renewable energies, wind power, and smart grids are her research areas of interest.