**ISTANBUL TECHNICAL UNIVERSITY ★ ENERGY INSTITUTE**

# WIND SPEED PREDICTION USING LINEAR PREDICTION METHODS

**M.Sc. THESIS**

**Zafer CANAL**

**Energy Science and Technology Division**

**Energy Science and Technology Program**

**MAY 2015**

**ISTANBUL TECHNICAL UNIVERSITY ★ ENERGY INSTITUTE**

# WIND SPEED PREDICTION USING LINEAR PREDICTION METHODS

**M.Sc. THESIS**

**Zafer CANAL**

**Energy Science and Technology Division**

**Energy Science and Technology Program**

**Thesis Advisor: Assist. Prof. Dr. Burak BARUTÇU**

**MAY 2015**

**<u>İSTANBUL TEKNİK ÜNİVERSİTESİ ★ ENERJİ ENSTİTÜSÜ</u>**

# LİNEER ÖNGÖRÜ METODLARI İLE RÜZGAR HIZI ÖNGÖRÜSÜ

**YÜKSEK LİSANS TEZİ**

**Zafer CANAL**
**(301111028)**

**Enerji Bilim ve Teknoloji Anabilim Dalı**

**Enerji Bilim ve Teknoloji Programı**

**Tez Danışmanı: Yar. Doç. Dr. Burak BARUTÇU**

**MAYIS 2015**

**Zafer Canal** is a **M.Sc.** student of **ITU Energy Institute** student ID 301111028, successfully defended the **thesis** entitled "**WIND SPEED PREDICTION USING LINEAR PREDICTION METHODS** ", which he prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

**Thesis Advisor :** **Assist. Prof. Dr. Burak BARUTÇU** ..............................
İstanbul Technical University

**Jury Members :** **Prof. Dr. Sibel MENTEŞ** .............................
İstanbul Technical University

**Assist. Prof. Dr. Deniz YILDIRIM** ..............................
İstanbul Technical University

**Date of Submission : 04 May 2015**
**Date of Defense : 26 May 2015**

***To my family,***

**FOREWORD**

Initially, I would like to thank my academic advisor Assist. Prof. Dr. Burak BARUTÇU. This thesis would not have been possible without his inspiration and effort. I appreciate for his advice, encouragement and support throughout the sudy and writing this thesis.

I am deeply thankful to my family for their love, support, and endless trust during my life. Finally, I want to thank my wife Cansu DENİZ for her help, support, patience, encouragement and invaluable advices at hard times.

May 2015 Zafer CANAL
Mechanical Engineer

**TABLE OF CONTENTS**

**Page**

## ABBREVIATIONS

**ACF** : Autocorrelation Function
**AIC** : Akaike Informatıon Criteria
**AR** : Autoregressive
**ARIMA** : Autoregressive Integrated Moving Average
**ARIMAX** : Autoregressive Integrated Moving Average with exogenous input
**ARMA** : Autoregressive Moving Average
**ARMAX** : Autoregressive Moving Average with exogenous input
**ARX** : Autoregressive with exogenous input
**BIC** : Bayesian Informatıon Criteria
**f-ARIMA** : Fractional Autoregressive Integrated Moving Average
**FIR** : Finite Impulse Response
**FPE** : Final Prediction Eror
**IIR** : Infinite Impulse Response
**MA** : Moving Average
**MSE** : Mean Squared Error
**NN** : Neural Network
**PACF** : Partial Autocorrelation Function
**s-ARIMA** : Seasonal Autoregressive Integrated Moving Average
**UC** : Unit Commitment
**WPF** : Wind Power Forecasting
**WSS** : Wide Sense Stationary

## LIST OF TABLES

**Page**

## LIST OF FIGURES

**Page**

**WIND SPEED PREDICTION USING LINEAR PREDICTION METHODS**


**SUMMARY**

Short-term forecasting of wind speed is of great importance to wind turbine operation and efficient energy harvesting. In this thesis, one-step ahead wind speed forecasting is performed. Six approaches based on linear prediction methods are employed for this purpose. The first approach features the autoregressive process (AR) with the model order eight. Model order selection criterias, Akaike information criteria (AIC) and Bayesian information criteria (BIC), are used for optimal model order selection. These information criterias selected the same autoregressive model with an order of eight, which is shown as AR(8). Second approach employs the autoregressive moving average process (ARMA). In this case, AIC and BIC selected the autoregressive moving average model with different order. First model is defined as autoregressive moving average model with an autoregressive order of four and moving average order of three which can be shown as ARMA(4,3) and second model is defined as ARMA(14,13). Third approach features the autoregressive integrated moving average process (ARIMA). In this case, AIC and BIC pointed different model orders once again. In addition the notation of ARMA, an integration process with an order of one is added and shown as ARIMA (30,1,29) and ARIMA(3,1,2). Two different models are performed in this case. On the other hand fourth, fifth and sixth approaches involve employing an exogenous input to the first three approaches. In first case, autoregressive model with an exogenous input, which is denoted as ARX is featured. Depending on the model selection criterias, the order of autoregressive model with an exogenous input is selected as one, which is shown as ARX(1). In the next case, the criterias for model order selection pointed the same model order. Autoregressive order of two and moving average order of one with an exogenous input model, which denoted as ARMAX(2,1) is performed. In third case, AIC and BIC selected the first order integrated autoregressive order of eight and moving average order of seven with an exogenous input which is shown as ARIMAX(8,1,7). By employing these six approaches, one step ahead wind speed forecasting is performed. Wind speed data observed in Bursa-Gemlik location with a time interval of ten minutes. The results are compared using mean absolute error (MAE) and root mean square error (RMSE) as a measure for forecasting quality. The goodness of fit is checked by calculating r-square $R^2$ statistics. It is found that the AR, ARX, ARMA and ARIMA model is better at predicting the wind speed corresponding the $R^2$ statistics. MAE, RMSE and $R^2$ statistics also show that ARX model is the best for forecasting one step ahead wind speed. Moreover, ARMAX model is also good at forecasting wind speed whereas it's lower than AR, ARX, ARMA and ARIMA . Results also show that ARIMAX (8,1,7) model is the worse for forecasting one step ahead wind speed. In order to check the success of criterias for model order selection, various ARIMAX models are analyzed. It can be seen from the results that other ARIMAX models are better than ARIMAX(8,1,7). In other words AIC and BIC is not withstanding selecting the model order of autoregressive integrated moving average models with an exogenous input.

# LİNEER ÖNGÖRÜ METODLARI İLE RÜZGAR HIZI ÖNGÖRÜSÜ

## ÖZET

Rüzgar hızının ve yönünün değişkenliği ve buna bağlı olarak rüzgar enerji santralinden elde edilen enerjinin değişkenliği ve kontrol edilememesi, yüksek miktarda rüzgar enerjisine dayalı elektriğin bağlı olduğu şebekelerde güç kalitesi, üretim/tüketim dengesi, bakım onarım planlaması ve güç sisteminin güvenilirliği açısından sorunlara yol açmaktadır. Türkiye'nin mevcut rüzgar potansiyeli ve 2023 yılı sonunda elde edilmesi hedeflenen 20 GW kurulu rüzgar gücü düşünüldüğünde, şebekeye entegre edilen rüzgar enerjisinde yaşanacak artışla birlikte yukarıda bahsedilen sorunların ilerleyen yıllarda öneminin artacağını söyleyebilriz. Yüksek doğruluğa sahip ve uygulaması kolay rüzgar hızı öngörü yöntemleri bu sorunları en aza indirmek için kullanılacak en etkin çözümdür. Bu amaçla bu tez kapsamında böyle bir çalışma gerçekleştirilmiştir.

Bu çalışmada gerçekleştirilecek rüzgar hızı öngörüsü için lineer öngörü metodları otoregresif modeli (AR), otoregresif hareketli ortalamalar modeli (ARMA), otoregresif bütünleşik hareketli ortalamalar modeli (ARIMA), dışsal değişkenli otoregresif modeli (ARX), dışsal değişkenli otoregresif hareketli ortalamalar modeli (ARMAX), dışsal değişkenli otoregresif bütünleşik hareketli ortalamalar modeli (ARIMAX) kullanılmıştır. Öngörü, bir değişkenin belirli varsayımlar altında gelecekte alabileceği değerlerin önceden yaklaşık olarak belirlenmesi olarak tanımlanır. Zaman serisi analizi ile öngörü, incelenen bir değişkenin şimdiki ve geçmiş dönemdeki gözlem değerlerini kullanarak ve birtakım varsayımlar altında öngörü değerlerinin hangi sınırlar arasında gerçekleşebileceğini ortaya koymak için yapılan uğraşlardır. Lineer öngörü metotları seriye en iyi uyan, en az parametre içeren doğrusal modeli belirleyerek öngörüde bulunur.

Öngörüde kullanılacak rüzgar hızı ölçüm verileri, meteoroloji genel müdürlüğü'nün Gemlik'te bulunan otomatik gözlem istasyonundan alınmıştır. Rüzgar hızı ölçüm değerleri 10 dakikalık aralıklarla, kap anemometre ile 10 m yükseklikte kaydedilmiş 4320 (1 aylık) veriden oluşmaktadır.

AR, ARMA, ARIMA, ARX, ARMAX ve ARIMAX modelleri sırasıyla rüzgar hızı ölçüm veri serisinin tamamına uygulandı. Model mertebeleri 1'den 75'e kadar değiştirilerek her bir model mertebesi veri serisine uygulanıp, optimal model mertebesi seçme için kullanılan Akaike enformasyon kriteri ve Bayesian enformasyon kriteri hesaplandı, bunların grafikleri çizdirildi Figure (3.4, 3.5, 3.6, 3.7, 3.8, 3.9). Table 3.1. de görüldüğü üzere en düşük enformasyon kriter değerinin yakalndığı mertebeler modeller için en uygun model mertebesi seçildi.

AR modeli için, AIC ve BIC model mertebesi olarak sekizi gösterdi. AR(8) modeline en büyük olabilirlik kestirimi yöntemi kullanılarak model katsayıları hesaplandı. Hesaplanan bu değerler kullanılarak bir-adım ileri öngörüde bulunuldu

ve gerçek ölçüm değerleri ile karşılaştırıldı. Öngörünün başarımını ölçmek için kalan değerler hesaplandı. Kalan değerlerden ortalama mutlak hata ve karesel ortalama hata hesaplaması yapıldı. Örtüşme düzeyinin başarısını ölçmek için ise $R^2$ istatistiği hesaplandı. Yapılan öngörüler sonunda elde edilen kalanların grafiği çizdirildi. Kalanların rastgele dağılıp dağılmadığını gözlemlemek için dağılım grafiği çizdirildi. Bunlara ek olarak öngörü edilen ve gerçek gözlem değerleri aynı grafikte gösterildi, öngörünün gerçek değerlerin değişimine nasıl tepki verdiğinin daha net görebilmek için 2000. veri ile 2100. veri arası ayrı olarak çizdirildi. yine Daha sonra bu işlemler sırasıyla ARMA, ARIMA, ARX, ARMAX ve ARIMAX modellerinde tekrarlandı. AIC ve BIC metodları ARMA ve ARIMA modelleme türleri için ikişer farklı mertebeyi işaret etti. ARMA ve ARIMA modellemeleri için ikişer farklı mertebede model belirlendi

ARMA modeli için AIC model mertebesi ARMA (14,13)'ü BIC ise model mertebesi ARMA(4,3)'ü işaret etti. Örtüşme düzeyinin başarısına baktığımızda $R^2$ istatistiği ARMA (14,13) de daha yüksek çıktığını ve ARMA (4,3)'e göre daha iyi bir örtüşme gösterdiğini görüyoruz. Hata terimlerine baktığımızda yine ARMA(14,13)'ün ARMA(4,3)'e göre daha düşük öngörü hatası verdiği gözlemlendi.

ARIMA modeli için AIC model mertebesi ARIMA (30,1,29)'u BIC ise model mertebesi ARIMA(3,1,2)'yi işaret etti. Örtüşme düzeyinin başarısına baktığımızda $R^2$ istatistiği ARIMA (30,1,29) de daha yüksek çıktığını ve ARIMA (3,1,2)'e göre daha iyi bir örtüşme gösterdiğini görüyoruz. Hata terimlerine baktığımızda yine ARIMA(30,1,29)'un ARIMA(3,1,2)'e göre daha düşük öngörü hatası verdiği gözlemlendi. Bu iki farklı modelleme türü için yapılan analiz sonuçlar incelendiğinde AIC ve BIC metodlarının farklı değer gösterdiği iki model mertebesinde AIC'in işaret ettiği modelin daha başarılı olduğu gözlenmektedir.

Eksojen girişli otoregresif modeller için bu işlemler tekrarlandığında: ARX modeli için AIC ve BIC model mertebesi ARX(1)'i ARMAX için model mertebesi ARMAX(2,1)'i ve ARIMAX için model mertebesi ARIMAX(8,1,7) 'yi işaret etmekte. Bunların içinde ARX(1) modelinin örtüşme düzeyi başarısı $R^2$ istatistiği değerine bakıldıında, diğer modellere kıyasla en yüksek değeri verdiğini görüyoruz. Hata terimlerine baktığımızda da diğer modellemelere göre en düşük hatayı yine ARX(1) modelinde yakaladığımızı görmekteyiz.

ARMAX(2,1) modelinin örtüşme düzeyinin başarısına baktığımızda $R^2$ istatistiği değerleri otoregresif modellere ve ARX(1) modeline göre biraz düşük olduğunu söylenebilir. Hata terimlerine baktığımızda ARMAX(2,1)'in AR, ARMA, ARIMA ve ARX modellerine göre daha yüksek öngörü hatası verdiği gözlemlendi.

ARIMAX modeli için ise AIC ve BIC, model mertebesi olarak ARIMAX(8,1,7)'i işaret etmiştir. ARIMAX(8,1,7)'nin örtüşme düzeyinin başarısına baktığımızda $R^2$ istatistiği değerinin diğer modellere göre çok düşük olduğu görülmektedir. . Hata terimlerine baktığımızda da en yüksek hatayı yine ARIMAX(8,1,7) modelinde olduğunu görmekteyiz. Bu sonuçlar ARIMAX(8,1,7) modelinin öngörüde bulunmak için uygun bir model olmadığını göstermektedir.

ARIMAX modelinde yakalanan bu başarısız sonuçlar üzerine ek bir çalışma olarak farklı model mertebelerine sahip ARIMAX modelleri veri serisinin öngörüsünde kullanıldı. Bu çalışmada ARIMAX(1,1,0), ARIMAX(2,1,1), ARIMAX(20,1,19), ARIMAX(9,1,8) ve ARIMAX(7,1,6) modelleri öngörüde kullanılmak üzere seçilerek ARIMAX modellerinin başarımı ile ilgili bir yorumla getirildi. Denenen ARIMAX modellerinin ARIMAX(8,1,7) modeline göre örtüşme düzeylerinin çok daha iyi durumda olduğunu ve hata terimlerinin daha küçük seviyelerde olduğu görüldü. Burdan yola çıkarak AIC ve BIC metotlarının ARIMAX modelleri için optimal model mertebesini seçmede kullanılmasının yanlış sonuçlar doğurabileceği gözlendi.

Yapılan analiz sonuçlarına göre mevcut veri serisinin analizinde kullanılmaya en uygun model ARX(1) modelidir.

# 1. INTRODUCTION

Wind energy is considered one of the most rapidly growing energy resources all over the world. It is expected that about 20% of the Canada total electricity demand to be supplied from wind energy resources by 2025. About 17% of the European Union electricity needs are expected to be supplied from wind energy, by the year 2020 [Url-1, Url-2]. Turkey has a global target of reaching 30% renewable energy based electricity production share and 20 GW wind power production capacity in markets by 2023 [Url-3]. Due to this expected high penetration rates of wind energy generation, wind farms are required to operate as controllable power plants. This increase the necessity for more accurate and reliable techniques for wind farms output power prediction. Wind power forecasting (WPF) approaches are also essential process for wind farms units' maintenance, optimal power flow between conventional units and wind farms, electricity marketing bidding, power system generators scheduling and energy reserves and storages planning and scheduling [1]. Reliable forecasting techniques lead to reliable power system by achieving the best schedule between the plants for day ahead and even on the short term for economical dispatch problem. In liberal electricity markets, having an accurate WPF models will considerably reduce the penalties imposed on such deviations in scheduling of power share of wind farms [2].

Considering the time scales of WPF, there are three classifications; very short term, short term and medium term WPF. Very short term is in the range of few seconds to 30 minutes ahead that helps in economical dispatch purpose. Short-term forecasting concerns from 30 minutes to 6 hours which is interesting in trading in day-ahead markets and UC (unit commitment). Medium term forecasting is in the range of 6 hours to a day that is very helpful in maintenance scheduling for conventional and wind plants and long term forecasting from a day to a week [3].

Since the issue of WPF arises, many researchers tried to get the state of the art of WPF techniques. There are two forecasting approaches, physical based approach and statistical approach [4-5]. Statistical models that are the subject of this thesis uses only historical wind speed data recorded at the site.

The statistical approach is based on training with measurement data and uses difference between the forecasted and the observed wind speeds in immediate past to tune model parameters. It is easy to model, inexpensive and provides timely prediction. It is not based on any predefined mathematical model  and rather it is based on patterns. Errors are minimized if patterns are met with historical ones. Sub-classification of this approach is: Time-series based models, and neural network (NN) based methods [6].

Auto regressive integrated moving average (ARIMA) also known as Box and Jenkins methods are the most popular type in time-series based approach to predict future values of wind speed or power. It can be used as autoregressive model (AR), moving average (MA), autoregressive moving average (ARMA). Several variations are seasonal ARIMA (s-ARIMA), and fractional ARIMA (f-ARIMA), AR with exogenous input (ARX), ARMA with exogenous input (ARMAX), and ARIMA with exogenous inpt (ARIMAX). Few other time–series models are grey predictors, linear predictors, exponential smoothing, etc. [7].

## 1.1 Purpose of Thesis

In this thesis, wind speed measurements time series data which recorded in Bursa-Gemlik weather measurement station during the April 2013 is analyzed and linear prediction methods AR, ARMA, ARIMA, ARX, ARMAX, and ARIMAX methods are used to forecast one step ahead forecasting. Corresponding the goodness of fit criterias, results are compared. Depending on the comparison results it’s aimed to see the best fit linear prediction model for the wind speed measurements.

## 1.2 Methodology

In this study, wind speed forecasting was done by using following procedure. Firstly, data series analyzed and checked for stationarity. The set of model was identified. After that, model parameter was estimated and checked for goodness of fit using information criteria AIC and BIC. Finally, optimal model was used for forecasting. These steps are shown as the scheme in Figure 1.1.

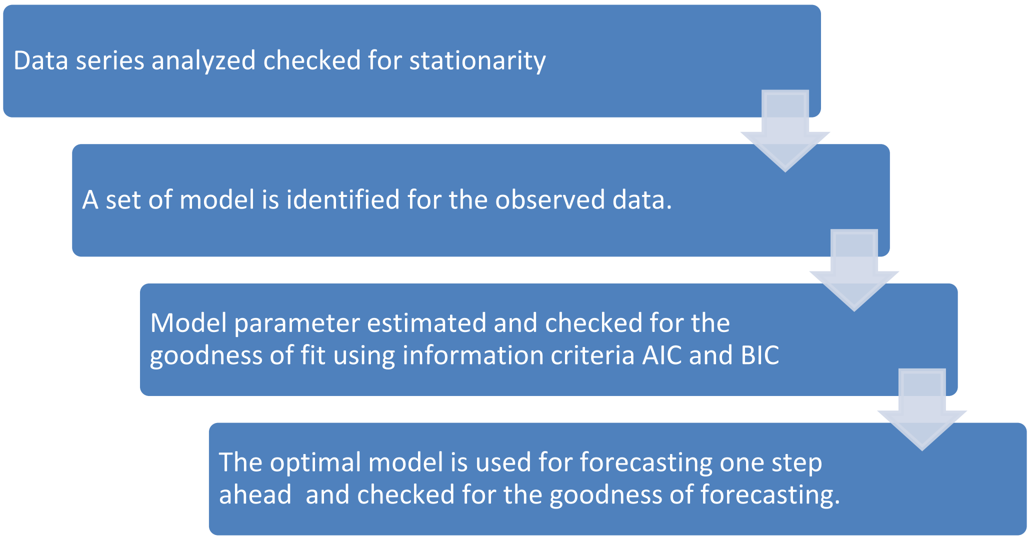


**Figure 1.1:** Scheme for the study of forecasting wind speed.

## 2. LINEAR PREDICTION METHODS

### 2.1 General Formulation

Let x(n) be a stationary random process. The value of the sample x(n) can be predicted using a linear combination of N most recent past samples. The estimate can be shown

$$\hat{x}_N^f(n) = -\sum_{i=1}^{N} \alpha_{N,i}^* \, x(n-i) \qquad \textbf{(2.1)}$$

Here N is the prediction order. The superscript * shows that $\alpha_{N,i}^*$ is optimum predictor with a prediction order N. The superscript f on the left is a reminder that we are discussing the ‘‘forward’’ predictor. The estimation error has the form

$$e_N^f(n) = x(n) - \hat{x}_N^f(n)\,) \qquad \textbf{(2.2)}$$

that is,

$$e_N^f(n) = x(n) + \sum_{i=1}^{N} a_{N,i}^* \, x(n-i) \qquad \textbf{(2.3)}$$

So we define “mean squared error (MSE)” as $\varepsilon_N^f$.

$$\varepsilon_N^f \triangleq E[\left|e_N^f(n)\right|^2] \qquad \textbf{(2.4)}$$

When we think wide sense stationary (WSS) process, MSE is independent of time. The optimum predictor with the optimum set of coefficients minimizes this MSE [8].

### 2.2 Autoregressive Processes (AR)

Linear predictive coding of a random process reveals a model for the process, called the autoregressive (AR) model. This model is very useful both conceptually and for approximating the process with a simple model.

If a WSS random process $\omega(n)$ can be generated by using recursive difference equation it can be defined as autoregressive (AR) and its formulation is:

$$\omega(n) = -\sum_{i=1}^{N} d_i^* \ \omega(n-i) + e(n) \qquad \textbf{(2.5)}$$

Here there are two assumptions

- $e(n)$ is a zero-mean white noise, and
- All zeros of the polynomial $D(z) = 1 + \sum_{i=1}^{N} d_i^* \ z^{-i}$ are inside the unit circle.

If $d_N \neq 0$, the process is AR(N), that is AR of order N. Because $e(n)$has zero mean, the AR process has zero mean according to the above definition.

Given a WSS process $x(n)$, let us assume that we have Nth-order optimal predictor polynomial $A_N(z)$. We know, we can then represent the process as the output of an infinite impulse response (IIR) filter as shown in formula:

$$e_N^f(n) \rightarrow \boxed{{}^{1}\!/_{A_N(z)}} \rightarrow e_N^f(n) \qquad \textbf{(2.6)}$$

IIR inverse filter

The input to this filter is the prediction error $e_N^f(n)$. In the time domain, we can write

$$x(n) = -\sum_{i=1}^{N} a_{N,i}^* \ x(n-i) + e_N^f(n) \qquad \textbf{(2.7)}$$

If the error stalls, that is, $\varepsilon_m^f$ does not decrease anymore as $m$ increases beyond some value $N$, then $e_m^f(n)$ is white (assuming $x(n)$ has zero mean). Thus, the stalling phenomenon iplies that $x(n)$ is $AR(N)$.

Summarizing, suppose the optimal predictors of various orders for a zero-mean process $x(n)$ are such that the minimized mean square errors satisfy

$$\varepsilon_m^f \geq \varepsilon_m^f \geq \cdots \geq \varepsilon_m^f = \varepsilon_m^f \ , \qquad m > N \qquad \textbf{(2.8)}$$

Then, $x(n)$ is $AR(N)$.

## 2.3 Moving Average (MA) and Autoregressive Moving Average (ARMA) Processes

We know that a WSS random process $x(n)$ is said to be AR if it satisfies a recursive (IIR) difference equation of the form

$$\mathrm{x(n)} = -\sum_{\mathrm{i=1}}^{\mathrm{N}} d_i^* \ \mathrm{x(n-i)} + \mathrm{e(n)}, \qquad \textbf{(2.9)}$$

where $\mathrm{e(n)}$ is a zero-mean white WSS process, and the polynomial $\mathrm{D(z)} = 1 +$

$\sum_{i=1}^{N} d_i^* z^{-i}$ has all zeros inside the unit circle. We say that a WSS process $x(n)$ is a moving average (MA) process if it satisfies a nonrecursive (FIR) difference equation of the form

$$x(n) = \sum_{i=0}^{N} \rho_i^* \; e(n-i) \tag{2.10}$$

where $e(n)$ is a zero-mean white WSS process. Finally, we say that a WSS process $x(n)$ is an ARMA process if

$$x(n) = -\sum_{i=1}^{N} d_i^* \; x(n-i) + \sum_{i=0}^{N} \rho_i^* \; e(n-i) \tag{2.11}$$

where $e(n)$ is a zero-mean white WSS process. Defining the polynomials

$$D(z) = 1 + \sum_{i=1}^{N} d_i^* z^{-i} \text{ and } P(z) = \sum_{i=0}^{N} \rho_i^* z^{-i} \tag{2.12}$$

we see that the above processes can be represented as in Figure. 2.1.

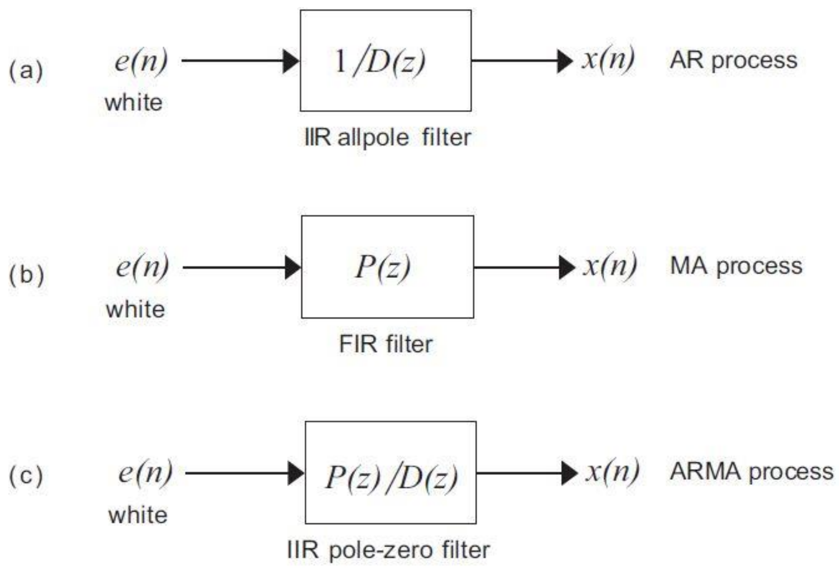


**Figure 2.1: (**a) AR, (b) MA, and (c) ARMA processes.

In each of the three cases, $x(n)$ is the output of a rational discrete time filter, driven by zero-mean white noise. For the AR process, the filter is an all-pole filter. For the MA process, the filter is FIR. For the ARMA process, the filter is IIR with both poles and zeros [8].

## 2.4 Autoregressive Integrated Moving Average (ARIMA) Processes

ARIMA models are extensions of ARMA class in order to include more realistic dynamics, in particular, respectively, non-stationarity in mean. In practice, many time series are nonstationary in mean and they can be modelled only by removing the nonstationary source of variation. Often this is done by differencing the series.

Suppose $X_t$ is nonstationary in mean, the idea is to build an ARMA model on the series $w_t$, defnible as the result of the operation of differencing the series d times (in general d = 1): $w_t = \Delta^d X_t$.

Hence, ARIMA models (where I stays for integrated) are the ARMA models defined on the d-th diference of the original process:

$$\Phi(B)\Delta^d X_t = \theta(B)a_t \tag{2.13}$$

where $\Phi(B)\Delta^d$ is called generalized autoregressive operator and $\Delta^d X_t$ is a quantity made stationary through the differentiation and can be modelled with an ARMA.
For example:

- ARIMA (0,1,1) is $\Delta X_t = a_t - \theta_1 a_{t-1}$ → the first difference of $X_t$ is modelled as MA(1).
- ARIMA (1,1,0) is $(1 - \Phi_1 B)\Delta X_t = a_t$ → the first difference of $X_t$ is modelled as AR(1).

Note that in this case:

$$(1 - \Phi_1 B)(1 - B)X_t = a_t \tag{2.14}$$

$$(1 - B - \Phi_1 B + \Phi_1 B^2)X_t = a_t \tag{2.15}$$

$$[1 - (1 + \Phi)B + \Phi B^2] = a_t \tag{2.16}$$

The last equation shows that ARIMA(1,1,0) is like an AR(2) where $\Phi_2 = -\Phi_1$ and $\Phi_2 + \Phi_1 = 1$. This reveals that, as we knew in advance, the stationary constraint does not hold [9].

### 2.4.1 The role of the constant term in ARMA and ARIMA

Suppose that a constant term is included in the ARMA model:

$$\Phi(B)X_t = \theta_0 + \theta(B)a_t \tag{2.17}$$

Taking the expected value on both sides (note that $E(X_t) = \mu$ due to stationarity of the process):

$$\Phi(B)\mu = \theta_0 \qquad \textbf{(2.18)}$$

Hence;

$$\theta_0 = \mu\left(1 - \Phi_1 - \cdots - \Phi_p\right) \qquad \textbf{(2.19)}$$

that represents the relationship between the constant term and the expected value of the process. From this, we can make two conclusions:

- If $X_t$ is MA (the AR component does not exist), the possible constant present in the model coincides with the mean of the process $X_t$ itself ($\mu = \theta_0$).
- If $X_t$ is only AR, then the aforementioned relationship holds. It is interesting to observe that for AR(1) in case $\Phi_1 \rightarrow 1$ (in general when the autoregressive process tends to nonstationarity) the constant tends to disappear.
- If $X_t$ is ARIMA, by including $\theta_0$ term:

$$\Phi(B)\Delta X_t = \theta_0 + \theta(B)a_t \qquad \textbf{(2.20)}$$

This is called ARIMA model *with drift*. The simplest case is the *Random Walk plus drift* process:

$$\Delta X_t = \theta_0 + a_t \qquad \textbf{(2.21)}$$

By taking the first diference of $X_t$ one obtains a quantity $\Delta X_t$ whose mean is not zero: a drift $\theta_0$ adjusts for this. The idea is that the constant $\theta_0$ was originally the slope of a deterministic trend, that after diferencing d = 1 times, disappears leaving only a level ($\theta_0$) around which $\Delta X_t$ moves with stationary oscillations. The random walk with drift is characterized by both stochastic and deterministic trend. Every time a diference is taken, a trend is removed whose nature (stochastic or deterministic) is clear only by checking whether the differences uctuate around zero (stochastic trend) or not (deterministic trend, whose slope remains in the form of the constant) [9].

### 2.4.2 Characteristics of ARIMA processes

- $d = 0$ stationary process
- $d = 1$ nonstationary process: the level changes in time, but the increase is constant → level is nonstationary, but its increments are
- $d = 2$ nonstationary process: both level and increments are stationary

When Xt is nonstationary, its theorethical ACF is not defined (only the empirical ACF is). However, by observing the behaviour of processes that are nearly stationary we can put in evidence the following regularities:

- The ACF decreases extremely slowly to zero, the decrease is not exponential by linear.
- The PACF takes value 1 for $k = 1$ and zero elsewhere.

These characteristics of ACF and PACF are motivated by the dominance of the trend on the other dynamics in the series. Unless the trend is removed, nothing else (e.g. other MA or AR components) can be recognized from ACF and PACF [9].

### 2.5 Autoregressive Models with Exogenous Input (ARX)

Probably the most simple input-output relationship is obtained by describing it as a linear difference equation:

$$y(t) + a_1 y(t-1) + \cdots + a_{n_a} y(t-n_a) = \\ b_1 u(t-1) + \cdots + b_{n_b} u(t-n_b) + e(t) \qquad \textbf{(2.22)}$$

Since the white-noise term $e(t)$ here enters as a direct error in the difference equation, the model (2.22) is often called an equation error model (structure). The adjustable parameters are in this case

$$\theta = \left[a_1 a_2 \cdots a_{n_a} b_1 \cdots b_{n_b}\right]^T \qquad \textbf{(2.23)}$$

If we introduce

$$A(q) = 1 + a_1 q^{-1} + \cdots + a_{n_a} q^{-n_a} \qquad \textbf{(2.24)}$$

and

$$B(q) = b_1 q + \cdots + b_{n_b} q^{-n_b} \qquad \textbf{(2.25)}$$

We see that (2.22) corresponds to (2.13) with

$$G(q,\theta) = \frac{B(q)}{A(q)} \qquad \textbf{(2.26)}$$

$$H(q,\theta) = \frac{1}{A(q)} \qquad \textbf{(2.27)}$$

Remark: It may seem annoying to use $q$ as an argument of $A(q)$. Being a polynomial in $q^{-1}$. The reason for this is, however, simply to be consistent with the conventional definition of the z-transform.

We shall also call the model (2.22) an ARX model, where AR refers to the autoregressive part $A(q)y(t)$ and X to the estra input $B(q)u(t)$ (called the exogenous variable in econometrics). ARX model structure is shown in Figure 2.2. In the special case where $n_a = 0$, $y(t)$ is modeled as a finite impulse response (FIR). Such model sets are particularly common in signal-processing applications [10].

The signal flow can be depicted as in Figure From that picture we see that the model (2.22) is perhaps not the most natural one from a physical point of view: the white noise is assumed to go through the denominator dynamics of the system before being added to the output. Nevertheless, the equation error model set has a very important property that makes it a prime choice in many applications: The predictor defines a linear regression.

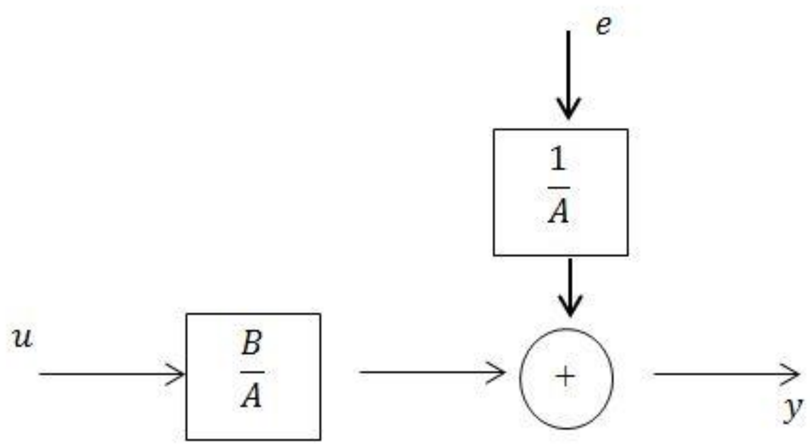


**Figure 2.2**: ARX Process.

### 2.6 Autoregressive Moving Average Models with Exogenous Input (ARMAX) and Autoregressive Integrated Moving Average Models with Exogenous Input (ARIMAX)

The basic disadvantage with the simple model (2.22) is the lack of adequate freedom in describing the properties of the disturbance term. We could add flexibility to that by describing the equation error as a moving average of white noise. This gives the model;

$$y(t) + y(t-1) + \cdots + a_{n_a} y(t-n_a) \ = b_1 u(t-1) + \cdots$$

$$+ b_{n_b} u(t-n_b) + e(t) + \ c_1 e(t-1) + \cdots + c_{n_c} e(t-n_c) \qquad \textbf{(2.28)}$$

with

$$C(q) = 1 + c_1 q^1 + \cdots + c_{n_c} q^{n_c} \qquad \textbf{(2.29)}$$

It can be rewritten

$$A(q)y(t) = B(q)u(t) + C(q)e(t) \qquad \textbf{(2.30)}$$

And clearly corresponds to (2.13) with

$$G(q,\theta) = \frac{B(q)}{A(q)} \qquad \textbf{(2.31)}$$

$$H(q,\theta) = \frac{C(q)}{A(q)} \qquad \textbf{(2.32)}$$

where now

$$\theta = \left[a_1 \cdots a_{n_a} b_1 \cdots b_{n_b} c_1 \cdots c_{n_c}\right]^T \qquad \textbf{(2.33)}$$

In the view of the moving averae (MA) part $C(q)e(t)$, the model (4.28) will be called ARMAX. The ARMAX model has become a standard tool in control and econometrics for both system description and control design. A version with an enforced integration in the noise description is the ARIMA(X) model (I for integration with or without the X-variable $u$) which is useful to describe systems with slow disturbances [10]. It is obtained by replacing $y(t)$ and $u(t)$ in (2.30) by their differences $\Delta y(t) = y(t) - y(t-1)$ and is further discussed.

## 2.7 Parameter Estimation

### 2.7.1 Yule-Walker method

Depending on the orthogonality principle the optimum value of $\alpha_{N,i}$ can be defined when the error $\mathrm{e}_{\mathrm{N}}^{\mathrm{f}}(\mathrm{n})$ is orthogonal to $\mathrm{x}(\mathrm{n}-\mathrm{i})$, that is,

$$\mathrm{E}\left[\mathrm{e}_{\mathrm{N}}^{\mathrm{f}}(\mathrm{n})\mathrm{x}^*(\mathrm{n}-\mathrm{i})\right] = 0,\ 1 \le \mathrm{i} \le \mathrm{N} \qquad \textbf{(2.34)}$$

There are N equations occurs and all can be shown in a special form, because of the condition.

$$[\mathrm{R}]_{\mathrm{im}} = \mathrm{E}[\mathrm{x}(\mathrm{n}-1-\mathrm{i})\mathrm{x}^*(\mathrm{n}-1-\mathrm{m})], \quad 0 \le \mathrm{i,m} \le \mathrm{N}-1 \qquad \textbf{(2.35)}$$

The autocorrelation sequence of WSS process x(n)can be defined as R(k),

$$R(k) = E[x(n)x^*(n-k)] \qquad \textbf{(2.36)}$$

Using the fact $R(k) = R^*(-k)$, we can simplfy Eq. (2.34) to obtain:

$$\underbrace{\begin{bmatrix} R(0) & R(1) & \dots & R(N-1) \\ R^*(1) & R(0) & \dots & R(N-2) \\ \vdots & \vdots & \ddots & \vdots \\ R^*(N-1) & R^*(N-2) & \dots & R(0) \end{bmatrix}}_{R_N} \begin{bmatrix} a_{N,1} \\ a_{N,2} \\ \vdots \\ a_{N,N} \end{bmatrix} = \underbrace{-\begin{bmatrix} R^*(1) \\ R^*(2) \\ \vdots \\ R^*(N) \end{bmatrix}}_{-r} \qquad \textbf{(2.37)}$$

For example, with N=3, we get

$$\underbrace{\begin{bmatrix} R(0) & R(1) & R(2) \\ R^*(1) & R(0) & R(1) \\ R^*(2) & R^*(1) & R(0) \end{bmatrix}}_{R_3} \begin{bmatrix} a_{3,1} \\ a_{3,2} \\ a_{3,3} \end{bmatrix} = -\begin{bmatrix} R^*(1) \\ R^*(2) \\ R^*(3) \end{bmatrix}. \qquad \textbf{(2.38)}$$

These are called Yule-Walker equations, normal equations and Wiener-Hopf equations in literature [11]. Assuming that the $n \times n$ symmetric matrix R is invertible, the coefficient $a_{N,i}$ are estimated by $a_{N,1} = R^{-1}r$. Once the coefficients are estimated, the linear prediction model can be applied to predict future samples, with $x_k = \sum_{i=1}^{N} a_i x_{k-i}$ . This technique is easy to implement, it is not adapted for nonlinear systems [12].

### 2.7.2 Levinson-Durbin algorithm

The method is recursive in nature and makes particular use of the Toeplitz structure having constant entries along diagonals of the correlation matrix of the tap inputs of the filter. It is known as the Levinson-Durbin algorithm, so named in recognition of its use first by Levinson (1947) and then its independent reformulation at a later date by Durbin (1960) [13]. The property of a forward prediction error filter operating on a stationary discrete time stochastic process is intimately relate to the autoregressive (AR) modelling of the process. The prediction error filter is an all zero filter with an impulse response of finite duration. On the other hand, the inverse of prediction error yields the AR model that is an all zero filter with an impulse response if infinite duration. From this relation the levinson durbin algorithm is adopted to compute the estimate AR coefficients which is:

Initialize the algorithm by setting;

$$a_{0,0} = 1 \quad \textbf{(2.39)}$$

$$P_0 = r(0) \quad \textbf{(2.40)}$$

Hence, compute for $m = 1,2,3, \dots \dots \dots.., M$:

$$K_m = -(1/P_{m-1}). \sum_{i=1,m-1} r(i-m) a_{m-1,i} \quad \textbf{(2.41)}$$

$$a_{m,i} = 1 \qquad for\ i = 1$$

$$a_{m,i} = a_{\ m-1,i} + K_m a^*{}_{m-1,i-1} \qquad for\ i = 1,2,..,m-1$$

$$a_{m,i} = K_m \qquad for\ i = m \quad \textbf{(2.42)}$$

$$P_m = P_{m-1}(1 - |K_m|^2) \quad \textbf{(2.43)}$$

where $a_{M,k}, k = 1,2, \dots., M$ denotes the estimated AR coefficients. $K_m$ is the reflection coefficient. $P_m$ is the prediction error power [14].

As the order m of the prediction error filter increases, the corresponding value of the prediction error power normally decreases or else remains the same. In addition to this, $P_m$ can never be negative. Hence,

$$0 \le P_m \le P_{m-1} \qquad m \ge 1 \quad \textbf{(2.44)}$$

For the elementary case of a prediction-error filter of order zero,

$$P_0 = r(0) \quad \textbf{(2.45)}$$

where $r(0)$ is the autocorrelation function of the input process for lag.

Starting with $m = 0$ and increasing the filter order by 1 at a time, repetition of equation occurs. Moreover, the prediction error power for a prediction error filter of final order M equals [13].:

$$P_M = P_0 \prod_{m=1}^{M} (1 - |K_m|^2) \quad \textbf{(2.46)}$$

### 2.7.3 Burg algorithm

The Yule–Walker coefficients $\hat{\phi}_{p1}, \cdots, \hat{\phi}_{pp}$ are precisely the coefficients of the best linear predictor of $X_{p+1}$ in terms of $\{Xp, \dots, X1\}$ under the assumption that the ACF of $\{Xt\}$ coincides with the sample ACF at lags $1, \dots, p$. Burg's algorithm estimates the PACF $\{\phi_{11}, \phi_{22}, \dots\}$ by successively minimizing sums of squares of forward and backward one-step prediction errors with respect to the coefficients $\phi_{ii}$ . Given observations $\{x_1, \dots, x_n\}$ of a stationary zero-mean time series $\{Xt\}$ we define

$u_i\ (t\ ),\ t =\ i + 1,\dots,n,\ 0\ \leq\ i\ <\ n$, to be the difference between $x_{n+1+i-t}$ and the best linear estimate of $x_{n+1+i-t}$ terms of the preceding $i$ observations. Similarly, we define $v_i\ (t\ ),\ t =\ i + 1,\dots,n,\ 0\ \leq\ i\ <\ n$, to be the difference between $x_{n+1-t}$ and the best linear estimate of $x_{n+1-t}$ in terms of the subsequent i observations. Then it can be shown that the forward and backward prediction errors $\{u_i\ (t)\}$ and $\{v_i\ (t)\}$ satisfy the recursions;

$$u_0\ (t) = v_0\ (t) = x_{n+1-t} \qquad \textbf{(2.47)}$$

$$u_i\ (t) = u_{i-1}\ (t-1) -\ \phi_{ii}v_{i-1}(t) \qquad \textbf{(2.48)}$$

and

$$v_i = v_{i-1}(t) - \phi_{ii}u_{i-1}(t-1) \qquad \textbf{(2.49)}$$

Burg's estimate $\phi_{11}^{(B)}$of $\phi_{11}$is found by minimizing

$$\sigma_1^2 = \frac{1}{2(n-1)}\sum_{t=2}^{n}[u_1^2(t) + v_1^2(t)] \qquad \textbf{(2.50)}$$

with respect to $\phi_{11}$. This gives corresponding numerical values for $u_1\ (t)$ and $v_1(t)$ and $\sigma_1^2$ that can then be substituted into together with $i =\ 2$. Then we minimize;

$$\sigma_2^2 = \frac{1}{2(n-1)}\sum_{t=3}^{n}[u_2^2(t) + v_2^2(t)] \qquad \textbf{(2.51)}$$

with respect to $\phi_{22}$ to obtain the Burg estimate $\phi_{22}^{(B)}$ of $\phi_{22}$ and corresponding values of $u_2\ (t)$, $v_2\ (t)$, and $\sigma_2^2$ . This process can clearly be continued to obtain estimates $\phi_{pp}^{(B)}$ and corresponding minimum values, $\sigma_p^{(B)2}$, $p\ \leq\ n\ -\ 1$. Estimates of the coefficients $\phi_{pj}$ , $1\ \leq\ j\ \leq\ p\ -\ 1$, in the best linear predictor

$$P_pX_{p+1} =\ \phi_{p1}X_p + \cdots + \phi_{pp}X_1 \qquad \textbf{(2.52)}$$

are then found by substituting the estimates $\phi_{ii}^{(B)}$ , $i = 1,\dots,p$, for $\phi_{ii}$ in the recursions . The resulting estimates of $\phi_{pj}$, $j\ = 1,\dots,p$, are the coefficient estimates of the Burg $AR(p)$ model for the data $\{x_1,\dots,x_n\}$. The Burg estimate of the white noise variance is the minimum value $\phi_p^{(B)2}$ found in the determination of $\phi_{pp}^{(B)}$ . The calculation of the estimates of $\phi_{pp}$ and $\sigma_p^2$ described above is equivalent to solving the following recursions:

*Burg's Algorithm:*

$$d(1) = \sum_{t=2}^{n}[u_0^2(t-1) + v_0^2(t)], \tag{2.53}$$

$$\phi_{ii}^{(B)} = \frac{1}{d(i)}\sum_{t=i+1}^{n} v_{i-1}(t)u_{i-1}(t-1), \tag{2.54}$$

$$d(i+1) = \left(1-\phi_{ii}^{(B)2}\right)d(i) - v_i^2(i+1) - u_i^2(n), \tag{2.55}$$

$$\sigma_i^{(B)2} = \left[\left(1-\phi_{ii}^{(B)2}\right)d(i)\right]/[2(n-i)] \tag{2.56}$$

The large-sample distribution of the estimated coefficients for the Burg estimators of the coefficients of an $AR(p)$ process is the same as for the Yule–Walker estimators, namely, $N\left(\phi, n^{-1}\sigma^2\Gamma_p^{-1}\right)$. Approximate large-sample confidence intervals for the coefficients can be found by substituting estimated values for $\sigma^2$ and $\Gamma_p$ [15].

### 2.7.4 Maximum likelihood estimation

Suppose that $X_t$ is a Gaussian time series with mean zero and autocovariance function $\kappa(i,j) = E\left(X_iX_j\right)$. Let $X_n = (X_1, \cdots, X_n)'$ and let $\hat{X}_n = \left(\hat{X}_1, \cdots, \hat{X}_n\right)'$, where $\hat{X}_1 = 0$ and $\hat{X}_j = E\left(X_j \middle| X_1, \cdots, X_{j-1}\right) = P_{j-1}X_j$ , $j \geq 2$. Let $\Gamma_n$ denote the covariance matrix $\Gamma_n = E(X_nX_n')$, and assume that $\Gamma_n$ is nonsingular.

The likelihood of $|X_n$ is

$$L(\Gamma_n) = (2\pi)^{-n/2}(\det\Gamma_n)^{-1/2}\exp\left(-\frac{1}{2}X_n'X_n^{-1}X_n\right) \tag{2.57}$$

As we shall now show, the direct calculation of $\det\Gamma_n$ and $\Gamma_n^{-1}$ can be avoided by expressing this in terms of the one-step prediction errors $X_j - \hat{X}_j$ and their variances $\upsilon_{j-1}$, $j = 1, \cdots, n$, both of which are easily calculated recursively from the innovations algorithm.

Let $\theta_{i,j}$, $j = 1, \cdots, i$ ; $i = 1,2, \cdots$, denote the coefficients obtained when the innovations algorithm is applied to the autocovariance function $\kappa$ of $\{X_t\}$, and let $C_n$ be the $n \times n$ lower triangular matrix. From above equation, we have the identity

$$X_n = C_n\left(X_n - \hat{X}_n\right)' \tag{2.58}$$

We also know components of $X_n - \hat{X}_n$ are uncorrelated. Consequently, by the definition of $\upsilon_j$ , $X_n - \hat{X}_n$ has the diagonal covariance matrix

$$D_n = diag\{\upsilon_0, \cdots, \upsilon_{n-1}\} \tag{2.59}$$

we conclude that

$$\Gamma_{\mathrm{n}} = C_n D_n C_n' \quad \textbf{(2.60)}$$

we see that

$$X_n' X_n^{-1} X_n = \left(X_n - \hat{X}_n\right)' D_n^{-1}\left(X_n - \hat{X}_n\right) = \sum_{j=1}^{n}\left(X_j - \hat{X}_j\right)^2 / \, v_{j-1} \quad \textbf{(2.61)}$$

and

$$det\Gamma_{\mathrm{n}} = (detC_n)^2(detD_n) = v_0 v_1 \cdots v_{n-1} \quad \textbf{(2.62)}$$

The likelihood of the vector $X_n$ therefore reduces to

$$L(\Gamma_n) = \frac{1}{\sqrt{(2\pi)^n v_0 \cdots v_{n-1}}} exp\left\{-\frac{1}{2}\sum_{j=1}^{n}\left(X_j - \hat{X}_j\right)^2 / v_{j-1}\right\} \quad \textbf{(2.63)}$$

If $\Gamma_{\mathrm{n}}$ expressible in terms of a finite number of unknown parameters $\beta_1, \dots, \beta_r$ (as is the case when $\{Xt\}$ is an ARMA(p, q) process), the maximum likelihood estimators of the parameters are those values that maximize $L$ for the given data set. When $X_1, X_2, \dots, X_n$ are iid, it is known, under mild assumptions and for n large, that maximum likelihood estimators are approximately normally distributed with variances that are at least as small as those of other asymptotically normally distributed estimators.

Even if $\{Xt\}$ is not Gaussian, it still makes sense to regard as a measure of goodness of fit of the model to the data, and to choose the parameters $\beta_1, \dots, \beta_r$ in such away as to maximize (5.2.6).We shall always refer to the estimators $\hat{\beta}_1, \dots, \hat{\beta}_r$ so obtained as "maximum likelihood" estimators, even when $\{Xt\}$ is not Gaussian. Regardless of the joint distribution of $X_1, \dots, X_n$, we shall refer to .. and its algebraic equivalent .. as the "likelihood" (or "Gaussian likelihood") of $X_1, \dots, X_n$. A justification for using maximum Gaussian likelihood estimators of ARMA coefficients is that the large-sample distribution of the estimators is the same for {Zt} $\sim$ IID $(0, \sigma 2)$, regardless of whether or not {Zt } is Gaussian.

The likelihood for data from an $\mathrm{ARMA(p, q)}$ process is easily computed from the innovations form of the likelihood by evaluating the one-step predictors $\hat{X}_{i+1}$ and the corresponding mean squared errors $v_i$. These can be found from the recursions;

$$\hat{X}_{n+1} = \begin{cases} \sum_{j=1}^{n} \theta_{nj}\left(X_{n+1-j} - \hat{X}_{n+1-j}\right), & 1 \le n \le m, \\ \phi_1 X_n + \cdots + \phi_p X_{n+1-p} + \sum_{j=1}^{q} \theta_{nj}\left(X_{n+1-j} - \hat{X}_{n+1-j}\right), & n \ge m, \end{cases} \quad \textbf{(2.64)}$$

$$E\left(X_{n+1}-\hat{X}_{n+1}\right)^2=\sigma^2E\left(W_{n+1}-\widehat{W}_{n+1}\right)^2=\sigma^2r_n \quad \textbf{(2.65)}$$

where $\theta_{nj}$ and $r_n$ are determined by the innovations algorithm with $\kappa$ and $m = max(p, q)$. Substituting in the general expression, we obtain the following:

*The Gaussian Likelihood for an ARMA Process***:**

$$L(\phi,\theta,\sigma^2)=\frac{1}{\sqrt{(2\pi\sigma^2)^n r_0 \dots r_{n-1}}}exp\left\{-\frac{1}{2\sigma^2}\sum_{j=1}^{n}\frac{\left(X_j-\hat{X}_j\right)^2}{r_{j-1}}\right\} \quad \textbf{(2.66)}$$

Differentiating ln $L(\phi, \theta, \sigma^2)$ partially with respect to $\sigma^2$ and noting that $\hat{X}_j$ and $r_j$ are independent of $\sigma^2$, we find that the maximum likelihood estimators $\hat{\phi}, \hat{\theta}$ *and* $\hat{\sigma}^2$ satisfy the following equations:

*Maximum Likelihood Estimators:*

$$\sigma^2=n^{-1}S\left(\hat{\phi},\hat{\theta}\right), \quad \textbf{(2.67)}$$

where

$$S\left(\hat{\phi},\hat{\theta}\right)=\sum_{j=1}^{n}\left(X_j-\hat{X}_j\right)^2/\,r_{j-1}, \quad \textbf{(2.68)}$$

and $\hat{\phi}, \hat{\theta}$ are the values of $\phi, \theta$ that minimize

$$l(\phi,\theta)=ln\left(n^{-1}S(\phi,\theta)\right)+n^{-1}\sum_{j=1}^{n}ln\, r_{j-1}. \quad \textbf{(2.69)}$$

Minimization of $l(\phi, \theta)$ must be done numerically. Initial values for $\phi$ and $\theta$ can be obtained from ITSM. The program then searches systematically for the values of $\phi$ and $\theta$ that minimize the reduced likelihood and computes the corresponding maximum likelihood estimate of $\sigma^2$ [15].

## 2.8 Methods and Criterias for Optimal Model Order Selection

Model selection is an important part of any statistical analysis and is central to the pursuit of science in general. Moreover, in statistical modeling, choosing a suitable model from among a collection of viable candidates bring problem for an investigator [16]. The visual inspection of autocorrelation function (ACF) and partial autocorrelation function (PACF) provides a useful way to construct an ARMA(p,q) model. However, the more objective way to chose the orders of p and q of an ARMA(p,q) process is to use objectively defined criterions such as AIC, BIC and FPE. These information criterions are statistical model fit measures. They quantify the relative goodness of fit of various previously derived statistical models, given a sample of data. Each of these criterions has their own merits and demerits [17]. Therefore, in this study, the decision of choosing model has been made after carefully examining all these criterions.

### 2.8.1 Akaike information criterion (AIC)

The Akaike information criterion, AIC, was introduced by Hirotogu Akaike who was former Director General of the Institute of Statistical Mathematics and a Kyoto PrizeWinner in 1973 with paper named as "Information Theory and an Extension of the Maximum Likelihood Principle" [18]. AIC was the first model selection criterion to attract widespread notice in statistical community. AIC continues to be the most widely used model selection tool among practitioners. The traditional maximum likelihood paradigm provides a mechanism for estimating the unknown parameters of a model having a specified dimension and structure. Akaike developed this paradigm by considering a framework in which the model dimension is also unknown, and must be determined from the data. In this way, Akaike proposed a framework wherein both model predicting and selection could be simultaneously achieved [19].

The information criterion I (f0:f1) that measures the deviation of a model specified by the probability distribution f1 from the true distribution f0 is defined by formula

$$I(f_0\colon f_1) = E_{f_0}.logf_0 - E_{f_0}.logf_1 \qquad \textbf{(2.70)}$$

It is known that $I(f_0\colon f_1) \geq 0$ with equality if $f_0 = f_1$. Therefore the smaller $I(f_0\colon f_1)$, the better model is $f_1$ as an estimator for $f_0$. However, since $E_{f_0}.logf_0$ is constant for all estimators $f_1$, the larger $E_{f_0}.logf_1$, the better is model $f_1$. Since $f_1$ is unknown, if we have a random sample x1,….,xn, $E_{f_0}.logf_1$ is estimated by the moment estimator,

$$\frac{1}{n}\sum_{j=1}^{n} logf_1(x_j, \theta) \qquad \textbf{(2.71)}$$

where $\theta$ is the vector of parameters of the model under $f_1$. However, $\theta$ may be unknown, so another estimator is needed, namely

$$\frac{1}{n}\sum_{j=1}^{n} logf_1(x_j, \hat{\theta}) \qquad \textbf{(2.72)}$$

where $\hat{\theta}$ is the maximum likelihood estimator for $\theta$. It is clear that this estimator is the average of the maximum log-likelihood under $f_1$. Therefore, it seems "reasonable" to say that the larger the maximum-log-likelihood, the better is the model [20].

To have an asymptotically unbiased estimator for the negative of twice the expected likelihood, the maximum likelihood estimates of the parameters that give the minimum of Akaike information criteria defined by Akaike

$$AIC\ (\hat{\theta}) = -2.\, log(maximum\ likelihood) + 2.\, k \qquad \textbf{(2.73)}$$

where k is the number of independently adjusted parameters to get $\hat{\theta}$. Hence the smaller the AIC, the better is the model [24].

### 2.8.2 Bayesian information criterion (BIC)

The Bayesian information criterion, was introduced by Schwarz (1978) as a competitor to the Akaike (1973, 1974) information criterion. An extension of the maximum likelihood principle is suggested by Akaike[21] for the slightly more general problem of choosing among different models with different numbers of parameters [23]. BIC is one of the most widely known and widespread used tools in statistical model selection. The computational simplicity and effective performance in many modelling frameworks including Bayesian applications where prior distributions may be elusive, bring its popularity. In Bayesian applications, pairwise comparisons between models are often based on Bayes factors. Assuming two candidate models are regarded as equally probable a priori, a Bayes factor represents the ratio of the posterior probabilities of the models. The model which is a posteriori most probable is determined by whether the Bayes factor is less than or greater than one [16]. The Bayesian information criterion is often called the Schwarz information criterion. Bayesian information criterion is defined as the formula:

$$BIC == -2.\, log(maximum\ likelihood) + k.\log(n) \qquad \textbf{(2.74)}$$

$maximum\ likelihood$ is the fitted model n is the number of observations and $k$ denotes the dimension of selected model. AIC and BIC share the same goodness-of-fit term, but the penalty term of BIC $k.\log(n)$ is potentially much more stringent than the penalty term of AIC (2k). Thus, BIC tends to choose fitted models that are more parsimonious than those favored by AIC [23].

### 2.8.3 Final prediction error (FPE)

The use of autoregressive representation of a stationary time series in the analysis of time series has been attracting attentions of many research workers. It is expected that this time domain approach will give answers to many problems. The main difficulty in fitting an autoregressive model;

$$X(n) = \sum_{m=1}^{M} a_m x(n-m) + a_0 + \varepsilon(n) \quad \textbf{(2.75)}$$

X(n) is the process being observed and $\varepsilon(n)$ is its innovation that is not correlated with X(l) and is forming a white noise, lies in the decision of the order M. the mutual independence and strict stationary of $\varepsilon(n)$ is assumed. To surmount this problem, the decision theoretic approach where a figure of merit is defined for each model being fitted is adopted and the one with the best figure is chosen as predictor. This figure of merit that is called as the final prediction error (FPE) is defined as the expected variance of the prediction error when an autoregressive model fitted to the present series of X(n). It is applied to another independent realization of X(n), or to the process with one and the same covariance characteristic as that of X(n) and to make a one-step prediction, it is independent of the present X(n). The study over final prediction error was investigated by Akaike. The estimation of FPE of each autoregressive model within a prescribed sufficiently wide range of possible orders was computed and the one that gives the minimum of the estimates was chosen. Akaike called this procedure as FPE scheme. The definition of final prediction error (FPE) of the autoregressive model of order M is given by the relation;

$$FPE = \left(1 + \frac{M+1}{N}\right) . r_M \quad \textbf{(2.76)}$$

where $r_M$ is the minimum of $E(X(n) - \sum_{m=1}^{M} a_m^{(M)} X(n-m) - a_0^{(M)})^2$ with respect to $\left\{a_m^{(M)}; m = 0,1, \dots \dots, M\right\}$. Obviously $r_M$ is equal to the variance of the innovation $\varepsilon(n)$ when X(n) is generated from $\varepsilon(n)$ by a finite auto regression of order equal to or less than M. FPE tends to be large when unnecessarily large value of M is adopted. When M is less than the true order of the process, $r_M$ and its estimate include, beside the contribution of the innovation variance, the contribution of the inevitable bias of the model. Therefore it tends to be significantly large when a too small value of M is adopted [24].

# 3. MODEL IDENTIFICATION

## 3.1 Data Specification

Wind speed datas are taken from the Turkish State Meteorological Service, Data Control and Statistical Division. Measurements are recorded during the April 2013 by Gemlik automatic meteorological observation station as shown Figure 3.1. The file has a 4320 data for 30 days and 10 minutes time interval. The station is on the location of 40.4401 latitute and 29.1504 longtitude that can be seen as a red star in Figure 3.2.

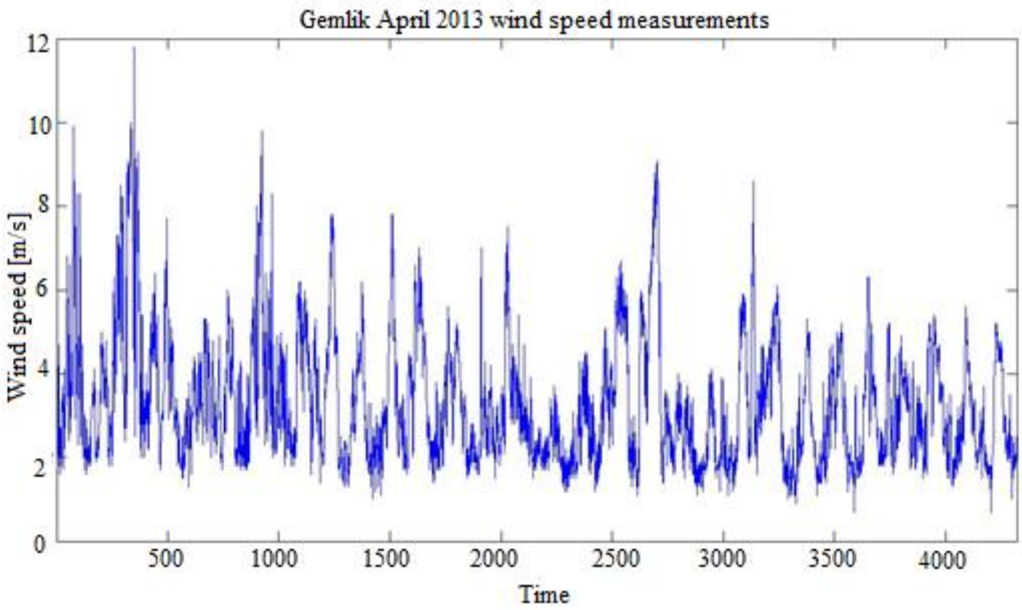


**Figure 3.1:** Gemlik meteorological station wind speed measurements on April 2013.

Automatic meteorological observation station has various sensors for different purpose. Wind speed measurement sensor is located at the top of the 10 m height bar and used for measuring the wind speed. The sensor is cup anemometer, wind rotates the anemometer and magnitude is measured by the number of rotation per unit time [Url-4].

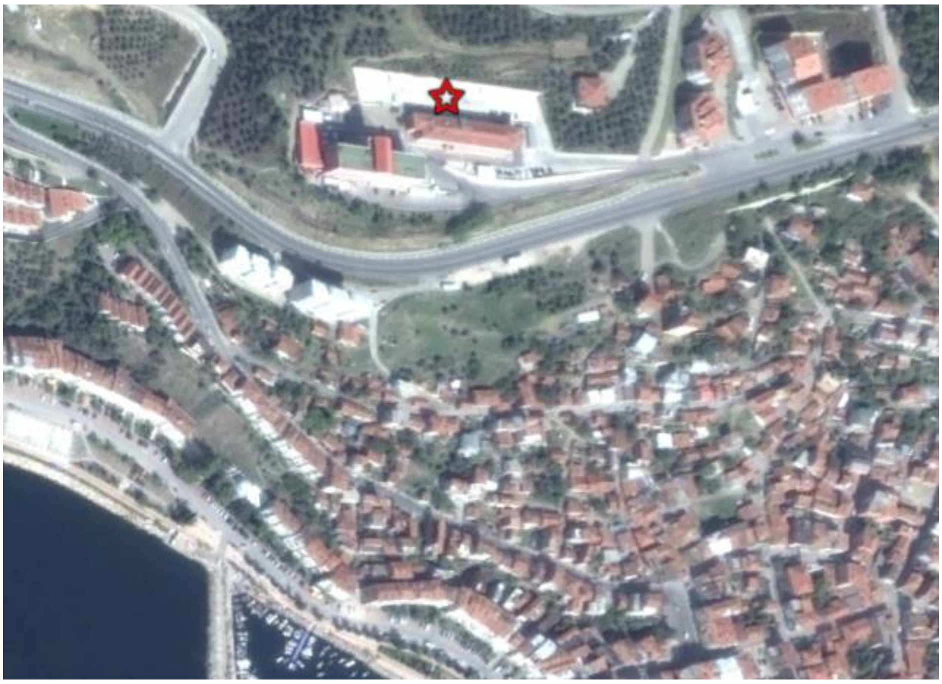

**Figure 3.2:** Location of the Automatic Meteorological Observation Station.

## 3.2 Autocorrelation Function (ACF) and Partial Autocorrelation (PACF) tests

In time series analysis, observed value at one point can be compared with observed value at one or more time points earlier. Such prior values are known as lagged values.The correlation between the original time series values and the corresponding k-lagged values is called autocorrelation of order k. The ACF provides the correlation between the serial correlation coefficients for consecutive lags. Figure 3.3 displays graphically the ACF and PACF. Autocorrelations for consecutive lags are formally dependent. If the first element is closely related to second, and the second to third, then the first element must also be somewhat related to the third one, etc. The serial dependencies can change considerably after removing the first order autocorrelation. By removing serial dependency, we can identify the hidden nature of seasonal dependencies in the time series and we can make the series stationary which is necessary for ARIMA and other techniques. Serial dependency for a particular lag of k can be removed by differencing the series, that is converting each element I of the series into its difference from the element i-k. Another useful method to examine serial dependencies is to examine the Partial Autocorrelation Function (PACF), an extension of autocorrelation where the dependence on the immediate elements (those with in the lag) is removed. For time series data, ACF and PACF measure the degree of relationship between observations k time periods, or lags, apart. These plots provide valuable information to help you identify an appropriate ARIMA model.

In a sense, the partial autocorrelation provides a cleaner picture of serial dependencies for individual lags [25].

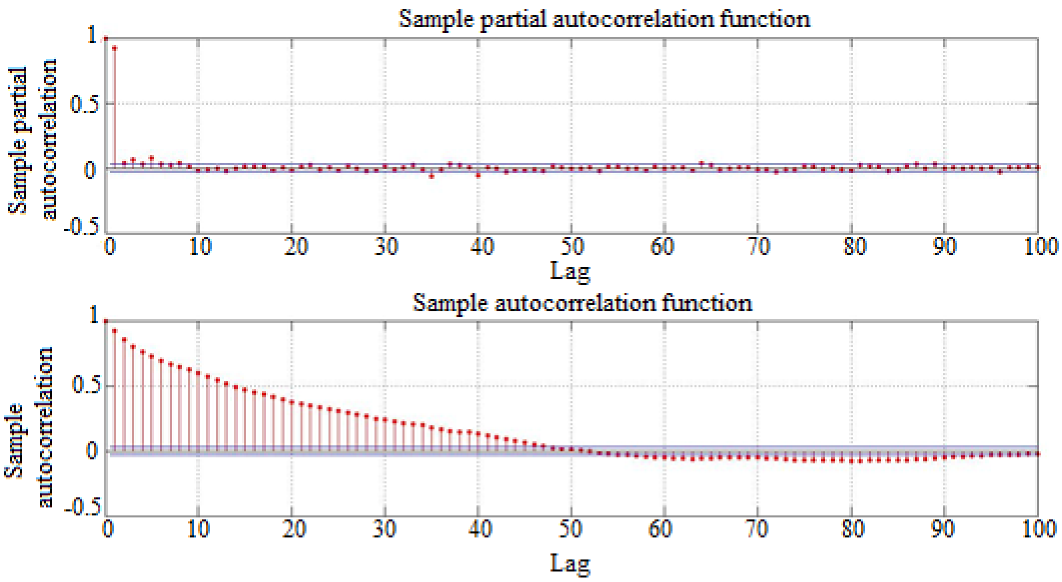


**Figure 3.3:** ACF and PACF plot of the data series.

In this wind speed measurements, time series data can be seen from Figure 3.3 that each observation is most similar (closest) to the adjacent observation (lag=1), also series do not follow any recurring seasonal pattern.

The PACF graph shows a large partial autocorrelations at lags 1.

### 3.3 AIC and BIC analysis

There are various criterion has been developed in order to select the optimal model order of linear prediction models. Commonly used Akaike information criteria, Bayesian information criteria and final prediction error are explained in Chapter 2. In order to see the best-fitted model to data series, AIC and BIC criterions are used for this study. FPE is not used in because FPE points the same order with AIC most of the time. In another word FPE value is nearly equal to the logarithm of AIC value. In order to calculate AIC and BIC values of AR, ARMA, ARIMA, ARX, ARMAX and ARIMAX MATLAB® 2013b with Econometrics Toolbox used and syntax *aicbic* is run. Here the results are given in Figure 3.4, 3.5, 3.6, 3.7, 3.8 and 3.9

respectively in the model order range of 1 to 75. As mentioned in the Section 2.8, optimal model order is selected where the Akaike and Bayesian information criterias take the minimum value.


**Figure 3.4:** AIC and BIC results of AR models from AR(1) to AR(75).

The optimal model results can be seen in Table 3.1 . Optimal models will be fitted to the data and corresponding results are presented in the Section 4 . Some of the information criteria values could not calculated beacuse of being noninvertible. These missing values can be seen in Figure 3.5, 3.6, 3.8 and 3.9.

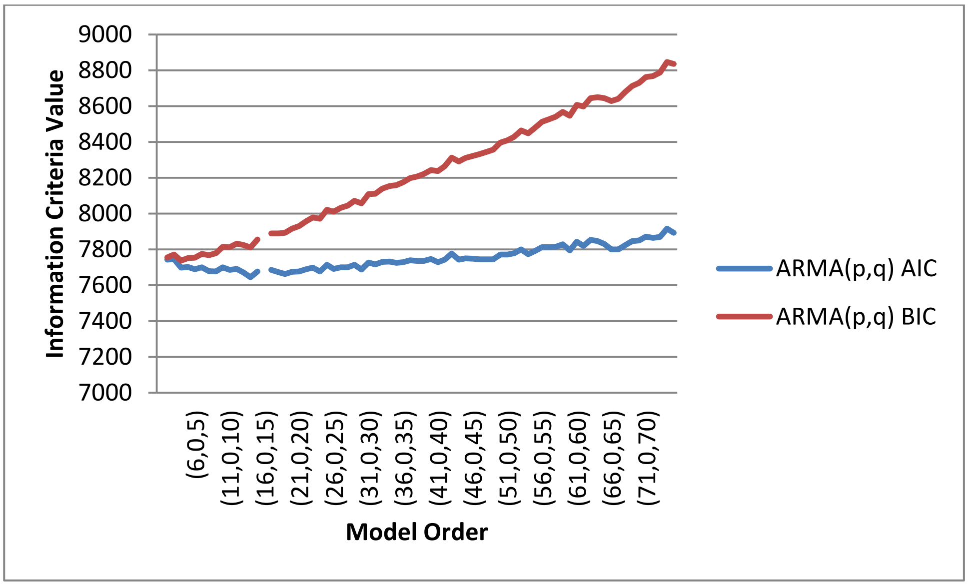


**Figure 3.5:** AIC and BIC results of ARMA models from ARMA(2,1) to ARMA(75,74).

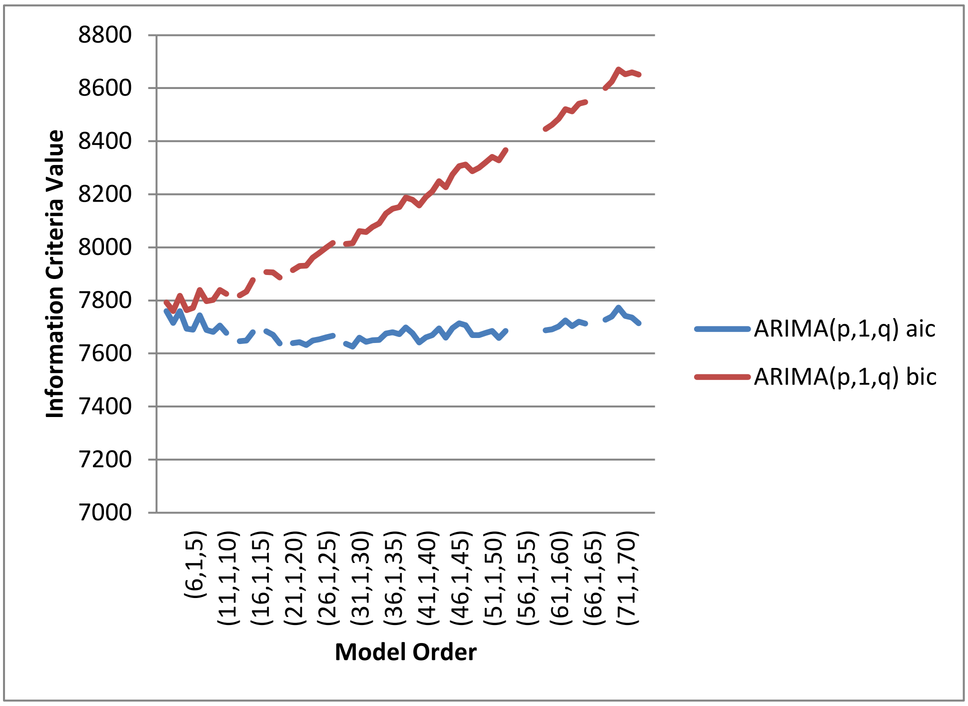


**Figure 3.6:** AIC and BIC results of ARIMA models from ARIMA(2,1,1) to ARIMA(75,1,74).

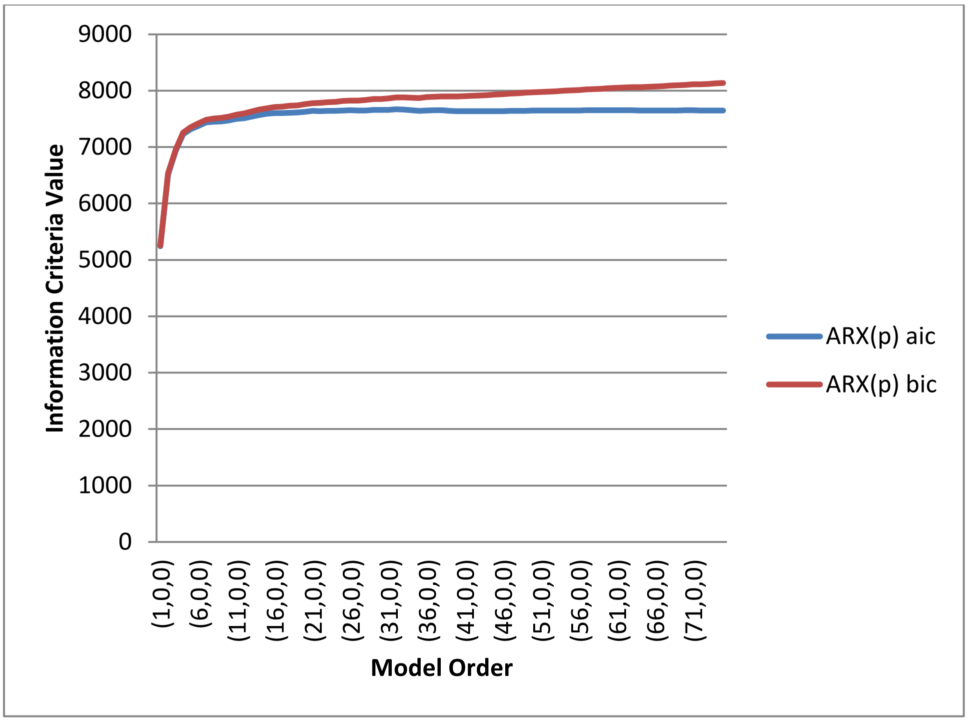


**Figure 3.7:** AIC and BIC results of ARX models from ARX(1) to ARX(75).

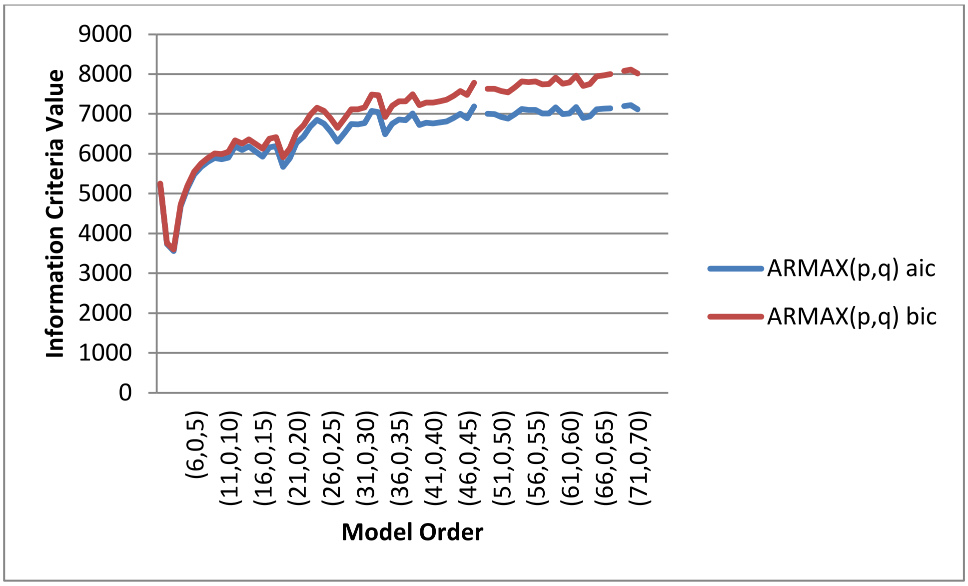


**Figure 3.8:** AIC and BIC results of ARMAX models from ARMAX(2,1) to ARMAX(75,74).

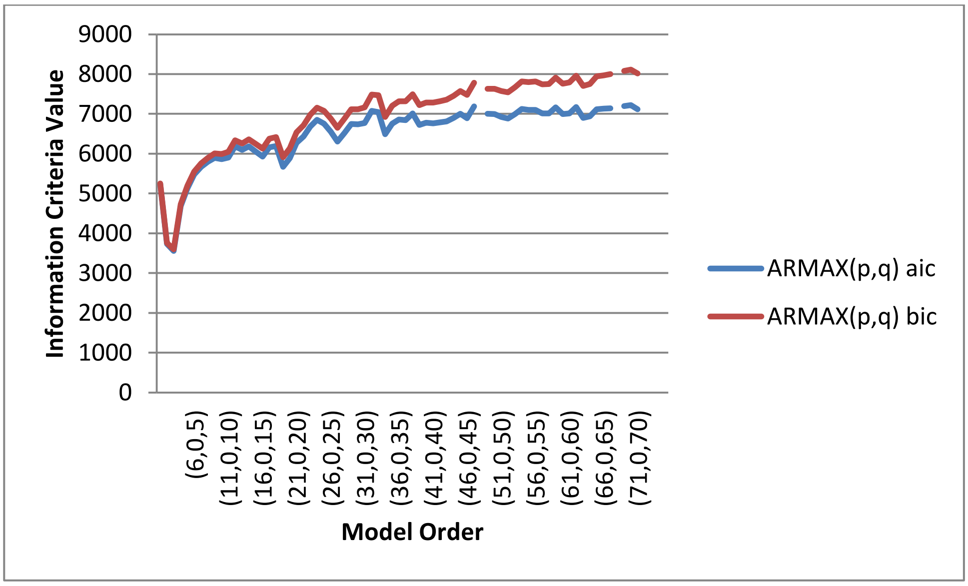


**Figure 3.9:** AIC and BIC results of ARIMAX model from ARIMAX(2,1,1) to ARIMAX(75,1,74)

**Table 3.1:** AIC and BIC optimal model order results.

| | AR | ARMA | ARIMA | ARX | ARMAX | ARIMAX |
|---|---|---|---|---|---|---|
| AIC | AR(8) | ARMA(14,13) | ARIMA(30,1,29) | ARX(1) | ARMAX(2,1) | ARIMAX(8,1,7) |
| BIC | AR(8) | ARMA(4,3) | ARIMA(3,1,2) | ARX(1) | ARMAX(2,1) | ARIMAX(8,1,7) |

# 4. RESULTS

## 4.1 AR Results

### 4.1.1 AR(8)

In this section, results of the one step ahead forecasting using the models that is selected by the optimal model order selection criterias in the previous section are plotted. Firts of all, autoregressive model with an order of eight AR(8) is performed for forecasting and goodness of forecasting can be seen in Figure 4.1 and Figure 4.2. In order to see the model fit better , data interval between 2000 and 2100 is plotted seperately. The red line represents the observed data and the black line represents the forecasted data.

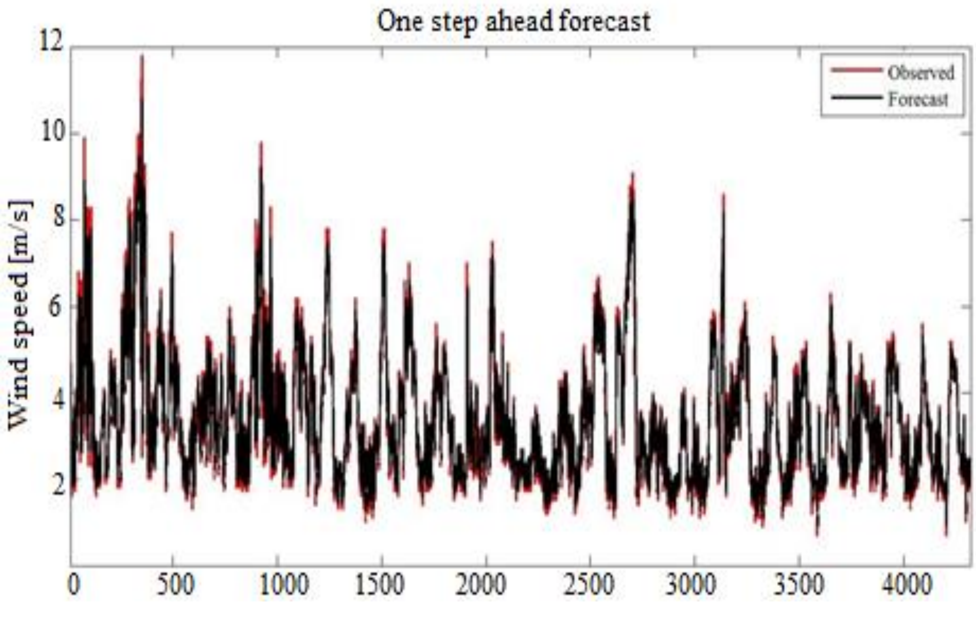


**Figure 4.1:** AR(8) Observed and Forecasted Datas from 9 to 4320.

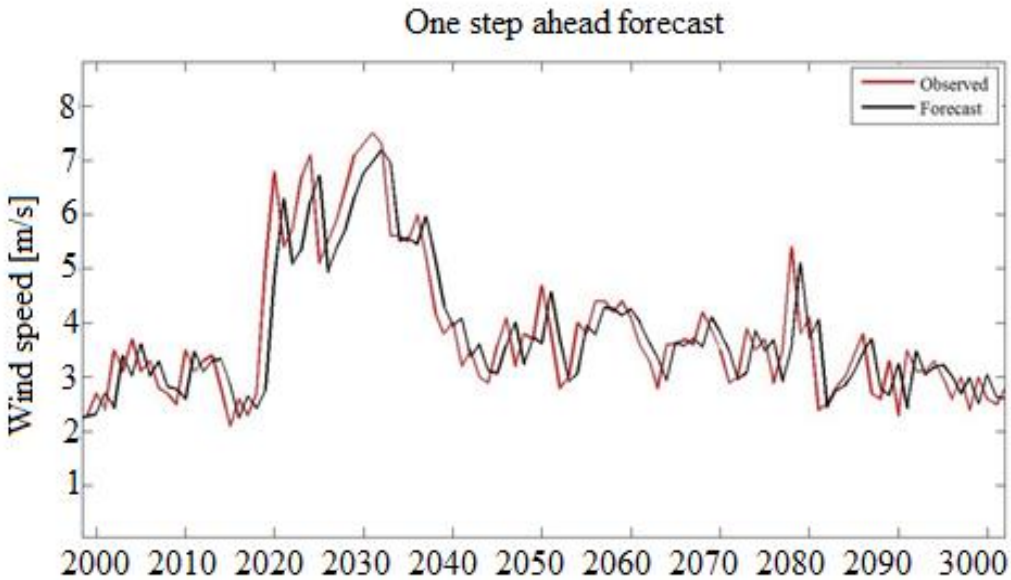


**Figure 4.2:** AR(8) Observed and Forecasted Datas from 2000 to 2100.

Residuals which mean the difference between forecasted value and actual value are calculated . In order to assess model fit by examining residuals, Figure4.3 and Figure 4.4 are plotted. Smooth residuals graph means a good model fit.

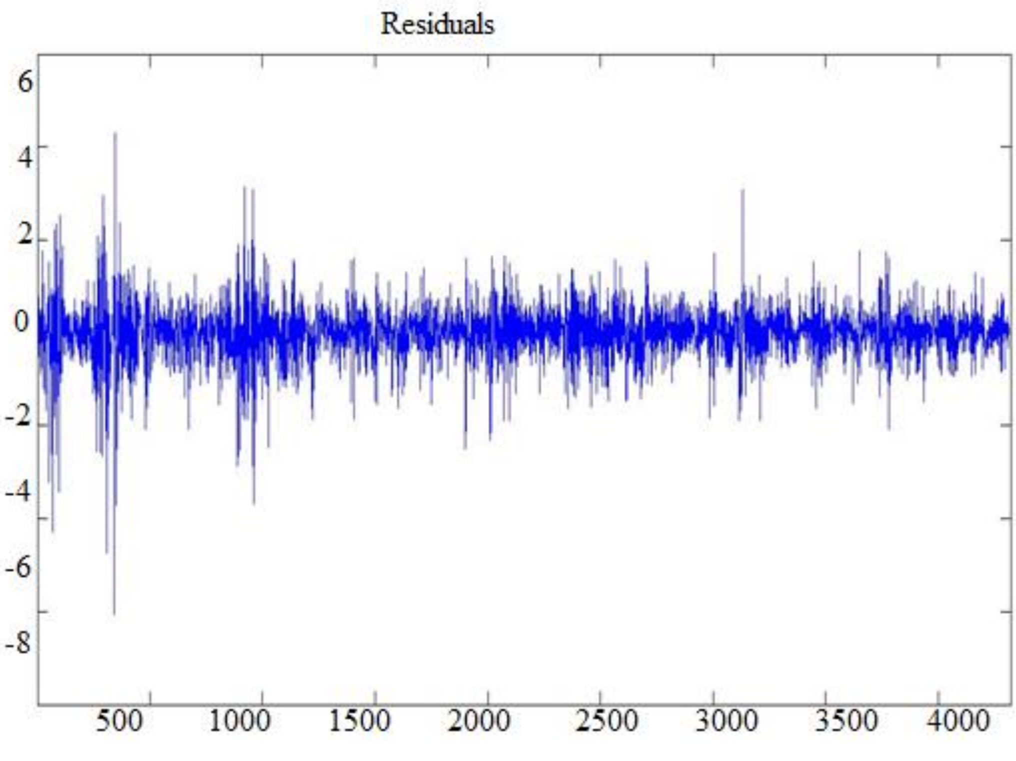


**Figure 4.3:** AR(8) Residuals.

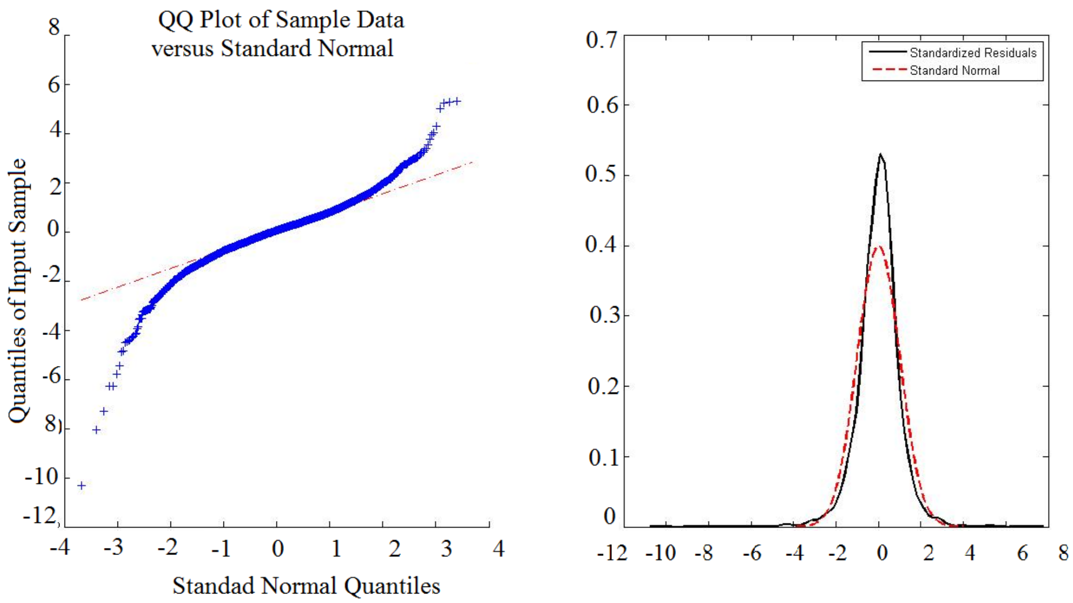


**Figure 4.4:** AR(8) Plot and Normality Plot of Standart Residuals.

## 4.2 ARMA Results

### 4.2.1 ARMA(4,3)

In another case, ARMA(4,3) and ARMA(14,13) are performed. Depending on the results, actual data versus forecasted data graphs are plotted. Goodness of forecasting can be seen in Figure 4.5 and Figure 4.9 respectively. More detailed graph also created. They can be seen in Figure 4.6 and Figure 4.10 respectively.

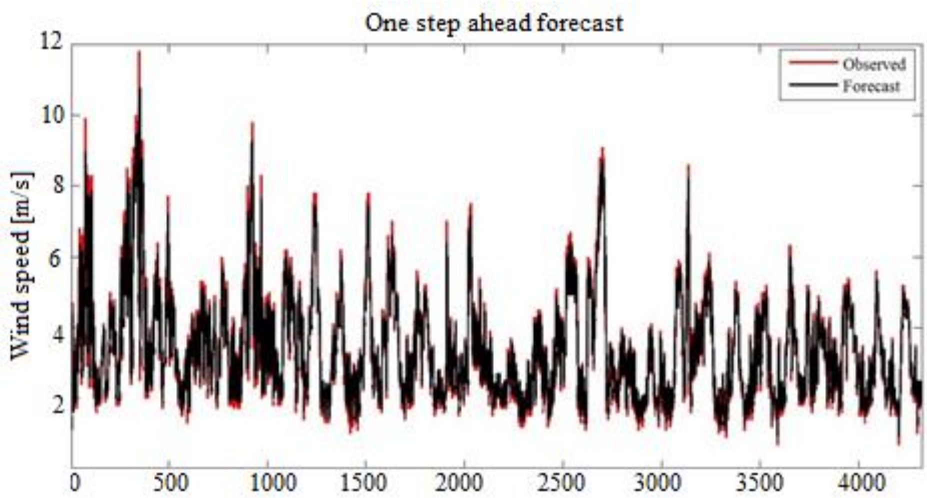


**Figure 4.5:** ARMA(4,3) Observed and Forecasted Datas from 5 to 4320.

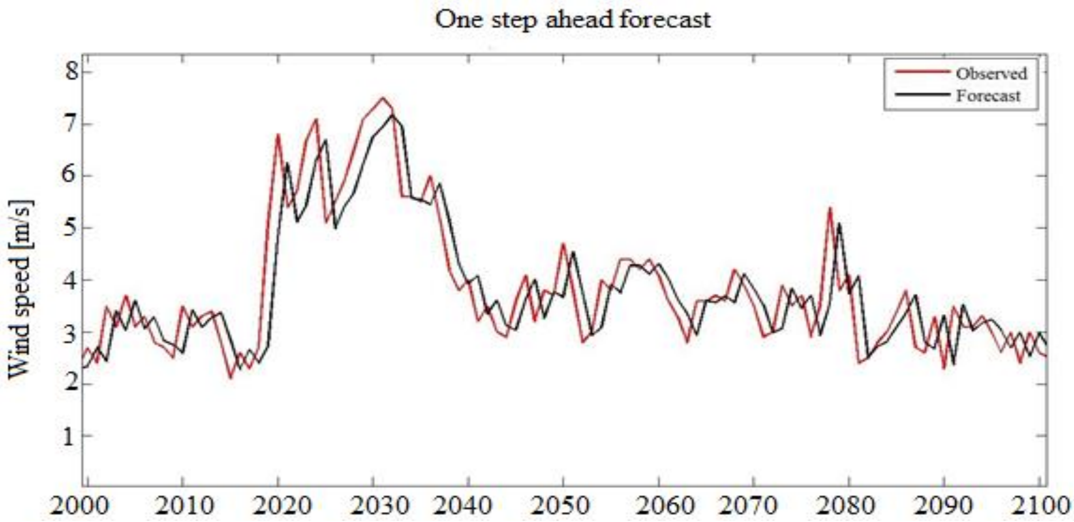


**Figure 4.6:** ARMA(4,3) Observed and Forecasted Datas from 2000 to 2100.

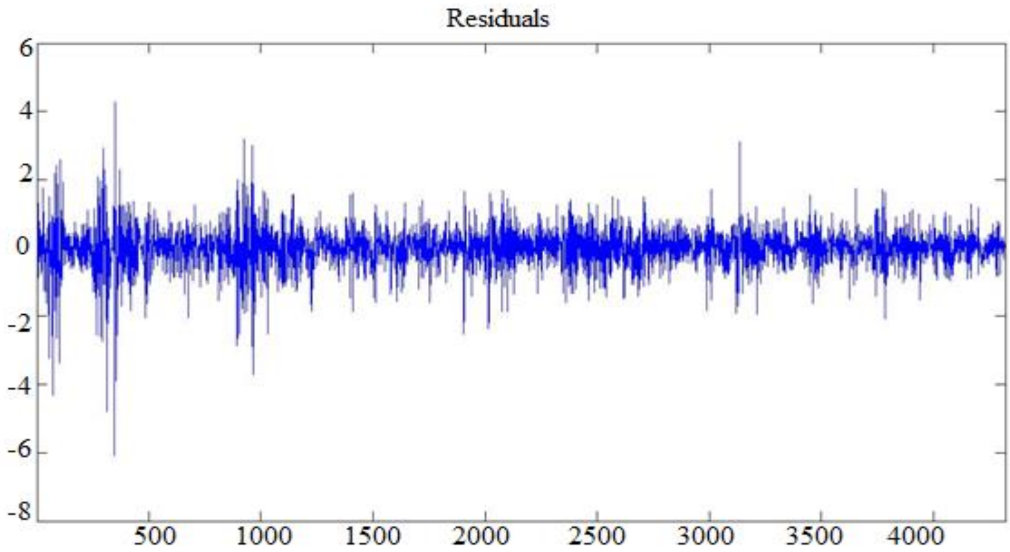


**Figure 4.7:** ARMA(4,3) Residuals.

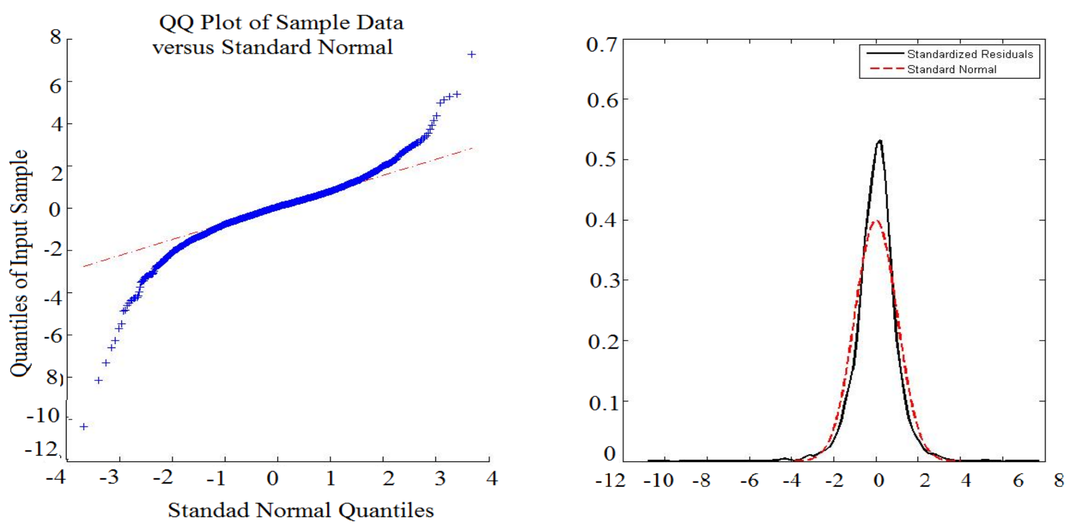


**Figure 4.8:** ARMA(4,3) Q-Q Plot and Normality Plot of Standart Residuals.

### 4.2.2 ARMA(14,13)

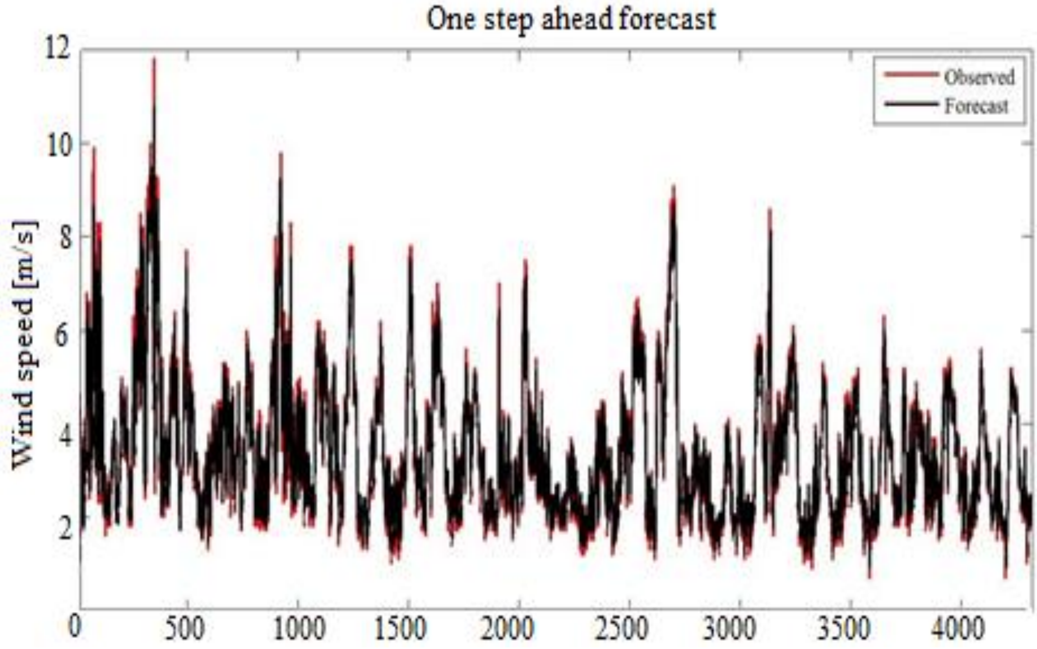


**Figure 4.9:** ARMA(14,13) Observed and Forecasted Datas from 15 to 4320.

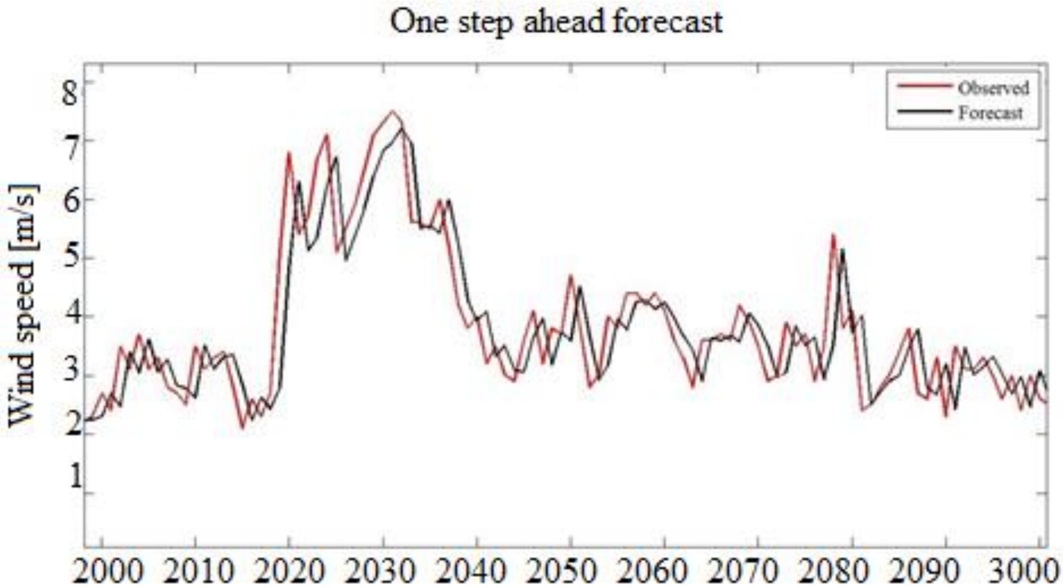


**Figure 4.10:** ARMA(14,13) Observed and Forecasted Datas from 2000 to 2100.

Goodness of forceasting graph for ARMA(4,3) and ARMA (14,13) seem identical. Moreover, residuals analysis appear similar. In this stiuation, numerical results of comparison parameters which are presented in the Section 4.7 have to be compared to see the performance of forecasting of a model with respect to the others.

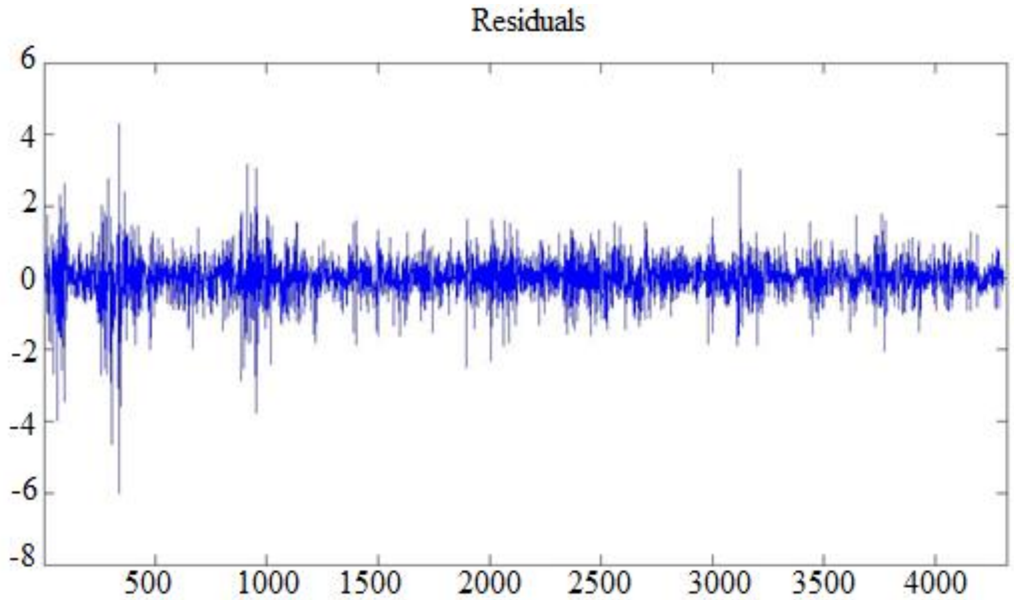


**Figure 4.11:** ARMA(14,13) Residuals.

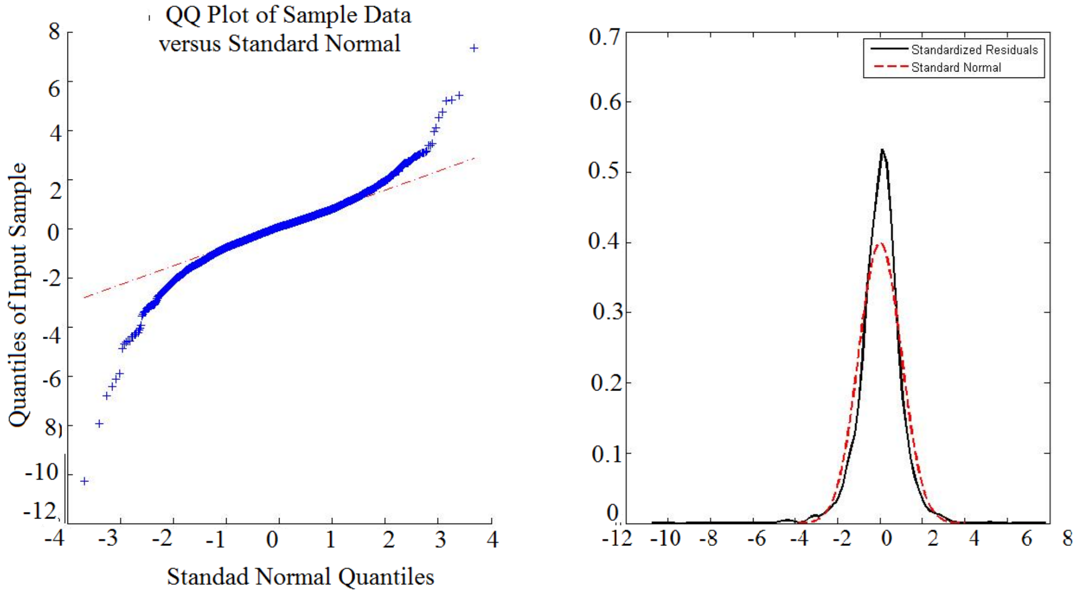


**Figure 4.12:** ARMA(14,13) Q-Q Plot and Normality Plot of Standart Residuals.

### 4.3 ARIMA Result

#### 4.3.1 ARIMA (3,1,2)

As presented in the Section 3.3, two different autoregressive integrated moving average model are presented as an optimal model depending on the model order selection criterias. These models are ARIMA (3,1,2) and ARIMA (30,1,29). Depending on the results, actual data versus forecasted data graphs are plotted.

Goodness of forecasting can be seen in Figure 4.13 and Figure 4.17 respectively. More detailed graph also created. They can be seen in Figure 4.14 and Figure 4.18 respectively.

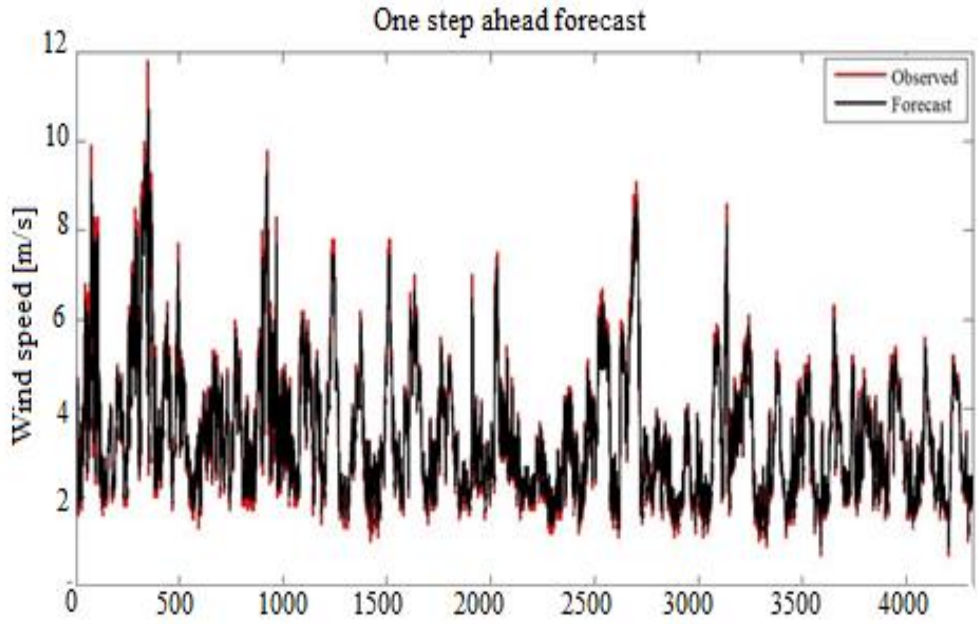


**Figure 4.13:** ARIMA(3,1,2) Observed and Forecasted Datas from 5 to 4320.

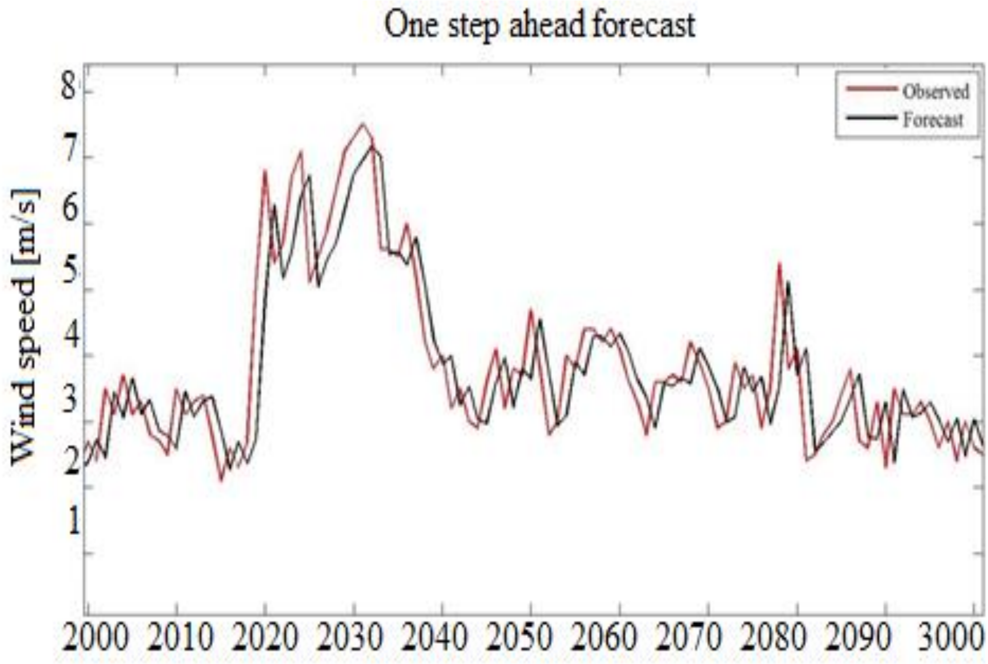


**Figure 4.14:** ARIMA(3,1,2) Observed and Forecasted Datas from 2000 to 2100.

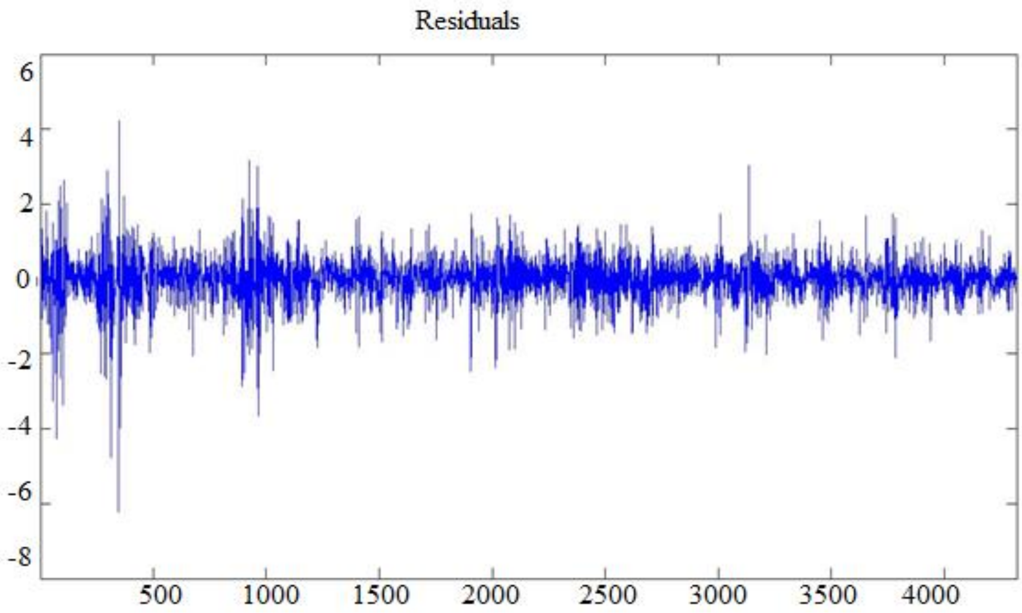


**Figure 4.15:** ARIMA(3,1,2) Residuals.

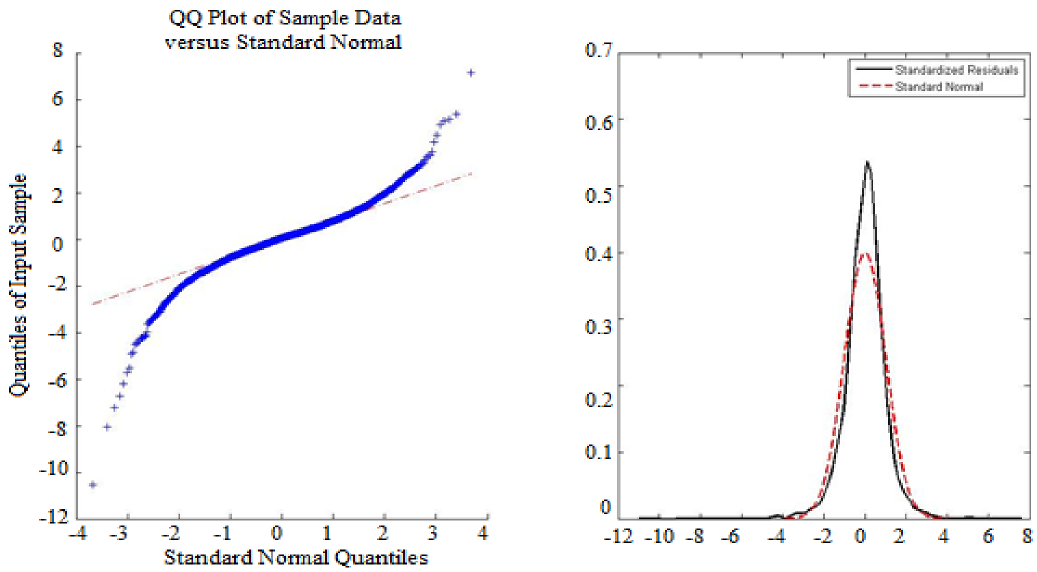


**Figure 4.16:** ARIMA(3,1,2) Q-Q Plot and Normality Plot of Standart Residuals.

### 4.3.2 ARIMA (30,1,29)

In addiditon to previous case ARMA, residuals analysis appear similar for ARIMA CASE too. Numerical results in the Section 4.7 have to be compared to see the performance of forecasting of an ARIMA model in itself and with respect to the other models. Residulas graph for ARIMA models are presented in Figure 4.15, 4.16, 4.19 and Figure 4.20.

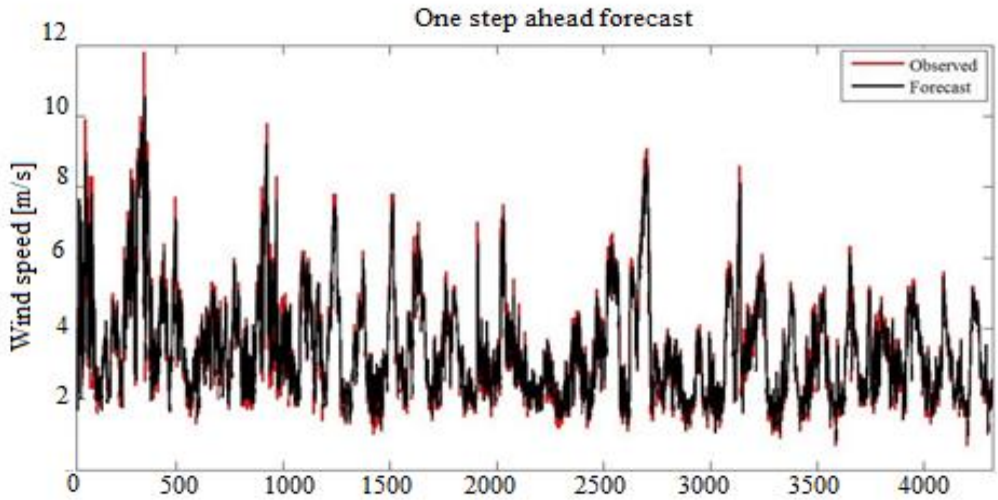


**Figure 4.17:** ARIMA(30,1,29) Observed and Forecasted Datas from 32 to 4320.

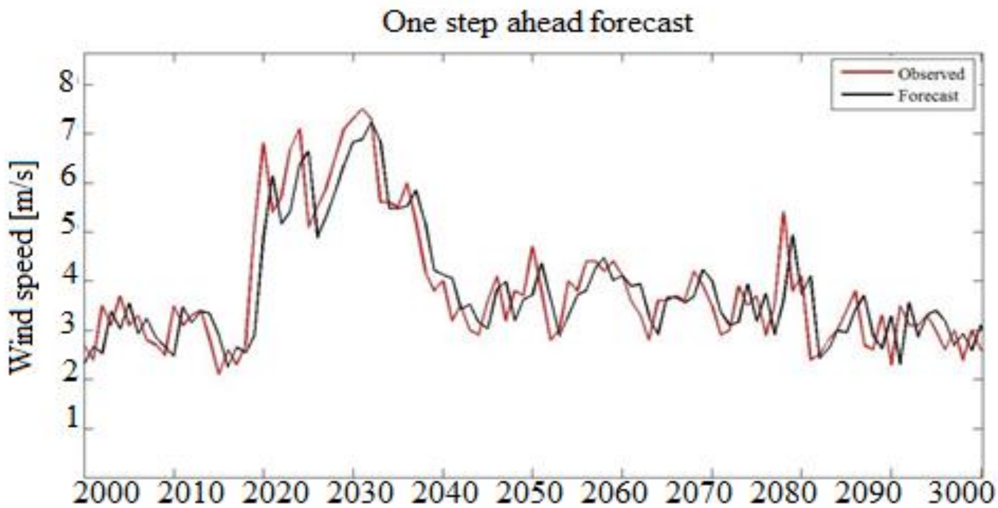


**Figure 4.18:** ARIMA(30,1,29) Observed and Forecasted Datas from 2000 to 2100.

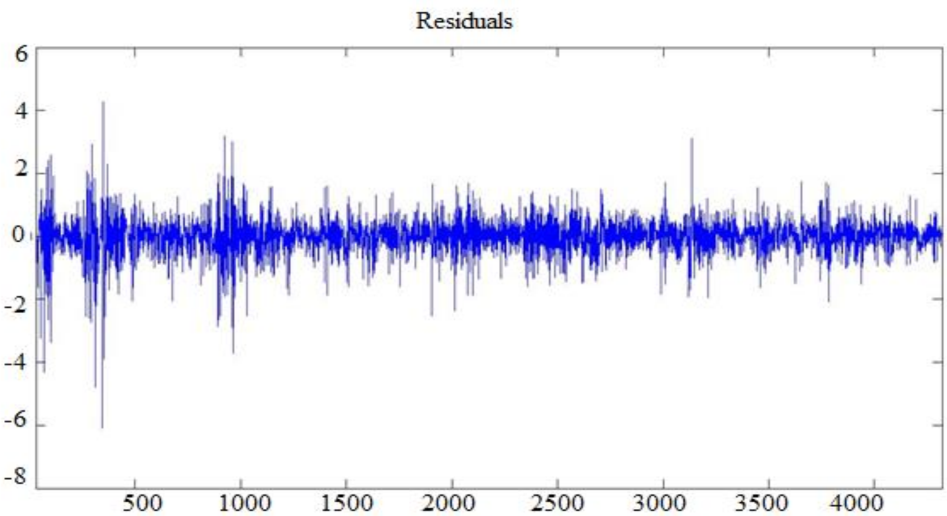


**Figure 4.19:** ARIMA(30,1,29) Residuals.

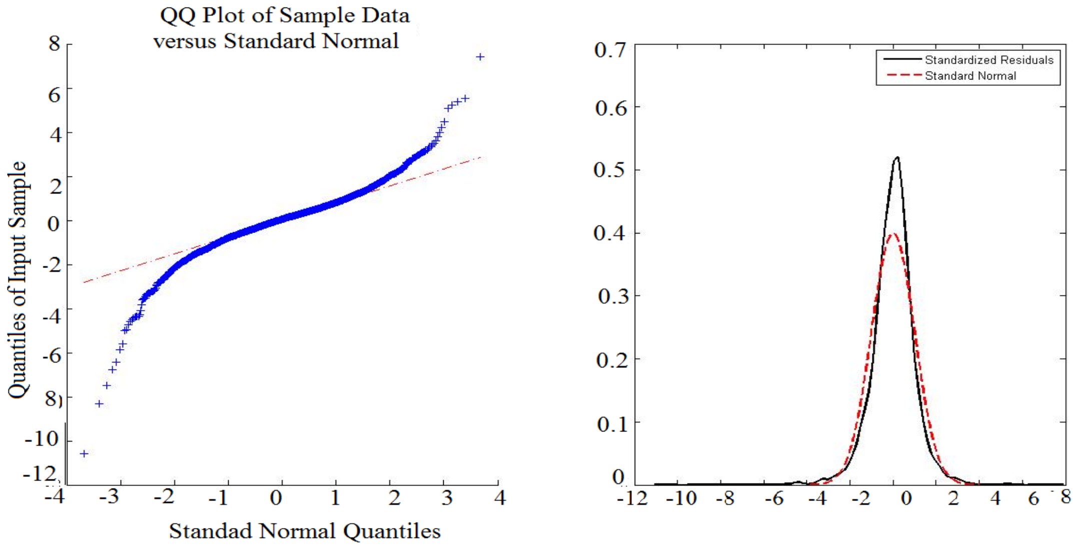


**Figure 4.20:** ARIMA(30,1,29) Q-Q Plot and Normality Plot of Standart Residuals.

### 4.4 ARX Results

#### 4.4.1 ARX(1)

One step ahead forecasting performance of autoregressive with exogenous input model presented in Figure 4.21, 4.22, 4.23 and 4.24 . Success of ARX(1) model fit corresponding to the other models can be seen clearly by examining the goodness of fit graph which is presented in Figure 4.21 and Figure 4.22 . Black line follows red line quite good. Residuals also seem more smother, quantile and density graphs are almost in an ideal shape.

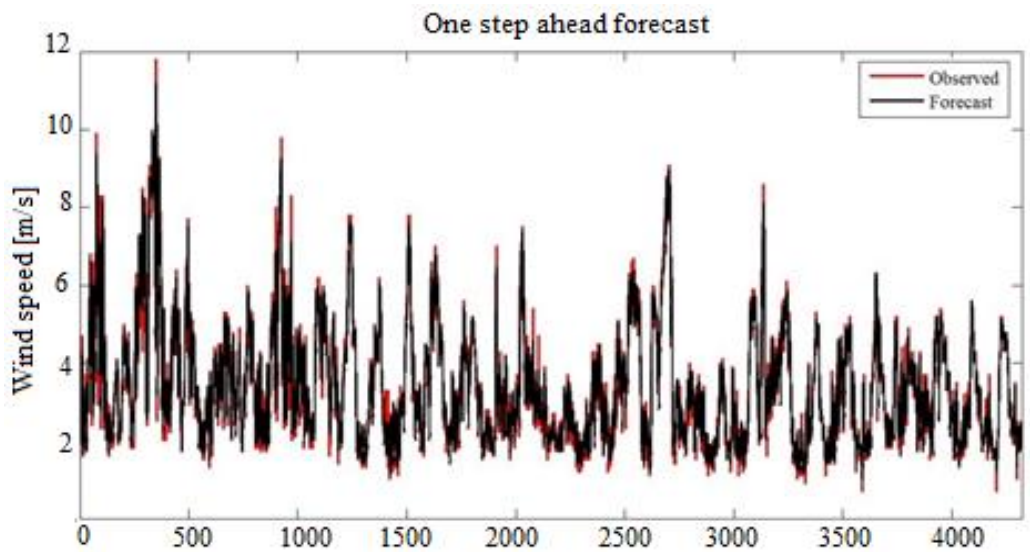


**Figure 4.21:** ARX(1) Observed and Forecasted Datas from 2 to 4320.

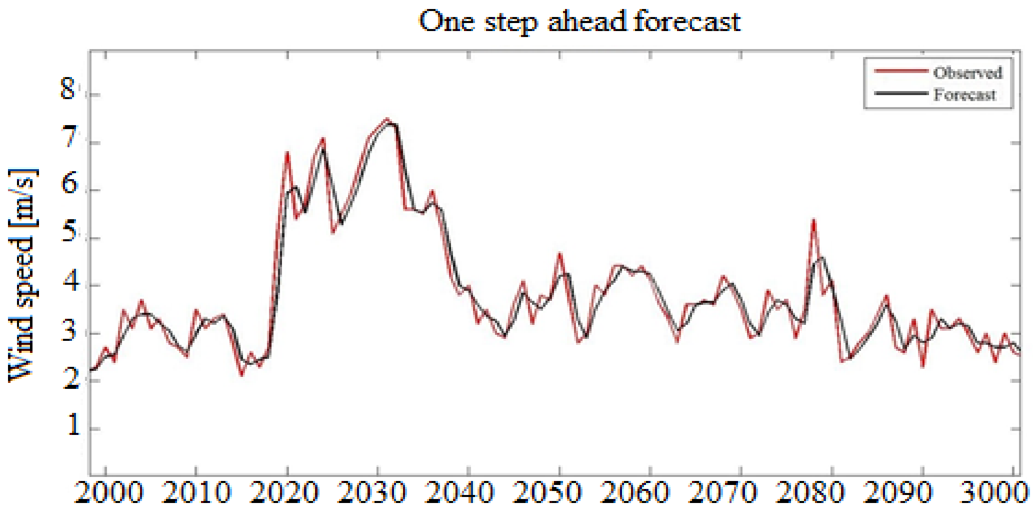


**Figure 4.22:** ARX(1) Observed and Forecasted Datas from 2000 to 2100.

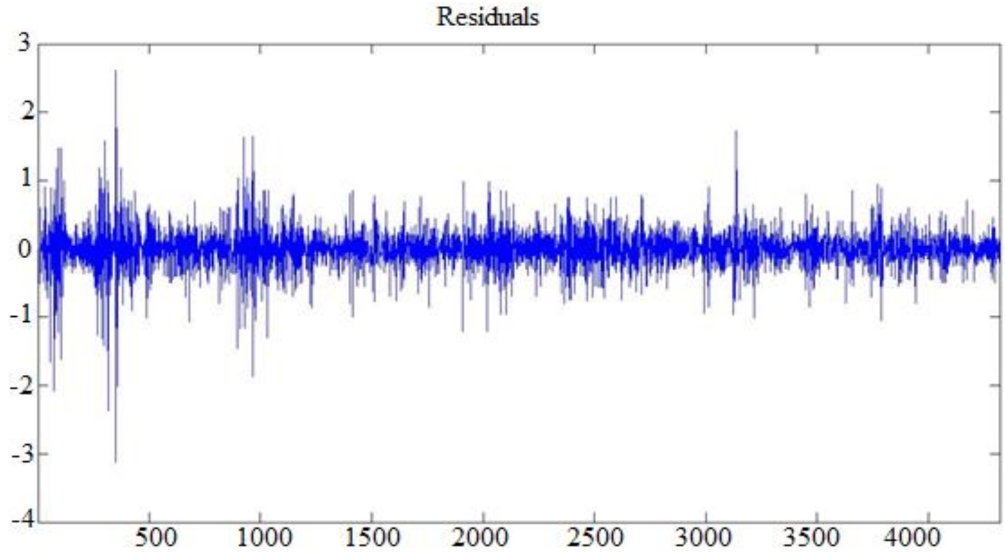


**Figure 4.23:** ARX(1) Residuals.

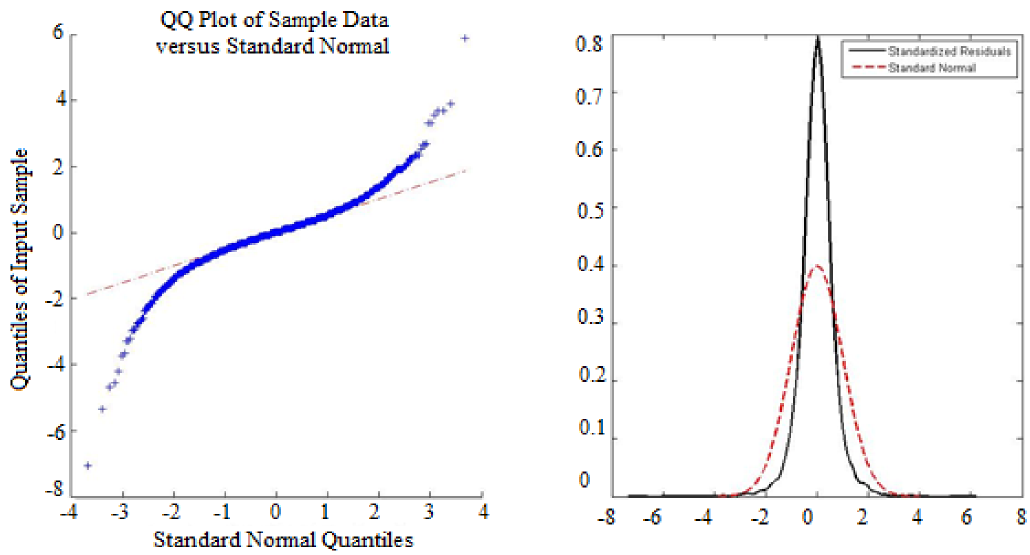


**Figure 4.24:** ARX(1) Q-Q Plot and Normality Plot of Standart Residuals.

### 4.5 ARMAX Result

#### 4.5.1 ARMAX(2,1)

In the autoregressive moving average with exogenous input model case, one step ahead forecasting performance which are presented in Figure 4.25, 4.26, 4.27 and 4.28 is less success comparing the AR, ARMA ARIMA and ARX model results. ARMAX (2,1) model fit can be examined in the goodness of fit graph which is presented in Figure 4.21 and Figure 4.22 . Red is more apparent in the big picture. Residuals also prove the failure of this model.

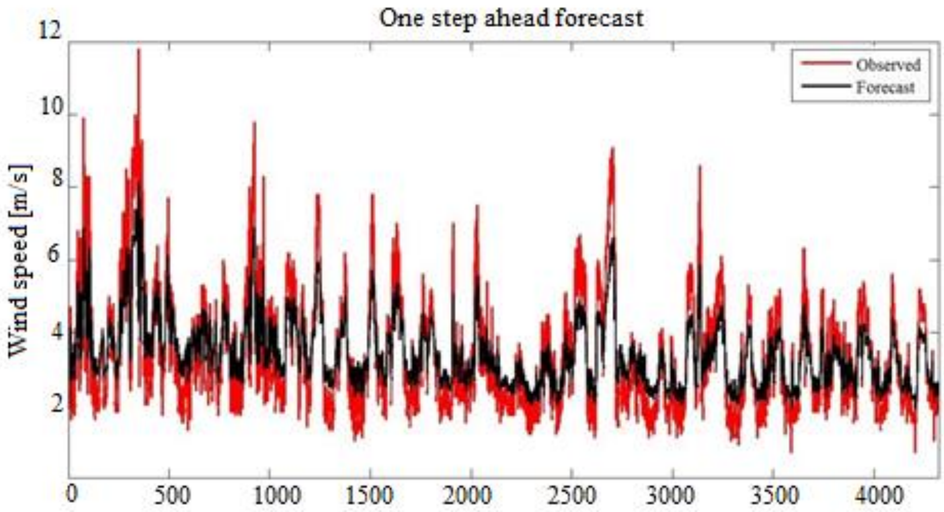


**Figure 4.25:** ARMAX(2,1) Observed and Forecasted Datas from 3 to 4320.

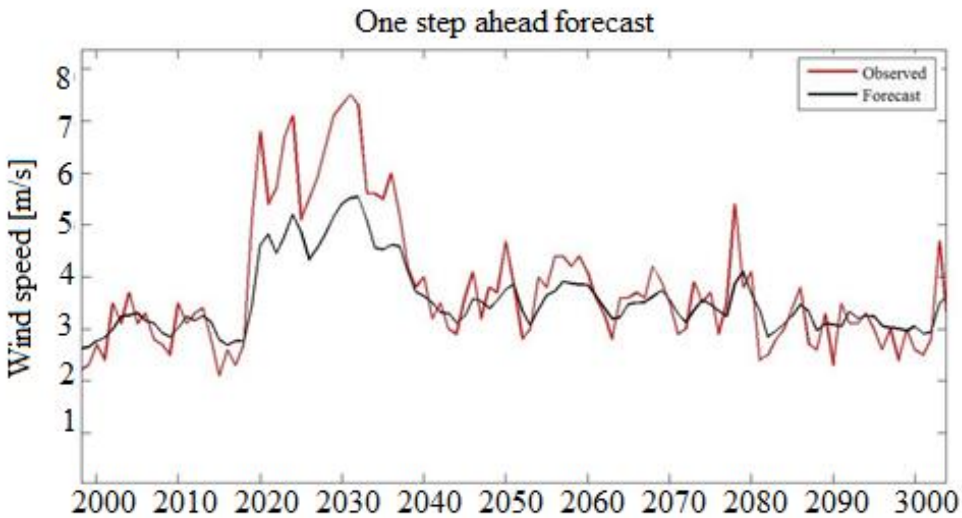


**Figure 4.26:** ARMAX(2,1) Observed and Forecasted Datas from 2000 to 2100.

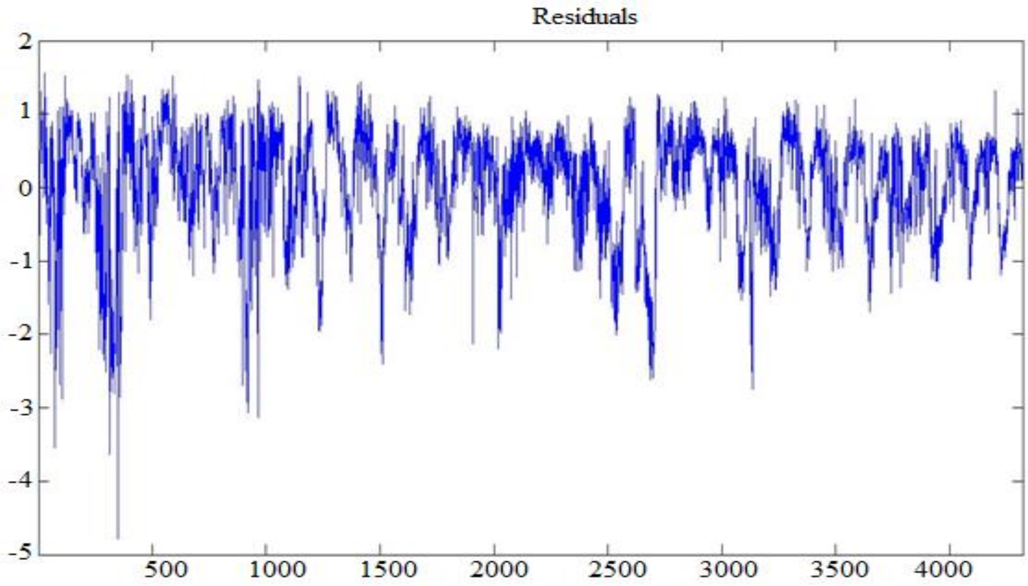


**Figure 4.27:** ARMAX(2,1) Residuals.

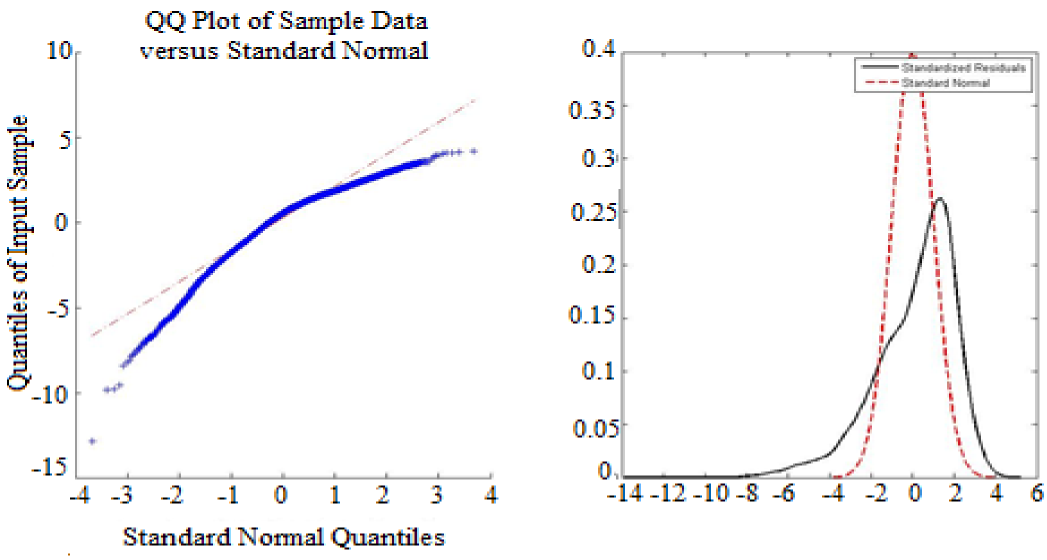


**Figure 4.28:** ARMAX(2,1) Q-Q Plot and Normality Plot of Standart Residuals.

## 4.6 ARIMAX Result

### 4.6.1 ARIMAX(8,1,7)

In the last case, one step ahead forecasting performance of ARIMAX (8,1,7) is unexpected. Model fit can be examined in the goodness of fit graph which is presented in Figure 4.29 and Figure 4.30. Black line is represents the forecasted values. Black line is far away from following red line and it is simply fluctuating. Residuals also prove the failure of this model. These results prove the failure of ARIMAX (8,1,7) model.

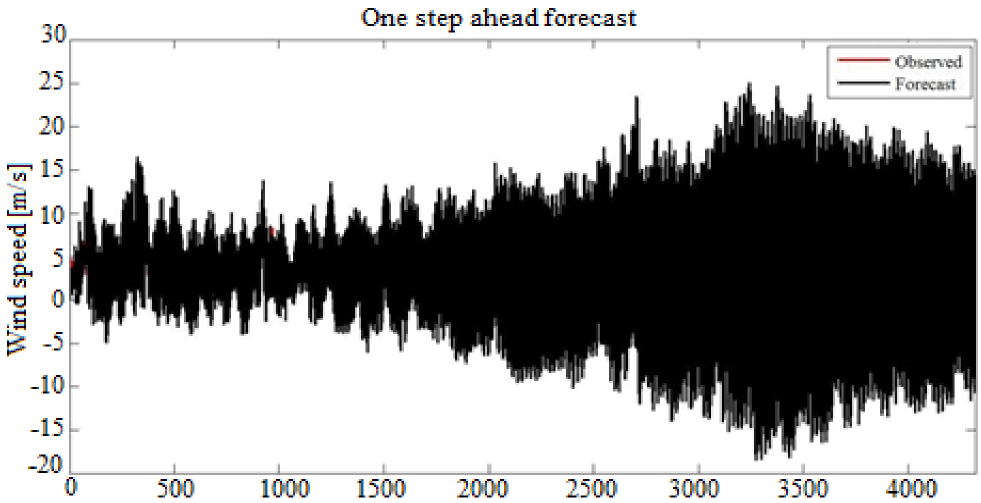


**Figure 4.29:** ARIMAX(8,1,7) Observed and Forecasted Datas from 3 to 4320.

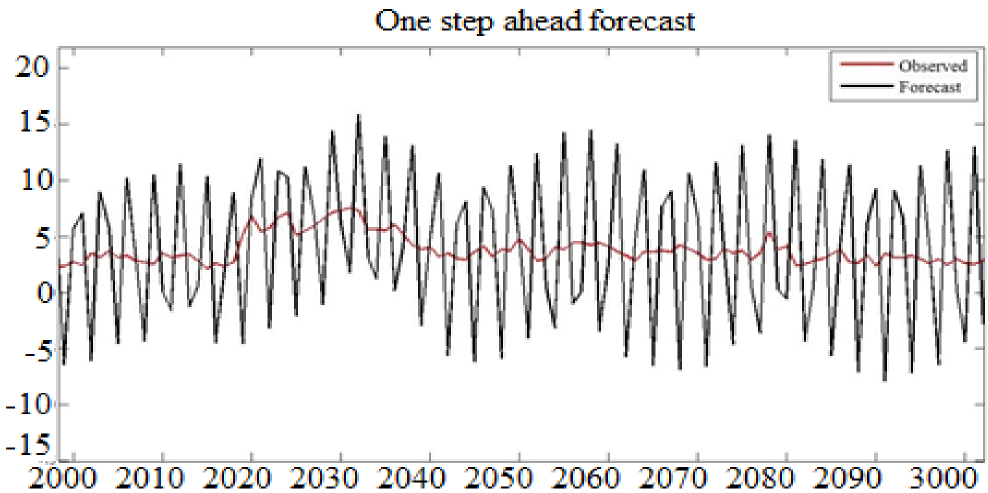


**Figure 4.30:** ARIMAX(8,1,7) Observed and Forecasted Datas from 2000 to 2100.

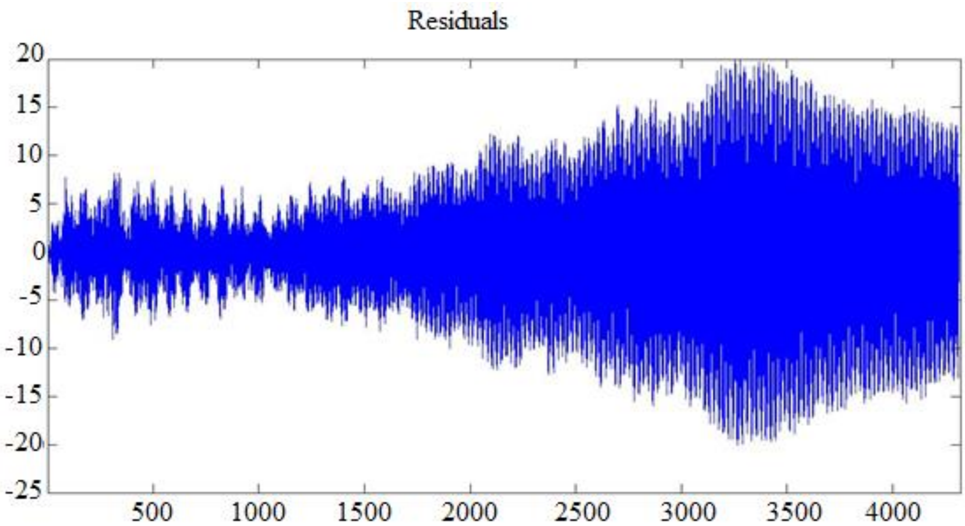


**Figure 4.31:** ARIMAX(8,1,7) Residuals.

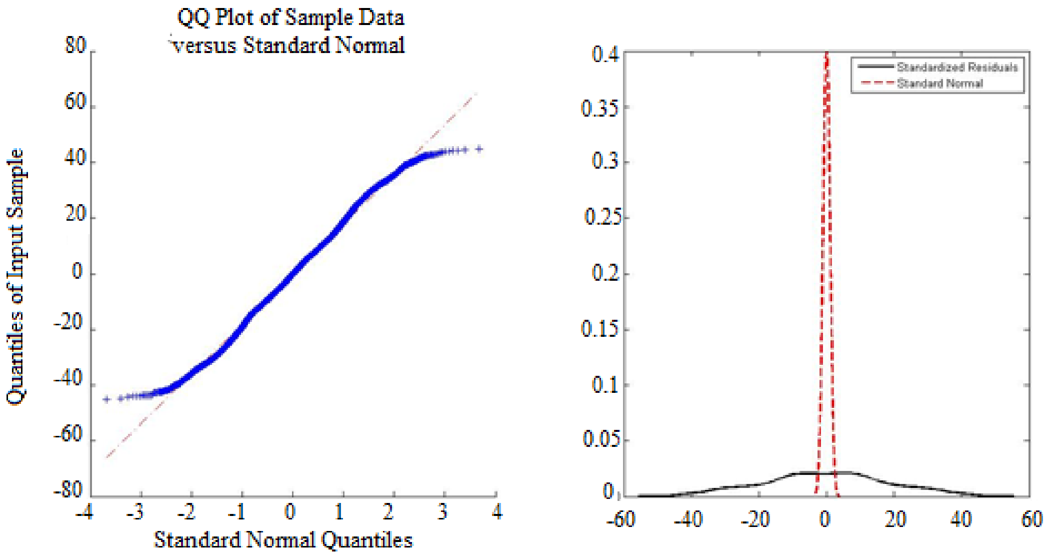


**Figure 4.32:** ARIMAX(8,1,7) Q-Q Plot and Normality Plot of Standart Residuals.

### 4.7 Comparison of the Results

Forecasting performance of the selected ARIMA/ARIMAX model have been evaluated on the basis of one-step ahead forecasts against the parameters MAE (mean absolute error), RMSE (root mean square error). In each case, Akaike information criterion (AIC) and Bayesian information criterion (BIC) have selected optimal model see Table 4.1. For the Autoregressive (AR) model, criterions have selected the AR(8). The MAE and RMSE values are 0.413804 and 0.587939 respectively. $R^2$statistic is 0.973441 see Table 4.2.

**Table 4.1:** Different ARIMAX model results.

| Model | R2 ($R^2$) | MAE | RMSE |
|---|---|---|---|
| ARIMAX(1,1,0) | 0.974308 | 0.411985 | 0.590469 |
| ARIMAX(2,1,1) | 0.973341 | 0.420488 | 0.599816 |
| ARIMAX(20,1,19) | 0.943843 | 0.663519 | 0.887354 |
| ARIMAX(9,1,8) | 0.954528 | 0.637523 | 0.799486 |
| ARIMAX(7,1,6) | 0.933439 | 0.755664 | 0.980566 |

**Table 4.2:** Comparison of the results.

| Model | R2 ($R^2$) | MAE | RMSE |
|---|---|---|---|
| ARIMA(30,1,29) | 0.973356 | 0.413699 | 0.58868 |
| ARIMA(3,1,2) | 0.973277 | 0.414104 | 0.589653 |
| ARMA(14,13) | 0.973712 | 0.41337 | 0.585437 |
| ARMA(4,3) | 0.973356 | 0.413699 | 0.58868 |
| AR(8) | 0.973441 | 0.413804 | 0.587939 |
| ARIMAX(8,1,7) | 0.163349 | 6.577232 | 8.143865 |
| ARMAX(2,0,1) | 0.956163 | 0.576348 | 0.728387 |
| ARX(1) | 0.993089 | 0.209998 | 0.30246 |

For autoregressive moving average (ARMA) case AIC and BIC pointed the different models, Akaike information criteria has selected the model ARMA(14,13) and Bayesian information criteria has selected the model ARMA(4,3). For ARMA(14,13), the MAE and RMSE values are 0.41337 and 0.585437 respectively. $R^2$statistic is 0.973712. In case of ARMA (4,3) MAE and RMSE values which are 0.413699 and 0.58868 a few higher. $R^2$statistic is 0.973356 and lower than higher order model statistics value. In ARIMA case AIC and BIC values have also selected the different models. AIC has selected the higher order ARIMA(30,1,29) again and BIC selected the lower order ARIMA(3,1,2). MAE and RMSE values for the ARIMA(30,1,29) are 0.413699 and 0.58868. On the other hand ARIMA(3,1,2) results are higher that is 0.414104 and 0.589653 respectively. $R^2$statistic for the ARIMA(3,1,2) is 0.973277 and for ARIMA(30,1,29), it is 0.973356.

In cases of model with exogenous input, AIC and BIC have selected the same model orders. For ARX, criterions have selected ARX (1). The MAE and RMSE values are 0.209998 and 0.30246 respectively. $R^2$statistic is 0.993089. These are the lowest values. For ARMAX, criterions have selected ARMAX (2,0,1). The MAE and RMSE values are 0,576348 and 0,728387respectively. $R^2$statistic is 0.956163 . For ARIMAX, criterions have selected ARMAX (8,1,7). The MAE and RMSE values are 6.576565 and 8.143116 respectively. $R^2$statistic is 0.163309.

# 5. CONCLUSION

The present thesis has successfully applied AR, ARMA, ARIMA, ARX, ARMAX and ARIMAX modelling procedure on wind speed measurements and found satisfactory results. Two of the information-criterions (AIC and BIC ) are evaluated to choose correct orders of AR(p), ARMA(p,q), ARIMA(p,d,q), ARX(p), ARMAX(p,q) and ARIMAX(p,d,q) models. By examining Figure 4.10 to Figure 4.41 and Table 4.1 , it can be seen that regarding to the wind speed, the AR, ARMA and ARIMA models have almost identical forecasting performance. Their MAE and MAPE values are very close. On the other hand, ARMAX models have almost identical forecasting performance but it is worse comparing AR, ARMA and ARIMA results. ARX model performs the best. It has the lowest mean absolute error MAE and root mean square error RMSE values and highest goodness of fit statistics. ARIMAX model forecasting performance is the worst. Various ARIMAX models forecasting performans can be seen in Table 4.1. Other ARIMAX model performs quite better than ARIMAX(8,1,7), It can be seen that second best model is ARIMAX(1,1,0) at this point that proves AIC or BIC are not sufficient selecting the optimal model order of autoregressive integrated moving average with exogenous input (ARIMAX). As a result ARX(1) and ARIMAX(1,1,0) procedures in the present thesis has worked quite good in forecasting wind measurements and can be effectively utilized for the electricity production of a wind power plant or achieving the best maintain schedule etc. In addition to this, these models can be used for longer-term wind speed forecasting and checked the goodness of fit.

## REFERENCES


[1] **El-Fouly, T. H. M., El-Saadany, E. F., Salama, M. M. A.** (2008). One day ahead prediction of wind speed and direction. *IEEE Trans. Energy Convers., vol. 23*, pp. 191–201.

[2] **Abdelaziz, A. Y., Rahman, M. A., El-Khayat, M. M., Hakim, M. A.** (2013). Short Term Wind Power Forecasting Using Autoregressive Integrated Moving Average Approach. *Journal of Energy and Power Engineering, 7*, 2089-2095, Cairo, Egypt.

[3] **Soman, S.S., Zareipour, H., and Malik, O.** (2010). A review of wind power and wind speed forecasting methods with different time horizons. *In: IEEE Conference*, 1-8.

[4] **Giebel, G., Brownsword, R., Kariniotakis, G., Denhard, M., Draxl, C.** (2011). *The State of the Art in Short-Term Prediction of Wind Power*, A Literature Overview, 2nd edition.

[5] **Monteiro, C., Bessa, R., Miranda, V., Botterud, A., Wang, J., Conzelmann, G.** (2009). *Wind Power Forecasting: State-of-the-Art*. Argonne National Laboratory Report, **ANL/DIS-10-1**.

[6] **Contreras J., Espinolar R., Nogales, F.J., and Conejo, A.J.** (2003). ARIMA Models to Predict Next-Day Electricity Prices. *IEEE Trans. Power Syst.*, ***18***, 1014-1020

[7] **Landberg, L.** (2009). Short-term prediction of the power production from wind farms, *Journal of Wind Engineering and Industrial Aerodynamics*, ***80***, 207-220.

[8] **Vaidyanathan, P. P.** (2008). *The Theory of Linear Prediction*, Morgan & Claypool Publishers.

[9] **Gerolimetto, M.** (2010). *ARIMA and SARIMA models (lecture notes)*. Addresses: http://www.dst.unive.it/~margherita/TSLectureNotes6.pdf.

[10] **Ljung, L.** (1999). *System Identification: Theory for the User*, $2^{nd}$ edition, Prentice-Hall, Inc.

[11] **Box, G. E. P., and Jenkins, J. M.** (1976). *Time Series Analysis: Forecasting and Control*. Holden-Day, San Francisco, CA.

[12] **Kallas, M., Honeine, P., Francis, C., and Amoud, H.** (2013). Kernel autoregressive models using yule-walker equations. Signal Processing, *Signal Proceeesing*, ***93***, 3053–3061.

[13] **Haykins, S.** (1996). *Adaptive Filter Theory*. Prentice Hall,Inc., Englewood Cliffs, NJ, Third edition, ISBN 0-13-322760-x.

[14] **Tharateeraparb, W.** (2008). *EE635: Digital Signal Processing II*. Computer Assignment-1, University of New Haven.

[15] **Brockwell P.J. and R.A. Davis.** (2002). *Introduction to Time Series and Forecasting*, 2nd Edition, Springer.

[16] **Neath, A.A., and Cavanaugh, J.E.** (2012). The bayesian information criterion: background, derivation, and applications. *WIREs comput stat*, *4*, 199-203.

[17] **Kumar, U., and Jain, V.K.** (2010). ARIMA forecasting of ambient air pollutants ($O_3$, NO, $NO_2$ and CO). *Environmental Res. Risk Assessment*, ***24***, 751–760.

[18] **Petrov, B. N., Csaki, F.** (1971). *2nd International Symposium on Information Theory*, Akademia Kiado, Budapest, pp. 267{281), September 2-8.

[19] **Lovric, M.** (2011). *International Encyclopedia of statistical science*. Springer.

[20] **Awad, A. M.** (1996). Properties of The Akaike Information Criterion, Department of Mathematics, University of Jordan, Amman, Jordan, *Microelectronic Reliab.*, ***36***, pp. 457-464.

[21] **Akaike, H.** (1974). A new look at the statistical identification model. *ZEEE Trans. Auto. Control*, ***19***, 716-723

[22] **Schwarz, G.** (1978). Estimating The Dimension of a Model. *The Annals of Statistics*, ***6***, 461-464.

[23**] Cavanaugh, J.E. and Neath, A.A.** (1999) Generalizing the derivation of the Schwarz information criterion, *Communications in Statistics – Theory and Methods*, ***28***, 249-266.

[24] **Akaike, H.** (1969). Fitting autoregressive models for prediction. *Annals of the Institute of Statistical Mathematics*, ***21***, 243-247.

[25] **Juan, A. A. and Serrat, C.** (2008). *Part II – Time Series Analysis C4 Autocorrelation Analysis (lecture notes).* Escola Polytechnical Superior Barcelona, UPC. Adresses: http://www.uoc.edu/

**Url-1** < *http://www.ewea.org/fileadmin/files/library/publications/scenarios/EWEA-Wind-energy-scenarios-2020.pdf* >, 10.04.2015.

**Url-2** < *http://canwea.ca/pdf/windvision/Windvision_summary_e.pdf* >, 10.04.2015.

**Url-3** < *http://www.enerji.gov.tr/yayinlar_raporlar/Arz_Guvenligi_Strateji_ Belgesi .pdf* >, 12.04.2015.

**Url-4** < http://www.mgm.gov.tr/genel/meteorolojikaletler.aspx?a=j >, 06.05.2015.

**CURRICULUM VITAE**

**Name Surname:** Zafer CANAL

**Date of Birth:** 30.08.1987

**Address:** Üsküdar/İstanbul

**E-Mail:** zafercanal@gmail.com

**B.Sc.:** Marmara University, Mechanical Engineering

**Professional Experience and Rewards:**

Çolak Petrol Mühendislik, Project Engineer, 2011-2013
Bursa Technical University, Research Assistant, 2013-2014
Ipragaz, Quality Control and Maintenance Supervisor, 2014-