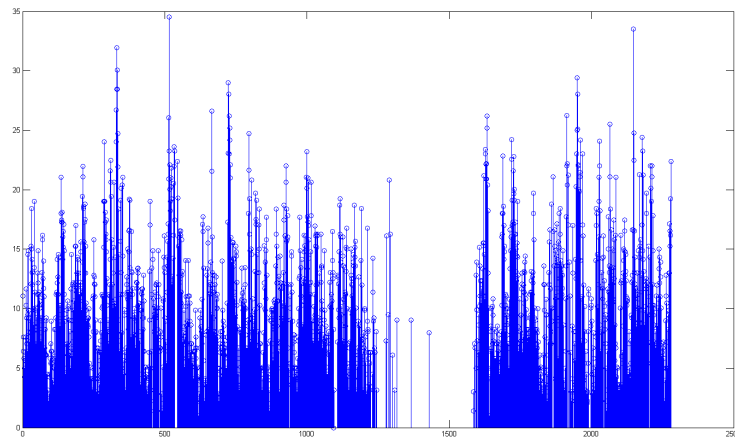


LUND UNIVERSITY

MASTER THESIS

Statistical Analysis of Wind Components



Author:
Joakim SAHLBERG

Supervisor:
Andreas JAKOBSSON

January 3, 2013

Nomenclature

ACF Autocorrelation function

GAPES Gapped-data amplitude and phase estimation

GRIB General Regularly-distributed Information in Binary form

ICAO International Civil Aviation Organization

ISA International Standard Temperature

ISCS International Satellite Communications System

NOAA National Oceanic and Atmospheric Administration

PACF Partial autocorrelation function

SADIS SAtellite DIstribution System

WAFS World Area Forecast System

WMO World Meteorological Organization

Contents

1	Abstract	4
2	Background	5
2.1	ICAO	5
2.2	WAFS	5
2.2.1	GRIB	6
2.3	General Wind Aspects	6
2.3.1	The Coriolis force	7
2.3.2	Geostrophic and Surface Winds	7
2.4	Aim	8
3	Wind speed	8
3.1	Reconstructing missing samples	8
3.2	Transforming the data	11
3.2.1	Log transformation	13
3.2.2	Square root transformation	14
3.2.3	Box-cox power transformation	15
3.3	Autocorrelation function and partial autocorrelation function	19
3.4	Initial model estimation	21
3.5	Recursive estimation of parameters	27
3.6	Testing our model on different pressure levels	28
3.7	A different approach - differentiation	29
4	Wind direction	33
5	Spatial Dependency	43
6	Conclusion and discussion	45
6.1	Additional Sources	46
7	Topics of further research	46
A	GAPES images	48

B	Wind Speed Model Testing For Different Pressure Levels	51
B.0.1	Pressure Level 700mb	51
B.0.2	Pressure Level 600mb	51
B.0.3	Pressure Level 500mb	56
B.0.4	Pressure Level 400mb	58
B.0.5	Pressure Level 300mb	60
B.0.6	Pressure Level 200mb	62
C	Interpolation Algorithm	64
D	Altitude to Pressure Conversion Table	65

1 Abstract

In the world of aviation, the need for accurate weather data is very important in order to plan and conduct a safe flight. Due to technology, there are ways to receive accurate weather data live today, but also, due to technology, this data might not be available. Thus the need for accurate statistical data is desirable. Is this possible to achieve using the SADIS data we currently have available? We will try to find the answer to this question by modelling the data as either an AR(p) or an ARMA(p,q) model using different types of estimation. It turns out that we can model the wind speed rather well using an AR(3) model and the wind direction using an ARMA(5,7) model, where the parameters are estimated using recursive LMS. In conclusion we can see that the type of data we have can be modelled using straightforward methods. However, we cannot simply take our predicted models and apply those everywhere. But what we can do, is take the methods which resulted in the given models and apply that to all desired locations.

2 Background

Even in the early days of aviation, it was realized that meteorological information was vital for the planning and a constant factor for the safe conduct of flights. In the 1950's a milestone was reached in aeronautical meteorology with the introduction of numerical weather prediction including the analysis and prediction of broad scale weather phenomena and upper winds. In the late 1960's computer generation of winds and forecasts began, minimizing unnecessary fuel uplift thus maximizing revenue payload.

At Flygprestanda we perform flight planning and many other avia-tional services and there are weather report services in place that provide weather forecasts for a period of up to 48h in the future. Problem occur when the flight planning is made several months in advance and a method including statistical weather data will greatly improve the accuracy of the costs for any particular flight.

The economic and financial impact of the weather on the aviation industry is difficult to quantify. Documented impacts in one country indicates that unexpected weather conditions at arrival times can result in a loss of as much as €265.000 in additional operating costs per day¹.

Furthermore, 80% of all reported delays over 15 min are caused by the weather, resulting in an estimated economic loss of over €750.000 million per year world wide.

According to planecrashinfo.com² about 30% of all accidents ranging from the 1950's up to present day, were caused by weather or weather related phenomena.

2.1 ICAO

A specialized agency of the United Nations, ICAO was created in 1944 to promote the safe and orderly development of international civil aviation throughout the world. It sets standards and regulations necessary for aviation safety, security, efficiency and regularity, as well as for aviation environmental protection. The Organization serves as the forum for cooperation in all fields of civil aviation among its 191 Member States.

2.2 WAFS

Within the ICAO network, the provider of aeronautical weather information is known as the World Area Forecast System or WAFS. This system provides meteorological authorities and other authorized users with information required for international air navigation.

There are two World Area Forecast Centers in the world that provide this information. One being the NOAA in Washington, and the other

¹There is no indication to what country this refers to.

²A website that keeps an updated database about accidents in aviation.

the UK Met Office in Exeter. These two publish the same information in order for one to be able to replace the other should the need occur.

Each of these two services operates its own satellite-based broadcast system to distribute data to airports all over the world. The U.S. NOAA broadcast system is ISCS and mainly covers America and the Pacific Ocean. The UK Met Office is called SADIS and mainly covers Europe, Asia, Indian Ocean and Africa. At Flygprestanda AB, we use SADIS.

2.2.1 GRIB

The format in which the distribution systems transmits their data is called GRIB. The GRIBs contain alot of data and the manual is 73 pages long, we will not cover all of it, we will simply explain the parts that are relevant to this paper.

First off we can give a better view of what kind of forecasts the GRIB contains.

```

90N-----
| I | J | K | L |
0 -----
| M | N | O | P |
90S--+-----
      330E  60E  150E  240E  330E

```

In the figure the + sign indicates where 0 degrees lie and the coordinates indicate the location of the octants of the globe, the numbers are the corresponding grid identification numbers, and the letters are the grid identification used in the WMO heading.

The grid point separation is 1.25x1.25 degrees on a latitude/longitude array, but the grid is 'thinned' by reducing the number of points in each row as one move away from the equator.

The latitudinal increment is always 1.25 degrees, resulting in 73 rows including the pole.

The longitudinal spacing on the equator is also 1.25 degrees, thus there will be 73 grid points at the equator in each octant.

The number of points on each latitudinal row, other than the equator, is calculated as the integer number of $2.0 + (90.0/1.25) * \cos(\text{latitude})$.

This formula results in there being 2 points at the pole, one on each meridian that deliniate the edges of the octant.

When all this is put together the result is that there are 3447 points of data actually transmitted in any individual GRIB bulletin containing one octant of the globe.

2.3 General Wind Aspects

By general agreement in the meteorological and aviation communities, the term "wind direction" is defined as the compass heading FROM

which the wind is blowing. This is in contrast to the term "bearing" which indicates the direction toward which an object is moving.

2.3.1 The Coriolis force

Since the globe is rotating, any movement on the Northern hemisphere is diverted to the right, if we look at it from our own position on the ground. (In the southern hemisphere it is bent to the left). This apparent bending force is known as the Coriolis force (Named after the French mathematician Gustave Gaspard Coriolis 1792-1843).

In the Northern hemisphere, the wind tends to rotate counterclockwise (as seen from above) as it approaches a low pressure area. In the Southern hemisphere, the wind rotates clockwise around low pressure areas.

The wind rises from the equator and moves north and south in the higher layers of the atmosphere. Around 30 degrees latitude in both hemispheres the Coriolis force prevents the air from moving much farther. At this latitude there is a high pressure area, as the air begins sinking down again. As the wind rises from the equator there will be a low pressure area close to ground level attracting winds from the North and South.

At the poles, there will be high pressure due to the cooling of the air. Keeping in mind the bending force of the Coriolis force, we thus have the following general results for the prevailing wind direction.

Prevailing wind directions	
Latitude	Direction
90-60N	NE
60-30N	SW
30-0N	NE
0-30S	SE
30-60S	NW
60-90S	SE

Prevailing wind directions can be a helpful tool in the initial prediction of the wind direction.

2.3.2 Geostrophic and Surface Winds

The geostrophic winds are largely driven by temperature differences, and thus pressure differences, and are not very much influenced by the surface of the earth. The geostrophic wind is found at altitudes above 1000 metres (3300 ft.) above ground level. The geostrophic wind speed may be measured using weather balloons.

Unlike geostrophic winds, the surface winds will be affected by the ground surface at altitudes up to 100 metres (330 ft). The wind will be slowed down by the earth's surface roughness and obstacles. Wind directions near the surface will be slightly different from the direction of the geostrophic wind because of the Coriolis force. When dealing with

aviation we are generally not concerned with the surface winds. The only time the surface winds are inflicting on aviation is during landing and take-off.

2.4 Aim

The aim of the thesis is to build a foundation for automatic generation of statistical models for time series in the Java programming language for Flygprestanda AB, to be used in their flight planning software. The main interest lies in wind data such as speed and direction. But there is also a general interest in the ability to generate statistical models for different time series.

3 Wind speed

The geographic position we have chosen for the data is suitingly enough the 1.25° by 1.25° square that houses Sturup airport, or ESMS as it is known as in the aviation industry. ESMS is its corresponding ICAO code. The location of ESMS is $55^\circ 32' 10'' \text{N} / 13^\circ 22' 23'' \text{E}$. The pressure level of our chosen data is 850mb, and according to the altitude to pressure conversion table in appendix D, this corresponds to approximately 4500 ft. This is not really an altitude that you continuously fly at, but this is the highest occurring pressure that the WAFS contain.

The forecast used is a 6h forecast, meaning that the received values will be valid from the point of receipt and 6h forward. The same transmission also contains 12h/18h/24h/30h/36h forecasts, correlation between these can be investigated at a later time.

The interval that the data is received is 6h, meaning there are 4 samples/day starting at 0200 (0000 UTC). There are some irregularities occurring with the times that the data is received, during some intervals they start at 0100. This is believed to have with daylight savings to do and is ignored.

In figure 1 we can clearly see that the data series is not ideal. There are many missing values and starting at approximately sample 1200 there is a huge chunk missing. If we are to use the data as it is, we will not be able to make a fair analysis.

3.1 Reconstructing missing samples

We need to reconstruct the missing values in some way if we can. The first thing is to pick a smaller sample size that is faster to work with and has less missing values, see figure 2.

An easy way to replace the single missing values is to use interpolation, which simply takes the mean of two adjacent values. For the purpose of this report the use of interpolation would have sufficed, had it not

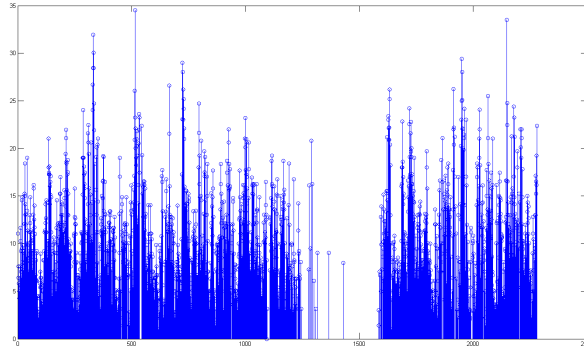


Figure 1: Wind speed data

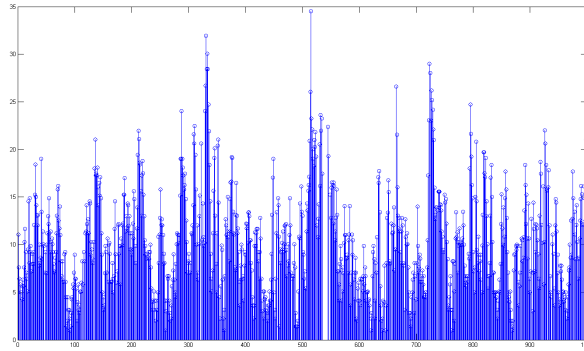


Figure 2: Wind speed data sample 1-1000

been for the missing chunk of samples in the original data. For the sake of trying to reconstruct the chunk of missing data we interpolate our smaller sample size and use that as our reference, resulting in figure 3.

Matlab code for the algorithm used to fill in the missing values can be found in appendix C.

We now remove a chunk from the data, as seen in figure 4, to resemble what we actually have in order to see if it is possible to reconstruct with satisfactory result.

For the reconstruction we will be using Gapped-data amplitude and phase estimation, reference [7]. The reason for this approach is because GAPES uses an approach that estimates the spectrum for the signal based on the values that do exist and continues to fill in the gaps via least squares fit.

Using the GAPES algorithm with default parameters, we try to reconstruct the missing values. The result is seen in figure 5 and a comparison between the interpolated version and the one filled with GAPES can be

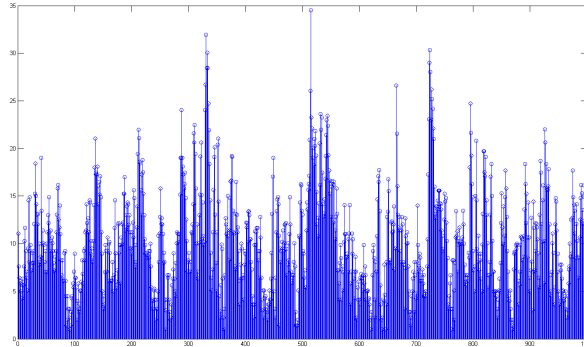


Figure 3: Wind speed data sample 1-1000 interpolated

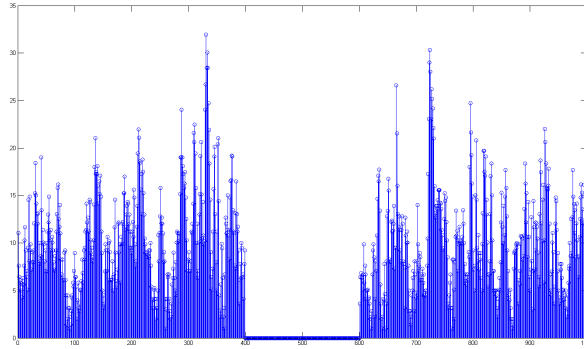


Figure 4: Wind speed data sample 1-1000 interpolated and gapped

seen in figure 6, the result is not quite satisfactory and along with the execution time being insanely long we should try a different approach.

To further examine this method of reconstructing missing chunks of data we need to improve the calculation speed alot. By decreasing the sample size even more, down to a number of 128 samples, figure 7, we should be able to achieve a more respectable calculation speed. We then proceed to remove some values corresponding to roughly the same relative amount missing in the original data, figure 8.

We then start things off by running the algorithm with defaults parameters and the result is seen in figure 9. In figures 64-67 found in Appendix A, the result of using different filter lengths as a percentage of the sample lengths as a percentage of the sample size can be seen. And in figures 68-71 we have also modified the initial filter length as percentage of the sample size.

In conclusion the decision is taken not to use the GAPES algorithm to fill in the gap of missing samples. As seen on the data of 128 samples, the result is not very accurate. There might be some way of getting a better result but with the calculation time of close to 3 days for the 2280

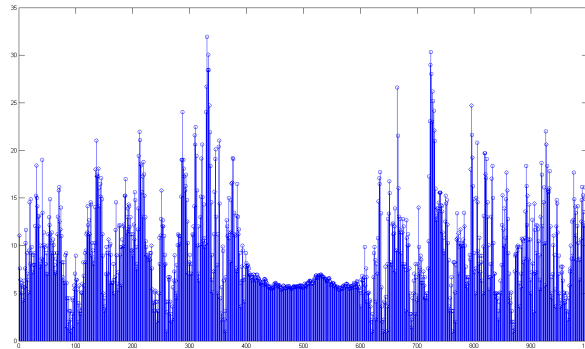


Figure 5: Wind speed data sample 1-1000 interpolated and gapped, filled with GAPES

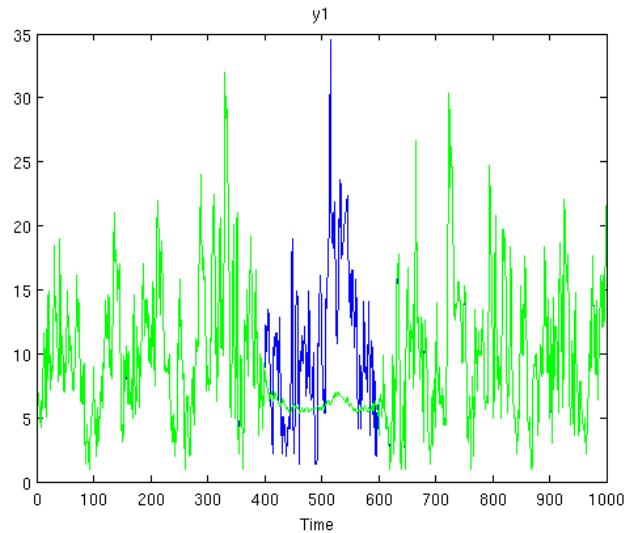


Figure 6: Comparison between interpolated and GAPES

sample size data on a regular workstation provided with 70Gb swap file, it would simply take too long to produce a satisfactory result.

Instead we will be using the interpolated subset of the original data set consisting of the first 1000 samples where interpolation has been used to fill in the missing values as shown in figure 3.

3.2 Transforming the data

As stated in reference [9] p.36-38 as well as reference [10], it is known that the average wind speed is Weibull distributed with the shape parameter close to 2 in Europe. This is a special case of the Weibull distribution

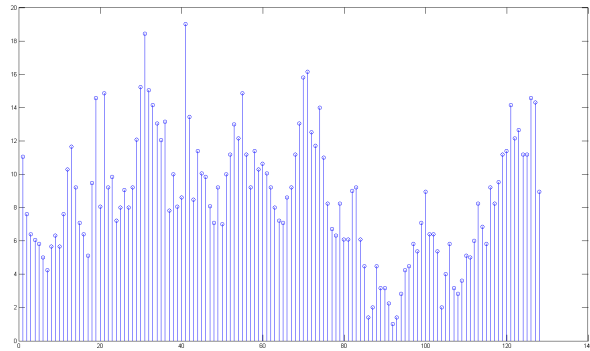


Figure 7: Wind speed data sample 1-128

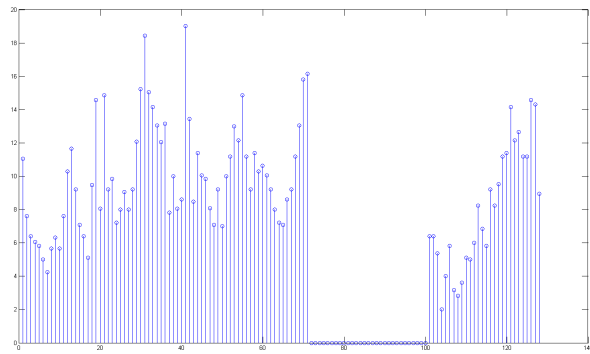


Figure 8: Wind speed data sample 1-128 gapped

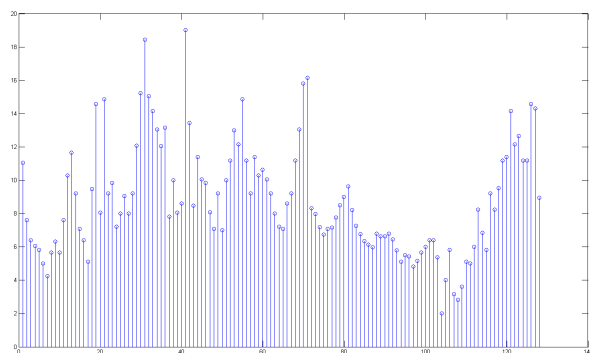


Figure 9: GAPES

and is known as a Rayleigh distribution.

"A Rayleigh distribution is often observed when the overall magnitude of a vector is related to its directional components."

This quote is taken from the website of the Danish Wind Industry Association, and is also stated in reference [10], and matches very well the kind of data we currently are working with. Since the values received are the orthogonal 2-dimensional vector components of the wind speed. And from these they are translated to wind speed in meters per second and wind direction in degrees.

If we take a look at the histogram for the data, shown in figure 10, we can see that the data is not Gaussian distributed. This is further confirmed with a normality plot, figure 11.

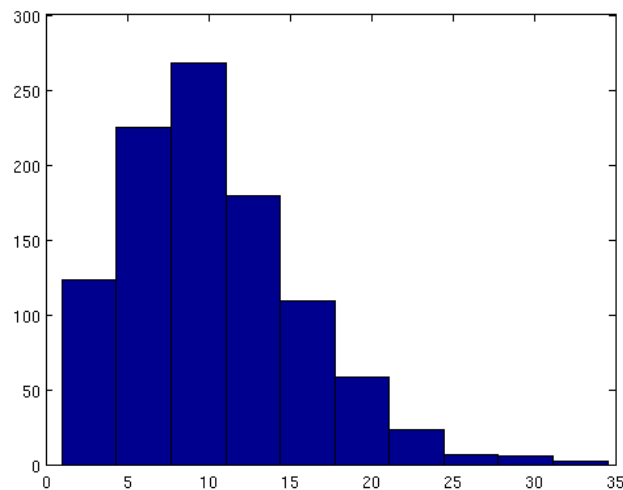


Figure 10: Histogram for wind speed data, sample 1 through 1000

If we take into considering what we think we know, the wind speed should be Weibull distributed with a shape parameter $\lambda = 2$, i.e. Rayleigh distributed. Figure 12 shows a Weibull plot of the data and makes us happy since it confirms what we suspected.

For further confirmation we generate a random Weibull distributed sample size with shape parameter $\lambda = 2$ for comparison. The plots are shown in figures 13 - 15 with the random data on the left hand side and our wind speed data on the right hand side.

3.2.1 Log transformation

The log function has the defining property that $\log(x * y) = \log(x) + \log(y)$ Therefore, using the logarithm tends to convert multiplicative relationships to additive relationships, also it tends to convert exponential

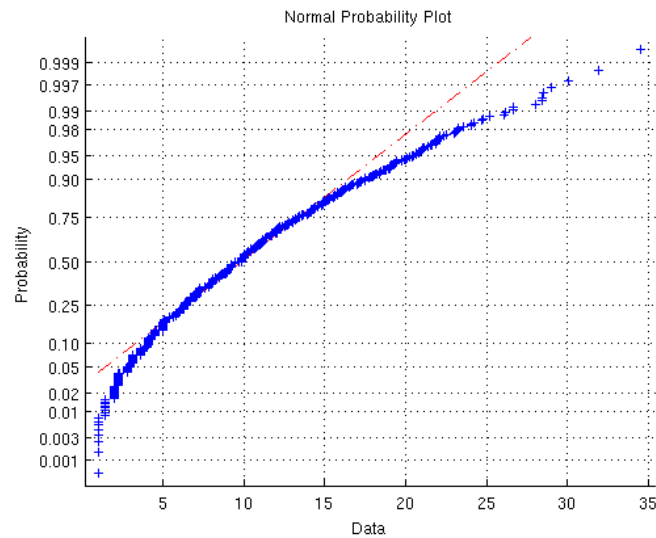


Figure 11: Normality plot for wind speed data, sample 1 through 1000

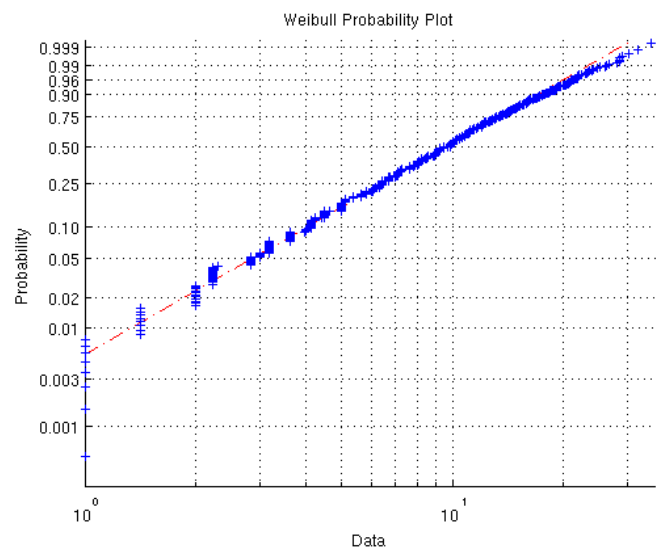


Figure 12: Weibull plot for wind speed data, sample 1 through 1000

trends to linear trends. By taking logarithms of variables which are multiplicatively related and/or growing exponentially over time, we can often explain their behavior with linear models.

3.2.2 Square root transformation

This transformation of data is appropriate for the data sets where the variance is proportional to the mean. Here, the data consists of small

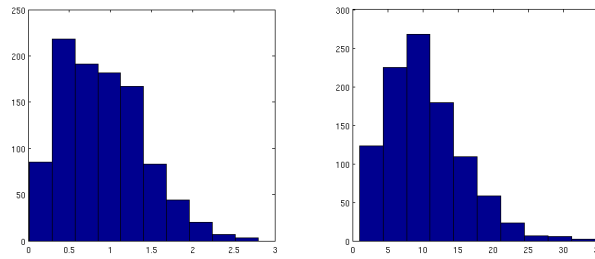


Figure 13: Histogram comparison for wind speed data

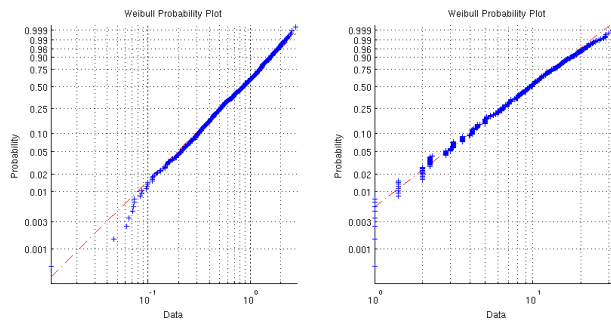


Figure 14: Weibull plot comparison for wind speed data

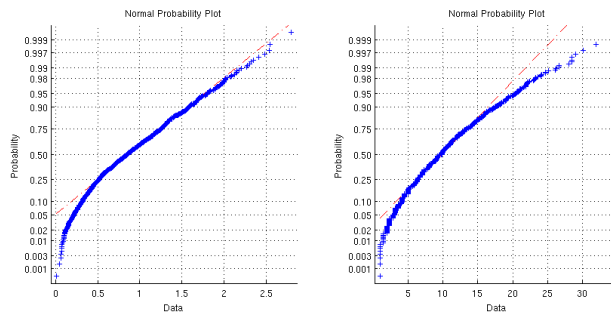


Figure 15: Normality plot comparison for wind speed data

whole numbers, for example, data obtained in counting rare events. This data set generally follows the Poisson distribution and square root transformation approximates Poisson to normal distribution. Furthermore, the square-root transformation is just a special case of Box-Cox power transformation with $\hat{\lambda}=0.5$ and omitting some centering.

3.2.3 Box-cox power transformation

The logarithm and square root transformations are commonly used for positive data. The power transform is a family of transformations parametrized by a non-negative value $\hat{\lambda}$ that includes the logarithm,

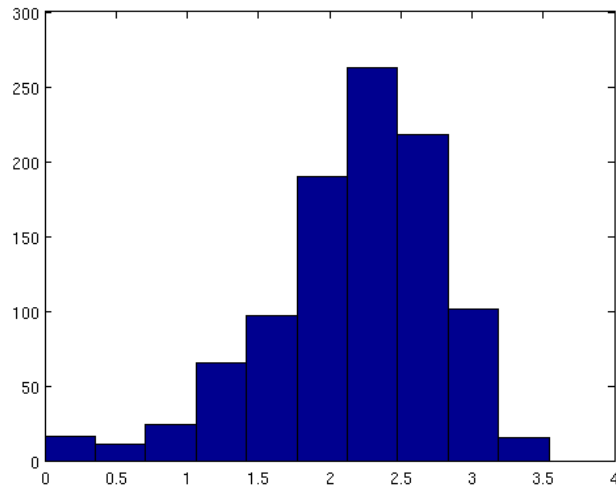


Figure 16: Histogram after log transformation

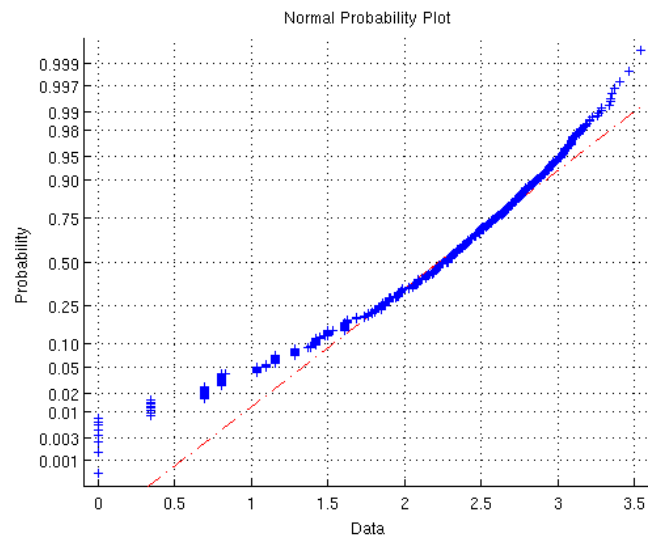


Figure 17: Normality plot after log transformation

square root, and multiplicative inverse as special cases. To approach data transformation systematically, it is possible to use statistical estimation techniques to estimate the parameter λ in the power transform, thereby identifying the transform that is approximately the most appropriate in a given setting. In regression analysis, this approach is known as the Box-Cox technique.

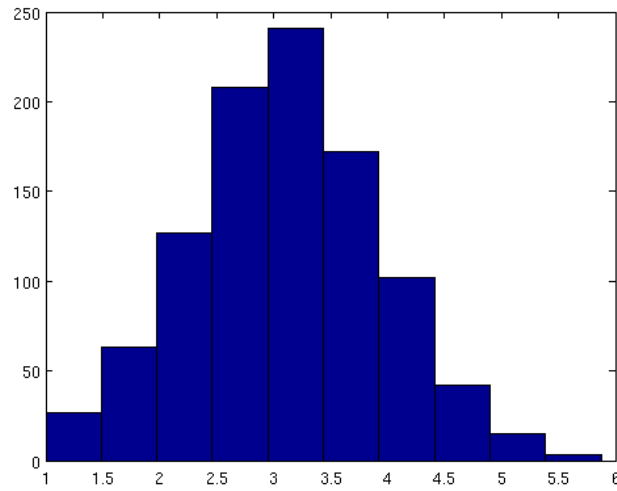


Figure 18: Histogram after square root transformation

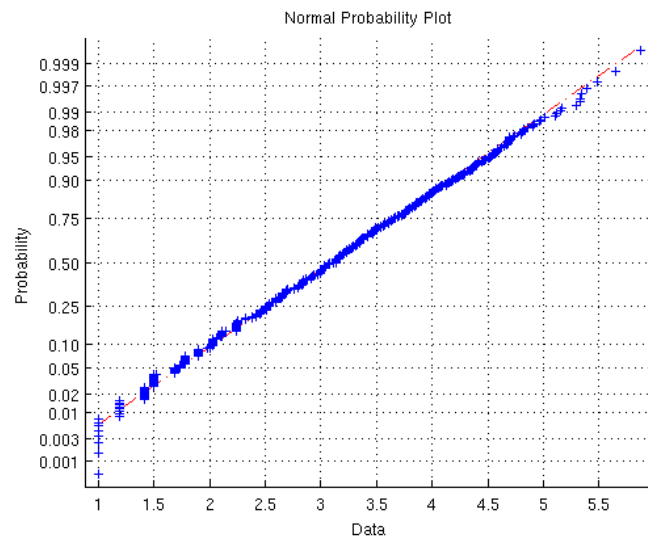


Figure 19: Normality plot after square root transformation

Box-cox transformation reference [3] p.179;

$$y_i^{(\lambda)} = \begin{cases} \frac{y_i^\lambda - 1}{\lambda} & \text{if } \lambda \neq 0 \\ \log(y_i) & \text{if } \lambda = 0 \end{cases} \quad (1)$$

With $\lambda = 0.4528$ the cost function is minimized and a satisfactory result is achieved. The resulting histogram and normality plot can be seen in figures 20 and 21.

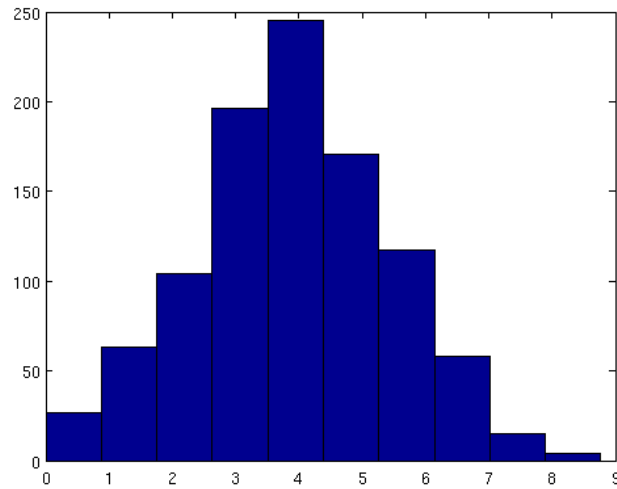


Figure 20: Histogram after Box-cox transformation

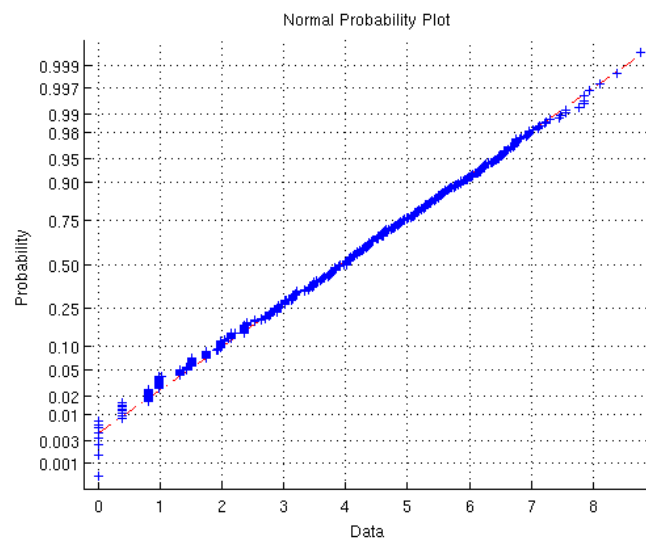


Figure 21: Normality plot after Box-cox transformation

Seeing the resulting $\lambda = 0.4528$ after the Box-cox transformation and knowing that the square root transformation is a special case of the Box-cox transformation with $\lambda = 0.5$, we can say that the square root transformation is the better choice here since the result is very similar and the square root is a simpler transformation.

3.3 Autocorrelation function and partial autocorrelation function

Now when we achieved Gaussian distributed data, we want some clues about the structure of it. In order to get a clue about the model structure we look at the autocorrelation function and the partial autocorrelation function, namely ACF and PACF for the series.

We will restrict ourselves to wide-sense stationary (WSS) processes here, since this is what we aim to achieve using transformation. A process is considered WSS if

1. The mean of the process is constant.
2. The auto-covariance Cy_s, y_t^* only depends on the difference $(s-t)$, and not on the actual values of s and t .
3. The variance of the process is finite, i.e., $E|y_y|^2 < \text{inf}$.

The moving average process.

Definition 2.8 from reference [1] p.24. *The process y_t is called a moving average (MA) process if*

$$y_t = e_t + c_1 e_{t-1} + \dots + c_q e_{t-q} \triangleq C(z)e_t, \quad (2)$$

where $C(z)$ is a monic polynomial of order q , i.e.,

$$C(Z) = 1 + c_1 z^{-1} + \dots + c_q z^{-q}, \quad (3)$$

where $c_q \neq 0$, and e_t is a zero-mean white noise process with variance σ_e^2 . The process is always stable, and is invertible if and only if all the zeros to the generating polynomial $C(z)$ are strictly within the unit circle. Which allows us to express the MA(q) process as

$$y_t = C(z)e_t \quad (4)$$

The autoregressive process.

Definition 2.9 from reference [1] p.28. *The process y_t is called a autoregressive (AR) process if*

$$A(z)y_t \triangleq y_t + a_1 y_{t-1} + \dots + a_p y_{t-p} = e_t, \quad (5)$$

where $A(z)$ is a monic polynomial of order p , i.e.,

$$A(z) = 1 + a_1 z^{-1} + \dots + a_p z^{-p}, \quad (6)$$

where $a_p \neq 0$, and e_t is a zero-mean white noise process with variance σ_e^2 , being uncorrelated with y_{tl} , for $l > 0$. The process is stationary (and thus an AR-process) if and only if all the zeros of the generating polynomial $A(z)$ are strictly within the unit circle. An AR-process is always invertible.

The autoregressive moving average process.

Definition 2.10 from reference [1] p.34-35. *The process y_t is called an autoregressive moving average (ARMA) process if*

$$A(z) = C(z)e_t, \tag{7}$$

where $A(z)$ and $C(z)$ are monic polynomial of order p and q , respectively, i.e.,

$$A(z) = 1 + a_1z^{-1} + \dots + a_pz^{-p}, \tag{8}$$

$$C(z) = 1 + c_1z^{-1} + \dots + c_qz^{-q}, \tag{9}$$

and e_t is a zero-mean white noise process with variance σ_e^2 .

Definition 2.4 from reference [1] p.11. *The auto-covariance function for y_t is defined as*

$$r_y(k) = Cy_t, y_{t-k}^* = E[y_t - m_x][y_{t-k} - m_y]^* \tag{10}$$

$$= Ey_t y_{t-k}^* - m_y m_y^* \tag{11}$$

Similarly, we define the cross-covariance of the WSS processes x_t and y_t as

$$r_{x,y}(k) = Cx_t, y_{t-k}^* = E[y_t - m_y][y_{t-k} - m_y]^*, \tag{12}$$

where m_x and m_y denote the means of the respective processes.

Definition 2.5 from reference [1] p.11. *The auto-correlation function for y_t is defined as*

$$\rho_y(k) = \frac{r_y(k)}{r_y(0)} \tag{13}$$

and will therefore be bounded such that $|\rho_y(k)| \leq 1$, with equality for $k = 0$, as well as, possibly, for $k = l$, with $l > 0$, if the signal is periodic with period. Similarly, we define the cross-correlation of the processes x_t and y_t as

$$\rho_{x,y}(k) = \frac{r_{x,y}(k)}{\sqrt{r_x(0)r_y(0)}} \tag{14}$$

which will be bounded as $|\rho_{x,y}(k)| \leq 1$

The partial auto-correlation function, reference [1] p.39. recalling definition 2.9, consider the AR(p)-process,

$$y_t = \phi_{k,1}y_{t-1} + \dots + \phi_{k,k}y_{t-p} + e_t, \tag{15}$$

where the notation $\phi_{k,l}$, denotes the l :th (negative) AR-coefficient of the k :th order AR model. Clearly, using our earlier notation, $\phi_{k,l} = a_l$. We may express the Yule-Walker equations, definition from reference [1] p.29, [2] p.357-358, [3] p.122, [4] p.111, [5] p.57 and [6] p.875, for the correlation function of (15) as

$$\begin{bmatrix} 1 & \rho_y(1) & \dots & \rho_y(1-k) \\ \rho_y(-1) & 1 & \dots & \rho_y(2-k) \\ \vdots & \vdots & \ddots & \vdots \\ \rho_y(k-1) & \rho_y(k-2) & \dots & 1 \end{bmatrix} \begin{bmatrix} \phi_{k,1} \\ \phi_{k,2} \\ \vdots \\ \phi_{k,k} \end{bmatrix} = \begin{bmatrix} \rho_y(1) \\ \rho_y(2) \\ \vdots \\ \rho_y(k) \end{bmatrix} \tag{16}$$

Table 3.3 shows the characteristics of the ACF and PACF for different processes, reference [1] p.41. This is what we will be using for the initial guess of the structure and order of the process.

Process	ACF	PACF
$AR(p)$	Damped exponential and/or sine functions	$\phi_{k,k} = 0$ for $k > p$
$MA(q)$	$\rho(k) = 0$ for $k > q$	Damped exponential and/or sine functions
$ARMA(p, q)$	Damped exponential and/or sine functions after lag $p - q$	Damped exponential and/or sine functions after lag $p - q$

3.4 Initial model estimation

Looking at the ACF, eq.13, seen in figure 22 we can suspect an exponentially decaying trend. Together with the PACF, eq.16, figure 23, it is looking much like the presence of a low order $AR(p)$ process according to the table 3.3 above.

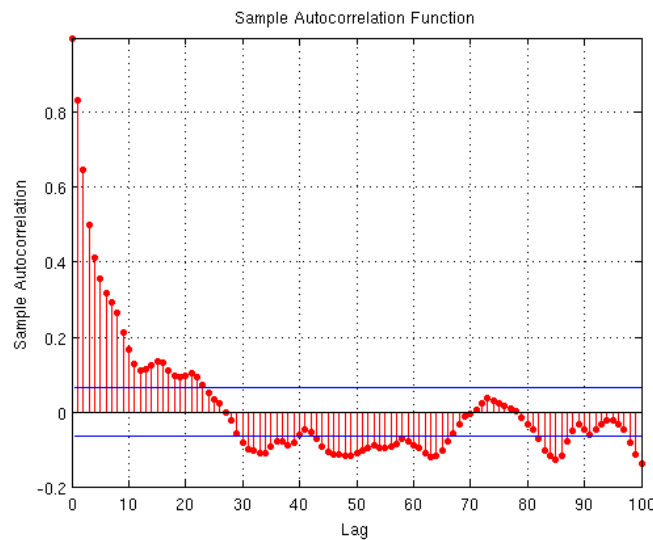


Figure 22: Autocorrelation function for wind speed data

At this point we are not fully confident that we have extracted all the information we can from the data. We are fairly certain that there is need to do a seasonal filtering due to the nature of the wind and the fact that the ACF is showing some correlation at greater lags. Due to the samples in the data corresponding to the wind speed every 6 hours we might think that a low level $AR(p)$ process can describe this seasonality well, such as an $AR(1)$ or perhaps $AR(2)$.

Considering real world physics and how the wind behave suggests the use of an $AR(1)$ process and at the maximum an $AR(2)$. Thinking

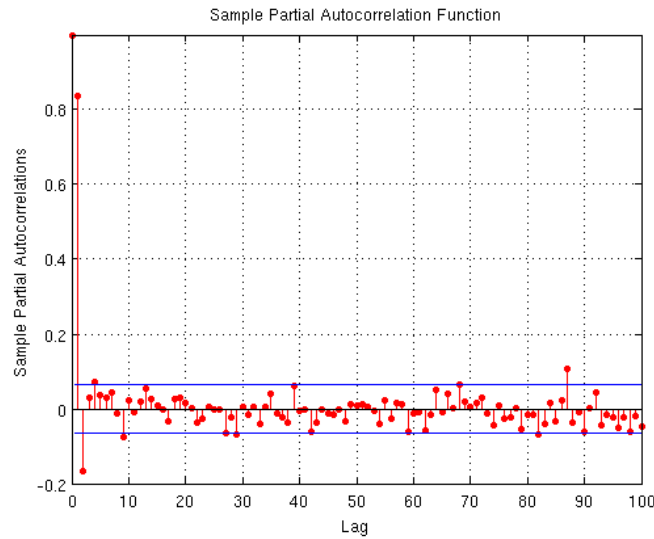


Figure 23: Partial autocorrelation function for wind speed data

that the wind speed occurring 6 hours prior and our current is correlated somehow seems likely, but imagining the same 12 hours prior is harder.

With this in mind we tried different approaches using the prediction error method, PEM, to estimate a 24h the seasonal model. We have a strong feeling that it will be a pure AR(p) process of low order, but the need to add MA(q) terms to create ARMA(p,q) processes is not unlikely. But as suspected, the loss function was minimal for an AR(1) process.

Seasonal model.

Discrete-time IDPOLY model: $A(q)y(t) = B(q)u(t) + e(t)$

$A(q) = 1 - 0.6905 q^{-1}$

$B(q) = 0.02915$

Estimated using PEM using SearchMethod = Auto on data set data_s

Loss function 1.45413 and FPE 1.45997

Sampling interval: 1

After seasonal filtering we examine the ACF and PACF. Figure 24 show these. Looking at these the desire to use an AR(3) is arising, even though there is still some correlation at higher lags present into the ACF.

When estimating our model we are using approximately 75% of our data set. Leaving the last 25% for evaluation. Even though everything points to an AR(3) process, we still play around with our parameters and examine what happens when using higher order processes such as AR(15) as well as adding low level MA(q) terms. In the end the AR(3) process resulting in the best fit is shown below along with the achieved

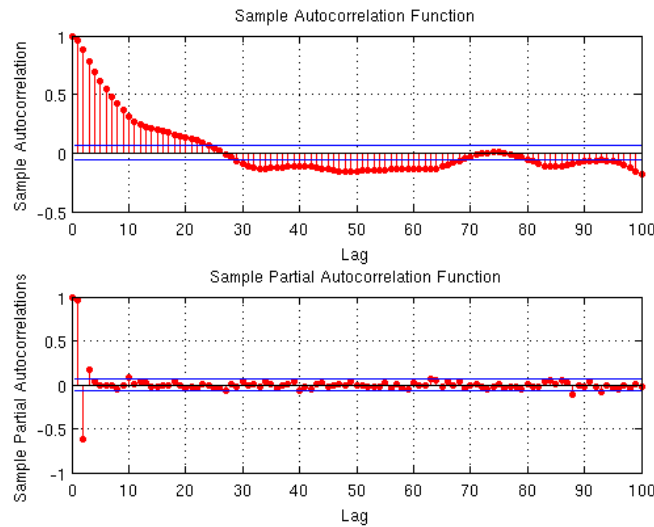


Figure 24: ACF and PACF after seasonal 24h filtering

fit which is shown in figure 25. What we see in figure 25 is the last 250 samples of our data that were not used to create our model, compared to 250 samples predicted by our model.

Estimated model.

Discrete-time IDPOLY model: $A(q)y(t) = e(t)$

$A(q) = 1 - 1.682 q^{-1} + 0.7825 q^{-2} + 0.03399 q^{-3} - 0.1315 q^{-4}$

Estimated using ARMAX

Loss function 0.714577 and FPE 0.722271

Sampling interval: 1

This looks like a rather satisfying result. To help us decide if the predicted values are good enough we use a few so called whiteness tests. These are Ljung-Box-Pierce, McLeod-Li, Monti and the sign change test as well as plotting the ACF for the residual.

Ljung-Box-Pierce

The test-statistic is

$$Q = n(n+2) \sum_{k=1}^h \frac{\hat{\rho}_k^2}{n-k} \quad (17)$$

where n is the sample size, $\hat{\rho}_k$ is the sample autocorrelation at lag k , and h is the number of lags being tested.

McLeod-Li

The test-statistic is

$$Q = n(n+2) \sum_{k=1}^h \frac{\hat{\rho}_k^2}{n-k} \quad (18)$$

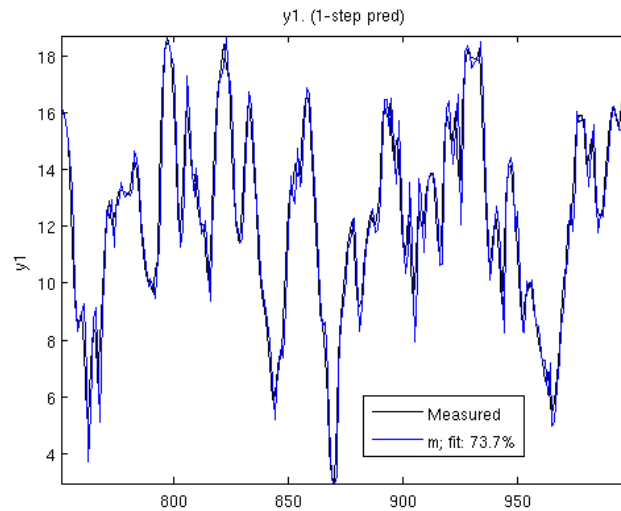


Figure 25: Fit for our predicted samples, AR(3)

where n is the sample size, $\hat{\rho}_k$ is the autocorrelation for the standard deviation at lag k , and h is the number of lags being tested.

Monti

The test-statistic is

$$Q = n(n+2) \sum_{k=1}^h \frac{\hat{\rho}_k^2}{n-k} \quad (19)$$

where n is the sample size, $\hat{\rho}_k$ is the sample partial autocorrelation at lag k , and h is the number of lags being tested.

In all the above cases, for significance level α , the critical region for rejection of the hypothesis of randomness is

$$Q > \chi_{1-\alpha, h}^2 \quad (20)$$

Where χ^2 is the Chi-squared distribution with $1 - \alpha$ degrees of freedom.

Sign change

The sign change test counts the number of sign changes in the data and the number seen when referred to is the ratio.

Whiteness test with 5% significance

Ljung-Box-Pierce test:	0 (white if 43.78 < 36.42)
McLeod-Li test:	0 (white if 52.32 < 36.42)
Monti test:	0 (white if 47.87 < 36.42)
Sign change test:	0 (white if 0.42 in [0.44,0.56])

The whiteness tests should return a 1 if the process is believed to be white. As we see above, none of the tests return a positive result. If we

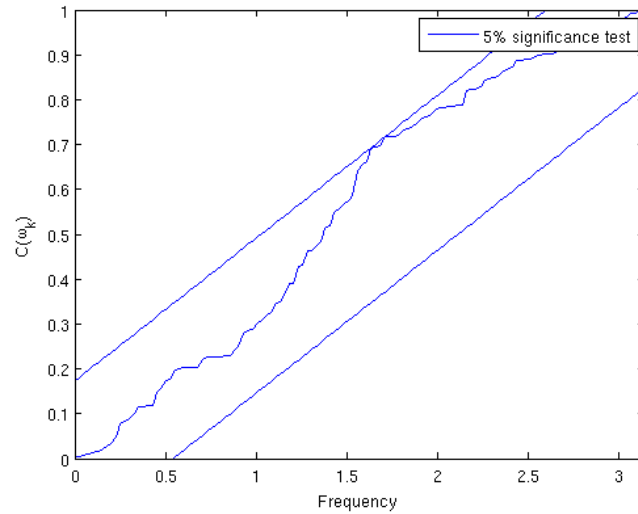


Figure 26: Whiteness plot for the residual, AR(3) model prediction

also take a look at the ACF for the residual it should resemble that of a random signal, in other words, it should be white. And by white we mean that the correlation at lag $r_k = 1$ when $k = 0$ otherwise 0.

Figure 27 shows the ACF for the residual and we can clearly see that the value at $k = 1$ is well outside the limit for zero. This, along with the whiteness tests, adds to our impression that the selected model is not optimal.

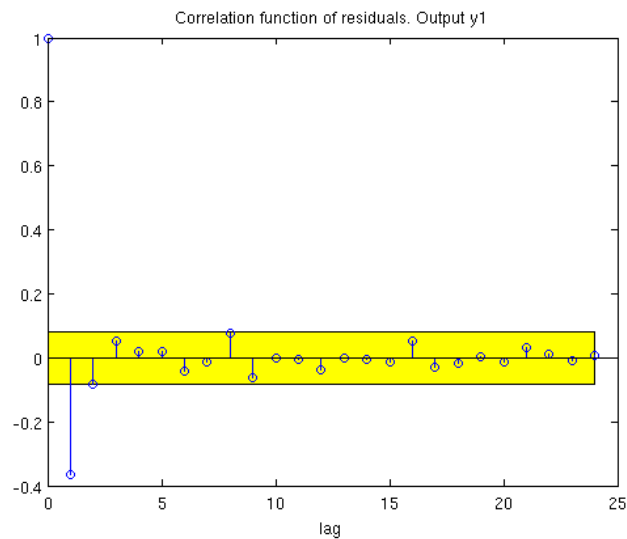


Figure 27: The ACF for the residual, AR(3) model prediction

One indication given by this is that we might want to try adding an MA(1) term or perhaps even an MA(2) term to account for this. We jump straight to the residuals to see if we managed to affect the ACF by adding MA terms. In figure 28 we can see that we haven't removed the term in $k = 1$ by adding an MA(1) term. Looking at figure 29 we can see that by adding an MA(2) term we add to the problem since the value in $k = 2$ starts drifting the wrong direction as well.

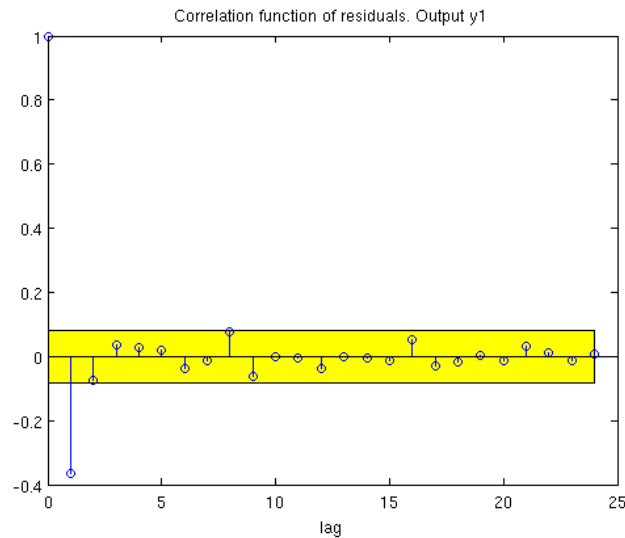


Figure 28: The ACF for the residual, ARMA(3,1) model prediction

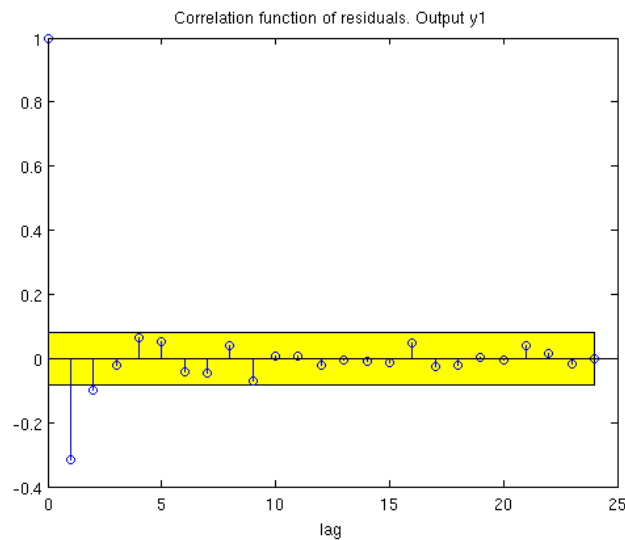


Figure 29: The ACF for the residual, ARMA(3,2) model prediction

This leads us to believe that the problem is not with the model structure

or even the order. The problem might be that the mean is not stationary and that it could cause some kind of drift. To try to account for this we will take a look at recursive estimation.

3.5 Recursive estimation of parameters

In this approach we will be using the forgetting factor algorithm, which in the linear regression case is known as recursive least squares or RLS. If a forgetting factor of $\lambda = 1$ is used it is the equivalent of the Kalman filter with $R_1 = 0$ and $R_2 = 1$. We will not be diving in to the Kalman filter much more but it is worth mentioning since it is very common in signal analysis.

Recursive least square filters are discussed in [5] p436-466, [4] p.541-551, [3] p.314, and the discrete Kalman filter is discussed in chapter 7.4 in [4] p.371-380, [5] p.466-506,554,638,656 and [3] p.289-296.

We go back to our AR(3) but this time using RLS with forgetting factor of $\lambda = 0.98$ to estimate the parameters.

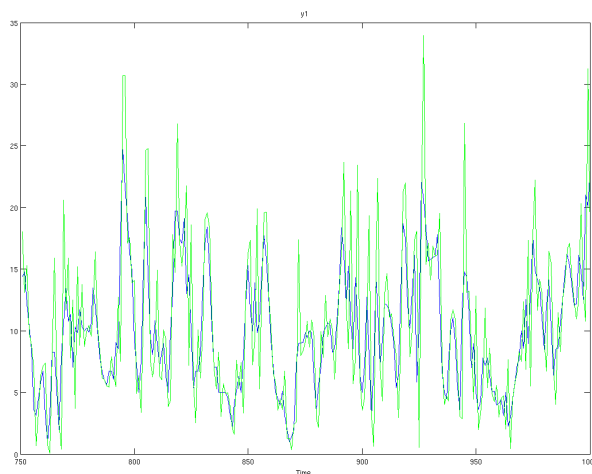


Figure 30: Prediction fit for recursive estimation, AR(3) model prediction

```
Whiteness test with 5% significance
Ljung-Box-Pierce test: 1 (white if 23.20 < 36.42)
McLeod-Li test:      0 (white if 47.14 < 36.42)
Monti test:          1 (white if 23.91 < 36.42)
Sign change test:    1 (white if 0.49 in [0.44,0.56])
```

After the recursive estimation we perform the whiteness tests once more and the results can be observed above. All except the McLeod-Li test are positive, this is very good, in fact this is good enough since the McLeod-Li test also measures higher order moments. Meaning it is hard to get that test to pass unless the residual is, in Gaussian terms, good.

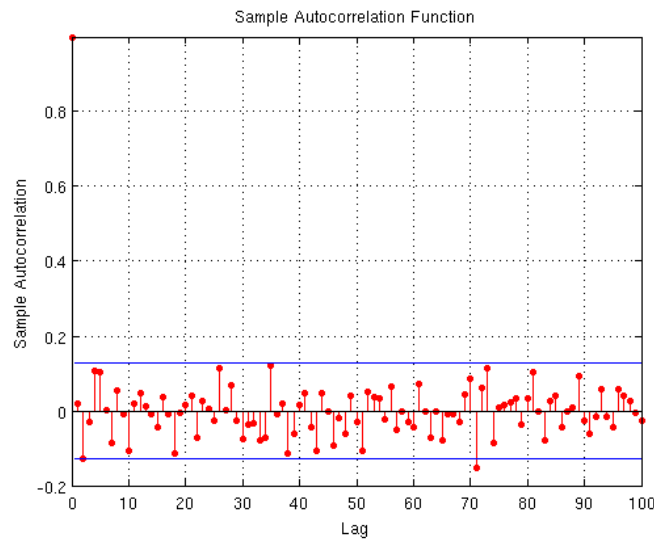


Figure 31: The ACF for the residual, recursive AR(3) model prediction

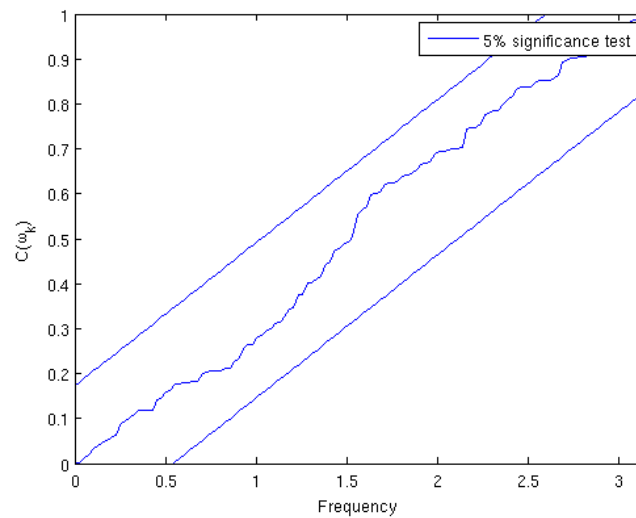


Figure 32: Whiteness plot for the residual, recursive AR(3) model prediction

3.6 Testing our model on different pressure levels

When we test our model structure for the other pressure levels present in our data we get a pretty good overall fit. Images for ACF and PACF after seasonal filtering, fit after recursive estimation, whiteness plot and ACF for the residual along with the results from the whiteness tests can be seen for all pressure levels in appendix B, even if not referenced here.

For the pressures 700, 600 and 500mb the only whiteness test that does

not go through is the sign change test. But when plotting the ACF for the residual in these cases, figures 74, 78 and 82, we can see that they are within the limit of a 95% confidence interval. For 400mb we can see that the Monti test is the one not giving a positive response, this is concerning since that should be easier to pass than the McLeod-Li. Arriving at 300mb, the only thing concerning us here is the fact that the residual ACF does not look as good as we would want, falling short with 6 out of 100 outside the allowed interval for our 95% confidence interval. This can be seen below in figure 33. At 200mb, the McLeod-Li test is returning negative as well as the sign change test, but once again looking at the residual ACF we are within our limits. Even though we are within our 95% confidence interval we can see a trend where the ACF for the residual is getting noisier.

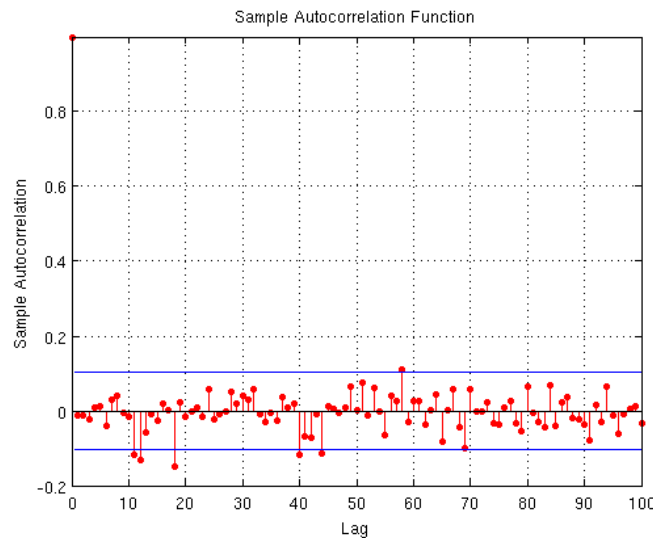


Figure 33: Residual ACF recursive AR(3), 300mb

Along with the residual ACF getting noisier we also have a trend in the PACF when the pressure decreases where the correlation coefficient at lag $k = 4$ is getting noticeable. This can be seen in figures 72, 76, 80, 84, 88, 92 in Appendix B.

This suggests that we might need to use an AR(4) model as we get to higher altitudes i.e, the pressure decreases. This is for our case of Malmoe airport and the same cannot be said about the rest of the world at this stage.

3.7 A different approach - differentiation

Another way to handle a drifting mean, other than the recursive approach, is to use differentiation.

After differentiation we take a quick look at the data, figure 34. It is looking more random now then before the differentiation, this is a good

sign, and the ACF, figure 35, and the PACF, figure 36, are suggesting the presence of an ARMA(p,q) model. We will be using PEM for the seasonal model as talked about earlier.

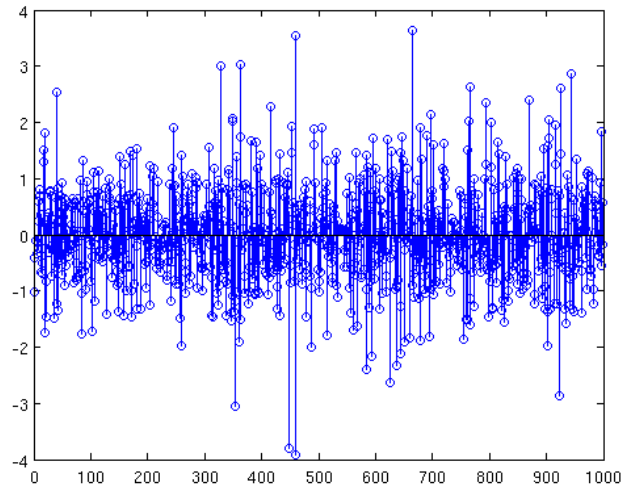


Figure 34: Differentiated wind speed data

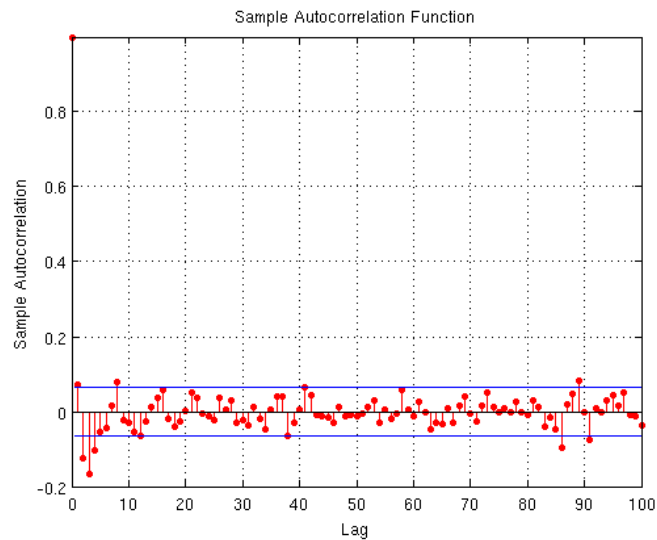


Figure 35: Autocorrelation function for wind speed data after differentiation

Corresponding seasonal model for fit 1.

Discrete-time IDPOLY model: $A(q)y(t) = B(q)u(t) + C(q)e(t)$

$A(q) = 1 - 0.09158 q^{-1} + 0.04503 q^{-2} + 0.1334 q^{-3} + 0.1208 q^{-4}$

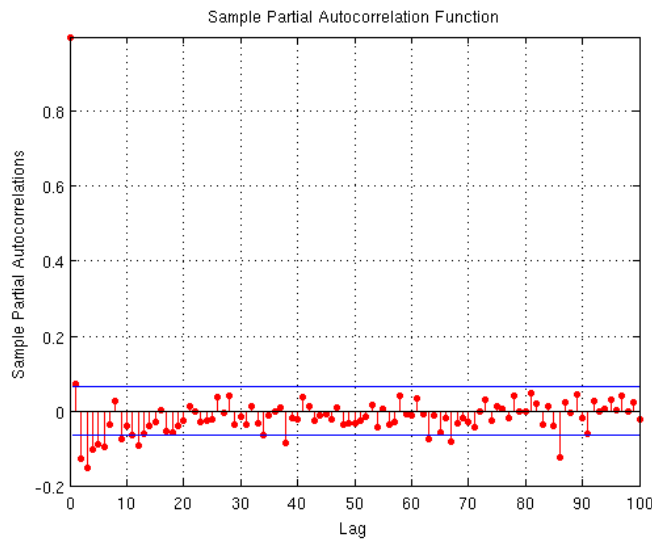


Figure 36: Partial autocorrelation function for wind speed data after differentiation

$$B(q) = 0.004772$$

$$C(q) = 1 - 0.05309 q^{-1} - 0.1122 q^{-2} - 0.01684 q^{-3} - 0.8179 q^{-4}$$

Estimated using PEM using SearchMethod = Auto on data set data_s

Loss function 0.767863 and FPE 0.781867

Sampling interval: 1

Predicted model for fit 1.

Discrete-time IDPOLY model: $A(q)y(t) = C(q)e(t)$

$$A(q) = 1 - 0.8384 q^{-1} + 0.12 q^{-2} + 0.2225 q^{-3} + 0.12 q^{-4} - 0.1822 q^{-5}$$

$$C(q) = 1 - 0.748 q^{-1} - 0.1428 q^{-2} + 0.1059 q^{-3} - 0.6527 q^{-4} + 0.4253 q^{-5} + 0.01233 q^{-6}$$

Estimated using ARMAX

Loss function 0.702923 and FPE 0.723963

Sampling interval: 1

Whiteness test with 5% significance

Ljung-Box-Pierce test: 1 (white if $20.41 < 36.42$)

McLeod-Li test: 0 (white if $39.32 < 36.42$)

Monti test: 1 (white if $33.57 < 36.42$)

Sign change test: 1 (white if 0.49 in $[0.44, 0.56]$)

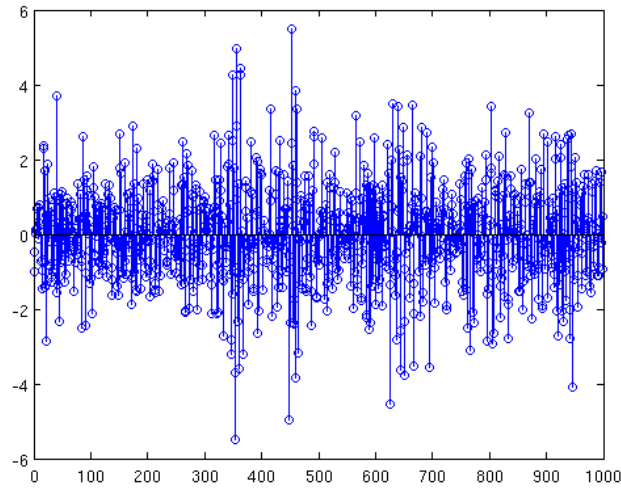


Figure 37: The differentiated wind speed data after seasonal filtering

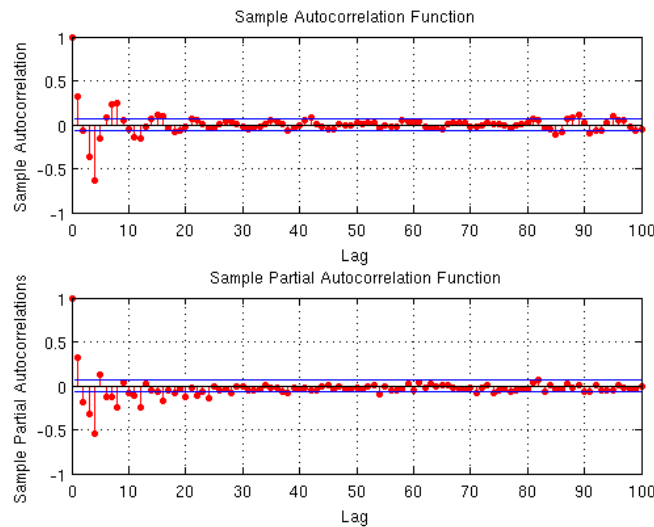


Figure 38: ACF and PACF for differentiated wind speed data after seasonal filtering

After a lot of testing we achieve the best fit of 33% with an ARMA(5,6) model, figure 39. If we take a look at the whiteness tests above it is yet again the McLeod-Li that is negative, but this is acceptable. Overall this approach with differentiation might be a viable option in some cases, but for us it resulted in a more complicated model with a worse fit for our predicted samples.

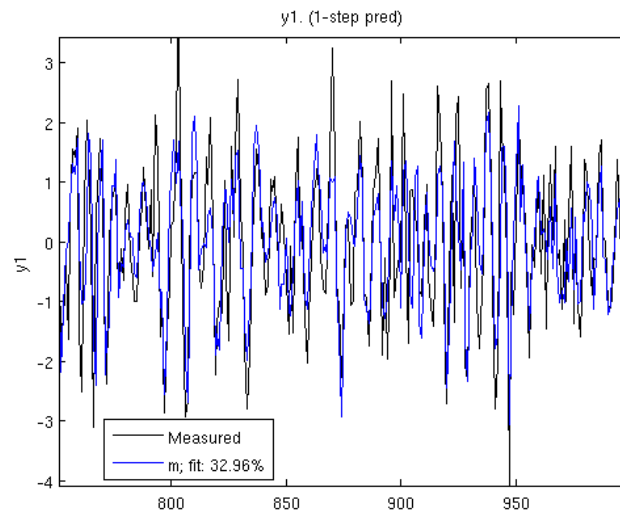


Figure 39: Prediction fit for differentiated wind speed, AR(3) model prediction

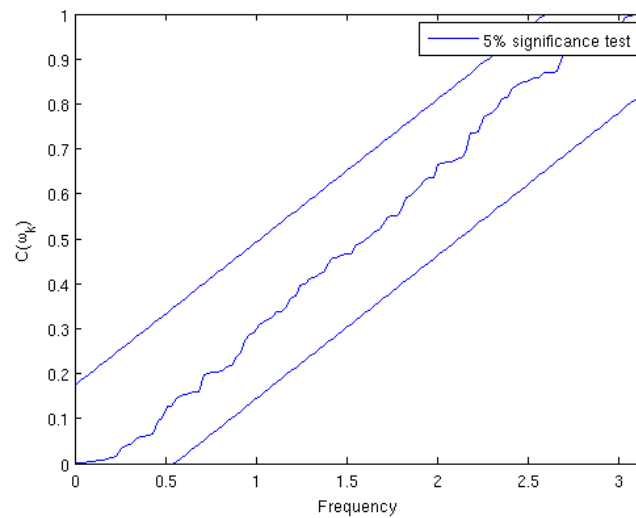


Figure 40: Whiteness plot for the residual, AR(3)

4 Wind direction

At first we take a look at the data for the wind direction and there are some gaps in this as well, this is expected since this data comes in the same transmission as the wind speed. Meaning that any missing wind speed value will result in a missing wind direction value as well. We use the same simple interpolation as before and figure 42 shows the corresponding 1000 values of wind direction for the wind speed data

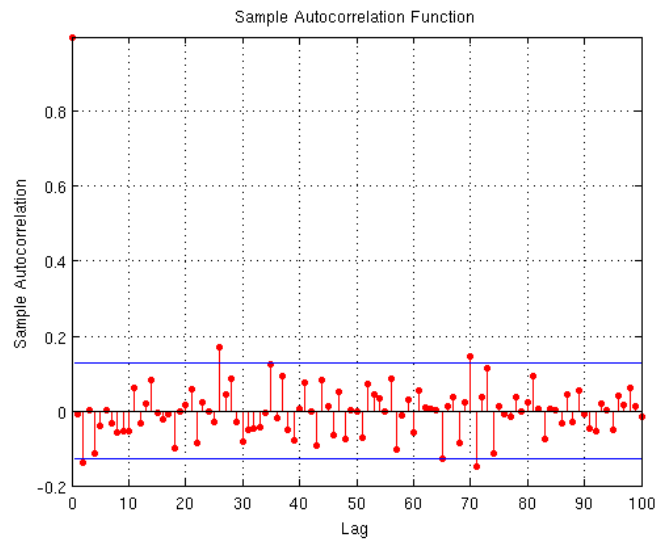


Figure 41: Residual ACF.

used earlier.

We will also introduce the rose plot here, which is commonly used when talking about wind direction. Since the values are between 0 and 360, the rose plot will give us a good overview of the actual direction the wind is coming from geographically. A rose plot for the data is seen in figure 43.

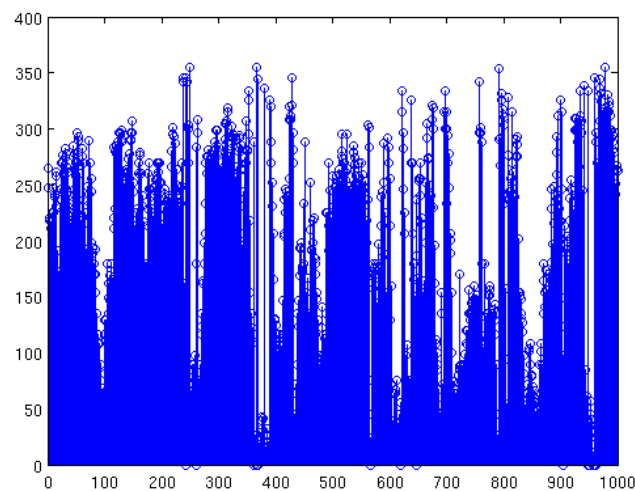


Figure 42: Wind direction sample 1 through 1000

As stated earlier the prevailing wind direction for our data would in fact

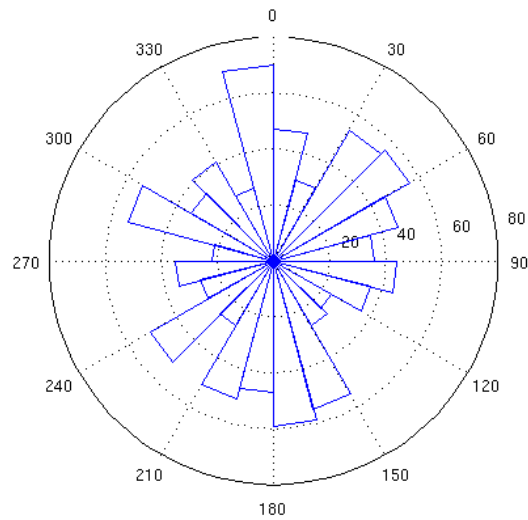


Figure 43: Rose plot

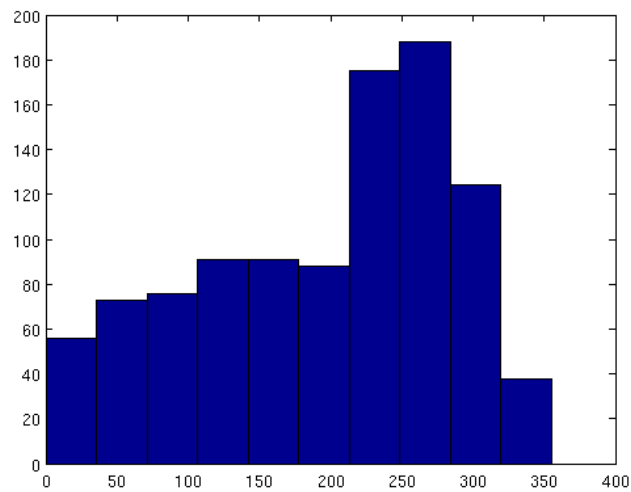


Figure 44: Histogram

be south west, meaning values of roughly 180 - 270, and if we take a look at the histogram in figure 44 we can see that it pretty much agrees with our prevailing wind direction.

We also take a look at the ACF and PACF for the data, figure 45 and figure 46.

If we apply the same approach for the wind direction as we did earlier for the wind speed, we can rather quickly see that a simple AR(3) process is a good starting point.

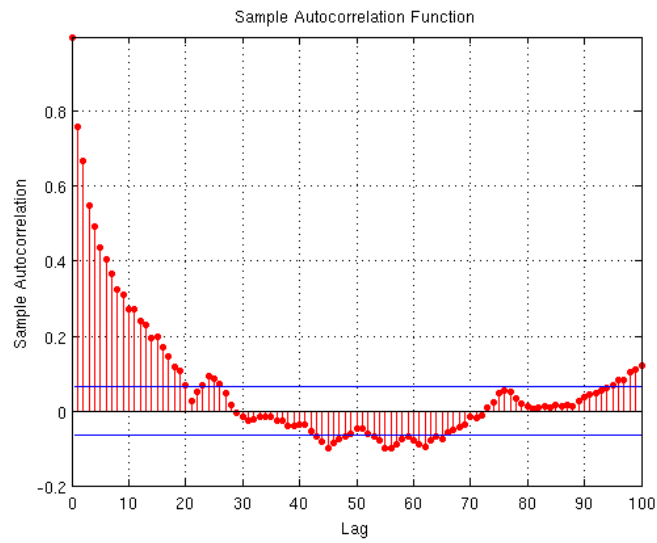


Figure 45: ACF

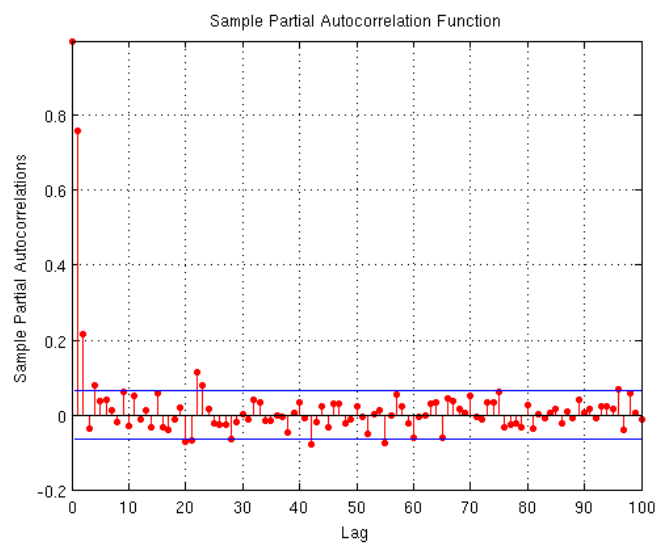


Figure 46: PACF

Discrete-time IDPOLY model: $A(q)y(t) = e(t)$
 $A(q) = 1 - 0.6472 q^{-1} - 0.3289 q^{-2} + 0.001983 q^{-3}$

Estimated using ARMAX
 Loss function 3168.52 and FPE 3194.04
 Sampling interval: 1

The fit is not very good, telling us that our model might not be suitable.
 We take a look at the ACF for the residual and perform our whiteness

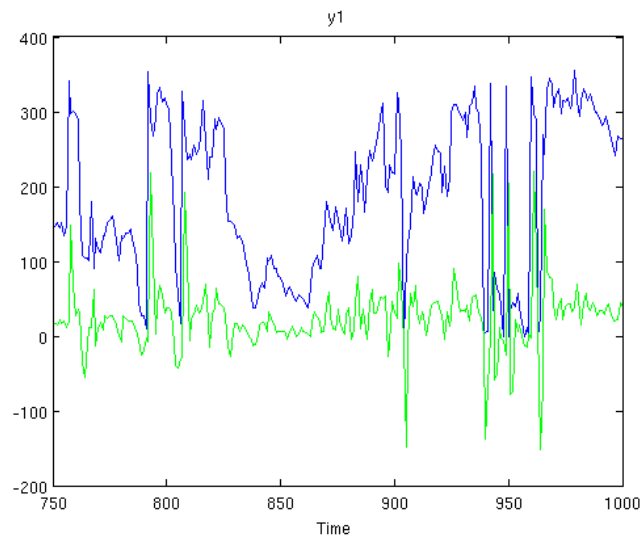


Figure 47: Fit for our predicted samples

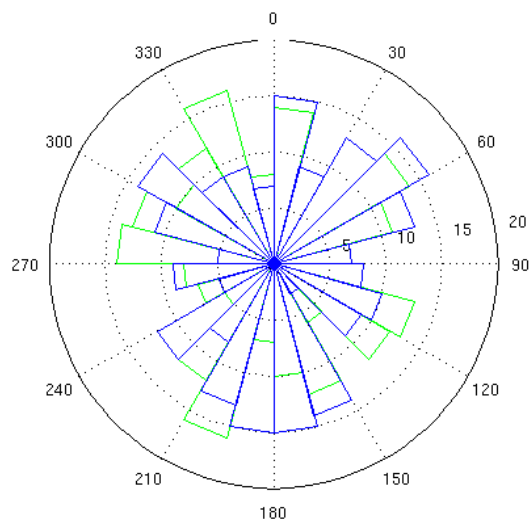


Figure 48: Rose plot comparison

tests nonetheless.

Whiteness test with 5% significance

Ljung-Box-Pierce test: 0 (white if $77.23 < 36.42$)

McLeod-Li test: 0 (white if $41.84 < 36.42$)

Monti test: 0 (white if $64.26 < 36.42$)

Sign change test: 0 (white if 0.41 in $[0.44, 0.56]$)

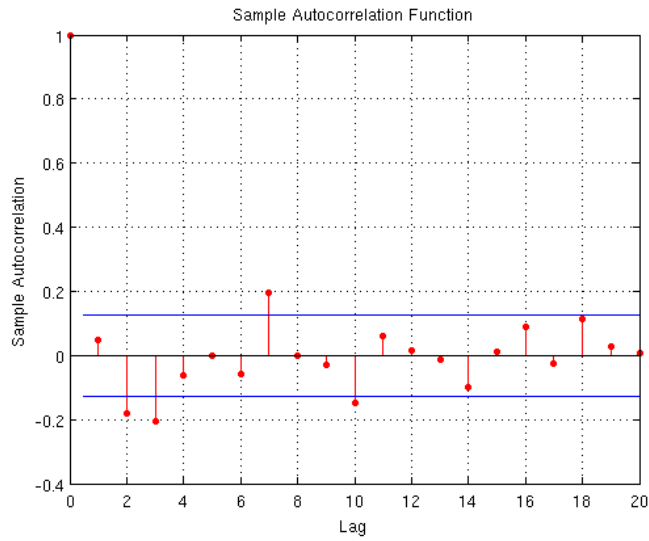


Figure 49: Residual ACF for AR(3)

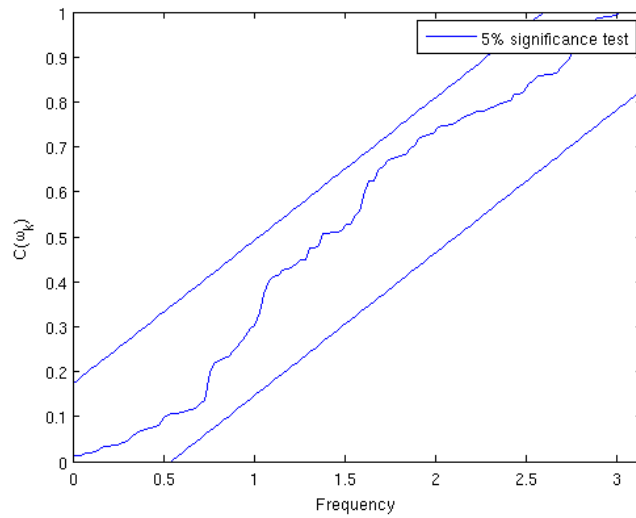


Figure 50: Whiteness plot, AR(3)

After looking at the residual we can confirm that our model is nowhere near as good as we would want. But first things first, as in the case with the wind speeds, the same sort of seasonal filtering might be a good idea to help with the initial fit of the prediction.

Without to much repetition of things, we perform the same kind of filtering, by estimating the seasonal parameters using PEM.

After seasonal filtering, using the same model structure and order AR(3), we get the signal fit seen in figure 51.

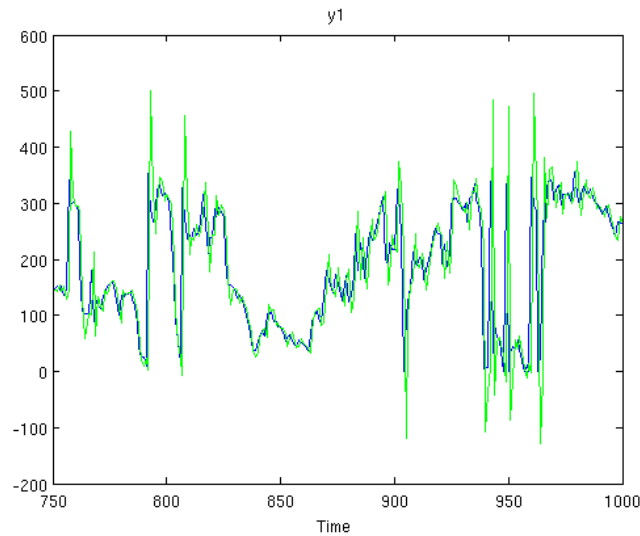


Figure 51: Fit after seasonal filtering, AR(3)

This is looking more like something we can expect. Looking at the ACF for the residual after seasonal filtering in figure 52 we see that the residual is nowhere near white.

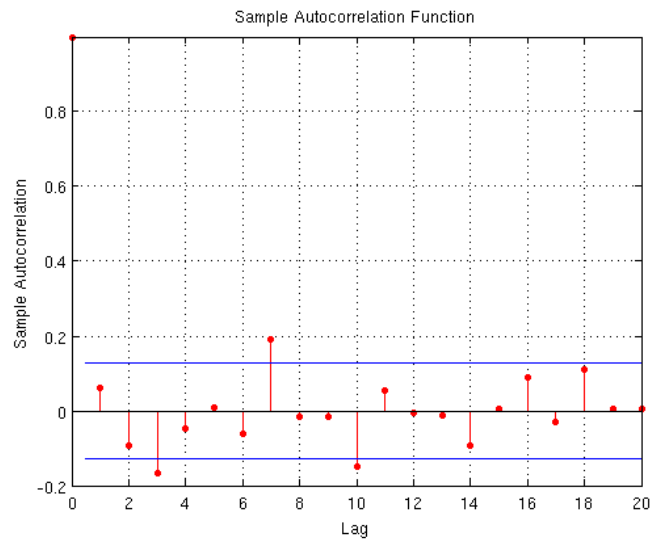


Figure 52: Residual ACF, AR(3)

Looking at the residual we get the indication that adding some MA-terms might be a good idea. But before we start complicating our model, we might want to go back to the ACF for our data and consider the correlation indicated in $k = 5$. If we simply focus on the resulting fit for our predicted model and the ACF for the residual we can power through

a couple of model variations quite fast.

Trying the AR(5) we get the fit seen in figure 53 and the residual ACF seen in figure 54.

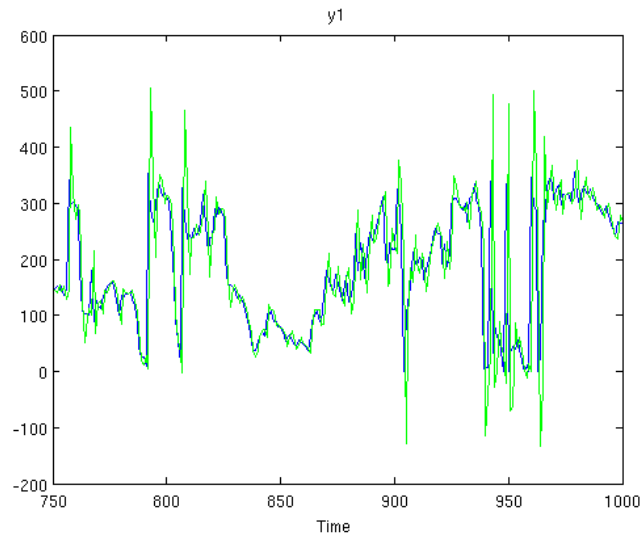


Figure 53: Fit after seasonal filtering, AR(5)

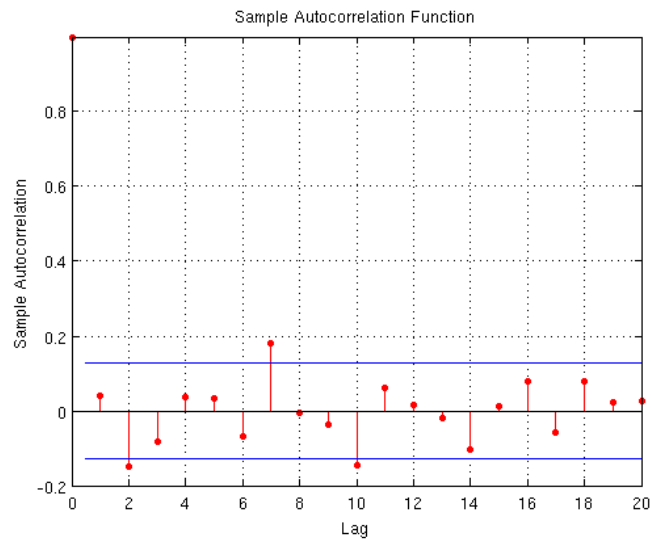


Figure 54: Residual ACF, AR(5)

Not much did change in the residual when adding to the order of the model. We are still left with the feeling that MA-terms are needed.

After some testing we are left with two good candidates, ARMA(5,2) and ARMA(5,7). We get similar results with both these, but the whiteness

tests are not going through and there is still a problem at $k = 7$ where the value is outside the boundaries.

A similar situation that we had with the wind speed that the mean is not stationary might be the issue here as well. Running the RLS for the wind direction as well expecting the ARMA(5,2) and the ARMA(5,7) to perform the best with a slight upperhand for the latter.

The best result was obtained with the ARMA(5,7) model, or more correct, the only model that performed well with the whiteness tests was this model.

Whiteness test with 5% significance

Ljung-Box-Pierce test: 1 (white if $29.81 < 36.42$)

McLeod-Li test: 0 (white if $54.35 < 36.42$)

Monti test: 1 (white if $34.31 < 36.42$)

Sign change test: 1 (white if 0.45 in $[0.44, 0.56]$)

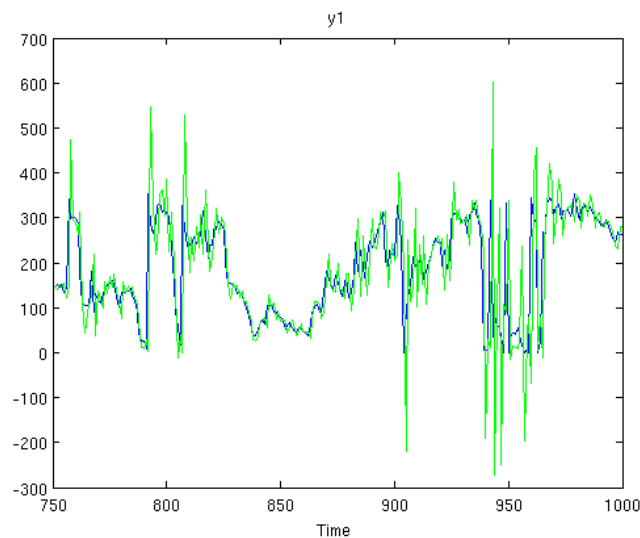


Figure 55: Fit ARMA(5,7)

Figure 55 shows the fit for the model and figure 56 shows a rose plot comparison for the last 250 samples. The residual ACF is seen in figure 57 and by looking at that we can conclude that this model is satisfactory in terms of the residual, we also have the whiteness plot in figure 58 which does not lead us to believe otherwise. On the other hand, looking at the predicted fit in figure 55 once more we can see that we have a problem with the result set not taking into consideration that the wind direction is circular. In short, all values not within the bounds of $0 < x < 360$ is simply not a valid wind direction.

Taking a look at figure 59 where we have used modular 360 on our predicted samples to represent the circular behavior of the wind direction. At a first glance it seems to be very off in its prediction. But after taking

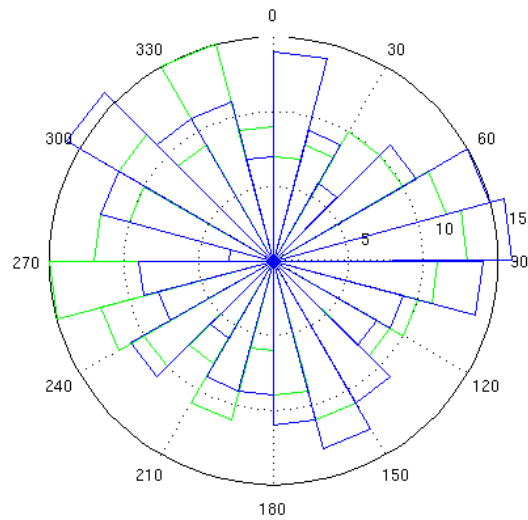


Figure 56: Rose plot comparison, ARMA(5,7)

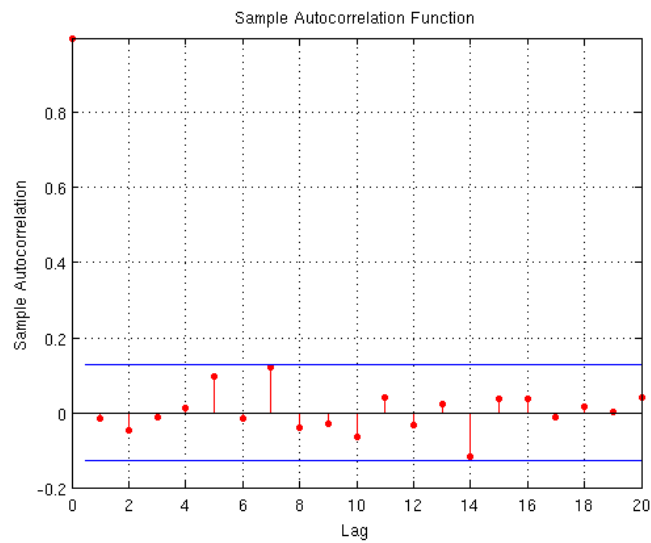


Figure 57: Residual ACF, ARMA(5,7)

a closer look we can see that the directional values are not that far off, the distance from 350 degrees over to 10 degrees is only 20 degrees, but in the plot it looks like a lot more.

Looking back at our initial conclusion that the model was satisfactory in terms of the whiteness of the residual, we can add to this that we are satisfied with our predicted fit as well.

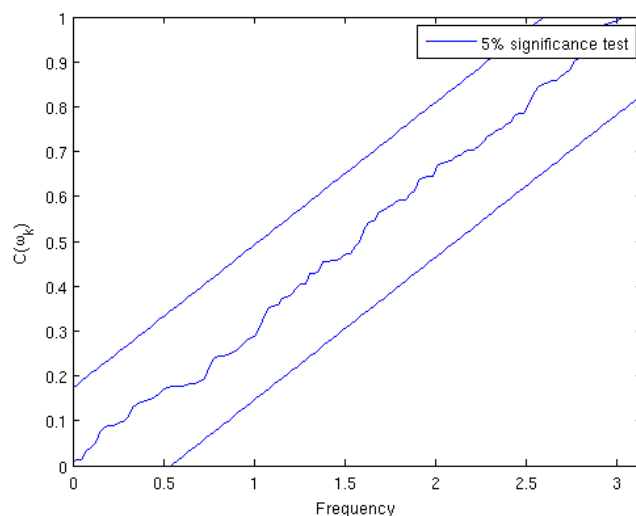


Figure 58: Whiteness plot

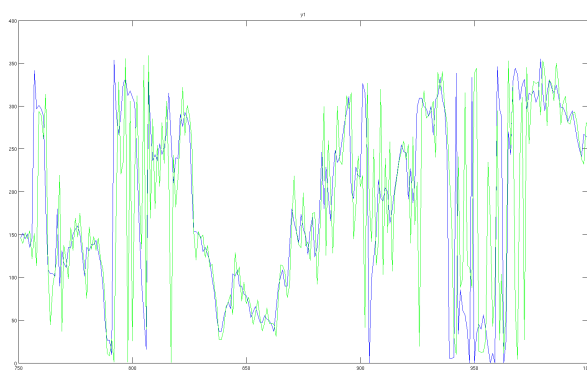


Figure 59: Fit for recursive ARMA(5,7) modular 360

5 Spatial Dependency

If we look at the wind speed data for both Kastrup airport and Sturup airport for the year 2011 at pressure level 850mb, both seen in figure 60, we can clearly see some similarity. Looking at the cross correlation between the two, figure 61, we can see that the data for Kastrup seems to be pretty much identical to that of Sturup, except it is shifted approximately 35 samples. 35 samples is roughly 9 days.

Since the the grid point separation is 1.25° by 1.25° , which is roughly 140km by 140km on the equator, the use of spatial dependency seems less useful the closer you are to the equator.

On the other hand, when looking at the correlation between Copenhagen

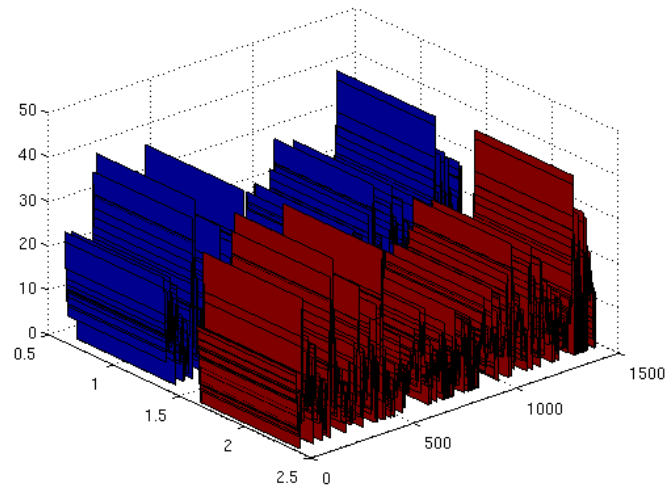


Figure 60: Wind speed for Sturup and Kastrup

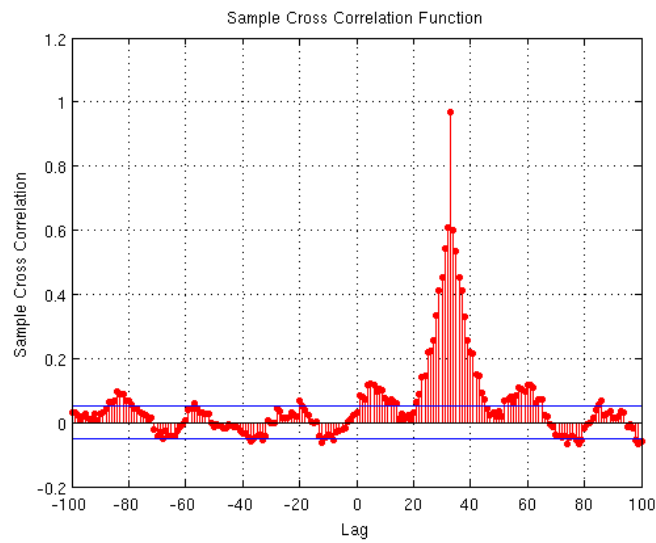


Figure 61: Cross correlation between Sturup and Kastrup

and Malmoe, as well as Los Angeles and San Diego, figures 62 and 63, we can come to the conclusion that spatial dependency is already accounted for in the models used to generate the forecasts. As a result of this, there should not be any need for us to account for spatial dependency a second time if we can achieve a good enough model.

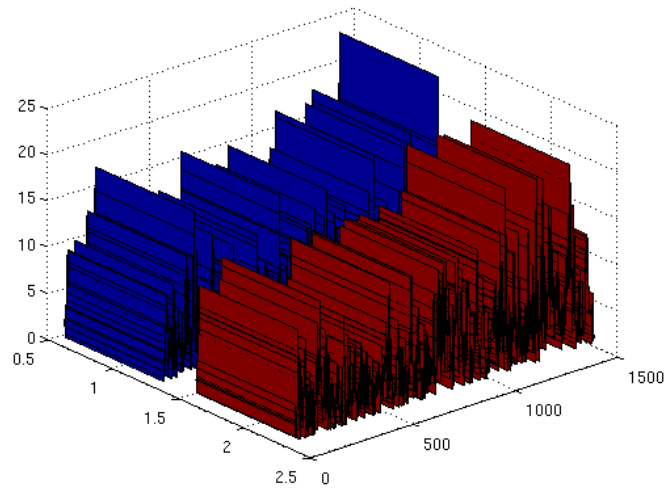


Figure 62: Wind speed for Los Angeles and San Diego

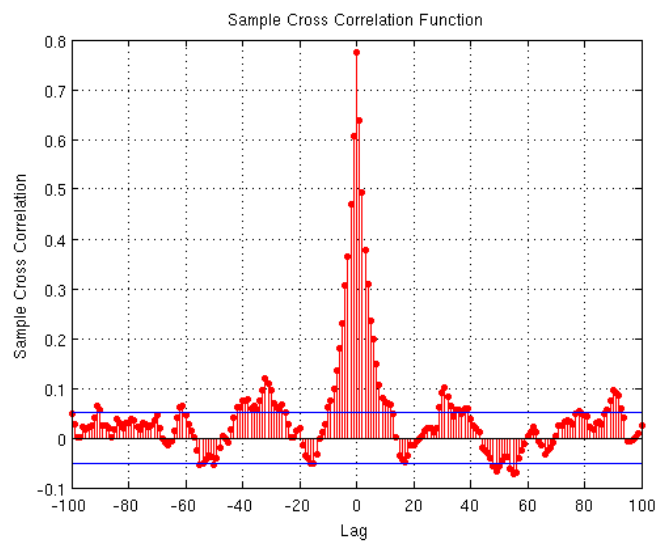


Figure 63: Cross correlation between Los Angeles and San Diego

6 Conclusion and discussion

It is known that predicting long term wind speed with less than a year of data can be subject to large errors.

Recursive estimation seems like the way to go for both the speed and direction, since the mean will not be stationary for any of those.

As conclusion we can say that a recursive model, initially generated from the approximately 2 years of data available at the company, that

updates whenever there is new samples available will give a more than satisfactory result for the purpose of the companys use. As well as the fact that the models will contingiously improve, there is also no need to retrieve any new sources of data.

6.1 Additional Sources

Given the purpose of the company use of the data, a seasonal mean data set would be very good since this will provide reliable models in case the live weather sevices are not operational. One big problem with this is the availability of such data that can be used to create a seasonal mean data set with the resolution we require. We have been in contact with MetOffice and by the time of writing (mid 2012) they are 12-18 months away from being able to provide historical GRIB data.

Another option is to use the current normal period of statistical weather data. The problem here will most likely lie in the resolution, but this is unconfirmed since it has proven difficult to recieve a sample from that data.

7 Topics of further research

- How does a multivariate approach affect the models.
- Is there an algorithm or a numerical approach suitable for automating the generation of models based on the method described in this paper.

References

- [1] Andreas Jakobsson, *Time Series Analysis and Signal Modeling*. Lund, Version 110511, 2011.
- [2] Daniel S Wilks, *Statistical Methods in the Atmospheric Sciences*. Oxford, Academig Press, 2nd Edition, 2006.
- [3] Henrik Madsen, *Time Series Analysis*. Boca Raton, Chapman & Hall/CRC, 2008.
- [4] Monson H. Hayes, *Statistical Digital Signal Processing and Modelling*. Hoboken, John Wiley & Sons, INC, 1996.
- [5] Simon Haykin, *Adaptive Filter Theory*. New Jersey, Prentice Hall, 4th Edition, 2002.
- [6] Sanjit K. Mitra, *Digital Signal Processing, A Computer-based Approach*. New York, McGraw-Hill, 3rd Edition, 2006.
- [7] Yanwei Wang, Jian Li and Petre Stoica, *Spectral Analysis, The Missing Data Case*. Morgan & Claypooll, 2005.
- [8] *FM 92-XII Ext. GRIB - General Regularly-distributed Information in Binary form*. WMO, 2nd Edition, 2003.
- [9] Paul Gipe, *Wind Power: Renewable Energy for Home, Farm, and Business*. Chelsea Green Publishing, 2004.
- [10] Wim Bierbooms, *Wind and wave conditions*. Section Wind Energy, Delft University of Technology, DOWEC 47 rev. 2.

A GAPES images

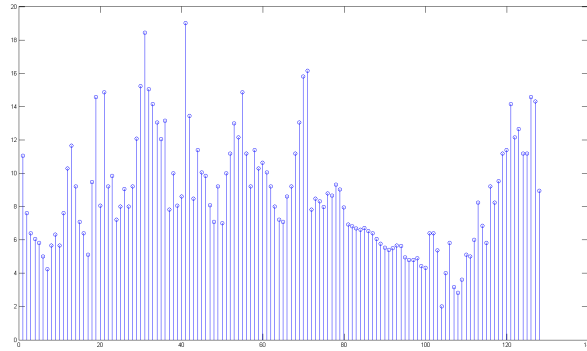


Figure 64: GAPES 10% filter length

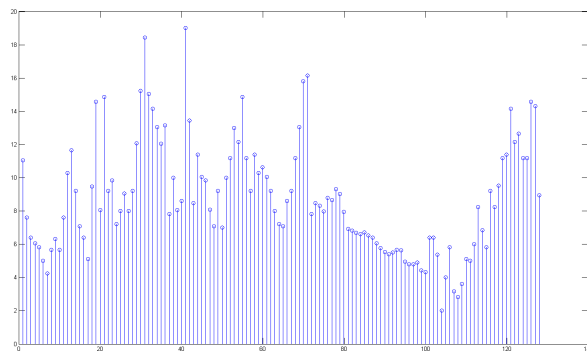


Figure 65: GAPES 20% filter length

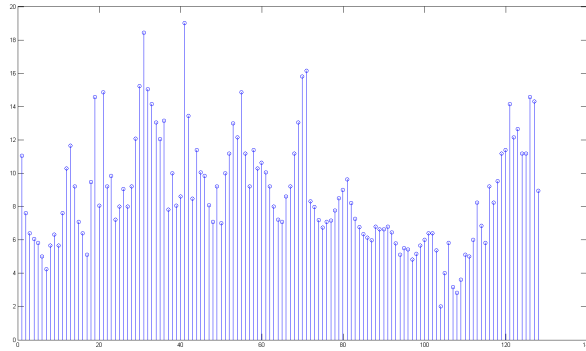


Figure 66: GAPES 30% filter length

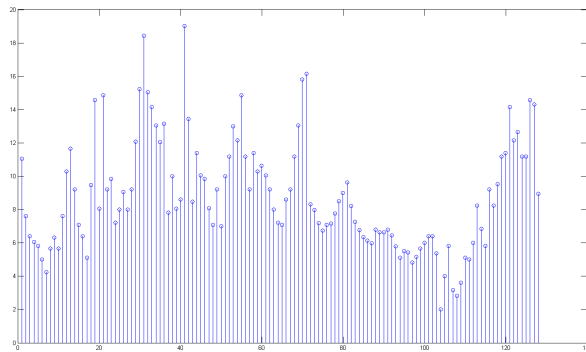


Figure 67: GAPES 50% filter length

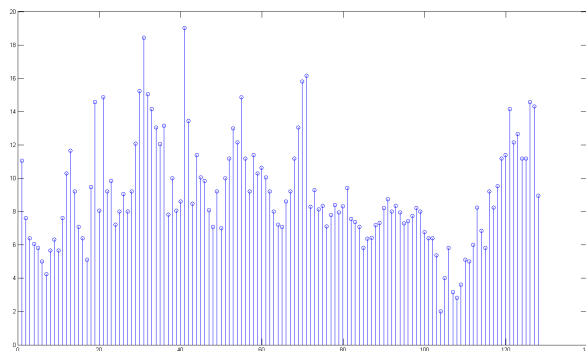


Figure 68: GAPES 32% filter length and 32% initial filter

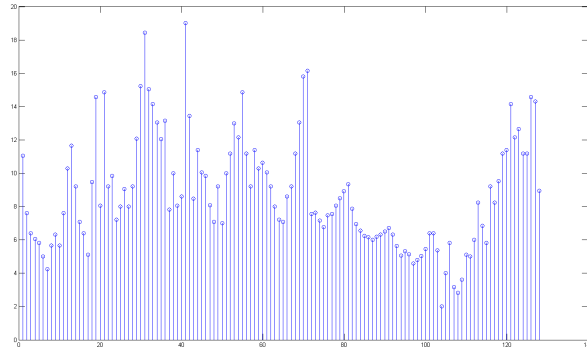


Figure 69: GAPES 50% filter length and 12% initial filter

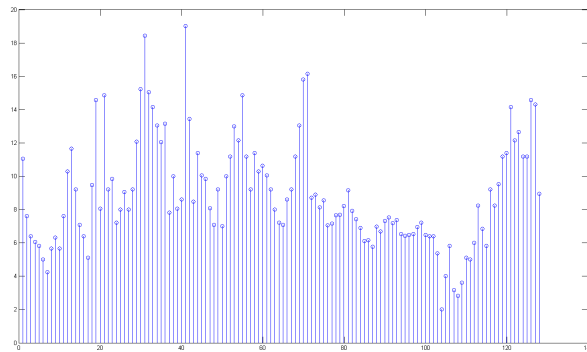


Figure 70: GAPES 50% filter length and 25% initial filter

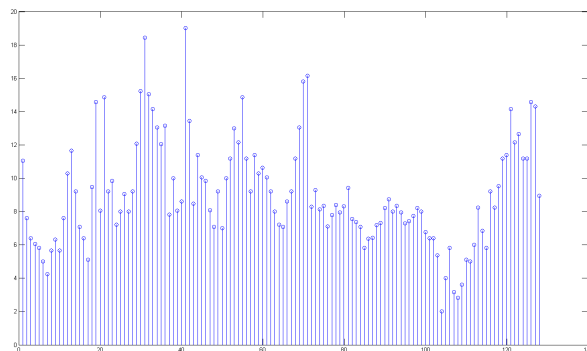


Figure 71: GAPES 50% filter length and 32% initial filter

B Wind Speed Model Testing For Different Pressure Levels

B.0.1 Pressure Level 700mb

Whiteness test with 5% significance

Ljung-Box-Pierce test: 1 (white if $26.94 < 36.42$)

McLeod-Li test: 1 (white if $27.19 < 36.42$)

Monti test: 1 (white if $30.67 < 36.42$)

Sign change test: 0 (white if 0.43 in $[0.45, 0.55]$)

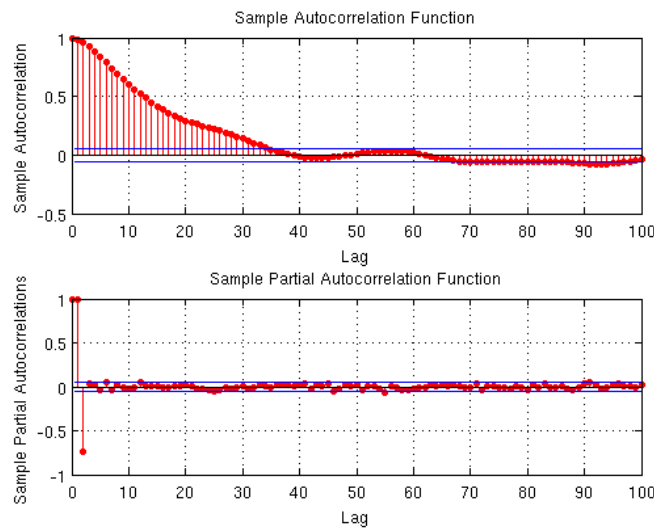


Figure 72: ACF and PACF for pressure 700mb

B.0.2 Pressure Level 600mb

Whiteness test with 5% significance

Ljung-Box-Pierce test: 1 (white if $21.50 < 36.42$)

McLeod-Li test: 1 (white if $30.55 < 36.42$)

Monti test: 1 (white if $26.39 < 36.42$)

Sign change test: 0 (white if 0.43 in $[0.45, 0.55]$)

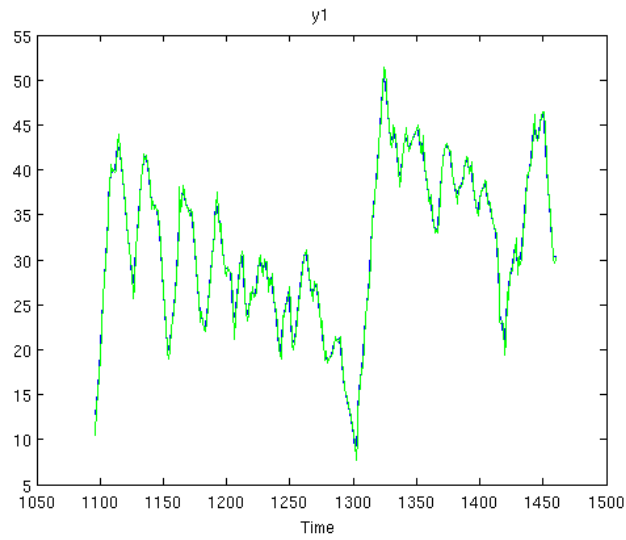


Figure 73: Fit after recursive estimation for 700mb, AR(3)

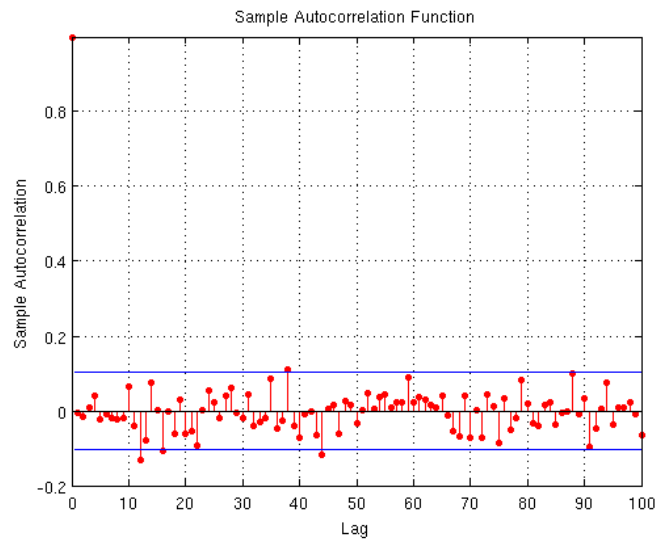


Figure 74: Residual ACF recursive AR(3), 700mb

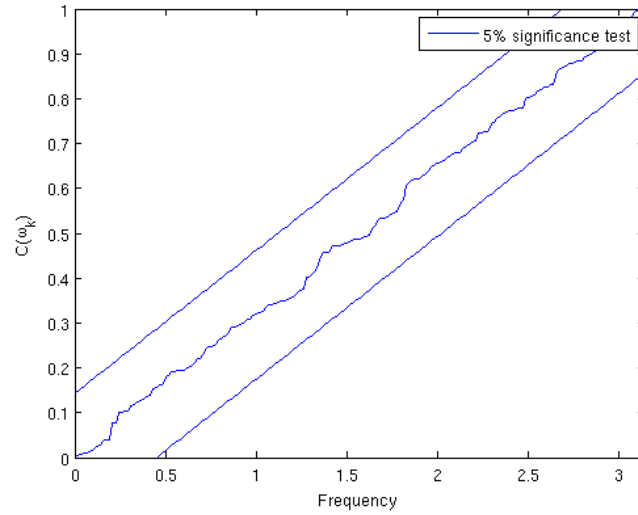


Figure 75: Whiteness plot for the residual, 700mb

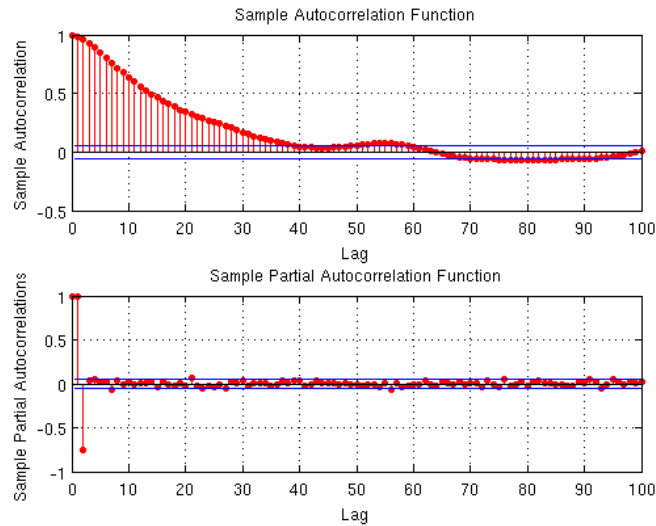


Figure 76: ACF and PACF for pressure 600mb

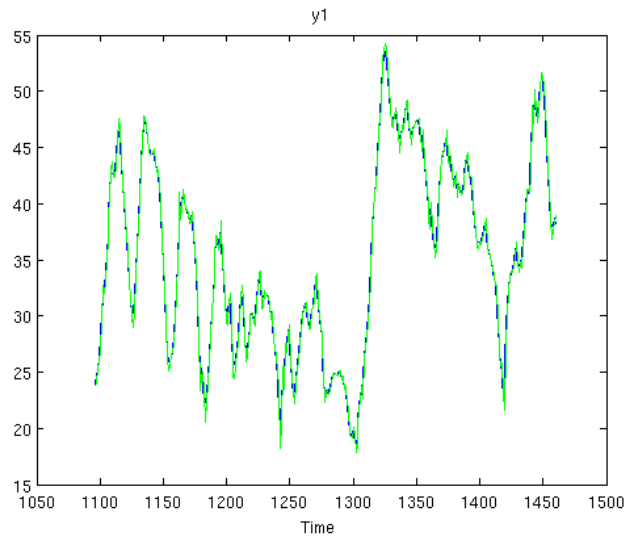


Figure 77: Fit after recursive estimation for 600mb, AR(3)

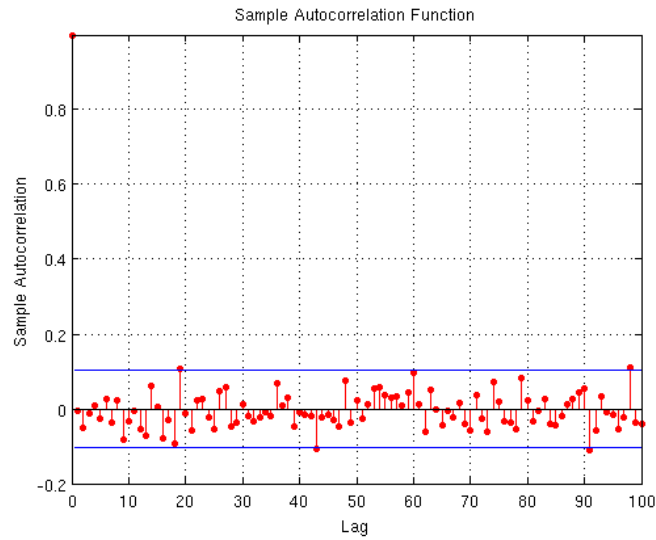


Figure 78: Residual ACF recursive AR(3), 600mb

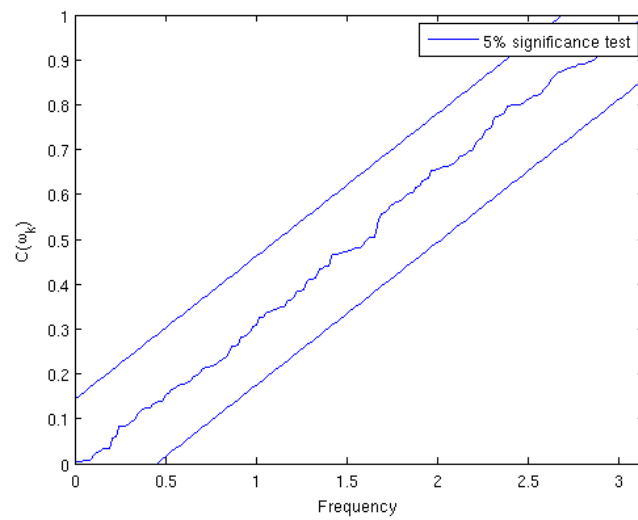


Figure 79: Whiteness plot for the residual, 600mb

B.0.3 Pressure Level 500mb

Whiteness test with 5% significance

Ljung-Box-Pierce test: 1 (white if $21.92 < 36.42$)

McLeod-Li test: 1 (white if $18.88 < 36.42$)

Monti test: 1 (white if $26.58 < 36.42$)

Sign change test: 0 (white if 0.43 in $[0.45, 0.55]$)

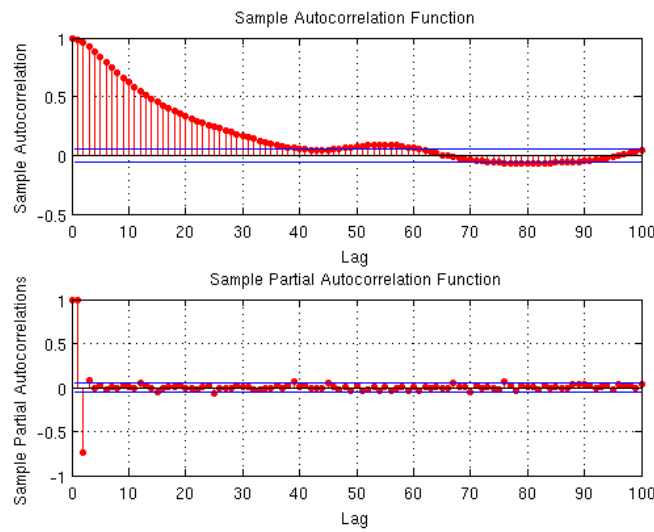


Figure 80: ACF and PACF for pressure 500mb

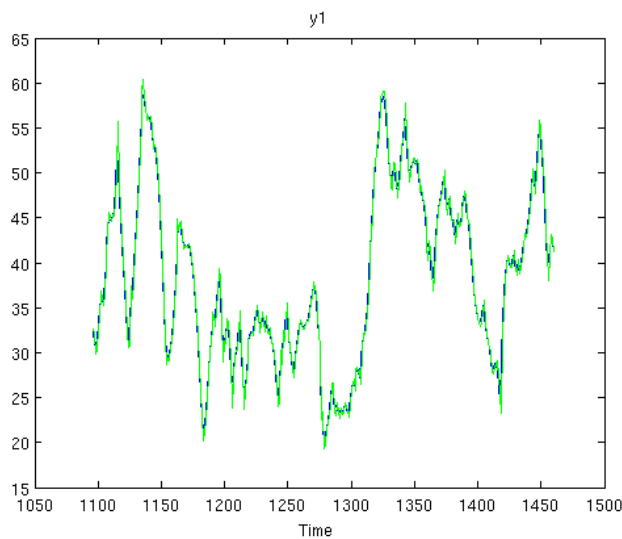


Figure 81: Fit after recursive estimation for 500mb, AR(3)

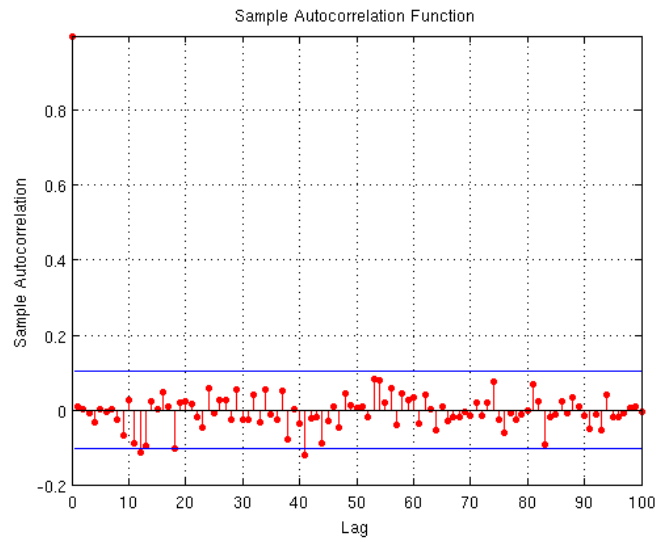


Figure 82: Residual ACF recursive AR(3), 500mb

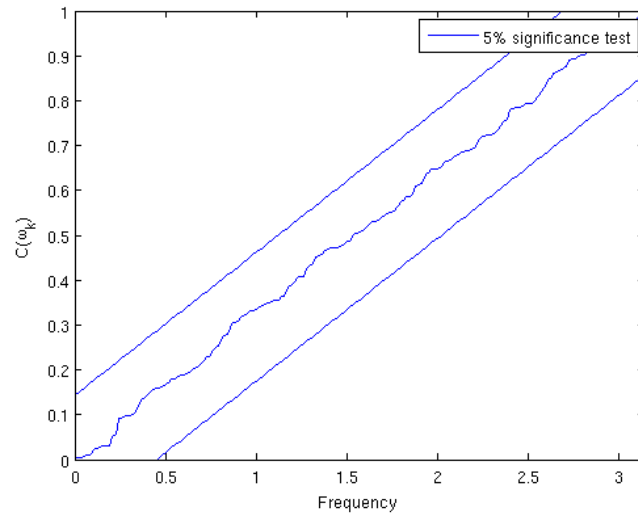


Figure 83: Whiteness plot for the residual, 500mb

B.0.4 Pressure Level 400mb

Whiteness test with 5% significance

Ljung-Box-Pierce test: 1 (white if $35.99 < 36.42$)

McLeod-Li test: 1 (white if $10.72 < 36.42$)

Monti test: 0 (white if $44.78 < 36.42$)

Sign change test: 1 (white if 0.45 in $[0.45, 0.55]$)

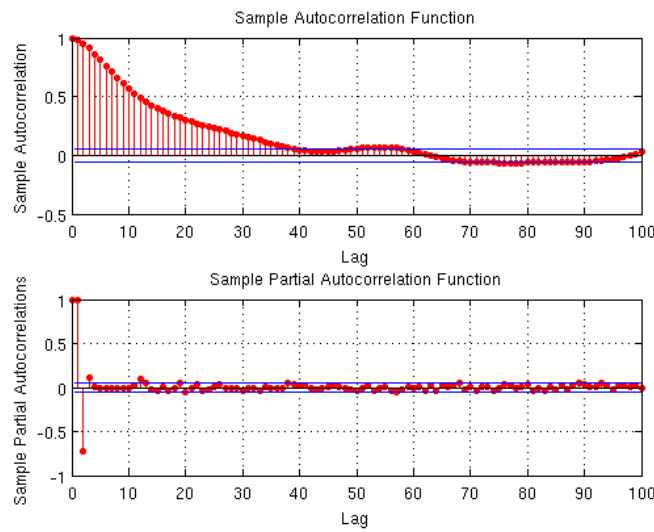


Figure 84: ACF and PACF for pressure 400mb

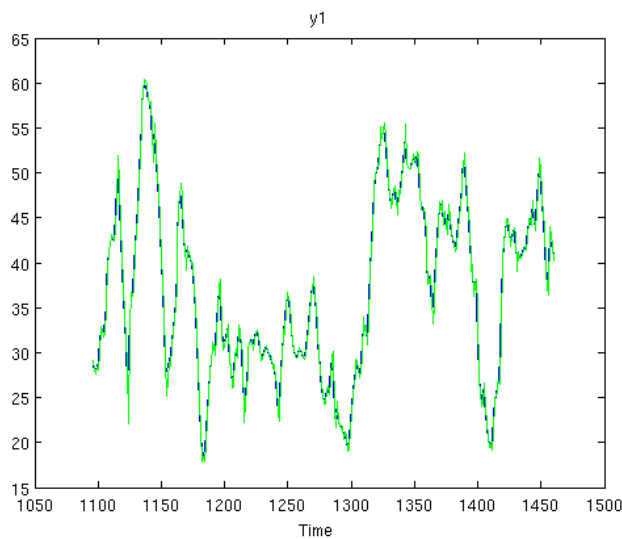


Figure 85: Fit after recursive estimation for 400mb, AR(3)

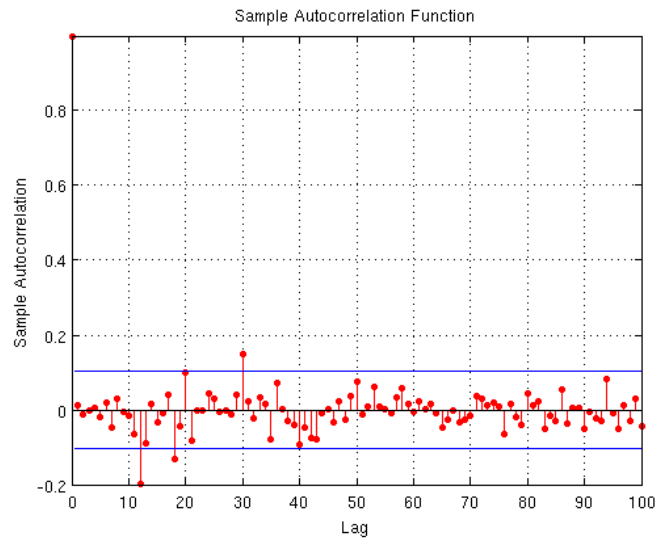


Figure 86: Residual ACF recursive AR(3), 400mb

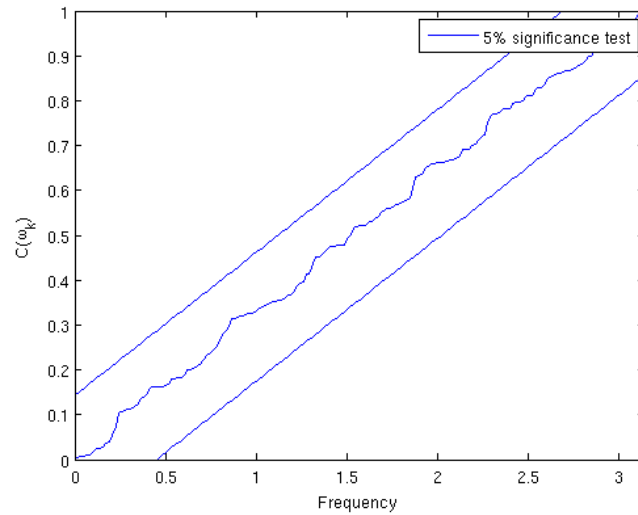


Figure 87: Whiteness plot for the residual, 400mb

B.0.5 Pressure Level 300mb

Whiteness test with 5% significance

Ljung-Box-Pierce test: 1 (white if $25.28 < 36.42$)

McLeod-Li test: 1 (white if $33.44 < 36.42$)

Monti test: 1 (white if $32.61 < 36.42$)

Sign change test: 0 (white if 0.41 in $[0.45, 0.55]$)

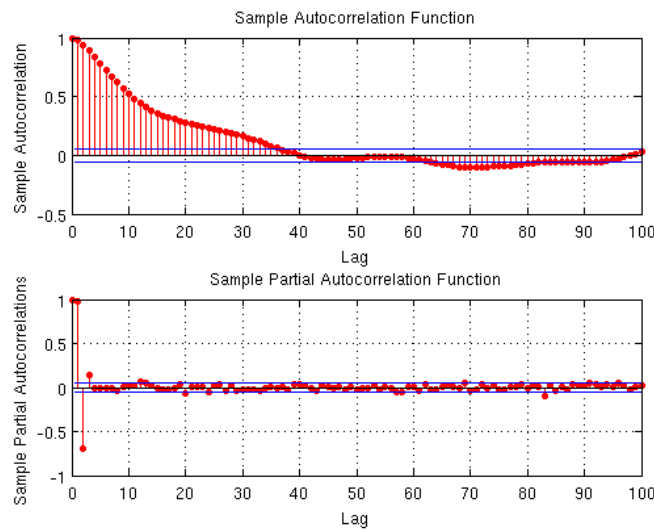


Figure 88: ACF and PACF for pressure 300mb

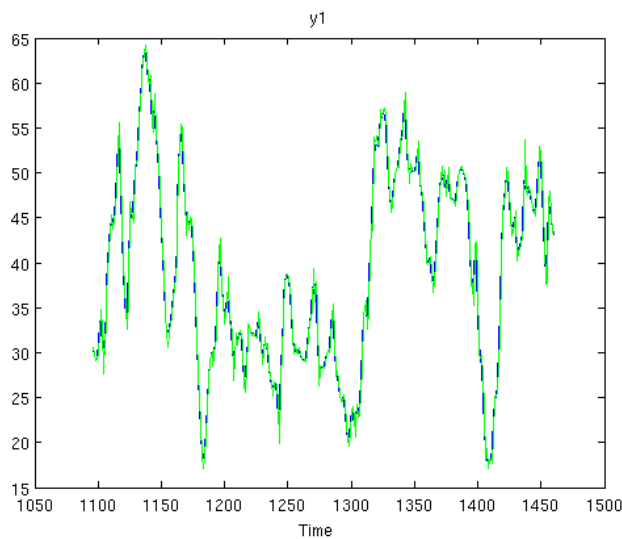


Figure 89: Fit after recursive estimation for 300mb, AR(3)

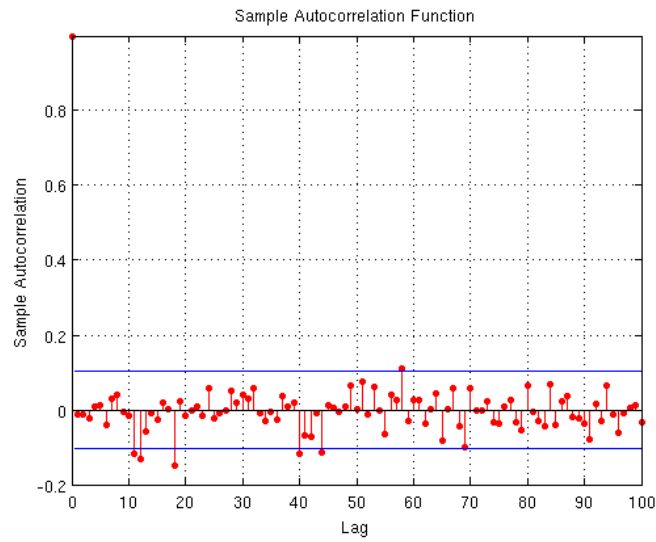


Figure 90: Residual ACF recursive AR(3), 300mb

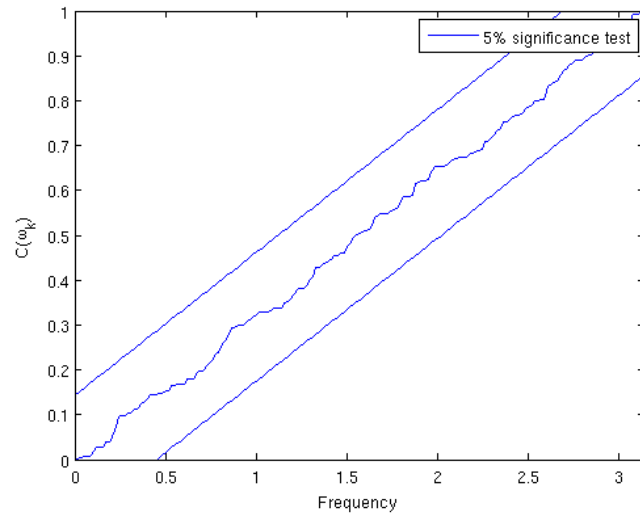


Figure 91: Whiteness plot for the residual, 300mb

B.0.6 Pressure Level 200mb

Whiteness test with 5% significance

Ljung-Box-Pierce test: 1 (white if $16.45 < 36.42$)

McLeod-Li test: 0 (white if $74.46 < 36.42$)

Monti test: 1 (white if $21.23 < 36.42$)

Sign change test: 0 (white if 0.41 in $[0.45, 0.55]$)

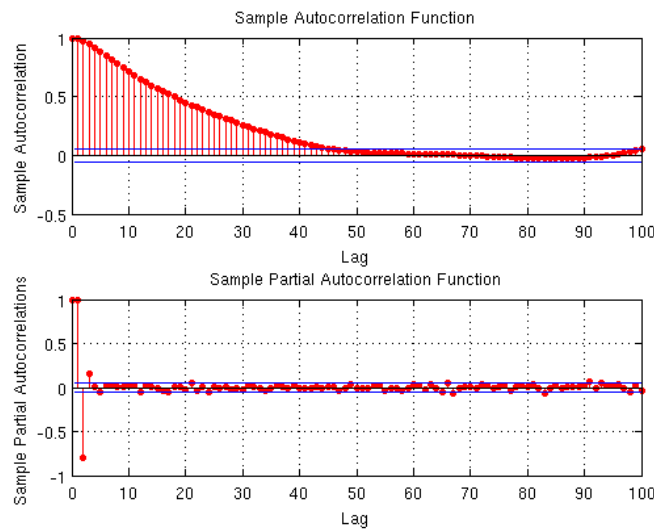


Figure 92: ACF and PACF for pressure 200mb

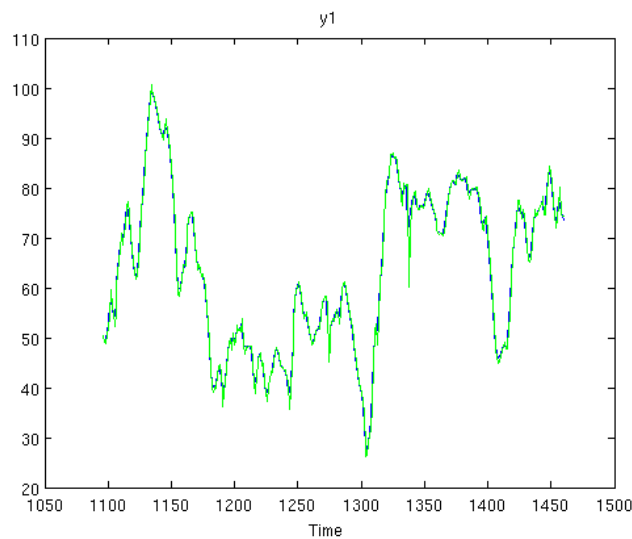


Figure 93: Fit after recursive estimation for 200mb, AR(3)

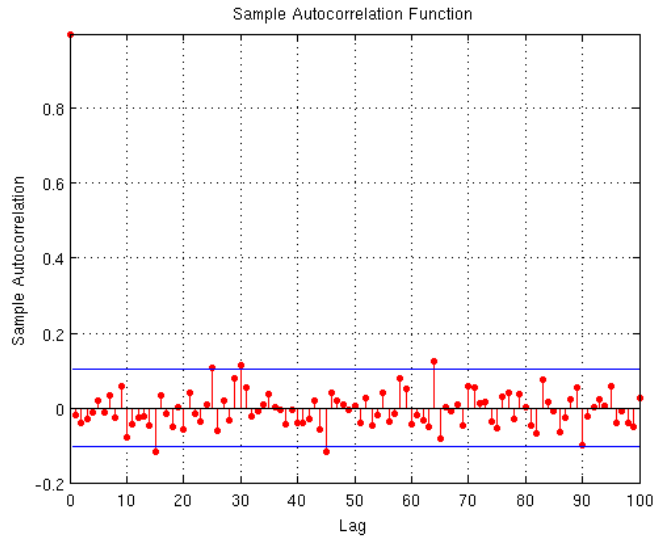


Figure 94: Residual ACF recursive AR(3), 200mb

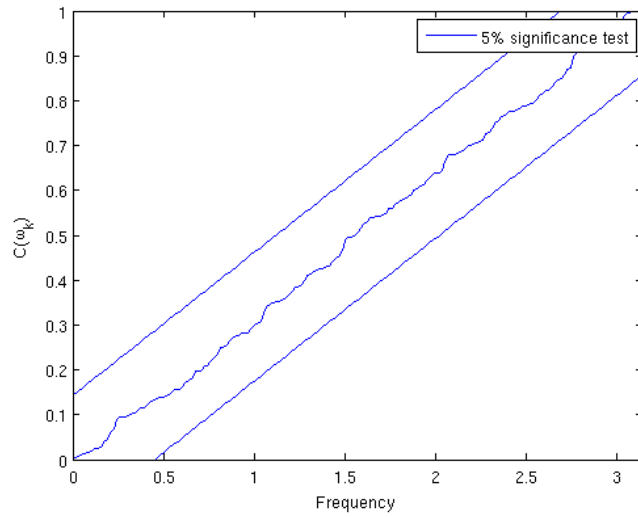


Figure 95: Whiteness plot for the residual, 200mb

C Interpolation Algorithm

```
% A counter for the number of 0-values to replace
counter = 0;
% The non-zero value that comes before a value to replace
nonZeroStart = 0;

% Loops for traversing the matrix column by column
for i=1:n
    for j=1:m-1
        % Case for when there is only one consecutive zero
        if ( data(j,i) == 0 && data(j+1,i) ~= 0 && counter == 0 )
            data(j,i) = (nonZeroStart + data(j+1,i)) / 2;
        % Several consecutive zeros
        elseif ( data(j,i) == 0 && data(j+1,i) == 0 )
            counter = counter + 1;
        % When reaching a non-zero value after several consecutive zeros
        elseif ( data(j,i) == 0 && data(j+1,i) ~= 0 )
            counter = counter + 1;
            nonZeroEnd = data(j+1,i);
            increment = (nonZeroEnd - nonZeroStart) / (counter+1);
            value = nonZeroStart;
            for k=1:counter
                data(j-counter+k,i) = value + increment;
                value = value + increment;
            end
            counter = 0;
        else
            nonZeroStart = data(j,i);
        end
    end
end
end
```

D Altitude to Pressure Conversion Table

Barometric Pressure

The altitude to pressure conversion table is derived using the US Standard Atmosphere 1976 barometric formula for geopotential altitude with the following physical constants and multi-valued constants for sea level up to a height of 36,089 ft (11,000m)

$$P = P_b \left[\frac{T_b}{T_b + L_b(h - h_b)} \right]^{\frac{g_0 M}{R^* L_b}} \quad (21)$$

P_b = Standard atmospheric pressure at sea level, 1013.25 mb

T_b = Standard temperature (K), 288.15K (15°C)

L_b = Standard temperature lapse rate $-0.0065(K/m)$ in ISA

h = Height above sea level (meters)

h_b = Height at bottom of layer b (meters; e.g., $h_1 = 11000$ meters)

R^* = Universal gas constant for air: 8.31432 Nm/(molK)

g_0 = Gravitational acceleration (9.80665 m/s²)

M = Molar mass of Earth's air (0.0289644 kg/mol)

ft	m	mb/hPa	psi	mmHg	inHg
0	0	1013.25	14.696	760.00	29.921
50	15.240	1011.42	14.669	758.63	29.867
100	30.480	1009.59	14.643	757.26	29.813
200	60.960	1005.95	14.590	754.52	29.706
300	91.440	1002.31	14.537	751.80	29.598
400	121.92	998.689	14.485	749.08	29.491
500	152.40	995.075	14.432	746.37	29.385
600	182.88	991.472	14.380	743.67	29.278
700	213.36	987.880	14.328	740.97	29.172
800	243.84	984.298	14.276	738.28	29.066
900	274.32	980.727	14.224	735.61	28.961
1000	304.80	977.166	14.173	732.93	28.856
2000	609.60	942.129	13.664	706.65	27.821
3000	914.40	908.117	13.171	681.14	26.817
4000	1219.2	875.105	12.692	656.38	25.842
5000	1524.0	843.073	12.228	632.36	24.896
6000	1828.8	811.996	11.777	609.05	23.978
7000	2133.6	781.854	11.340	586.44	23.088
8000	2438.4	752.624	10.916	564.51	22.225
9000	2743.2	724.285	10.505	543.26	21.388
10000	3048.0	696.817	10.106	522.66	20.577
15000	4572.0	571.820	8.2935	428.90	16.886
20000	6096.0	465.633	6.7534	349.25	13.750
25000	7620.0	376.009	5.4536	282.03	11.104
30000	9144.0	300.896	4.3641	225.69	8.8855
35000	10668	238.423	3.4580	178.83	7.0406