Stockholms universitet

SARIMA versus naive models, forecasting the weather with time series analysis

Jenny Holm

Kandidatuppsats i matematisk statistik
Bachelor Thesis in Mathematical Statistics

Kandidatuppsats 2020:19
Matematisk statistik
September 2020

www.math.su.se

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm

Mathematical Statistics
Stockholm University
Bachelor Thesis **2020:19**
http://www.math.su.se

SARIMA versus naive models, forecasting the weather with time series analysis

Jenny Holm*

June 2020

Abstract

In this thesis we apply univariate time series analysis on monthly mean weather data. By comparing SARIMA models with a naive approach, we evaluate the forecast performance for the air temperature, wind speed and precipitation in the south of Sweden. To stabilize the variance of the monthly mean wind speed and precipitation we apply Box-Cox transformations on the time series. It is concluded that the SARIMA models have better forecast performance during a 159 months forecast horizon than the naive models.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: jeho4048@student.su.se. Supervisor: Mathias Millberg Lindholm, Kristoffer Lindensjö och Felix Wahl.

1 Acknowledgements

This is a Bachelor's thesis of 15 ECTS in Mathematical Statistics at the Department of Mathematics at Stockholm University. I would like to thank my supervisor Mathias Millberg Lindholm for his encouragements and guidance throughout the Bachelor thesis.

I would also like to thank my partner Johannes Borg for his fantastic support, Dasha Medvedeva for her inputs and my rescue dog Digger for being cute.

Contents

2 Introduction

Forecasting the weather is a highly relevant subject nowadays with the current awareness of climate change. There is extensive research in this field, for example the Rossby Centre, a climate model research unit at the Swedish Meteorological and Hydrological Institute (SMHI), creates climate scenarios using a supercomputer and very complex models. The models are based upon emission or radiation scenarios which impacts the climate. The climate scenarios are not weather forecasts however, they are more focused on the long-term changes. Weather forecasts on the other hand are relevant for short-term predictions *(SMHI, n.d (a))*. In this study we utilize univariate time series analysis to examine three basic weather components: Temperature, wind speed, and precipitation.

When forecasting the weather, the naive approach is one of the most common. The air temperature in Sweden is very dependent on the current season, with temperatures below 0 Celsius in the winter and above 20 Celsius in the summer. The seasonal naive approach assumes on a mothly basis that the the conditions for a specific month are the same every year. With focus on the SARIMA class of models we explore if the more advanced models provide better forecasts than the naive approach.

When forecasting into the future the possibility to check the preciseness of the predicted values are lost. To measure the forecast performance of models we perform out-of-sample forecasts for periods between 1966 to 2019 where we have known observations to compare against. By using univariate time series analysis we explore the forecasting of the monthly mean weather of air temperature, wind speed and precipitation. How well does the considered forecast models perform, and are they better than the naive approach?

3 Theory

In this section definitions and other vocabulary applied in the method and analysis are explained. Unless stated otherwise, the theory comes from *(Tsay, Chapter 2, 2010)*.

3.1 Stationarity

A time series $\{X_t\}$ consists of a series of observations x_t ordered by the indexed time t. Stationary time series refers to time series for which the change over time are constant, for example if it grows linearly. For a strictly stationary time series $\{X_t\}$, the joint distribution of $(X_{t_1}, ..., X_{t_k})$ does not vary over time. More common is the assumption of the weaker stationarity, where the mean $\mathrm{E}[X_t] = \mu$ (μ constant) and the covariance $\mathrm{Cov}(X_t, X_{t-l}) = \gamma_l$ does not vary over time for some integer l. The covariance γ_l is called the lag-l autocorrelation of X_t. For $l = 0$ the lag-0 autocorrelation is $\gamma_0 = \mathrm{Var}(X_t)$ *(Tsay, 2010. p.30)*.

3.2 Transformations

Sometimes when the time series does not meet the criteria or assumptions of for example weakly stationarity then a transformation may be in order.

3.2.1 Differencing

Transforming non-stationary series to stationary can be done by differencing as many times as necessary. We introduce the backshift operator for a time series $\{X_t\}$ given by $BX_t = X_{t-1}$. We refer to $\Delta X_t = X_t - X_{t-1}$ as a regular differencing. For $d = 1, 2, ...$ times differencing, we can then express

$$\Delta^d X_t = (1 - B)^d X_t.$$

We are also able to difference at lags of higher order. For a differencing at lag-s where $s > 1$, we use $\Delta_s X_t = X_t - X_{t-s}$ which then for $D = 1, 2, ...$ times differencing is

$$\Delta_s^D X_t = (1 - B^s)^D X_t.$$

For example, for a yearly seasonal differencing on monthly data, one uses $s = 12$ and $D = 1, 2, ...$, for each new yearly differencing. In the SARIMA model described in *Section 3.8* the concept of seasonal differencing occurs.

3.2.2 Box-Cox transformation

One of the assumptions of stationarity is constant variance, thus time series which exhibit changing variance with time are non-stationary. Also then a transformation of the time series might be desirable *(Tommaso and Helmut, 2011)*. The Box-Cox family of power transformations applies a parameter λ such that

$$X_t(\lambda) = \begin{cases} \frac{X_t^\lambda - 1}{\lambda} & , \lambda \neq 0 \\ \ln X_t & , \lambda = 0 \end{cases} \quad , \text{for all } t \geq 0$$

so for $\lambda = 1$ the original data scale is used with a shift of -1. The parameter λ is estimated from the data, see *Appendix A: Section 8.1.1*.

The Box-Cox transformation may yield a stationary process, but if it still is non-stationary then other methods can be applied, for example differencing described in *Section 3.2.1*.

3.2.3 Seasonal adjustment

Data that is seasonal can be adjusted by removing the seasonal component. This is a way of handling the seasonality of data. The additive model is applied for data which does not increase/decrease over time, otherwise a multiplicative model is more appropriate. For the additive model, one can decompose the time series $\{X_t\}$ into $X_t = T_t + S_t + C_t + I_t$, where T_t represents the trend, S_t the seasonality, C_t the cycle and I_t the irregular decomposition *(Linde, 2005)*. This method will be discussed but not applied.

3.3 White noise

The white noise series $\{a_t\}$ is defined by a sequence of independent identically distributed (i.i.d) random variables with finite mean and variance *(Tsay, 2010)*. Further the Gaussian white noise series follows a normal distribution with zero mean and constant variance σ_a^2.

For *Section 3.4-3.7* the presentation of the terms in the AR and MA polynomials, may differ from what is used in *(Tsay, 2010)*. This notation was preferred since the `R` function `Arima` used in the methodology *Section 4* is theoretically based on the notations that are presented.

3.4 AR

We define the general Autoregressive model, denoted AR(p) with p parameters, by

$$X_t = c + \sum_{i=1}^{p} \phi_i X_{t-i} + a_t,$$

where $\{a_t\}$ is a white noise series and c is a constant *(Tsay, 2010. p.46)*.

With the backshift operator we can also write the AR(p) as

$$(1 - \sum_{i=1}^{p} \phi_i B^i) X_t = c + a_t.$$

3.5 MA

The moving averages model, MA(q) with q parameters, is generalized to

$$X_t = c + a_t + \sum_{j=1}^{q} \theta_j a_{t-j},$$

with $q > 0$ and where $\{a_t\}$ is a white noise series and c is a constant. With the backshift operator we can also write the MA(q) as

$$X_t = c + (1 + \sum_{j=1}^{q} \theta_j B^j) a_t.$$

The MA(q) model exhibits time invariant mean, $\mathrm{E}[X_t] = c$, and variance $\mathrm{Var}(X_t) = (1 + \sum_{j=1}^{q} \theta_j^2)\sigma_a^2$, thus the model is always weakly stationary *(Tsay, 2010. p.59)*.

3.6 ARMA

The combining of the AR and MA models gives way for the Autoregressive Moving Averages model, ARMA(p, q) which is generalized with the parameters $p \geq 0$ and $q \geq 0$ as

$$X_t = c + \sum_{i=1}^{p} \phi_i X_{t-i} + a_t + \sum_{i=1}^{q} \theta_i a_{t-i},$$

where $\{a_t\}$ is a white noise series. With the backshift operator we can write the generalized ARMA(p, q) model as:

$$(1 - \sum_{i=1}^{p} \phi_i B^i) X_t = c + (1 + \sum_{j=1}^{q} \theta_j B^j) a_t.$$

Thus the ARMA(p, q) model is set up by the respective polynomials of the AR(p) and MA(q) models. Note that for $p = 0$ we are left with a pure MA(q) model, and likewise if $q = 0$ then the ARMA($p, 0$) is the same as an AR(p) model.

The ARMA model meets the criteria of weak stationary when all solutions of the characteristic equation in absolute value are less than 1. We then obtain an unconditional mean, $\mathrm{E}[X_t] = c/(1 - \phi_1 - ... - \phi_p)$ *(Tsay, 2010. p.66)*.

3.7 ARIMA

The Autoregressive Integrated Moving Average model, ARIMA, utilizes differencing to obtain stationary time series before applying an ARMA model on the changed series. The model is generalized such that for non-negative integers p, d, q with the backshift operator the ARIMA(p, d, q) model is given by

$$(1 - \sum_{i=1}^{p} \phi_i B^i)(1 - B)^d X_t = (1 + \sum_{i=1}^{q} \theta_i B^i) a_t, \qquad (3.7.1)$$

where $\{a_t\}$ is a white noise series. The left hand side of *(3.7.1)* consists of the AR-polynomial of p parameters, the d times differencing of the $\{X_t\}$ time series and the right hand side shows the MA-polynomial of q parameters. Note that if we write the d times differenced $\{X_t\}$ series as $W_t = (1 - B)^d X_t$ the equation *(3.7.1)* match with the ARMA(p, q) model applied on the changed time series $\{W_t\}$ described in *(3.6.1)* in *Section 3.6*. The AR-polynomial contains an unit root and so the ARIMA model is called unit-root non-stationary, which can be managed by differencing *(Tsay, 2010. p.76)*. According to *(Hyndman and Athanasopoulos, 2014)* the differencing of a time series may contribute to a more stable mean while reducing trend and seasonality.

As an example we can describe the ARIMA$(1,1,1)$ model by the time series $\{X_t\}$, so the change series $W_t = \Delta X_t = X_t - X_{t-1}$ follows a stationary ARMA$(1,1)$ model *(Tsay, 2010. p.76)*. We can estimate the parameters using maximum likelihood or least squares method. Assume that we have a data set of N observations on which we fit the ARIMA$(1,1,1)$ model, and for which we want to estimate the unknown parameters $(\phi_1, \theta_1, \sigma_a^2)$. *(Box et al., 2016. p.209-214 & 262-264)* provides a parameter estimation on the generalized ARIMA(p,d,q) model which we have adjusted to the case $(p,d,q) = (1,1,1)$.

The ARIMA$(1,1,1)$ model can be described by

$$X_t = c + X_{t-1} + \phi_1(X_{t-1} - X_{t-2}) + \theta_1 a_{t-1} + a_t \qquad (3.7.2)$$

which for the change series $\{W_t\}$ of the once differenced $\{X_t\}$ series can be written as

$$W_t = c + \phi_1 W_{t-1} + \theta_1 a_{t-1} + a_t, \qquad (3.7.3)$$

where $\{a_t\}$ is a white noise process, c a constant, ϕ_1 the autoregressive parameter and θ_1 the moving averages parameter. For differenced series, if the mean $\mathrm{E}[\Delta X_t] = \mu \neq 0$ then $c = \mu(1-\phi_1)$ *(Shumway and Stoffer, 2011. p.141)*. For this example we assume a zero mean, thus $c = 0$, which is also suggested for differenced series by *(Box et al., 2016. p.211)*.

Let us rewrite the equation *(3.7.2)* in the following

$$a_t = X_t - X_{t-1} - \phi_1(X_{t-1} - X_{t-2}) - \theta_1 a_{t-1}, \qquad (3.7.4)$$

which written with the change series W_t from equation *(3.7.3)* is

$$a_t = W_t - \phi_1 W_{t-1} - \theta_1 a_{t-1}. \qquad (3.7.5)$$

The equation *(3.7.5)* involves past values, W_{t-1} and a_{t-1}, which at time t are unknown. This makes it difficult to obtain the parameter estimates of (ϕ_1, θ_1). However by including the unknown past values as parameters to be estimated, we can then use these estimates in the likelihood optimization to obtain the parameter estimates of (ϕ_1, θ_1).

If we let W_{t-1}^* and a_{t-1}^* be estimates of the unknown past values W_{t-1} and a_{t-1}. We could set the values of W_{t-1}^*, a_{t-1}^* as their unconditional expectations, thus $a_{t-1}^* = 0$ and if $E[W_t] = \mu \neq 0$ then we use the mean $W_{t-1}^* = \sum_{t=1}^{N-1} W_t/(N-1)$ otherwise $W_{t-1}^* = 0$. Other approximations are possible, and can be further explored in *(Box et al., 2016. p.211-212)*. Starting at $t = 0$ we can then recursively calculate a_t for $t = 1, 2, ..., N-1$ as a function of the observations $\{W_t\}$, the estimated values W_{-1}^*, a_{-1}^* and using some parameters ϕ_1, θ_1. Let us denote the set of values by $a_t(\phi_1, \theta_1 \mid W_{t-1}^*, a_{t-1}^*, W_t)$.

$$\begin{aligned}
a_0 &= W_0 - \phi_1 W_{-1}^* - \theta_1 a_{-1}^* \\
a_1 &= W_1 - \phi_1 W_0 - \theta_1 a_0 = W_1 - (\phi_1 + \theta_1) W_0 + \phi_1 \theta_1 W_{-1}^* + \theta_1^2 a_{-1}^* \\
&\vdots
\end{aligned}$$

If the series $\{a_t\}$ is a Gaussian white noise series, then for $t = 1, 2, ..., N-1$, a_t are normally distributed with density function

$$p(a_1, a_2, ...a_{N-1}) \propto (\sigma_a^2)^{-(N-1)/2} \exp\left[-\sum_{t=1}^{N-1} \frac{a_t^2}{2\sigma_a^2} \right]. \qquad (3.7.6)$$

The `R` package `forecast` function `Arima` which is employed in *Section 5* uses maximum likelihood or minimizing conditional sum-of-squares methods to estimate the parameters. The conditional loglikelihood function is given by

$$L^*(\phi_1, \theta_1, \sigma_a^2) = -\frac{N-1}{2}\ln(\sigma_a^2) - \frac{S^*(\phi_1, \theta_1)}{2\sigma_a^2} \qquad (3.7.7)$$

conditioned on the choise of values (W_{t-1}^*, a_{t-1}^*). By minimizing the conditional sum-of-squares estimates

$$S^*(\phi_1, \theta_1) = \sum_{t=1}^{N-1} a_t^2(\phi_1, \theta_1 \mid W_{t-1}^*, a_{t-1}^*, W_t) \qquad (3.7.8)$$

we obtain the parameter estimates. To stress that the conditional loglikelihood and sum-of-squares estimates are conditioned on the choise of the values (W_{t-1}^*, a_{t-1}^*) we use the asterisks for $L^*(\phi_1, \theta_1, \sigma_a^2)$ and $S^*(\phi_1, \theta_1)$.

From the equation *(3.7.7)* it is clear that the conditional sum-of-squares function $S^*(\phi_1, \theta_1)$ plays a big part in the calculation of the conditional loglikelihood. For a fixed value of σ_a^2, the conditional loglikelihood L^* is a linear function of the conditional sum-of-squares function S^*. The minimized conditional sum-of-squares parameter estimates are called conditional least-squares estimates. Ordinarily the conditional least-squares parameter estimates are in close proximity to the conditional maximum likelihood parameter estimates.

3.8 SARIMA

The seasonal ARIMA model, SARIMA, can be applied for data that exhibits seasonality and is generalized such that ARIMA$(p, d, q)(P, D, Q)_s$ with $s > 1$ being the number of periods in each season and p, q, P, Q non-negative integers. We can present the general SARIMA model by

$$(1 - \sum_{i=1}^{p} \phi_i B^i)(1-B)^d(1 - \sum_{i=1}^{P} \Phi_i B^{si})(1-B^s)^D X_t = (1 + \sum_{i=1}^{q} \theta_i B^i)(1 + \sum_{i=1}^{Q} \Theta_i B^{si})a_t,$$

where $\{a_t\}$ is a Gaussian white noise series.

Denoting the seasonal autoregressive part of the SARIMA model as SAR(P) and the seasonal moving averages part of the seasonal ARIMA model by SMA(Q), *(Shumway and Stoffer, 2011. p.154-157).*

3.9 The Naive model, a random walk

The naive model when forecasting time series states that the conditions today will be the same as yesterday. The naive model is also called the random walk, an ARIMA$(0, 1, 0)$ model, which can be described by

$$X_t = c + X_{t-1} + a_t,$$

where $a_t \sim N(0, \sigma^2)$ is a white noise series with zero mean and the constant $c = \mathrm{E}[X_t - X_{t-1}]$ is referred to as the drift of the model. We can easily compute the residual series $a_t = X_t - X_{t-1}$. The random walk series do not meet the requirement of weak stationarity *(Tsay, 2010. p.72).*

The random walk models a differencing at lag-1, while a seasonal differencing at lag-s will instead take the difference between observations x_t and the previous season's observation x_{t-s} for the period s. This is highly relevant for data with seasonality gives away to the seasonal naive model described by

$$X_t = c + X_{t-s} + a_t$$

for s periods and the white noise series $a_t \sim N(0, 1)$ with zero mean. The seasonal naive model can be written as an ARIMA$(0, 0, 0)(0, 1, 0)_s$ model at period s *(Hyndman and Athanasopoulos, 2014).*

3.10 Forecast

Let h be the forecast origin and let us denote the information set available at time h by $\mathcal{F}_h$. The 1-step ahead forecast of the series X_h can with general ARMA(p,q) model be described by

$$\hat{X}_h(1) = \mathrm{E}[X_{h+1} \mid \mathcal{F}_h] = c + \sum_{i=1}^{p} \phi_i X_{h+1-i} + \sum_{i=1}^{q} \theta_i a_{h+1-i},$$

with the forecast error $e_h(1) = X_{h+1} - \hat{X}_h(1) = a_{h+1}$ and its variance $\mathrm{Var}(e_h(1)) = \sigma_a^2$.

The l-steps ahead forecast can be described by

$$\hat{X}_h(l) = \mathrm{E}[X_{h+l} \mid \mathcal{F}_h] = c + \sum_{i=1}^{p} \phi_i \hat{X}_h(l-i) + \sum_{i=1}^{q} \theta_i a_h(l-i)$$

where

$$a_h(l-i) = \begin{cases} 0 & \text{if} \quad l-i > 0 \\ a_{h+l-i} & \text{if} \quad l-i \leq 0, \end{cases} \qquad \text{and} \quad \hat{X}_h(l-i) = X_{h+l-i} \quad \text{if} \quad l-i \leq 0.$$

Recursively we can calculate the 1 to l steps ahead forecast *(Tsay, 2010. p.68-69)*.

We describe the forecast process by an example in which we calculate the 1-step ahead forecast of the ARIMA$(1,1,1)$ model for the time series $\{X_t\}$. As explained in *Section 3.7* if the time series $\{X_t\}$ is modeled with the ARIMA$(1,1,1)$ model, then the $W_t = \Delta X_t$ series can be modeled with the ARMA$(1,1)$ model. Let $\hat{X}_h(1)$ denote the 1-step ahead forecast, which we calculate by the same theoretical approach described above, then

$$\begin{aligned}
\hat{X}_h(1) &= \mathrm{E}[X_{h+1} \mid \mathcal{F}_h] \\
&= \mathrm{E}[(X_{h+1} - X_h) + X_h \mid \mathcal{F}_h] \\
&= \mathrm{E}[W_{h+1} + X_h \mid \mathcal{F}_h] \\
&= \hat{W}_h(1) + X_h \\
&= \mathrm{E}[W_{h+1} \mid \mathcal{F}_h] + X_h \\
&= c + \phi_1 W_h + \theta_1 a_h + X_h \\
&= c + \phi_1 (X_h - X_{h-1}) + \theta_1 a_h + X_h
\end{aligned}$$

where we have used that for $i > h$ the mean of a_i is zero and the $\hat{W}_h(1)$ is the corresponding 1-step ahead forecast of the ARMA$(1,1)$ model. The forecast error of the 1 step ahead forecast, $e_h(1)$ is calculated by

$$\begin{aligned}
e_h(1) &= X_{h+1} - \hat{X}_h(1) \\
&= X_{h+1} - (\hat{W}_h(1) + X_h) \\
&= W_{h+1} - \hat{W}_h(1) \\
&= a_{h+1}.
\end{aligned}$$

A 95 % prediction interval of the 1 step ahead forecast is calculated by $\hat{X}_h(1) \pm 1.96\hat{\sigma}_a$ where $\hat{\sigma}_a$ is the residual standard deviation. For the multistep prediction interval of the ARIMA model the calculations are more complex and the theory out of the scope of this study. We will however compute prediction intervals in *Section 5.3.2* using the `R` package `forecast` for our future forecast plots.

3.11 Statistical tests

3.11.1 Autocorrelation function

A way to determine the linear dependency (autocorrelation) between X_t and X_{t-l} for an arbitrary l is by the Autocorrelation function (ACF) which is defined by

$$\rho_l = \frac{\mathrm{Cov}(X_t, X_{t-l})}{\sqrt{\mathrm{Var}(X_t)\,\mathrm{Var}(X_{t-l})}} = \frac{\mathrm{Cov}(X_t, X_{t-l})}{\mathrm{Var}(X_t)} = \frac{\gamma_l}{\gamma_0},$$

under the weak stationary assumption that $\mathrm{Var}(X_t) = \mathrm{Var}(X_{t-l})$ where ρ_l is X_t's lag-l autocorrelation coefficient between X_t and X_{t-l}.

For a weakly stationary time series $\{X_t\}$ satisfying $X_t = \mu + \sum_{i=0}^{q} \psi_i a_{t-i}$ where $\psi_0 = 1$ and $\{a_t\}$ is a sequence of independent identically distributed random variables with mean $\mathrm{E}[a_t] = 0$, the estimated lag-l autocorrelation $\hat{\rho}_l$ is asymptotically normally distributed $N(0, (1 + 1\sum_{i=1}^{q} \rho_i^2)/T)$ for $l > q$.

The estimated lag-l sample autocorrelation of X_t is

$$\hat{\rho}_l = \frac{\sum_{t=l+1}^{T}(X_t - \bar{X})(X_{t-l} - \bar{X})}{\sum_{t=1}^{T}(X_t - \bar{X})^2}$$

for $1 \leq l < T - 1$ *(Tsay, 2010. p.31-32).*

3.11.2 The Partial Autocorrelation Function

The Partial Autocorrelation Function, denoted PACF, is a function of ACF which calculates the amount of correlation between a time series $\{X_t\}$ and a lag l which is not explained by the correlation at lags $< l$.

For $k = 1, 2, ...$ number of parameters and $m = 1, 2, ..$ we let $\phi_{0,k}$ be constants, $\phi_{m,k}$ the coefficients of X_{t-m} and $\{a_{kt}\}$ be the error terms of the AR(k) models. Consider the following AR models which increase by the amount of parameters

$$\begin{aligned}
X_t &= \phi_{01} + \phi_{11}X_{t-1} + a_{1t} \\
X_t &= \phi_{02} + \phi_{12}X_{t-1} + \phi_{22}X_{t-2} + e_{2t} \\
X_t &= \phi_{03} + \phi_{13}X_{t-1} + \phi_{23}X_{t-2} + \phi_{33}X_{t-3} + e_{3t} \\
X_t &= \phi_{04} + \phi_{14}X_{t-1} + \phi_{24}X_{t-2} + \phi_{34}X_{t-3} + \phi_{44}X_{t-4} + e_{4t} \\
&\vdots
\end{aligned}$$

The parameter estimates $\hat{\phi}_{kk}$ for $k = 1, 2, ...$ can be estimated by a least squares method and are the lag-k sample PACF of X_t. The equations above calculated in a sequential order then shows the added contribution of X_{t-k} to X_t in an AR($k-1$) model. For an AR(p) model, the sample PACF lag-k should be close to zero for all $k > p$, which can be helpful when selecting the amount of AR parameters in the model selection *(Tsay, 2010. p.46-47).*

3.11.3 Ljung-Box Portmanteau test

The Ljung-Box Portmanteau test is a way of looking for serial correlation in the observations *(Tsay, 2010. p.32)*. Let ρ_l be the autocorrelation, then the test can be set up with the hypothesis

$$H_0 : \rho_1 = ... = \rho_m = 0$$
$$H_a : \rho_l \neq 0, \quad for \quad some \quad l \in (1, m).$$

The test statistic is calculated by

$$Q(m) = T(T+2)\sum_{l=1}^{m} \frac{\hat{\rho}_l^2}{T-l},$$

where T is the number of observations, m is the chosen lag and $\hat{\rho}_l$ is the sample autocorrelation defined in *Section 3.11.1*. Under the null hypothesis $Q(m)$ is asymptotically χ^2- distributed with m degrees of freedom.

If $Q(m) > \chi^2_m(\alpha)$ for the chosen level α the H_0 is rejected and the model can be considered inadequate. The test can be applied on the residuals of a fitted model, which under the null hypothesis $Q(m)$ is instead asymptotically χ^2_{m-p} distributed and p is the corresponding amount of AR parameters of the fitted model *(Tsay, 2010. p.50-51)*.

3.11.4 ADF

The Augmented Dickey-Fuller unit root test is a tool to discover an unit root in an AR(p) process, and is thus a way of evaluating the stationarity of a time series. The ADF test is set up by testing the null hypothesis $H_0 : \beta = 1$ against the alternative hypothesis $H_a : \beta < 1$. This is done by using the regression $X_t = c_t + \beta X_{t-1} + \sum_{i=1}^{p-1} \phi_i \Delta X_{t-i} + a_t$, such that c_t is the deterministic function of the time index t, the differenced series of $\{X_t\}$ is $\Delta X_j = X_j - X_{j-1}$. Let $\hat{\beta}$ be the least-squares estimate of β, then we have the t-statistic

$$\text{ADF-test} = \frac{\hat{\beta} - 1}{se(\hat{\beta})}.$$

When the null hypothesis is rejected, it signals a stationary time series *(Tsay, 2010. p.77)*.

3.11.5 AIC

Akaike's Information Criterion uses estimated maximum likelihood to determine the goodness of fit of a model. AIC is defined by

$$AIC = -2L(\hat{\theta}) + 2k,$$

where $L(\hat{\theta})$ is the estimated maximum loglikelihood of parameters θ and k is the number of parameters in θ. A model with a low AIC is then seen as a better fit to data than a model with high a AIC *(Held and Sabanés Bové, 2014. p.224)*.

3.11.6 Likelihood Ratio test

We can compare two models M_i and M_j, where $M_i \subset M_j$, by a Likelihood Ratio test (LR-test). By using the description of the LR-test from *(Held and Sabanés Bové, 2014. p.221-222)* we write the hypothesis

$$\begin{aligned} &H_0 : M_i \text{ holds} \\ &H_1 : M_j \text{ holds, but not } M_i, \end{aligned}$$

which we wish to test against each other. The test statistic can be calculated by

$$\begin{aligned} G^2(M_i \mid M_j) &= G^2(M_i) - G^2(M_j) \\ &= 2(L(\hat{\theta}_{M_i}) - L(\hat{\theta}_{M_j})), \end{aligned}$$

with the estimated maximum loglikelikoods $L(\hat{\theta}_{M_i})$ and $L(\hat{\theta}_{M_j})$.

$G^2(M_i \mid M_j)$ is asymptotically $\chi^2_{df}(\alpha)$- distributed for $df =$ number of additional parameters in M_j degrees of freedom and chosen level α. The corresponding p-value is given by

$$P(\chi^2_{df}(\alpha) \geq G^2(M_i \mid M_j)).$$

3.11.7 RMSE

The root mean square error is a measurement of the deviation of the predicted value $\hat{x}_t$ to the observed value x_t at time t. This comes in handy when determining how well a forecast model performs. For N observations the RMSE of is computed by

$$RMSE(\hat{X}) = \sqrt{MSE(\hat{X})} = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(\hat{x}_t - x_t)^2}.$$

That a model fits data and thus shows a good in-sample forecasting does not directly lead to a good out-of-sample forecasting. The RMSE will be used to evaluate the forecasting performance of fitted models.

4 Data

Originally daily weather data was collected from *(SMHI, n.d (b))* of air temperature, wind speed and precipitation. The data was distributed over 28 stations from different locations in Sweden between the dates 1966-01-01 and 2019-11-01. For this thesis the stations were divided into the categories South, Mid and North based on their geographical location, where the 15 chosen stations were selected after inspecting lowest amount of missing values in each data set.

The analysis was focused on the South region, which is represented by Falsterbo, Hallands Väderö, Hanö, Vinga and Visby, unless otherwise stated. The data was looked at on a monthly basis, and it exhibited strong seasonal patterns that is approached in the model selection. The mean over each month was taken, so the analysis is without consideration to the fluctuation that daily data may provide.

In general the monthly mean air temperatures varied between 20 to -20 degrees Celsius depending on in which part of Sweden the station was located. For example, the mean air temperatures varied between 20 and -5 degrees (Celsius) in the south (*Figure 1*), and there were some very clear seasonal components. When looking closer on smaller ranges of the data, the yearly periods were particularly distinct.

The monthly wind speed in the south varied between 3 and 12 meter per second (mps) as seen in *Figure 1*, whereas in the mid area and the north the interval was between 0 and 7 mps. The precipitation had monthly observations between 0 to 6 milliliter (mm) of precipitation for the whole country of Sweden. The data appeared to spike quite randomly but otherwise had a clear periodic behavior on a yearly basis as seen in *Figure 1*.

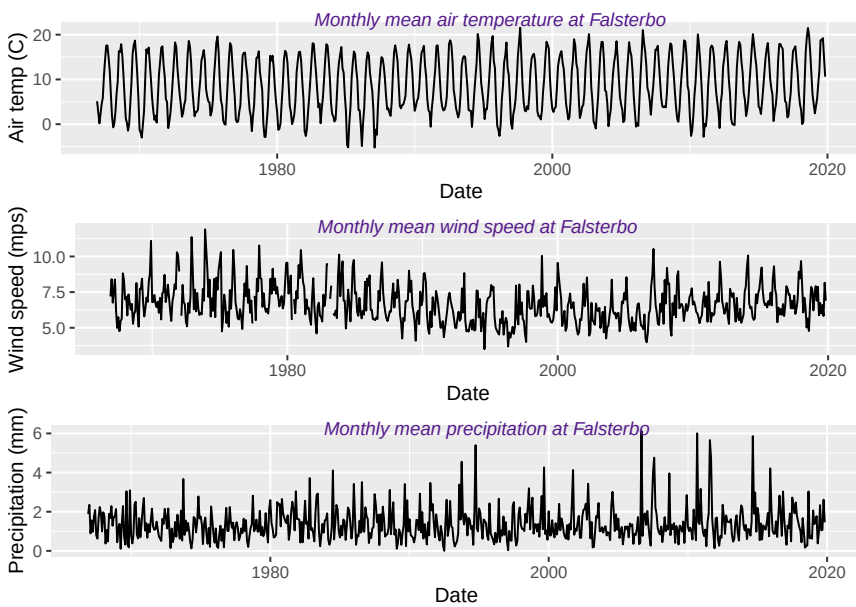

Figure 1: Representative plots of monthly mean weather data in Sweden

4.1 Transforming data

As illustrated in *Figure 1*, the wind speed and precipitation plots exhibits unstable variances. To remedy this, the Box-Cox transformations were applied to these time series. The method of transformation is more thoroughly presented in *Appendix A: Section 8.1.1*. The transformed time series are plotted in *Appendix A: Figure 8*.

An approach to handling the seasonality of a time series was to seasonally adjust the data before modelling, thus yielding seasonal stationary data. The weather data consisted of at least trends and seasons which in general did not change over time, so an additive model could have been applied as seasonal adjustment.

Another approach of handling the seasonality could be by seasonal differencing in the seasonal ARIMA (SARIMA) model. The SARIMA model was more effective than the seasonal adjusting since it allowed random changes in the seasonal component *(Linde, 2005)*. For example, there was a strong pattern of annual seasonality in the mean air temperature, so one seasonal differencing at appropriate lag-l would have been applied in the SARIMA model at least. For this paper the SARIMA model method of dealing with seasonality has been applied.

5 Methodology

In the forecasting model selection we looked at models using past data from the station, and also considered possible forecasting models obtained using past data from the other stations from the same category. Since we focused on the southern region of Sweden, the station located in Falsterbo was chosen to be used for prediction. The other stations in the region were thus not used in the forecast process, and were only considered in the model selection phase. A selection of tested models are presented in *Appendix B: Table 9*. We also evaluated the naive and seasonal naive models for the Falsterbo station data. The chosen models were then compared with the naive and seasonal naive models in the *Section 5.3* about forecast performance.

We followed the iterative approach to model building described in *(Box et al., 2016, p.16)*, which was manually conducted in different stages. After exploring the data the SARIMA class of models was considered. Then an identification of a potential SARIMA model was made. Diagnostic checking of the potential model was performed to see if it was adequate. If deemed a good fit the model was used in forecasting, otherwise if proved inadequate the potential model was discarded and the process starts over on identifying a new potential model.

In the second stage of model building, where we attempted to identify a potential model, an automatic model selection was initially considered. The `R` function `auto.arima` was considered, but the resulting models gave a poorer fit than the ones obtained in the manual model selection and was thus not included. When performing the manual model selection, the following approach and logic from *(Hyndman and Athanasopoulos, 2014)* was applied in identifying a potential model. The SARIMA model was fitted by first observing the data in plots to see if there were any seasonality, trends, fluctuating observations or change in variance. In the case of unstable variance, Box-Cox transformations were applied in the previous *Section 4.1*. If the data was non-stationary, seasonal and non-seasonal differencing were applied the required amount of times until stationarity was accomplished. ADF tests were performed in this phase to test the stationarity of the data using an appropriate lag produced by the chosen order of the function `ar` in the `R` package `stats` for the test. The amount of differencing was reflected in the d (non-seasonal) and D (seasonal) values of the $\text{ARIMA}(p, d, q)(P, D, Q)_s$ model.

The ACFs and PACFs can give indication on whether an AR(p), MA(q), SAR(P) or SMA(Q) term should be added to the model. For instance *(Shumway and Stoffer, 2011. p.108,155)* suggests that, for the non-seasonal terms we look for if the PACFs plot cuts off after lag-p and the ACFs plot tails off after lag p then we may consider adding AR(p) terms to the model. While if the ACFs plot cuts off after lag q and the PACFs plot tails off after lag q we instead consider MA(q) terms. For seasonal terms we consider a similar reasoning, from *(Hyndman and Athanasopoulos, 2014)*, which suggests that if at a specific lag s and further on $2s$, $3s$, ... the ACFs "spikes" and the PACFs cuts off at lag s then adding a SAR(P) may be appropriate. While if the reverse is true of the ACFs and PACFs then adding a SMA(Q) term to the model should be considered instead. Here the term "spikes" refers to the sample ACF presence outside of the standard error limits, which are depicted in blue in the ACFs plots. A spike in the ACFs plot tells us that the ACF at lag-l is significantly different from 0 at a 5 % level. The same applies to the PACF case.

Further in the next stages of model building the potential model was fitted and we calculated the AIC as a measure of the goodness of fit. The ACFs and PACFs plots of the potential model's residuals are presented, which for a white noise series that consists if i.i.d random variables should be approximately 0. When determining which models appeared to be adequate at this stage, all the sample ACFs should have been within their 5 % standard error limits to indicate them being not significantly different from 0 *(Tsay, 2010. p.32)*.

The Ljung-Box test was performed to test the independence between the lags and thus the absence of autocorrelation. We further checked if the residuals resembled white noise with residual diagnostics in the form of Q-Q plots, standardized histogram and scatter plot over residuals versus fitted values were evaluated in the diagnostic checking.

For mixed ARIMA models there are simulation studies suggesting a Likelihood Ratio test may perform better than a t-test when testing for unit roots according to *(Box et al., 2016, p.359)*. To test the significance of the whole models we conducted LR-tests (*Section 3.11.6*) in the model selection. Thus we had an additional way of comparing models. This was a way of determining if more parameters should have been included in the model for a better goodness of fit. The in-sample RMSE was also used to see how well the fitted values, the in-sample forecasts, had been calculated in comparance to the observed values. When the best model had been chosen, it was used for out-of-sample forecast in *Section 5.3*.

5.1 Model selection

We will present plots in a representative manner for the air temperature, whilst referring to the appropriate appendix for the other variables. This is applied for both the model selection and the evaluation of the naive models.

5.1.1 Air temperature model selection

In this section we wanted to find a model which fitted the data using the methodology described above. Let us denote the monthly mean air temperature in Falsterbo by the series $\{X_t\}$. We started the model selection phase by performing ADF tests on the monthly mean air temperature for all stations in the south region, with main focus on Falsterbo. Since Falsterbo station was chosen as a representative of the south of Sweden stations, it was important to see if the other stations in the region tested similar. The ADF test statistic for the monthly mean air temperature in Falsterbo was calculated using the `R` package `fUnitRoots` function `adfTest` to -5.11 and with a p-value of 0.01. Thus we could reject the null hypothesis (*Appendix A: Table 8*) on a 1 % level, meaning the series could be considered stationary and that no unit root was detected. Thus no regular differencing has to be applied for the monthly mean air temperature in the South.

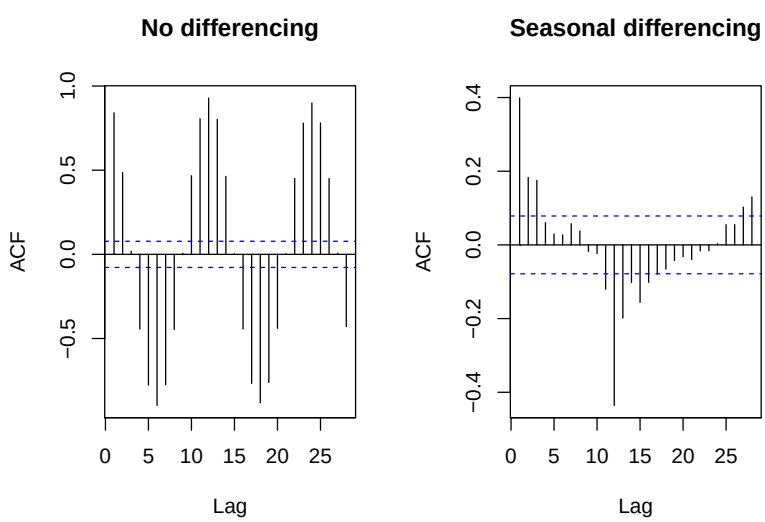

Figure 2: Sample ACFs plots of monthly mean air temperature in Falsterbo series $\{X_t\}$. The right ACF plot have been seasonal differenced once, $D = 1$ at the period $s = 12$.

The ACFs plot of the monthly mean air temperature without seasonal differencing displays a clear seasonal pattern in *Figure 2*. With the seasonal differencing the ACFs plot in *Figure 2* shows significant spike at lag-1 which suggests a non-seasonal MA-term, and a significant spike at lag-12 which suggests a seasonal MA-term *(Hyndman and Athanasopoulos, 2014)*. Thus the ARIMA$(0,0,1)(0,1,1)_{12}$ model was chosen as our starting point. We see the same plots and corresponding PACF in *Appendix A: Figure 9*.

By evaluating the ACFs and PACFs of the residuals of the fitted models, we continued to look for significant spikes which would have indicated that we should have added another component for a better fit. We also fitted some other models, as the ARIMA$(1,0,0)(1,1,0)_{12}$. This was also done on the other stations in the South to see which models fitted the data best, since we wanted the best model for the whole area. The ACFs and PACFs of the ARIMA$(1,0,0)(0,1,1)_{12}$ model´s residuals are found in *Appendix B: Figure 12*, where approximately all sample ACFs are within the 5 % standard error limits which indicates that they are not significantly different from 0.

The model with the lowest AIC of 2248.76 when applied to the Falsterbo data was ARIMA$(1,0,1)(0,1,1)_{12}$ with ARIMA$(1,0,0)(0,1,1)_{12}$ and ARIMA$(0,0,3)(0,1,1)_{12}$ close behind as seen in *Table 1*. After also checking these models as well as other similar models for the monthly mean air temperature of the other stations in the South, these models were also suggested as the best fits even when fitted on the other stations in the South. The models are summarized in *Appendix B: Table 9*. To determine which of these models to choose we wanted to evaluate the in-sample RMSE and the residuals, and also perform LR-tests.

Table 1: Selection of potential models for monthly mean air temperature in Falsterbo series $\{X_t\}$. The chosen model is presented in bold font.

Model	AIC	RMSE	LogLik
ARIMA$(0,0,1)(0,1,1)_{12}$	2284.27	1.439356	-1139.136
ARIMA$(0,0,2)(0,1,1)_{12}$	2267.2	1.46336	-1129.602
ARIMA$(0,0,3)(0,1,1)_{12}$	2253.6	1.419938	-1121.802
ARIMA$(1,0,0)(0,1,1)_{12}$	**2248.76**	**1.417881**	**-1121.381**
ARIMA$(1,0,1)(0,1,1)_{12}$	2248.45	1.414223	-1120.224

For the ARIMA$(0,0,3)(0,1,1)_{12}$ the in-sample RMSE measured up to 1.417881, which was slightly lower than 1.419938 for the ARIMA$(1,0,1)(0,1,1)_{12}$ model. The model with the lowest in-sample RMSE was however ARIMA$(1,0,1)(0,1,1)_{12}$ at 1.414223.

The residuals should have constant variance and hopefully be normally distributed. Before deciding which model to use in the forecasting process we looked at Q-Q plots and histograms of the residual series for the models. The residual plots for the three models were very similar. They all displayed some lower tails and also seemed to have some outliers at the bottom of the Q-Q plots, as seen for the ARIMA$(1,0,0)(0,1,1)_{12}$ model in *Figure 3*. The Residual versus fitted plot in *Figure 3* shows that the points are mostly evenly scattered along the estimated regression line depicted in red, but there appears to be some sort of clustering. With that in mind the residuals were considered approximately white noise, which was desireable. Both the AIC and RMSE indicated that the ARIMA$(1,0,1)(0,1,1)_{12}$ model was the best fit for the monthly mean air temperature in Falsterbo.

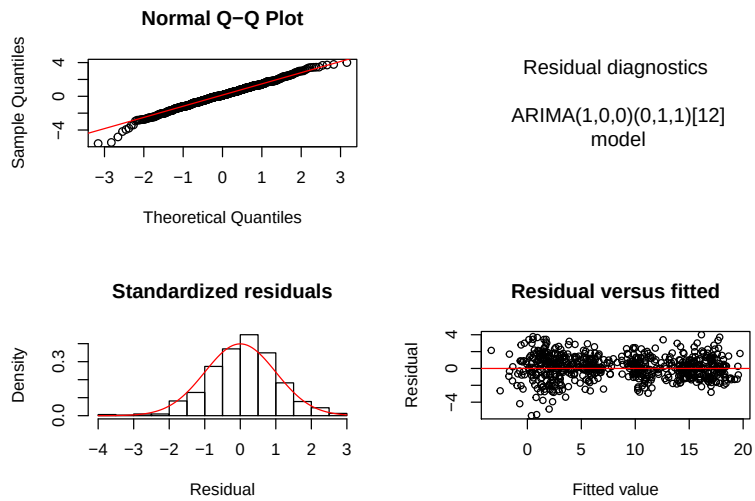

Figure 3: Residual diagnostics of the ARIMA$(1,0,0)(0,1,1)_{12}$ model of monthly mean air temperature in Falsterbo series $\{X_t\}$.

We performed LR-tests with one model against another model with more parameters. We used the null hypothesis that the smaller model holds against the alternative hypothesis that the bigger model holds but not the smaller model. Out of the three considered models the ARIMA$(0,0,3)(0,1,1)_{12}$ model had the most parameters and was thus the largest model, so we started by comparing that model to the ARIMA$(1,0,1)(0,1,1)_{12}$ model. The loglikelihood of the models are presented in *Table 1*.

For simplicity we let M_{A1} represent the ARIMA$(1,0,1)(0,1,1)_{12}$ model, M_{A2} the ARIMA$(1,0,0)(0,1,1)_{12}$ model and M_{A3} the ARIMA$(0,0,3)(0,1,1)_{12}$ model in the LR-tests.

For $M_{A1} \subset M_{A3}$, we got:

$$G^2(M_{A1} \mid M_{A3}) = G^2(M_{A1}) - G^2(M_{A3}) = -2(-1121.8 - (-1120.22)) = 3.16$$

With one additional parameter in M_{A3} we got that there was $df = 1$ degrees of freedom and p-value:

$$P(\chi_1^2 \geq G^2(M_{A1} \mid M_{A3})) = 0.0757$$

In comparison to $\chi_1^2(0.05) = 3.84$ we got a lower test statistic, which together with the p-value that barely exceeded 0.05 made us not reject the null hypothesis on a 5 % level. We continued by comparing the ARIMA$(1,0,1)(0,1,1)_{12}$ model with the simpler ARIMA$(1,0,0)(0,1,1)_{12}$ model.

For $M_{A2} \subset M_{A1}$, we got:

$$G^2(M_{A2} \mid M_{A1}) = -2(-1121.38 - (-1120.22)) = 2.31$$

With one additional parameter in M_{A1} we got that there was $df = 1$ degrees of freedom and p-value:

$$P(\chi_1^2 \geq G^2(M_{A2} \mid M_{A1})) = 0.1283$$

So we could not reject the null hypothesis that the M_{A2} model held on a 5 % level. Thus we chose the simpler ARIMA$(1,0,0)(0,1,1)_{12}$ model to use for forecasting.

We performed the Ljung-Box test on the ARIMA$(1,0,0)(0,1,1)_{12}$ residual series to look for serial correlation in the observations. This was set up with the null hypothesis that the lag-l for $l \in (1,m)$ were independent of each other against the alternative hypothesis that at least one of the lag-l in $l \in (1,m)$ were not independent. For $m = 12$ we got the following test result.

$$Q(12) = 17.932 < \chi^2_{12}(0.05) = 21.03,$$

with the p-value

$$P(\chi^2_{12}(0.05) \geq Q(12)) = 0.118.$$

Since the test statistic was lower than $\chi^2_{12}(0.05)$ and the p-value exceeded 0.05, we could not on a 5 % level reject the null hypothesis of independence between the observations in favor of the alternative hypothesis that there were serial correlation between the observations. For $m = 24$ we could once more not reject the null hypothesis, with the results being presented in *Appendix B: Table 10*. However for $m = 36$ we got the p-value $0.048 < 0.05$ for which we could reject the null hypothesis on a 5 % level. This meant that at higher lags there might exist serial correlation between the observations. We chose to keep the ARIMA$(1,0,0)(0,1,1)_{12}$ model as a the selected model for forecasting in *Section 5.3*.

The model selected for the monthly mean air temperature series $\{X_t\}$ could be described by:

$$(1-\phi_1 B)(1-B^{12})X_t = (1+\Theta_1 B^{12})a_t$$

So,

$$X_t = X_{t-12} + \phi_1(X_{t-1} - X_{t-13}) + \Theta_1 a_{t-12} + a_t,$$

where $\{a_t\}$ was a white noise series. With the coefficients estimates we could write the model as:

$$X_t = X_{t-12} + 0.5014(X_{t-1} - X_{t-13}) - 0.9299a_{t-12} + a_t,$$

where we used coefficients estimations computed by R. A summary of the model is found in the *Appendix B: Table 11*.

5.1.2 Wind speed model selection

Let us denote the Box-Cox transformed monthly mean wind speed by the series $\{V_t(\lambda_W)\}$. We performed the ADF test for the Box-Cox transformed monthly mean wind speed as described in the beginning of *Section 5.1.1*, for which the test results are found in *Appendix A: Table 8*. The p-value of the test statistic for the Falsterbo station data was 0.69 for which we could not reject the null hypothesis, meaning the test had detected an unit root in the series, and that the non-stationary time series needed regular differencing. There was only one stations in the South (Hallands Väderö) for which we could reject the null hypothesis on a 5 % level. Upon differencing the $\{V_t(\lambda_W)\}$ series one time, the null hypothesis in the ADF test could be rejected and thus the monthly mean wind speed could after a Box-Cox transformation and one differencing be seen as stationary *(Appendix A: Figure 10)*. So once again we went through the procedure of examining the sample ACFs and PACFs plots of the transformed data and manually proceeding in the same manner for the residual series. There are significant spikes at lag-1 and 12 in the sample ACF plot in *Appendix A: Figure 10*, and the base SARIMA models to be considered were then the ARIMA$(0,1,1)(0,1,1)_{12}$ and ARIMA$(1,1,0)(1,1,0)_{12}$ models.

After manually fitting different similar SARIMA models the one with the lowest AIC of -1629.314633 value was the ARIMA$(1,1,2)(0,1,1)_{12}$ model, with several similar models close behind presented in *Table 2* and also in *Appendix B: Table 9*.

Table 2: Selection of potential models for the Box-Cox transformed monthly mean wind speed in Falsterbo series $\{V_t(\lambda_W)\}$. The chosen model is presented in bold font.

Model	AIC	RMSE	LogLik
ARIMA$(0,1,1)(0,1,1)_{12}$	-1624.01	0.062277	815.005
ARIMA$(0,1,2)(0,1,1)_{12}$	-1626.75	0.061996	817.376
ARIMA$(0,1,3)(0,1,1)_{12}$	-1625.9	0.061937	817.952
ARIMA$(1,1,1)(0,1,1)_{12}$	-1627.21	0.061972	817.603
ARIMA$(1,1,2)(0,1,1)_{12}$	**-1629.31**	**0.061857**	**819.657**
ARIMA$(1,1,3)(0,1,1)_{12}$	-1624.23	0.061896	818.113

By looking at the in-sample RMSE in the same table we saw that the ARIMA$(1,1,2)(0,1,1)_{12}$ model's in-sample RMSE of 0.061857 was slightly lower than the other models. However considering the small difference between the models AIC and in-sample RMSE, we computed LR-tests and looked at residuals diagnostics before discarding any model. The best model candidate for the best fit was ARIMA$(1,1,2)(0,1,1)_{12}$, thus we performed LR-tests to compare it with other models. We let M_{W1} represent the ARIMA$(1,1,1)(0,1,1)_{12}$ model and M_{W2} the ARIMA$(1,1,2)(0,1,1)_{12}$ model which we compared in a LR-test. The null hypothesis $H_0 : M_{W1}$ holds, was tested against the alternative hypothesis $H_a : M_{W2}$ holds, but not M_{W1}. The loglikelihood of the models are presented in *Table 2*.

For $M_{W1} \subset M_{W2}$, we got:

$$G^2(M_{W1} \mid M_{W2}) = G^2(M_{W1}) - G^2(M_{W2}) = -2(817.6 - (819.66)) = 4.11$$

With there being one additional parameter in M_{W2} there was $df = 1$ degrees of freedom and p-value:

$$P(\chi_1^2 \geq G^2(M_{W1} \mid M_{W2})) = 0.0427$$

In comparison to $\chi_1^2(0.05) = 3.84$ we got a higher test statistic, which together with the significant p-value lesser than 0.05 made us reject the null hypothesis on a 5 % level. A LR-test between the smaller ARIMA$(1,1,2)(0,1,1)_{12}$ model and the bigger ARIMA$(1,1,3)(0,1,1)_{12}$ model was also performed with the null hypothesis that the smaller model would hold against the alternative hypothesis that the bigger model would hold but not the smaller model. The resulting test statistic indicated that null hypothesis could not be rejected, which lead us to consider the ARIMA$(1,1,2)(0,1,1)_{12}$ model as the best fit for the Box-Cox transformed and once differenced mean wind speed in Falsterbo.

The sample ACFs and PACFs of the ARIMA$(1,1,2)(0,1,1)_{12}$ model's residuals depicted in *Appendix B: Figure 13* shows that all sample ACFs except at lag-5 are well within the 5 % standard error limits. The ACF at lag-5 is almost at the upper limit, which lead us to approximately consider all the sample ACFs as not significantly different from 0 at a 5 % level. The residual diagnostics of the ARIMA$(1,1,2)(0,1,1)_{12}$ model are plotted in *Appendix B: Figure 14*. The residuals follow a normal distribution with some outliers on the tail, which explains the left shifted histogram of the standardized residuals. The residual vs fitted scatter plot indicates a constant variance.

We also performed the Ljung-Box test on the residuals of the ARIMA$(1,1,2)(0,1,1)_{12}$ model to evaluate if there was any serial correlation present in the residual series. The Ljung-Box test on the residuals results *(Appendix A: Table 10)* shows p-values larger than 0.05, which indicates that one can not reject the null hypothesis of independence between the observations for the alternative hypothesis that there are serial correlation in the observations. So we may consider the residual series white noise.

The selected ARIMA$(1,1,2)(0,1,1)_{12}$ model can for the Box-Cox transformed monthly mean wind speed $\{V_t(\lambda_W)\}$ be described by:

$$(1-\phi_1 B)(1-B^{12})(1-B)V_t(\lambda_W) = (1+\theta_1 B+\theta_2 B^2)(1+\Theta_1 B^{12})a_t$$

where $\{a_t\}$ is a white noise series.

Which can in terms of the Box-Cox transformer $\lambda_W = -0.4378918$ and by using the backshift operator defined by $B^i V_t = V_{t-i}$ for the time series $\{V_t\}$ be rewritten as

$$\begin{aligned} V_t(\lambda_W) = (1+\phi_1)(V_{t-1}(\lambda_W) + V_{t-12}(\lambda_W) - V_{t-13}(\lambda_W)) + \\ +\theta_1 a_{t-1} + \theta_2 a_{t-2} + \Theta_1 a_{t-12} + \theta_1\Theta_1 a_{t-13} + \theta_2\Theta_1 a_{t-14} + a_t. \end{aligned}$$

$$\begin{aligned} V_t(\lambda_W) = (1+0.8334)(V_{t-1}(\lambda_W) + V_{t-12}(\lambda_W) - V_{t-13}(\lambda_W)) - \\ -1.6706a_{t-1} + 0.6807a_{t-2} - 0.9576a_{t-12} + 1.5998a_{t-13} - 0.6518a_{t-14} + a_t. \end{aligned}$$

The model's coefficients estimates of the Box-Cox transformed monthly mean wind speed was computed by `R`. A summary of the model is found in the *Appendix B: Table 12*.

5.1.3 Precipitation model selection

Let us denote the Box-Cox transformed monthly mean precipitation by the series $\{Y_t(\lambda_P)\}$. We started by performing the ADF test in the same manner as earlier, which test results are found in *Appendix A: Table 8*. As a summary we obtained the p-value 0.01 for all the stations in the South in the ADF test which lead us to reject the null hypothesis and to consider the $\{Y_t(\lambda_P)\}$ series as stationary. Thus no regular differencing was applied for the Box-Cox transformed monthly mean precipitation in the South. The ACF of the Box-Cox transformed monthly mean precipitation without seasonal differencing displayed a clear seasonal pattern, leading us to proceed with one seasonal differencing *(Appendix A: Figure 11)*. The ACFs plot in *Appendix A: Figure 11* shows a significant spike at lag-12 which suggests a seasonal MA-term (SMA) *(Hyndman and Athanasopoulos, 2014)*. Thus the ARIMA$(0,0,0)(0,1,1)_{12}$ was our starting point for the manual model selection, and we also considered similar models.

Table 3: Selection of potential models for the Box-Cox transformed monthly mean precipitation in Falsterbo series $\{Y_t(\lambda_P)\}$. The chosen model is presented in bold font.

Model	AIC	RMSE	LogLik
ARIMA$(0,0,0)(0,1,1)_{12}$	1321	1.175032	-658.499
ARIMA$(0,0,1)(0,1,1)_{12}$	**1318.38**	**1.175942**	**-656.191**
ARIMA$(1,0,0)(0,1,1)_{12}$	1318.96	1.175582	-656.478
ARIMA$(1,0,1)(0,1,1)_{12}$	1323.49	1.172896	-657.745

After fitting different similar SARIMA models, the one with the lowest AIC value was ARIMA$(0,0,1)(0,1,1)_{12}$ closely followed by the ARIMA$(1,0,0)(0,1,1)_{12}$ presented in *Table 3* and also in *Appendix B: Table 9*. The in-sample RMSE did not add much clarity on which model that should be chosen, but further evaluation of the residuals indicated that the ARIMA$(0,0,1)(0,1,1)_{12}$ model was the better fit.

We performed LR-tests to compare the models and decide upon which model to chose. We let M_{P1} represent the ARIMA$(0,0,0)(0,1,1)_{12}$ model, M_{P2} the ARIMA$(0,0,1)(0,1,1)_{12}$ model and M_{P3} the ARIMA$(1,0,1)(0,1,1)_{12}$ model which we compared in LR-tests. The null hypothesis $H_0 : M_{P1}$ holds, was tested against the alternative hypothesis $H_a : M_{P2}$ holds, but not M_{P1}. The loglikelihood of the models are presented in *Table 3*.

For $M_{P1} \subset M_{P2}$, we got:

$$G^2(M_{P1} \mid M_{P2}) = G^2(M_{P1}) - G^2(M_{P2}) = -2(-658.5 - (-656.19)) = 4.62$$

With there being one additional parameter in M_{P2} there is $df = 1$ degrees of freedom and p-value:

$$P(\chi_1^2 \geq G^2(M_{P1} \mid M_{P2})) = 0.0317$$

In comparison to $\chi_1^2(0.05) = 3.84$ we got a higher test statistic, which together with the significant p-value lesser than 0.05 made us reject the null hypothesis on a 5 % level. We continued with comparing M_{P2} and M_{P3} in a LR-test with the null hypothesis $H_0 : M_{P2}$ holds, against the alternative hypothesis $H_a : M_{P3}$ holds, but not M_{P2}.

$$G^2(M_{P2} \mid M_{P3}) = G^2(M_{P2}) - G^2(M_{P3}) = -2(-656.19 - (-656.48)) = -0.57$$

With there being one additional parameter in M_{P3} we get that there is $df = 1$ degrees of freedom and p-value:

$$P(\chi_1^2 \geq G^2(M_{P2} \mid M_{P3})) = 0.0779.$$

We could thus not reject the hypothesis that the smaller model held on a 5% level since the p-value was higher than 0.05 and the test statistic was lower than $\chi_1^2(0.05) = 3.84$. Thus we prefered the smaller model, in this case the $\text{ARIMA}(0,0,1)(0,1,1)_{12}$ model.

LR-tests with the $\text{ARIMA}(0,0,1)(0,1,1)_{12}$ model replaced by the $\text{ARIMA}(1,0,0)(0,1,1)_{12}$ model were also performed, and provided the same results. Thus the LR-tests did not provide an explicit model that gave the best fit, but we knew that both the $\text{ARIMA}(0,0,1)(0,1,1)_{12}$ model and the $\text{ARIMA}(1,0,0)(0,1,1)_{12}$ model were adequate. Since $\text{ARIMA}(0,0,1)(0,1,1)_{12}$ had a slightly lower AIC we chose this model to continue with in the forecasting in *Section 5.3*.

The sample ACFs plot depicted in *Appendix B: Figure 18* of the $\text{ARIMA}(0,0,1)(0,1,1)_{12}$ model's residuals have all ACFs within the standard error limits, which indicates that they are at a 5 % level not significantly different from 0. We wanted the residuals to look approximately normally distributed to be a good fit and resemble white noise. The residual diagnostics of the $\text{ARIMA}(0,0,1)(0,1,1)_{12}$ model for the Box-Cox transformed monthly mean precipitation in Falsterbo is found in *Appendix B: Figure 19*. There we see that the histogram is clearly shifted to the left due to the outliers that are spotted in the Q-Q plot. We also detected a tail in the Q-Q plot, but we approximately treated the residuals as a white noise series. These outliers could possibly be explained by the extreme dips seen in the plot of the Box-Cox transformed monthly mean precipitation data in *Appendix A: Figure 8*, with the most prominent outlier occuring in June 1992.

The Ljung-Box test of the residuals results indicated that one could not reject the null hypothesis of independence between the observations (with p-values of approximately 0.8) as seen in *Appendix B: Table 10*.

The model selected can for the Box-Cox transformed monthly mean precipitation $\{Y_t(\lambda_P)\}$ be described by:

$$(1 - B^{12})Y_t(\lambda) = (1 + \theta_1 B)(1 + \Theta_1 B^{12})a_t$$

where $\{a_t\}$ is a white noise series. Thus in terms of the Box-Cox transformed $\lambda_P = 0.0823902$ and by using the backshift operator defined by $B^i Y_t = Y_{t-i}$ for time series $\{Y_t\}$, can be written as

$$Y_t(\lambda_P) = Y_{t-12}(\lambda_P) + \theta_1 a_{t-1} + \Theta_1 a_{t-12} + \theta_1 \Theta_1 a_{t-13} + a_t$$

which with the coefficients estimates are

$$Y_t(\lambda_P) = Y_{t-12}(\lambda_P) + 0.0918 a_{t-1} - 0.9858 a_{t-12} - 0.0905 a_{t-13} + a_t$$

The model's coefficients estimates for the Box-Cox transformed monthly mean precipitation data are computed by `R` and a summary of the model is found in the *Appendix B: Table 13*.

5.2 Naive and seasonal naive models

By the analysis of the ACF and PACF plots of the Box-Cox transformed Falsterbo data (*Appendix A: Figure 10,11*) we concluded that the Box-Cox transformed data could not be considered stationary without any differencing. For the case of the Box-Cox transformed monthly mean wind speed $\{V_t(\lambda_w)\}$, only the cases with either one regular differencing or with both regular and seasonal differencing are displayed. This was taken in mind when fitting the naive models on the transformed data, since it meant the Box-Cox transformed data may not meet the requirements of stationarity. All variables exhibited seasonality on a yearly basis and thus the seasonal naive model had some potential since it involved one seasonal differencing.

After fitting the naive and seasonal naive models to the monthly mean air temperature and the Box-Cox transformed monthly mean wind speed and precipitation in Falsterbo we see from the sample ACFs and PACFs plots clear indication on there being serial correlation present for the two naive models that are employed, the random walk and the seasonal naive model.

5.2.1 Air temperature

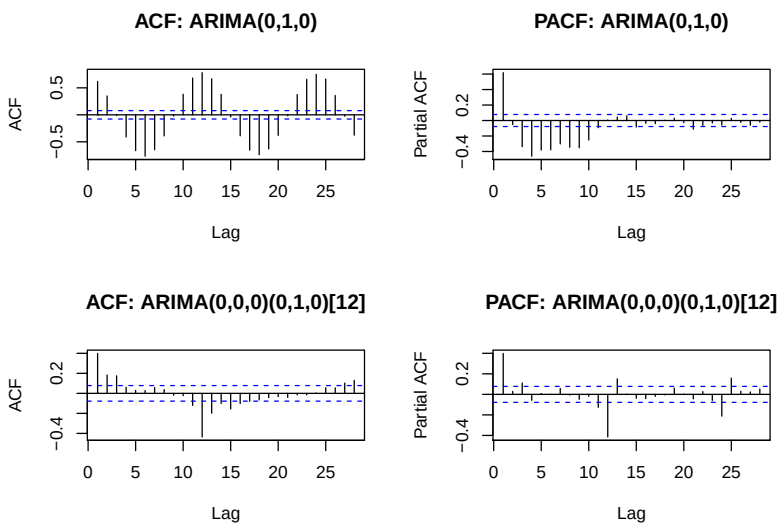

Figure 4: ACF and PACF plot of the naive models residuals of the monthly mean air temperature in Falsterbo series $\{X_t\}$. In these plots the naive model is denoted by ARIMA$(0,1,0)$ and the seasonal naive model is represented by ARIMA$(0,0,0)(0,1,0)[12]$.

The sample ACFs plot of the naive models' residuals for the monthly mean air temperature in *Figure 4* shows big positive spikes at the seasonal lags $12, 24$ among others. The corresponding sample PACFs also show a lot of significant spikes. The sample ACFs and PACFs of the seasonal naive models residuals for the monthly mean air temperature have spikes at lag-1 and 12, but most ACFs are within the standard error limits and thus not significantly different from 0 at a 5 % level.

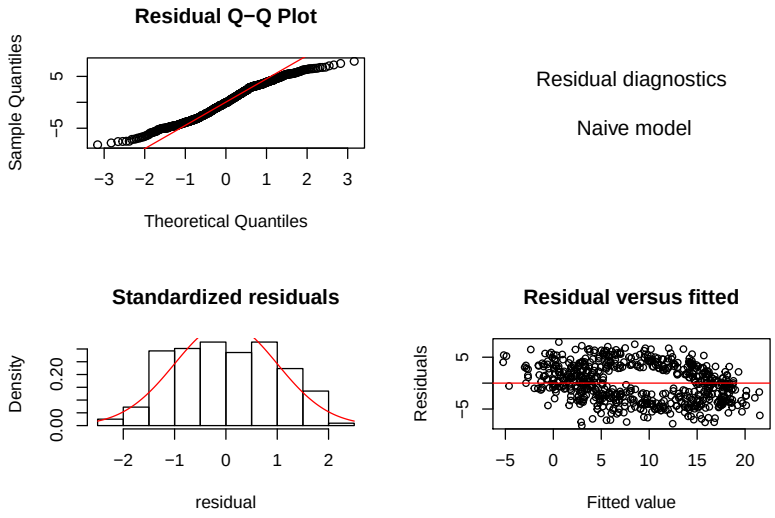

Figure 5: Residual diagnostics of the naive model for the monthly mean air temperature in Falsterbo series $\{X_t\}$.

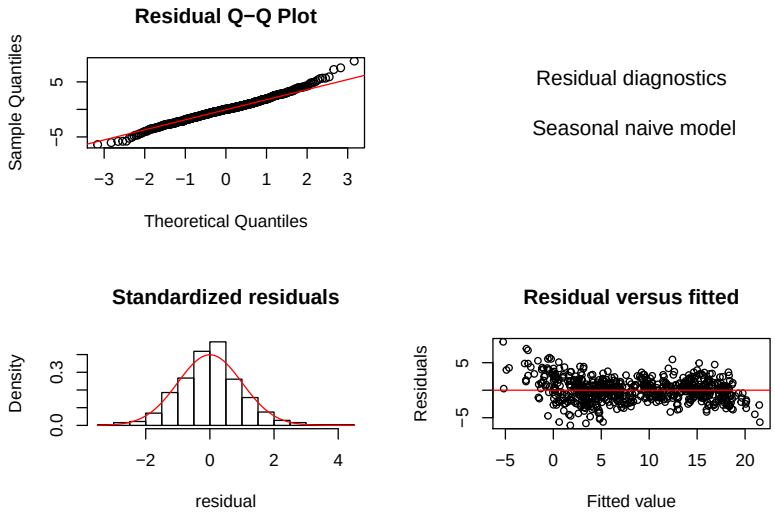

Figure 6: Residual diagnostics of the seasonal naive model for the monthly mean air temperature in Falsterbo series $\{X_t\}$.

The Q-Q plots of the naive models residuals for the monthly mean air temperature does not look to be following a normal distribution *(Figure 5-6)*. The Q-Q plot of the seasonal naive model's residuals for the monthly mean air temperature shows tails, but resembles normal distribution fairly well *(Figure 6)*. The points do not follow the standardized normal line as well as the selected SARIMA model does, but it is fairly close. The naive model's Residual versus fitted plot shows a scatter that clearly deviates from the estimated regression line (depicted in red in *Figure 5*) in a systematic way. The variance of the residuals is obviously not constant, so we see the presence of conditional heteroscedasticity in the residuals. The bars in the histogram in *Figure 5* does not follow the standardized normal curve depicted in red. The bars in the histogram of the seasonal naive model's residuals in *Figure 6* are evenly distributed across the standardized normal curve represented by the red curve in *Figure 6*. The seasonal naive model's Residual versus fitted plot shows a fairly even scatter of points along the estimated regression line.

5.2.2 Wind speed

The sample ACFs plot of the naive models' residuals for the Box-Cox transformed monthly mean wind speed have small spikes at lag-$12, 24$ and at other lags as seen in *Appendix B: Figure 15*. There is a seasonal behavior on both the sample ACFs and PACFs. The seasonal naive model's residuals sample ACFs plot and PACFs plot both has spikes at lag-1 and 12, but most lags are within the standard error limits and thus not significantly different from 0 at a 5 % level.

The Q-Q plots for both naive models' residuals of the Box-Cox transformed monthly mean wind speed follows the standardized normal line well, and the histograms are fairly evenly distributed as depicted in *Appendix B: Figure 16-17*. The Residual versus fitted plots shows points evenly scattered along the estimated regression line depicted in red in *Appendix B: Figure 16-17*.

5.2.3 Precipitation

The sample PACFs plot of the naive models residuals for the Box-Cox transformed monthly mean precipitation has many spikes but not in a clear seasonal pattern (*Appendix B: Figure 20*). The ACFs plot however has a spike at lag-1 and the rest of the lags have within the standard error limits and thus not significantly different from 0 at a 5 % level. The seasonal naive models residuals sample ACFs and PACFs plots have a spike at lag-12 which is thus significantly different from 0.

The Q-Q plots from the naive models' residuals for the Box-Cox transformed monthly mean precipitation in *Appendix B: Figure 21-22* seems to approximately follow the standardized normal line. There are two to three distinct outliers which skews mainly the histograms and the Residual versus fitted plots. Without these outliers, the points and bars in the plots are quite evenly distributed. The two Residual versus fitted plots show a cluster of points with a few outliers. There is however a clear tilt in the clusters, meaning they are only partially centered along the estimated regression lines depicted in red in *Appendix B: Figure 21-22*.

The Ljung-Box tests results on the residuals for the naive models for all variables all had a p-value of 0 which indicates that there may exist some serial correlation between the observations *Appendix B: Table 10*. The AIC and in-sample RMSE on the fitted naive models are also presented in *Appendix B: Table 9* with some of the models from the model selection in the previous section. For all variables the naive models have a distinctly higher AIC than the chosen model. For monthly mean **air temperature** the seasonal naive model has a lower AIC than the naive model, but for the monthly mean **Box-Cox transformed wind speed** and the monthly mean **Box-Cox transformed precipitation** the reverse is true.

5.3 Prediction

"The forecasting of the random walk model is considered not predictable" (*Tsay, 2010. p.73*)

At this point we were ready to start with the forecasting. To get a grasp of the out-of-sample forecasting performance of the models we used the chosen models and compared with the naive models. Since we used Box-cox transformations on the monthly mean wind speed and precipitation data when fitting the models, the estimated coefficients for those variables were on that scale. That meant we had to transform the forecast values back to the original scale before evaluating the forecast performance. For this the inverse of the respective transformations were employed, see *Appendix A: Section 8.1.1.1* for calculations.

For the forecast we divided the data into a train set on which we fitted the model, and a control set that we could compare the out-of-sample forecast of the model with. We used a recursive method of prediction, which went from 1 to n-steps ahead starting at the last observation of the time series at time t. The parameters of the models were fixed and thus not re-evaluated for each step of the forecasting.

Depending on where the split between the train and control set was made, the forecast could differ. This was because the data contained some abnormalities which could, depending on the location of the split and the choice of model, cause a very bad forecast. For example the naive model would only use the last observation of the train set and thus only forecast that value for all future points. Therefore we looked at the forecast performance for different splits of the data set. To get a fair comparison, the 1 to n-steps ahead forecast would be of the same length independently of the location of the split. By the choice of these splits the length of forecast was a quarter of the data, a 1 to 159- steps ahead forecast. For consistency, the same length of data (159 observations) were used in the train samples. For example if a split between the training and control set was put between February and Mars 1980, then the train set consisted of data points between January 1966 to February 1980 and the control set between Mars 1980 to May 1993 (which is also the out-of-sample forecast period).

As a bonus we also looked at the 1 to 60- steps ahead future forecasting for the chosen models. For this we had no real observations to check against, as we had used the whole data set to fit the models on.

5.3.1 Naive versus SARIMA

5.3.1.1 Air temperature

In *Table 4* it seems that the ARIMA$(1,0,0)(0,1,1)_{12}$ overall distinctly obtains the lowest values of the out-of-sample and in-sample RMSE in comparison to the naive models. So in this case, the model with the best performing in-sample forecast also performed the best in the out-of-sample forecast by our method of measuring the forecast performance. Something worth highlighting is that the out-of sample forecasted RMSE are low for the seasonal naive model and almost as good as the ARIMA$(1,0,0)(0,1,1)_{12}$ model's. This is also reflected in *Appendix C: Figure 23*, where the seasonal naive model appears to more or less follow the data.

Table 4: Forecasting performance of the monthly mean air temperature in Falsterbo series $\{X_t\}$. The AIC and RMSE highlighted in green are based on the in-sample forecasts, while the other five columns to the right are based on out-of-sample forecasts. The model chosen from the model selection is presented in bold font.

	In-sample forecast		RMSE: Out-of-sample forecast				
Model	AIC	RMSE	Feb/Mars 1980	Sep/Oct 1986	May/June 1993	Dec 1999/Jan 2000	Aug/Sep 2006
ARIMA$(1,0,0)(0,1,1)_{12}$	**2248.76**	**1.417881**	**1.795518**	**2.270690**	**1.602446**	**1.630145**	**1.798879**
ARIMA$(0,1,0)(0,0,0)_{12}$	3405.4	3.525834	11.295829	8.990851	6.691491	6.952178	12.885543
ARIMA$(0,0,0)(0,1,0)_{12}$	2680.35	2.049622	2.440139	3.005130	2.036506	1.711359	2.456042

The in-sample RMSE together with the AIC value gave us an idea of how well the model fit the data. So although the naive model had a low in-sample RMSE of 3.525834, the AIC at 3405.4 reflects the very poor fit to data versus the seasonal naive model and the ARIMA$(1,0,0)(0,1,1)_{12}$ model which both had AIC below 2700.

This lead us to reflect upon the low in-sample RMSE value of the naive model. The naive model's in-sample RMSE is calculated by

$$RMSE(\hat{X}) = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(\hat{x}_t - x_t)^2} = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(x_{t-1} - x_t)^2},$$

meaning we take the squared difference between successive months. So an interpretation is that there are approximately a standard deviation of 3.5 degrees Celsius in monthly mean air temperature between the successive months.

The naive model out-of-sample forecast is sensitive to its starting point which is reflected in the corresponding RMSE values in *Table 4*. As seen in *Appendix C: Figure 23* there is a clear association between the green lines showing the naive model's forecast and the RMSE values for those predictions. The more off-center the line is, the higher the RMSE value is. It becomes apparent that the naive model did not perform well at the out-of-sample forecasts of the monthly mean air temperature. The used data has observations in the range -10 to 20 and the naive model produced RMSE values between 6.691491 and 12.885543.

The seasonal naive model forecasts the monthly mean air temperature surprisingly well considering its simplicity. Unlike the basic naive model, it is not affected by which month of the year that is used as its starting point.

5.3.1.2 Wind speed

After transforming the data back on original scale by inverse Box-cox transformation we were able to evaluate the forecast performance for the monthly mean wind speed. The seasonal naive model had a higher AIC than the naive model and similar in-sample forecasting RMSE. So based on that alone it was possible that the naive model performed better on the in-sample forecasting than the seasonal naive model did.

From *Appendix C: Figure 24* it is clear that the naive model, depicted in green, provides a poor forecast of the monthly mean wind speed. The out-of-sample RMSE of the naive model in *Table 5* are between 1.194 and 3.051, which in comparison to the other considered models provide the worst out-of-sample forecast performance. The best out-of-sample forecasts are by the ARIMA$(1,1,2)(0,1,1)_{12}$ model and the seasonal naive model. The out-of-sample RMSE are slightly lower for the seasonal naive model in three out of the five of the presented out-of-sample RMSE in *Table 5*, and both vary between approximately 1 and 2.

From *Appendix C: Figure 24* we see that the seasonal naive model (in blue) and the ARIMA$(1,1,2)(0,1,1)_{12}$ model (in red) have very similar out-of-sample forecasting performance. On a closer look, the monthly mean wind speed ranges on an interval of approximately seven units, so a sample standard deviation of 1 mps is still quite large.

Table 5: Forecasting performance of the monthly mean wind speed in Falsterbo series $\{X_t\}$. The AIC and RMSE highlighted in green are based on the in-sample forecasts, while the other five columns to the right are based on out-of-sample forecasts. The model chosen from the model selection is presented in bold font.

	In-sample forecast		RMSE: Out-of-sample forecast				
Model	AIC	RMSE	Feb/Mars 1980	Sep/Oct 1986	May/June 1993	Dec 1999/Jan 2000	Aug/Sep 2006
ARIMA$(1,1,2)(0,1,1)_{12}$	**-1629.31**	**0.061857**	**1.801266**	**1.078409**	**1.382186**	**0.970540**	**1.986762**
ARIMA$(0,1,0)(0,0,0)_{12}$	-1324.01	0.084174	1.492305	1.364077	1.649253	1.194350	3.050572
ARIMA$(0,0,0)(0,1,0)_{12}$	-1277.19	0.085042	1.468920	1.382715	1.379295	1.090901	1.778216

5.3.1.3 Precipitation

After transforming the data back on original scale by inverse Box-cox transformation we are now able to evaluate the forecast performance for the monthly mean precipitation.

Overall the out-of-sample RMSE is the lowest for the ARIMA$(0,0,1)(0,1,1)_{12}$ model when comparing with the naive models in *Table 6* and ranges between 0.872 and 1.116. The ARIMA$(0,0,1)(0,1,1)_{12}$ model forecasts steadily with a similar performance result indifferent of the placement of the split between train and control set. Something worth noting however is that the scale of the y-axis in *Appendix C: Figure 25* is very large in comparance to the amplitude of the original data. With that in mind, an out-of-sample forecast RMSE of size 0.877883 mm as seen in *Table 6* is quite large for data that mostly varies between 0 and 6 mm.

The out-of-sample forecast RMSE in *Table 6* for the naive models are slightly higher in general. The naive model is sensitive to the starting point of the forecasting, which is reflected both in the out-of-sample RMSE that vary between 0.828 and 1.555, and in *Appendix C: Figure 25*. The green line representing the naive model in *Appendix C: Figure 25 (c-e)* is not as centered to the mean of that forecasting period as in the first two plots *(a-b)*, which in turn is reflected in the larger RMSE values for those out-of-sample forecasts. The monthly mean wind speed data show more large peaks after around the 1990´s, which clearly affects the performance of the seasonal naive model which needs a good representative year at its basis to perform out-of-sample forecast adequately. We see in *Appendix C: Figure 25* a large variety of how well the seasonal naive model, depicted in blue, can perform depending on the starting points. In *Appendix C: Figure 25 (b,d)* the predictions are based on years with distinctly large variation in the monthly mean precipitation in comparance to most of the following years.

The AIC of the ARIMA$(0,0,1)(0,1,1)_{12}$ model is distinctly better than for the naive models. Likewise as for the wind speed we have an AIC for the naive model that is slightly lower than for the seasonal naive model. This we can interpret as that the in-sample forecasting of the seasonal and nonseasonal naive model predict similarly poorly for the monthly mean precipitation in Falsterbo.

Table 6: Forecasting performance of the monthly mean precipitation in Falsterbo series $\{Y_t(\lambda_P)\}$. The AIC and RMSE highlighted in green are based on the in-sample forecasts, while the other five columns to the right are based on out-of-sample forecasts. The model chosen from the model selection is presented in bold font.

	In-sample forecast		RMSE: Out-of-sample forecast				
Model	AIC	RMSE	Feb/Mars 1980	Sep/Oct 1986	May/June 1993	Dec 1999/Jan 2000	Aug/Sep 2006
ARIMA$(0,0,1)(0,1,1)_{12}$	**1318.38**	**1.175942**	**0.877883**	**0.871667**	**0.903777**	**1.007906**	**1.115926**
ARIMA$(0,1,0)(0,0,0)_{12}$	1683.46	0.908696	0.827624	0.952827	1.315823	1.336915	1.555054
ARIMA$(0,0,0)(0,1,0)_{12}$	1716.48	0.946784	1.063988	1.252456	1.187473	1.130985	1.290487

5.3.2 Future forecasts

For the 1 to 60- step ahead forecasting of the future we also includes a 95% prediction interval. The certainty of the prediction interval relies on the residuals following a normal distribution and not being correlated, otherwise it may lead to incorrect prediction intervals *(Hyndman and Athanasopoulos, 2014)*. In *Figure 7* the prediction intervals for all time series future forecast are approximately within the length of 6 units (Celsius, mps and mm). The 95% prediction interval for the monthly mean air temperature future forecast gets wider at the top of the curves than in between the wave shaped curves (*Figure 7*).

So values closer to the mean show more certainty in its prediction, while the more extreme air temperatures occurring in the summer and winter are more uncertain. The forecast of the monthly mean precipitation have its lower limit of the confidence interval in close proximity to the forecast, but an upper limit farther away. Considering that the time series is quite zero heavy and it randomly peaks with higher amount of precipitation (*Figure 1*), the biggest variation of the precipitation occurs mostly for higher values than the forecasted points. The monthly mean wind speed forecast also shows an upper limit with room for more variation and uncertainty than the lower limit of the forecast.

Prediction intervals for multistep ahead forecasts tends to grow the larger the choice of the forecast horizon for ARIMA models. According to *(Hyndman and Athanasopoulos, 2014)* models with stationary data and no differencing will for large forecast horizons have prediction intervals that converges. But for a model with data that is originally non-stationary and is differenced, the prediction intervals increase the larger the forecast horizon. In *Figure 7* the prediction intervals do not appear to increase for the forecast horizon $h = 60$, but after testing larger forecast horizons ($h = 100$ and $h = 500$) it was noted that only the $\text{ARIMA}(1,1,2)(0,1,1)_{12}$ model of the monthly mean wind speed had increasing prediction intervals.

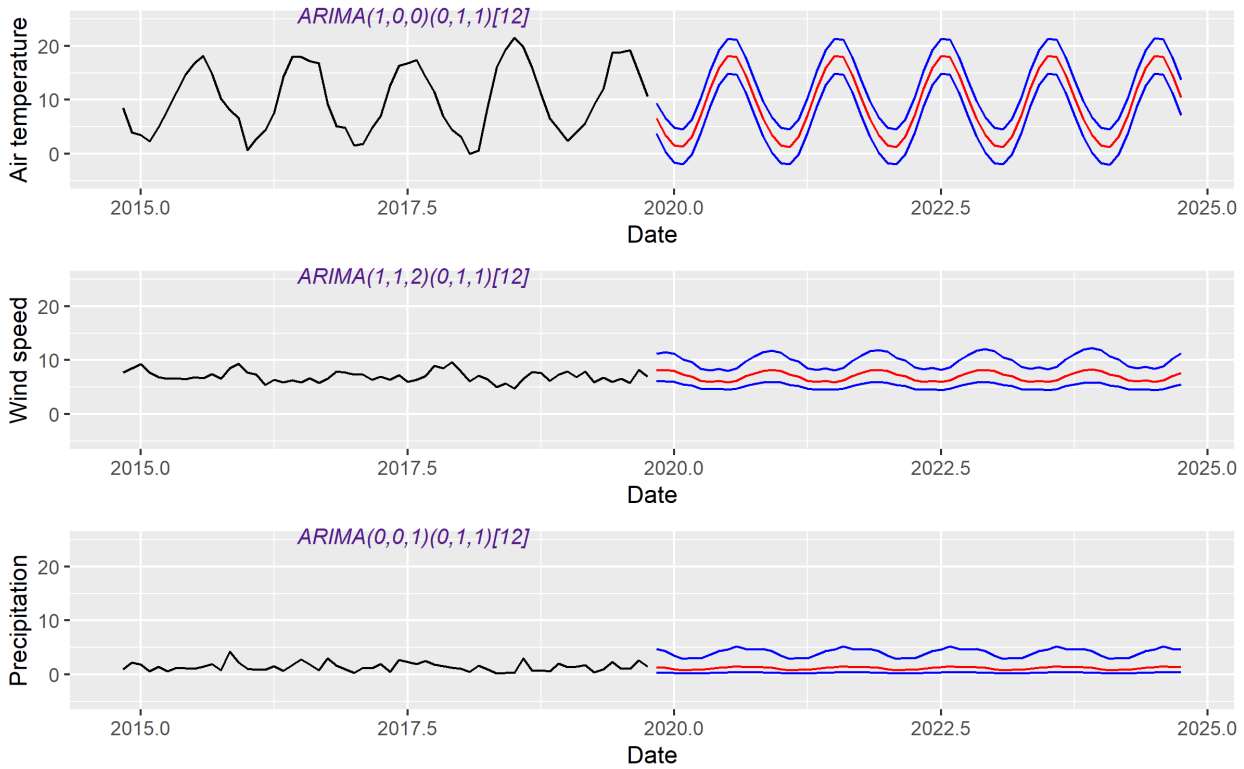

Figure 7: 1 to 60 steps ahead future forecasts of the monthly mean air temperature (top), wind speed (middle) and precipitation (bottom) in Falsterbo with 95 procent prediction interval depicted in blue. Presented in black is the original data which has been cut of from November 2014 to get a close up of the forecasting.

6 Discussion

The focus on this thesis has been on analyzing whether we can find time series methods, mainly focusing on the SARIMA model class, which predict the weather better than the naive approaches. It is quite clear that the SARIMA models do provide both better in-sample forecast and out-of-sample forecasts than the naive models of the monthly mean weather data in Falsterbo. However the seasonal naive model provided quite a good and almost adequate forecast performance, especially for the monthly mean air temperature and wind speed. What affected the out-of-sample forecasts performance the most was whether the train sample contained any major abnormalities, which could seriously worsen the forecasting performance. In *Section 5.3* where we analysed the out-of-sample forecasts of the monthly mean wind speed and precipitation *(Appendix C: Figure 24-25)*, we see that the naive models were particularly sensitive to their forecasts starting points.

The parameters of the models are not re-evaluated during the out-of-sample forecasting, but if we instead use a rolling window we would be able to re-fit the model in each step of the multistep ahead forecast. It would be interesting to see if the rolling window forecasting method would lead to better forecast performance, considering new observations would be included in the model re-fit of each step. In *Section 5.3* where we evaluate the forecast performance we chose to only do 1 to 159 steps ahead forecasts performance evaluation, thus we do not look closer on in particular the shorter lengths of forecasts and if that would result in lower out-of-sample RMSE. The amount of data that the models were fitted for the out-of-sample forecasts did not either vary.

It seems that using the seasonal naive model for predicting the monthly mean air temperature is not that bad of an approach, as seen by the out-of sample forecasts in *Appendix C: Figure 23-25*. It is likely that the naive models will perform worse on more granular data. This is due to the monthly data being just an average over daily observations which may fluctuate, but we only use a smooth mean.

We also performed Box-Cox transformations on some of the variables, and there are widely discussed consequences of how that affects the forecasting. *(Tommaso and Helmut, 2011)* investigated for macroeconomic data whether the Box-Cox transformation provided better forecasting performance. Most of the used series did not receive a better forecast performance after transforming the time series data. Also the larger forecast horizon h that was set, the less advantages the transformation yielded. We base our forecast evaluation on a large forecast horizon $h = 159$, so whether the Box-Cox transformations actually were necessary for a good forecast is for now left unknown and with need of further investigations.

There were cases when there was a clear existence of serial correlation. For example the sample ACFs plot of the naive models residuals of the monthly mean air temperature in *Figure 4* displayed spikes outside of the standard error limits which indicated that the ACFs are significantly different from 0. Together with the Ljung-Box tests results summarized in *Appendix B: Table 10* for the naive models, it is appearent that the residuals of these naive models are correlated. This can lead to bad computations in for example the forecasting process, since the forecast model is based on uncorrelated residuals. The presence of correlation in the residuals can thus lead to not trustable forecasts according to *(Hyndman and Athanasopoulos, 2014)*.

To avoid the problem of overfitting we used the LR-test to compare the likely best models against each other in the model selection. But for the initial evaluation we only considered one type of information criterion, AIC, while there are other measures possible. For example, it is possible that the Bayesian Information Criterion (BIC) would indicate the $\text{ARIMA}(1,0,0)(0,1,1)_{12}$ model as the best fit in the first place for the monthly mean air temperature which the AIC did not. Clearly there are some restrictions to this study and many aspects that are not explored. One of the most interesting aspect is if and how the variables affects each other, which could be explored by multivariate time series methods that are left for more advanced studies.

7 Bibliography

SMHI, the Swedish Meteorological and Hydrological Institute. n.d (a). https://www.smhi.se/klimat/framtidens-klimat/klimatscenarier/info/haag#mod (Accessed 2020-08-03)

Tsay, Ruey S. 2010. Analysis of financial time series. Third edition. Hoboken, New Jersey: John Wiley Sons, Inc.

Shumway, Robert H. and Stoffer, David S. 2011. Time Series Analysis and Its Applications: With R Examples. Third edition. New York: Springer Science+Buisness Media, LCC.

Hyndman, Rob J. and Athanasopoulos, George. 2018. Forecasting: principles and practice. Second edition. Melbourne, Australia: OTexts. https://www.otexts.org/fpp/ (Accessed 2020-02-28)

Held, Leonhard and Sabanés Bové, Daniel. 2014. Applied Statistical Inference: likelihood and bayes. Berlin, Heidelberg: Springer, Berlin, Heidelberg.

Box, G., Jenkins, G., Reinsel, G. and Ljung, G. 2016. Time Series Analysis: Forecasting and Control. Fifth edition. Hoboken, New Jersey: John Wiley Sons, Inc.

Tommaso, Proietti and Helmut, Luetkepohl. 2011. Does the Box-Cox transformation help in forecasting macroeconomic time series? https://mpra.ub.uni-muenchen.de/32294/ (Downloaded 2020-03-15)

Guerrero, Victor M. 1993. Time-series analysis supported by power transformations. Journal of Forecasting, Volume 12, p.37–48.

Linde, Peter. 2005. Seasonal Adjustment. Statistics Denmark. http://www.dst.dk/en/Statistik/dokumentation/metode/~/media/C7EE0A18506E44F694E209C789BF8699.pdf (Downloaded 2020-04-08).

SMHI, the Swedish Meteorological and Hydrological Institute. n.d (b). https://www.smhi.se/data/meteorologi/ladda-ner-meteorologiska-observationer (Downloaded 2020-02-09)

Thombre, Amit. 2016. Lambda (Box-Cox transformation parameter) value for forecasting using ARIMA. https://www.r-bloggers.com/lambda-box-cox-transformation-parameter-value-for-forecasting-using-arima/ (Accessed 2020-06-10)

8 Appendices

8.1 Appendix A: Initial diagnostics/evaluations

8.1.1 Box-Cox transformation

The `r` package `forecast` has a function `BoxCox.lambda` which employs two methods of calculating λ, the "guerrero" method or the "loglik" method. There is an experimental study exploring the methods available in `BoxCox.lambda` with the purpose to find out which method benefits the forecasting the best *(Thombre, 2016)*. From the conclusion that the "guerrero" method provided slightly better forecasting results, that was the method applied in this thesis. The method of calculating λ was presented by *(Guerrero, 1993)* and it chooses the λ that minimizes the coefficient of variation for time series. Analyzing the method further is outside the scope of this thesis.

The Box-Cox transformations are given by:

$$x_t(\lambda) = \begin{cases} \frac{x_t^{\lambda}-1}{\lambda} & , \lambda \neq 0 \\ \ln x_t & , \lambda = 0 \end{cases} ,$$

so for $\lambda = 1$ the original data scale is used. Note that the observations are required to be positive for the transformation to work.

By the `BoxCox.lambda` function we are provided with the values for λ presented in *Appendix A: Table 4*.

8.1.1.1 Inverse Box-Cox transformation Reversing the transformation to original scale can be accomplished by

$$x_t = \begin{cases} (\lambda x_t(\lambda) + 1)^{1/\lambda} & , \lambda \neq 0 \\ \exp(x_t(\lambda)) & , \lambda = 0 \end{cases} .$$

8.1.1.2 Box-Cox table of transformations

Table 7: The Box-Cox transformation values of the parameter λ for the monthly mean wind speed (λ_W) and precipitation (λ_P) for all stations.

	Wind speed, (λ_W)	Precipitation, (λ_P)
South		
Falsterbo	-0.4378918	0.0823902
Hallands Väderö	-0.1148000	0.1526751
Hanö	-0.9999242	0.0986390
Vinga	-0.4391843	0.2874954
Visby	-0.3965855	0.1721041
Mid		
Hagshult	0.1110818	0.1596284
Såtenär	-0.9021474	0.0593852
Sunne	0.7857286	0.1002408
Tullinge	0.1258573	0.0754168
Stockholm	0.0529460	0.0747601
North		
Krångede	0.4101825	-0.0711494
Örnsköldsvik	0.2952780	0.0714934
Lycksele	0.5071668	0.1369188
Rödkallen	-0.0416061	0.2016316
Arjeplog	0.5128968	0.1086850

8.1.1.3 Box-Cox transformations plots

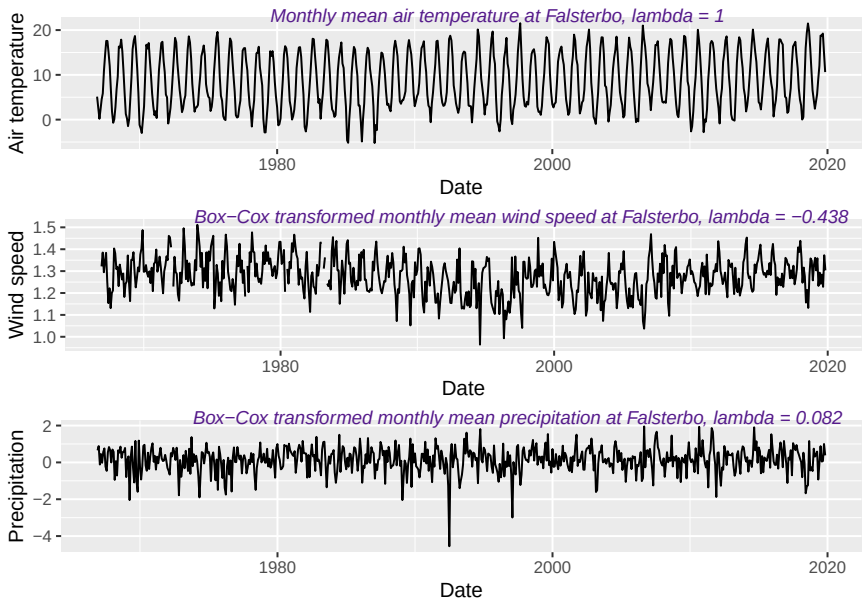

Figure 8: Plots of the monthly mean air temperature, $\{X_t\}$, the Box-Cox transformed monthly mean wind speed, $\{V_t(\lambda_P)\} = \{V_t(-0.438)\}$, and precipitation, $\{Y_t(\lambda_W)\} = \{Y_t(0.082)\}$, in Falsterbo.

8.1.2 ADF test results

Table 8: ADF-tests of the monthly mean air temperature series $\{X_t\}$, the Box-Cox transformed monthly mean wind speed series $\{V_t(\lambda_P)\}$ and precipitation series $\{Y_t(\lambda_W)\}$, in the south of Sweden. The test were computed on all the five stations in the category South.

	Air temperature		Transformed wind speed		Transformed precipitation	
	Statistic	p-value	Statistic	p-value	Statistic	p-value
Falsterbo	-5.110716	0.0100000	-1.732564	0.6915102	-17.035680	0.01
Hallands Väderö	-3.308740	0.0700543	-3.917002	0.0135719	-9.273514	0.01
Hanö	-4.638676	0.0100000	-3.195784	0.0886557	-16.924571	0.01
Vinga	-4.352447	0.0100000	-2.375166	0.4195061	-5.708321	0.01
Visby	-5.223918	0.0100000	-2.420918	0.4001391	-8.867496	0.01

8.1.3 ACF and PACF plots

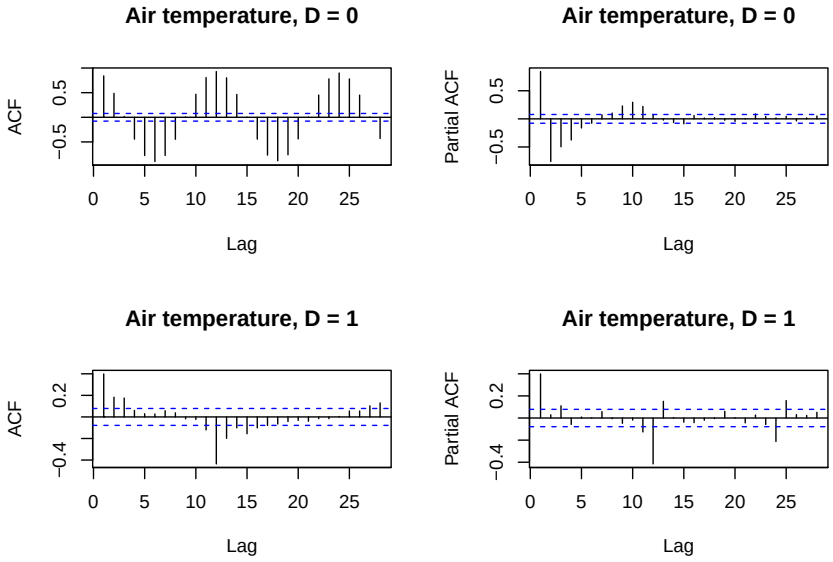

Figure 9: Sample ACF and PACF plots of monthly mean air temperature in Falsterbo series $\{X_t\}$. The seasonal differencing is denoted by D and is defined in Section 3.2.1.

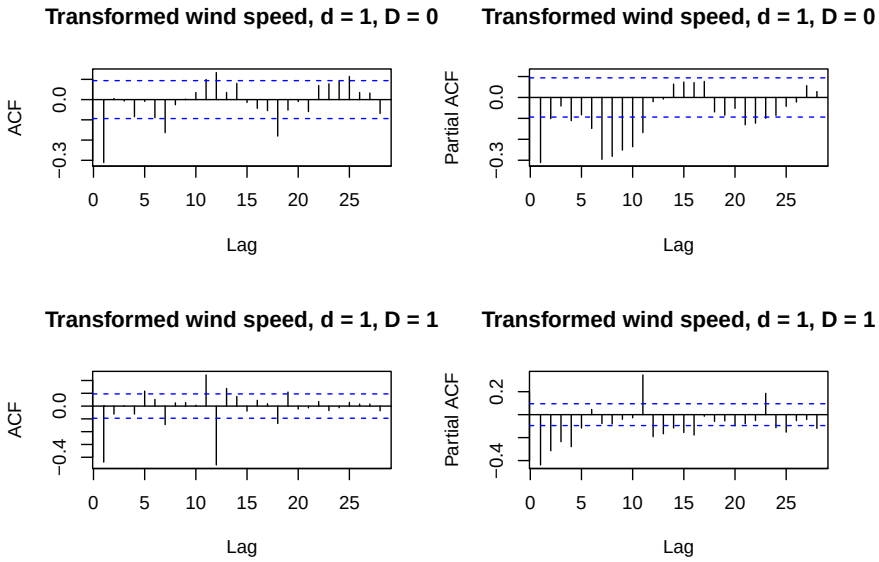

Figure 10: Sample ACF and PACF plots of the Box-Cox transformed monthly mean wind speed in Falsterbo series $\{V_t(\lambda_P)\}$. The regular differencing is denoted d and the seasonal differencing D at lag 12, which are defined in Section 3.2.1.

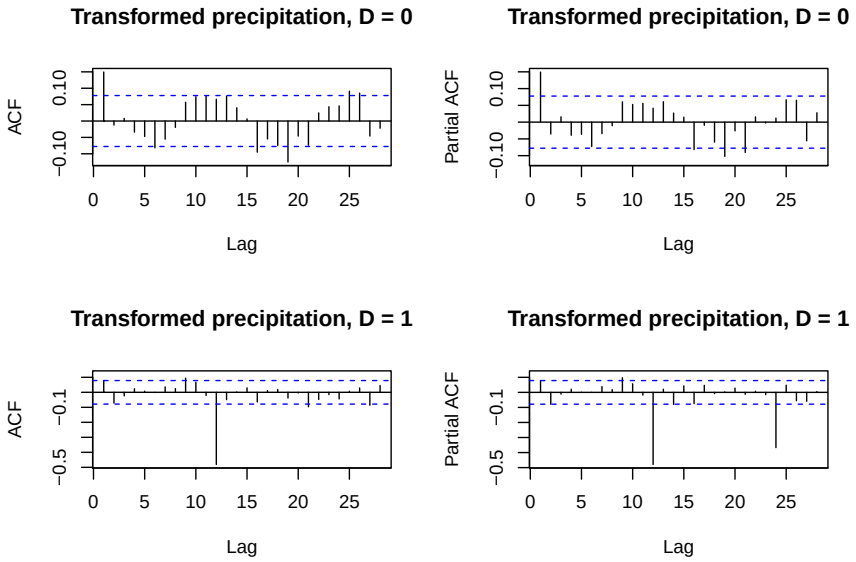

Figure 11: Sample ACF and PACF plots of the Box-Cox transformed monthly mean precipitation in Falsterbo series $\{Y_t(\lambda_W)\}$. The seasonal differencing is denoted by D and is defined in Section 3.2.1.

8.2 Appendix B: Model evaluation

8.2.1 Model selection

Table 9: Table of some of the considered models in the model selection in Section 5.1 for the monthly mean weather. The models selected for forecasting are higlighted in the table. The model chosen from the model selection is highlighted in light green, while the naive model and seasonal naive model are highlighted in progressively darker shades of green.

South		
Model	AIC	RMSE
Air temperature		
ARIMA$(0,0,1)(0,1,1)_{12}$	2284.27	1.439356
ARIMA$(0,0,2)(0,1,1)_{12}$	2267.2	1.46336
ARIMA$(0,0,3)(0,1,1)_{12}$	2253.6	1.419938
ARIMA$(1,0,0)(0,1,1)_{12}$	**2248.76**	**1.417881**
ARIMA$(1,0,1)(0,1,1)_{12}$	2248.45	1.414223
ARIMA$(0,1,0)(0,0,0)_{12}$	**3405.4**	**3.525834**
ARIMA$(0,0,0)(0,1,0)_{12}$	**2680.35**	**2.049622**
Box-Cox transformed wind speed		
ARIMA$(0,1,1)(0,1,1)_{12}$	-1624.01	0.062277
ARIMA$(0,1,2)(0,1,1)_{12}$	-1626.75	0.061996
ARIMA$(0,1,3)(0,1,1)_{12}$	-1625.9	0.061937
ARIMA$(1,1,1)(0,1,1)_{12}$	-1627.21	0.061972
ARIMA$(1,1,2)(0,1,1)_{12}$	**-1629.31**	**0.061857**
ARIMA$(1,1,3)(0,1,1)_{12}$	**-1624.23**	**0.061896**
ARIMA$(0,1,0)(0,0,0)_{12}$	**-1324.01**	**0.084174**
ARIMA$(0,0,0)(0,1,0)_{12}$	-1277.19	0.085042
Box-Cox transformed precipitation		
ARIMA$(0,0,0)(0,1,1)_{12}$	1321	1.175032
ARIMA$(0,0,1)(0,1,1)_{12}$	**1318.38**	**1.175942**
ARIMA$(1,0,0)(0,1,1)_{12}$	1318.96	1.175582
ARIMA$(1,0,1)(0,1,1)_{12}$	1323.49	1.172896
ARIMA$(0,1,0)(0,0,0)_{12}$	**1683.46**	**0.908696**
ARIMA$(0,0,0)(0,1,0)_{12}$	**1716.48**	**0.946784**

8.2.1.1 Ljung Box test

Table 10: Ljung-Box test on residuals of monthly mean air temperature, Box-Cox transformed monthly mean wind speed and precipitation. The test statistics $Q(m)$ are asymptotically $\chi^2_m(\alpha)$ distributed, which on a 5 percent level are $\chi^2_{12}(0.05) = 21.026$, $\chi^2_{24}(0.05) = 36.415$ and $\chi^2_{36}(0.05) = 50.998$ for the used lag-m.

	Air temperature		Transformed wind speed		Transformed precipitation	
lags	$Q(m)$	p-value	$Q(m)$	p-value	$Q(m)$	p-value
Models: ARIMA(1,0,0)(0,1,1)[12], ARIMA(1,1,2)(0,1,1)[12], ARIMA(0,0,1)(0,1,1)[12]						
$m = 12$	17.932	0.118	14.214	0.287	6.917	0.863
$m = 24$	30.557	0.167	22.298	0.561	17.222	0.839
$m = 36$	51.181	0.048	36.066	0.466	28.144	0.822
Naive model						
$m = 12$	2252.5	0	109.9	0	115.6	0
$m = 24$	4484.2	0	144.8	0	134.9	0
$m = 36$	6657.9	0	186.6	0	155.9	0
Seasonal naive model						
$m = 12$	282.6	0	225.5	0	168.4	0
$m = 24$	347.9	0	268.9	0	183.6	0
$m = 36$	410.3	0	282.7	0	196.6	0

8.2.2 Models

8.2.2.1 Air temperature

8.2.2.1.1 ARIMA$(1,0,0)(0,1,1)_{12}$ model

Table 11: Summary of the ARIMA$(1,0,0)(0,1,1)_{12}$ forecasting model for the monthly mean air temperature in Falsterbo series $\{X_t\}$.

	ARIMA(1,0,0)(0,1,1)[12]	
	Estimates	Standard errors
AR	0.5013920	0.0013019
SMA	-0.9298825	0.0004073

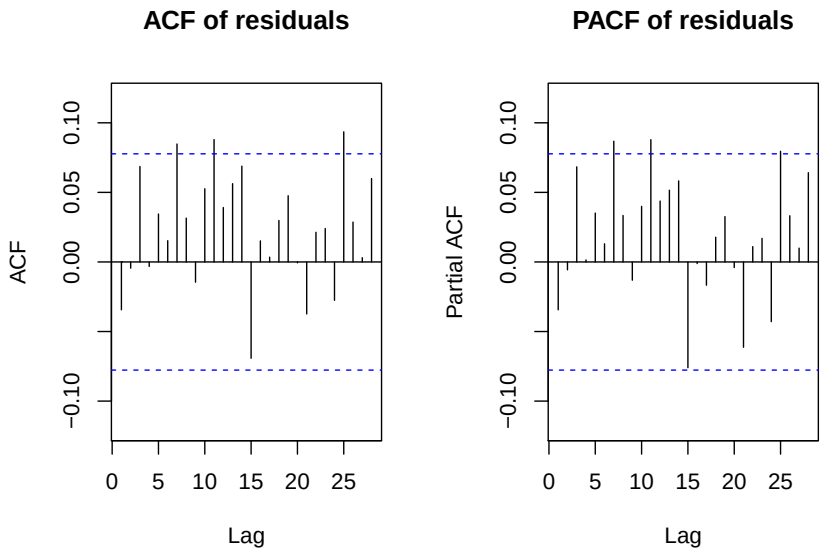

Figure 12: Sample ACF and PACF plots of the ARIMA$(1,0,0)(0,1,1)_{12}$ model's residuals of the monthly mean air temperature in Falsterbo series $\{X_t\}$.

8.2.2.1.2 Naive models (air temperature)

Found in *Section 5.2*.

8.2.2.2 Box-Cox transformed wind speed

8.2.2.2.1 ARIMA$(1,1,2)(0,1,1)_{12}$ model

Table 12: Summary of the ARIMA$(1,0,0)(0,1,1)_{12}$ forecasting model for the Box-Cox transformed monthly mean wind speed in Falsterbo series $\{V_t(\lambda_W)\}$.

	ARIMA(1,0,1)(0,1,1)[12]	
	Estimates	Standard errors
AR	0.8334250	0.0066899
MA1	-1.6706499	0.0093987
MA2	0.6806928	0.0083891
SMA	-0.9575664	0.0007132

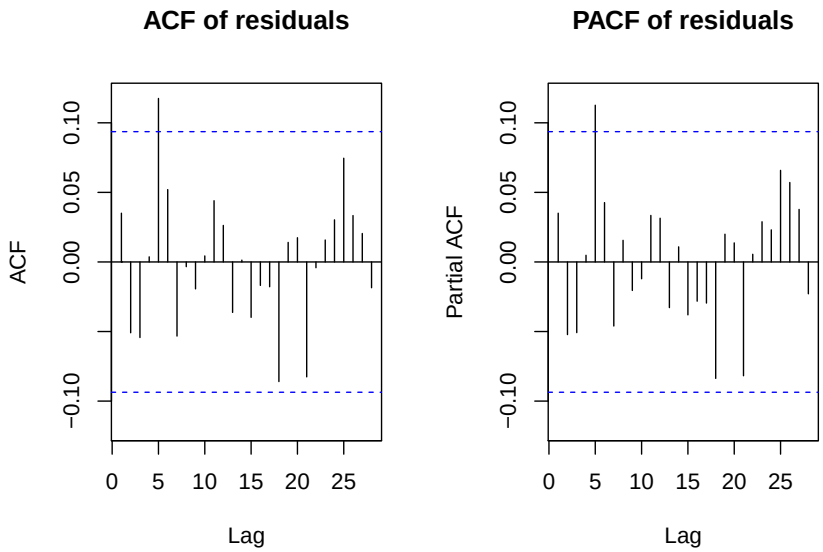

Figure 13: Sample ACFs and PACFs plots of the ARIMA$(1,1,2)(0,1,1)_{12}$ model's residuals of the Box-Cox transformed monthly mean wind speed in Falsterbo series $\{V_t(\lambda_W)\}$.

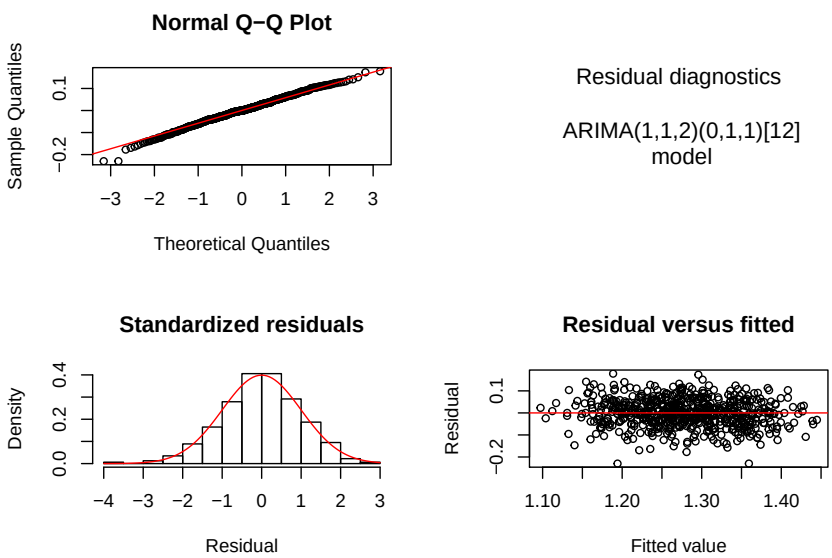

Figure 14: Residual diagnostics on the ARIMA$(1,1,2)(0,1,1)_{12}$ model for the Box-Cox transformed monthly mean wind speed in Falsterbo series $\{V_t(\lambda_W)\}$.

8.2.2.2.2 Naive models (Box-Cox transformed wind speed)

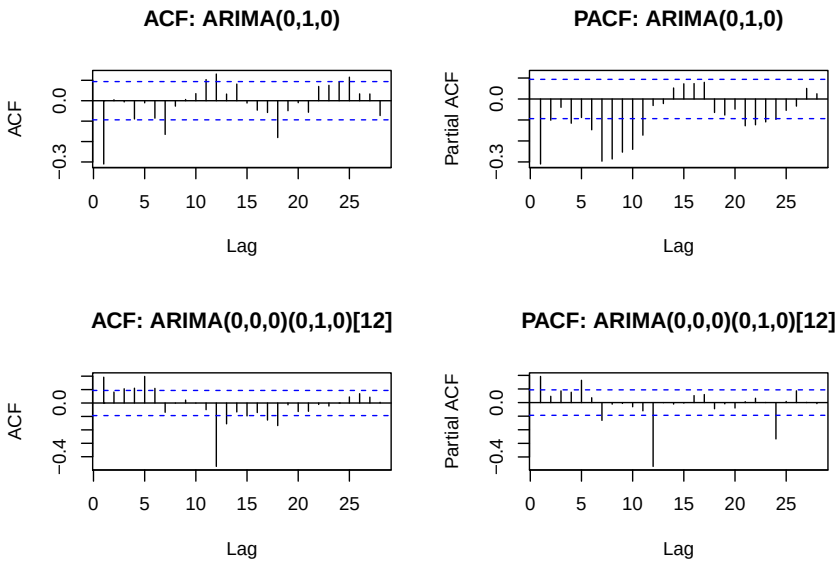

Figure 15: Sample ACFs and PACFs plots of the naive models residuals of the Box-Cox transformed monthly mean wind speed in Falsterbo series $\{V_t(\lambda_W)\}$. In these plots the naive model is denoted by ARIMA$(0,1,0)$ and the seasonal naive model is represented by ARIMA$(0,0,0)(0,1,0)[12]$.

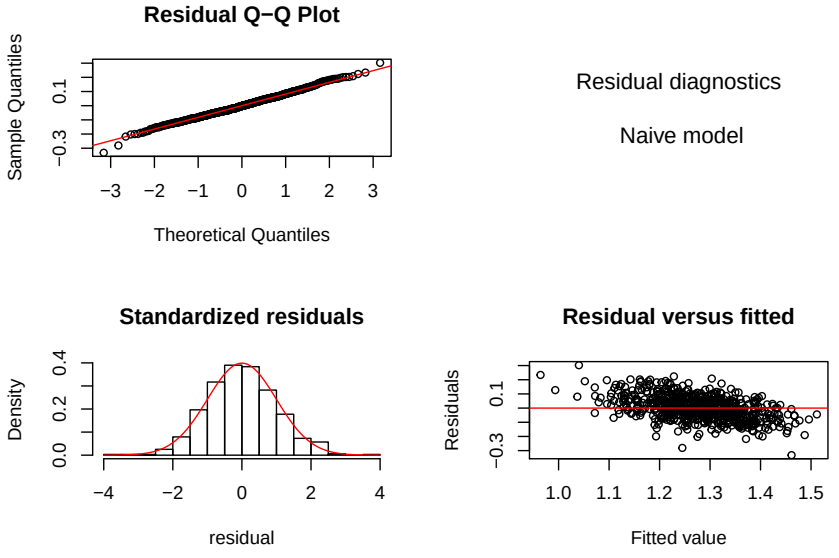

Figure 16: Residual diagnostics for the naive model of the Box-Cox transformed monthly mean wind speed in Falsterbo series $\{V_t(\lambda_W)\}$.

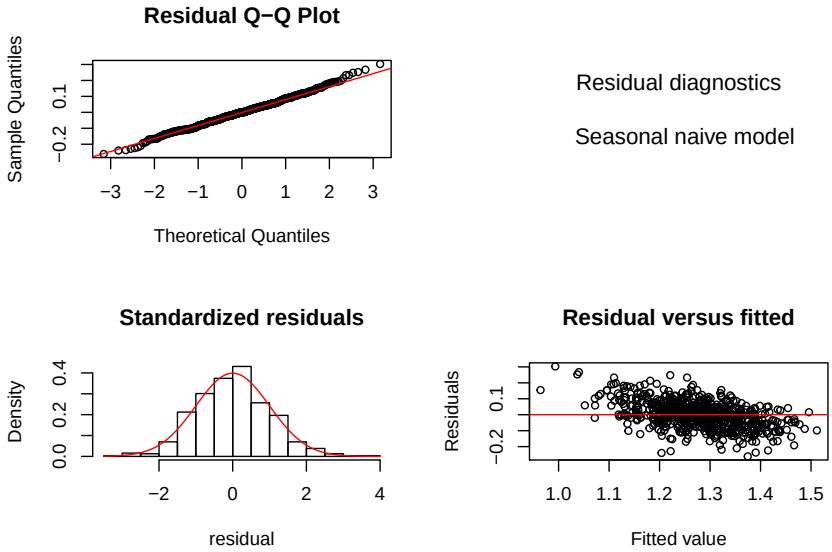

Figure 17: Residual diagnostics for the seasonal naive model of the Box-Cox transformed monthly mean wind speed in Falsterbo series $\{V_t(\lambda_W)\}$.

8.2.2.3 Box-Cox transformed precipitation

8.2.2.3.1 ARIMA$(0,0,1)(0,1,1)_{12}$ model

Table 13: Summary of the ARIMA$(1,0,0)(0,1,1)_{12}$ forecasting model for the Box-Cox transformed monthly mean precipitation in Falsterbo series $\{Y_t(\lambda_P)\}$.

	ARIMA(0,0,1)(0,1,1)[12]	
	Estimates	Standard errors
MA	0.0918550	0.0018120
SMA	-0.9858011	0.0020567

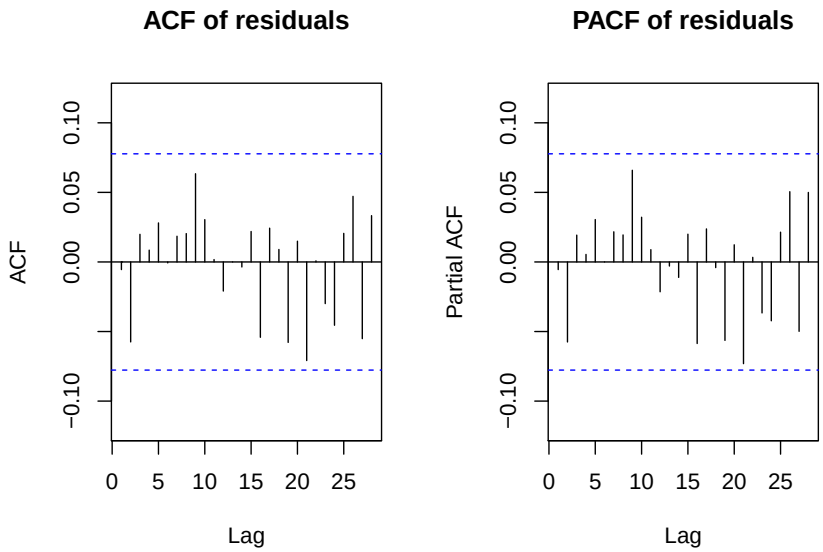

Figure 18: Sample ACFs and PACFs plots of the ARIMA$(0,0,1)(0,1,1)_{12}$ model's residuals of the Box-Cox transformed monthly mean precipitation in Falsterbo series $\{Y_t(\lambda_P)\}$.

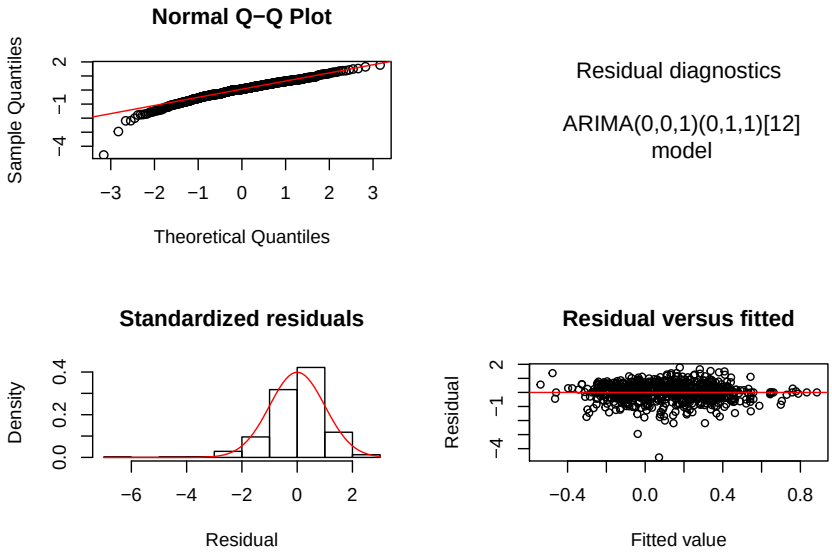

Figure 19: Residual diagnostics on the ARIMA$(0,0,1)(0,1,1)_{12}$ model for the Box-Cox transformed monthly mean precipitation in Falsterbo series $\{Y_t(\lambda_P)\}$.

8.2.2.3.2 Naive models (Box-Cox transformed precipitation)

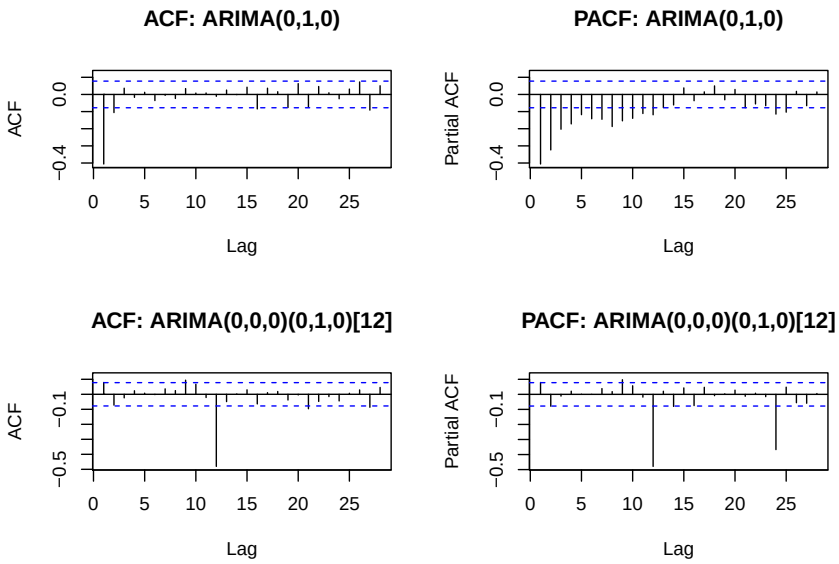

Figure 20: Sample ACFs and PACFs plots of the naive models residuals of the Box-Cox transformed monthly mean precipitation in Falsterbo series $\{Y_t(\lambda_P)\}$. In these plots the naive model is denoted by ARIMA$(0,1,0)$ and the seasonal naive model is represented by ARIMA$(0,0,0)(0,1,0)[12]$.

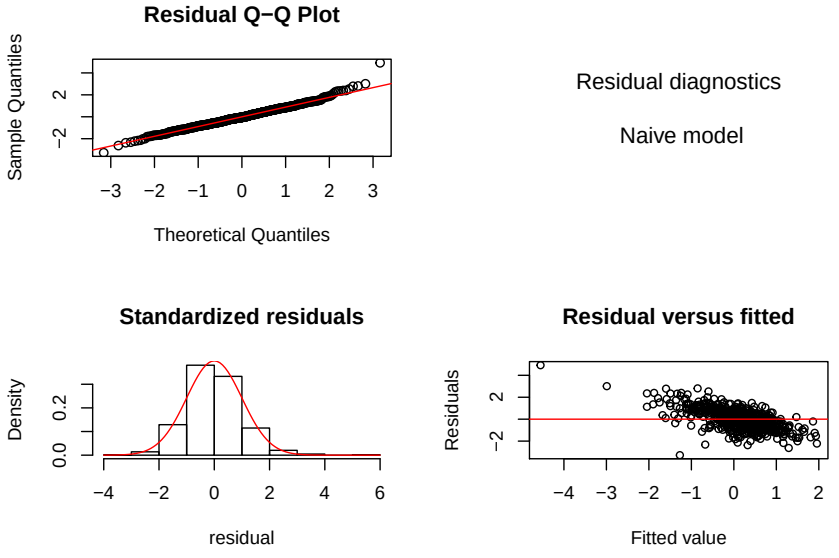

Figure 21: Residual diagnostics of the naive model for the Box-Cox transformed monthly mean precipitation in Falsterbo series $\{Y_t(\lambda_P)\}$.

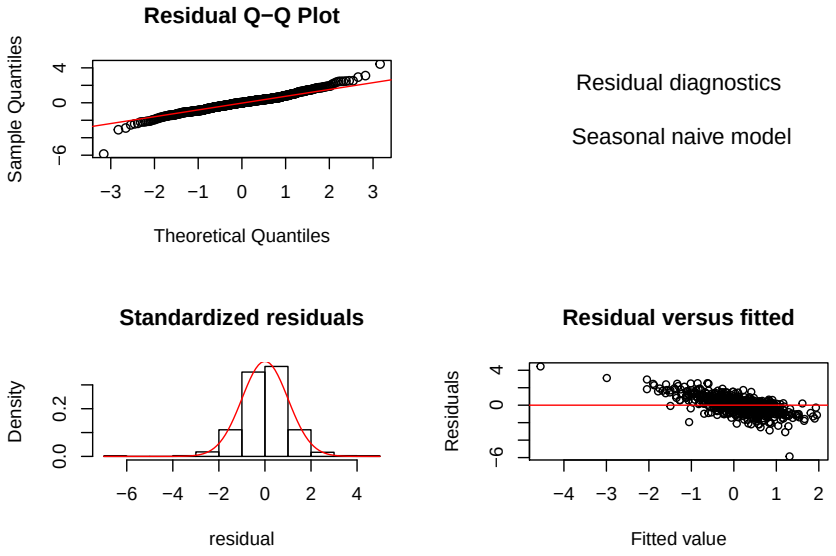

Figure 22: Residual diagnostics of the seasonal naive model for the Box-Cox transformed monthly mean precipitation in Falsterbo series $\{Y_t(\lambda_P)\}$.

8.3 Appendix C: Forecasts

8.3.1 Air temperature

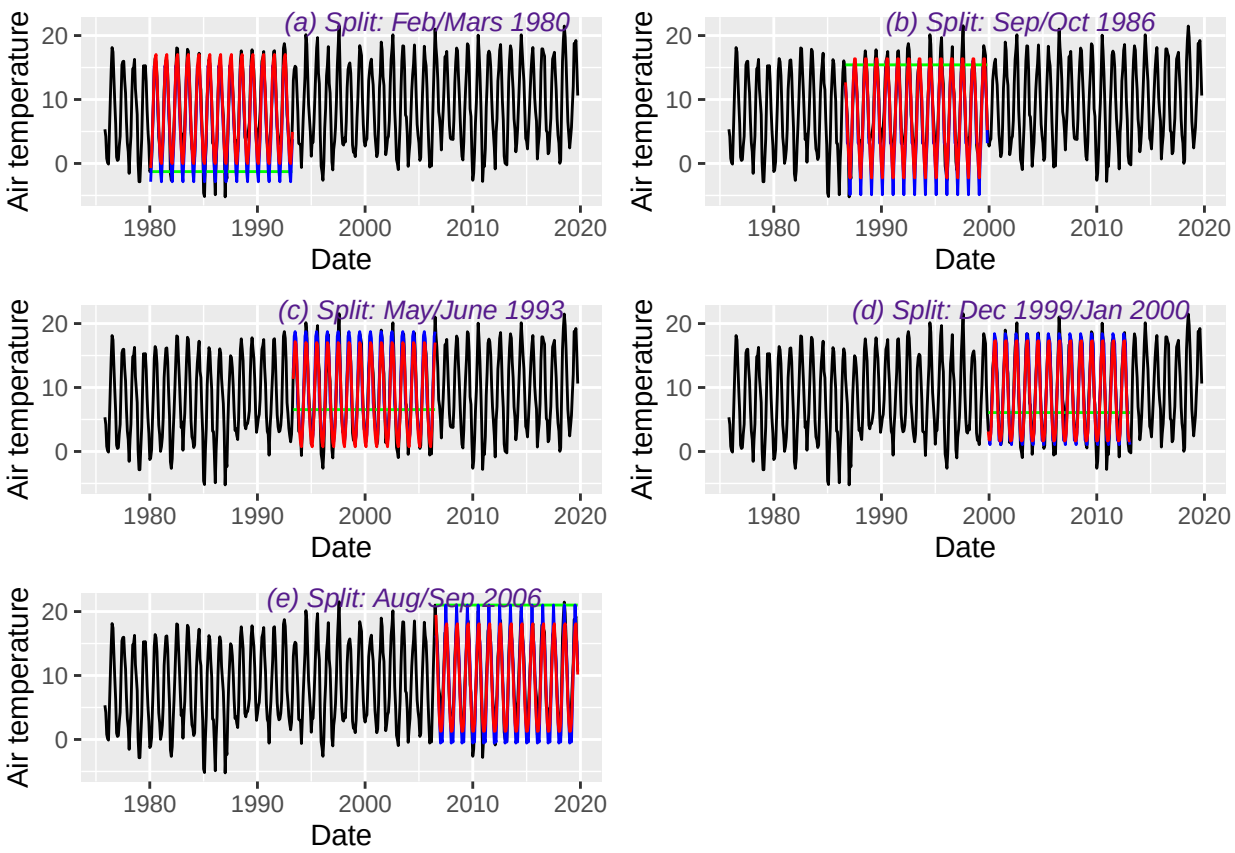

Figure 23: 1 to 159 steps ahead out-of-sample forecast for the monthly mean air temperature in Falsterbo. The naive model is depicted in green, the seasonal naive model in blue and the chosen model from the model selection, ARIMA$(1,0,0)(0,1,1)_{12}$ is in red. The split between the train and control samples was made between (a) February and Mars 1980, (b) September and October 1986, (c) May and June 1993, (d) December 1999 and January 2000, (e) August and September 2006. For all out-of-sample forecasts (a-e), 159 points in data prior to the location of the split was used to fit the models.

8.3.2 Wind speed

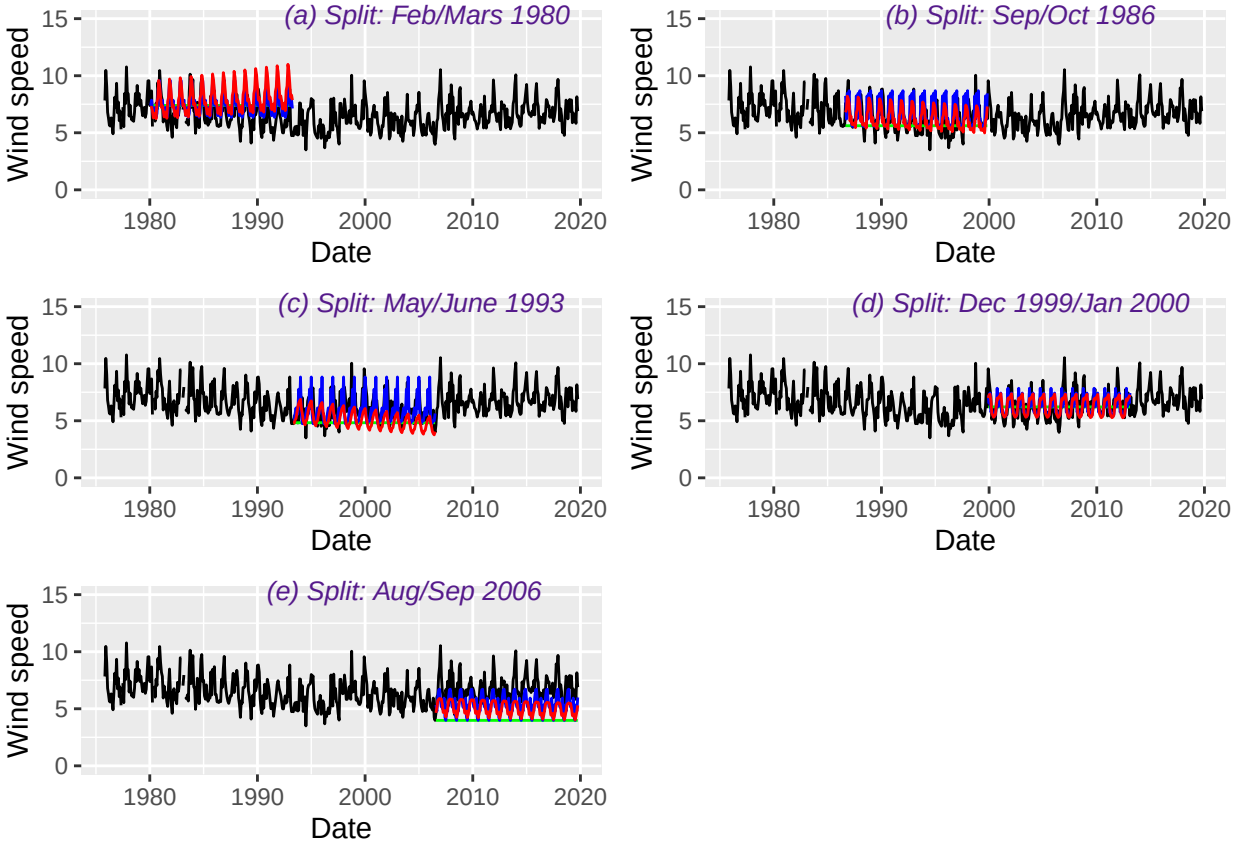

Figure 24: 1 to 159 steps ahead out-of-sample forecast for the monthly mean wind speed in Falsterbo. The naive model is depicted in green, the seasonal naive model in blue and the chosen model from the model selection, ARIMA$(1,1,2)(0,1,1)_{12}$ is in red. The split between the train and control samples was made between (a) February and Mars 1980, (b) September and October 1986, (c) May and June 1993, (d) December 1999 and January 2000, (e) August and September 2006. For all out-of-sample forecasts (a-e), 159 points in data prior to the location of the split was used to fit the models.

8.3.3 Precipitation

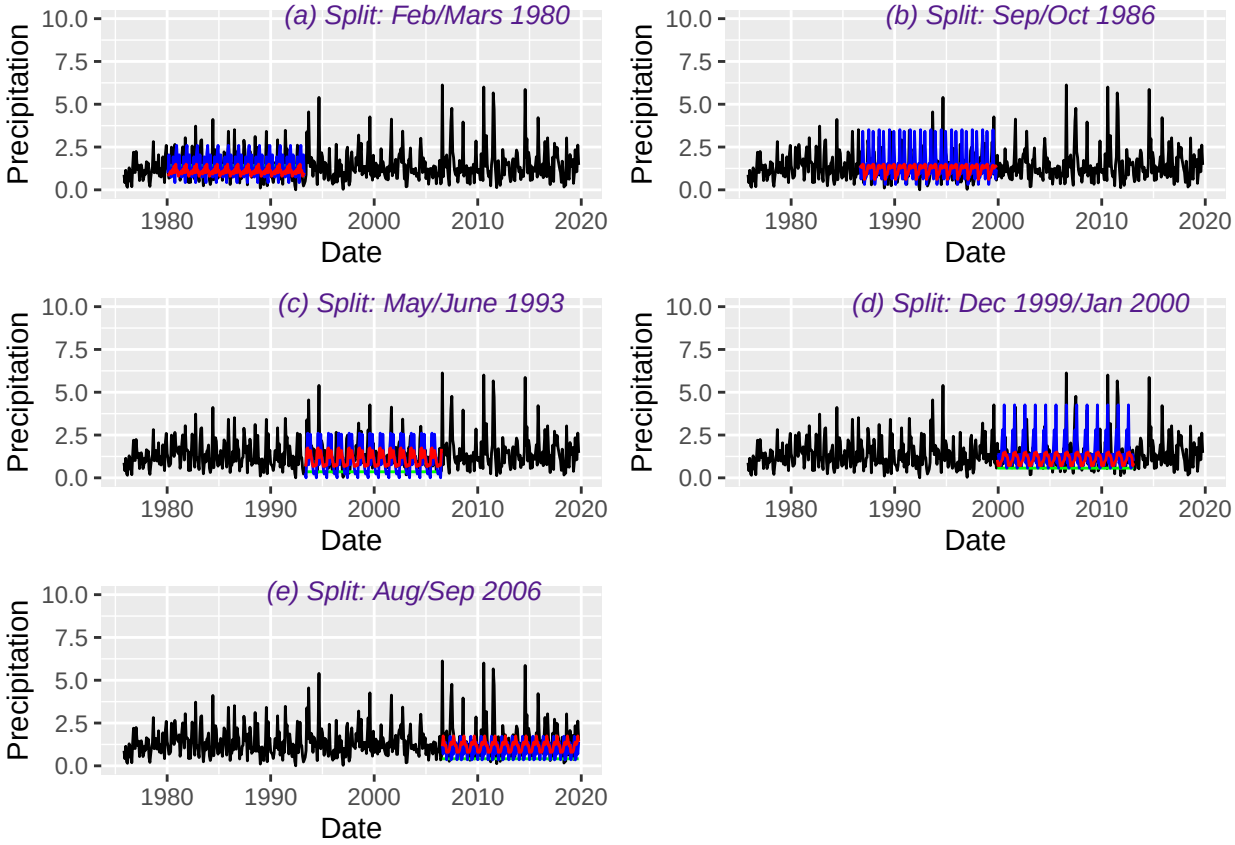

Figure 25: 1 to 159 steps ahead out-of-sample forecast for the monthly mean precipitation in Falsterbo. The naive model is depicted in green, the seasonal naive model in blue and the chosen model from the model selection, ARIMA$(0,0,1)(0,1,1)_{12}$ is in red. The split between the train and control samples was made between (a) February and Mars 1980, (b) September and October 1986, (c) May and June 1993, (d) December 1999 and January 2000, (e) August and September 2006. For all out-of-sample forecasts (a-e), 159 points in data prior to the location of the split was used to fit the models.