

Forecasting Electricity Consumption

November 30, 2020

Contents

Introduction	1
Forecast without using outdoor temperature	1
Exponential smoothing: seasonal HW smoothing	4
Seasonal ARIMA	5
Seasonal Neural Network Auto-Regression (NNAR)	7
Forecast using outdoor temperature	8
Multivariate seasonal ARIMA (Dynamic regression model)	9
VAR model	10
Forecasting electricity consumption on 2010-02-17.	12

Introduction

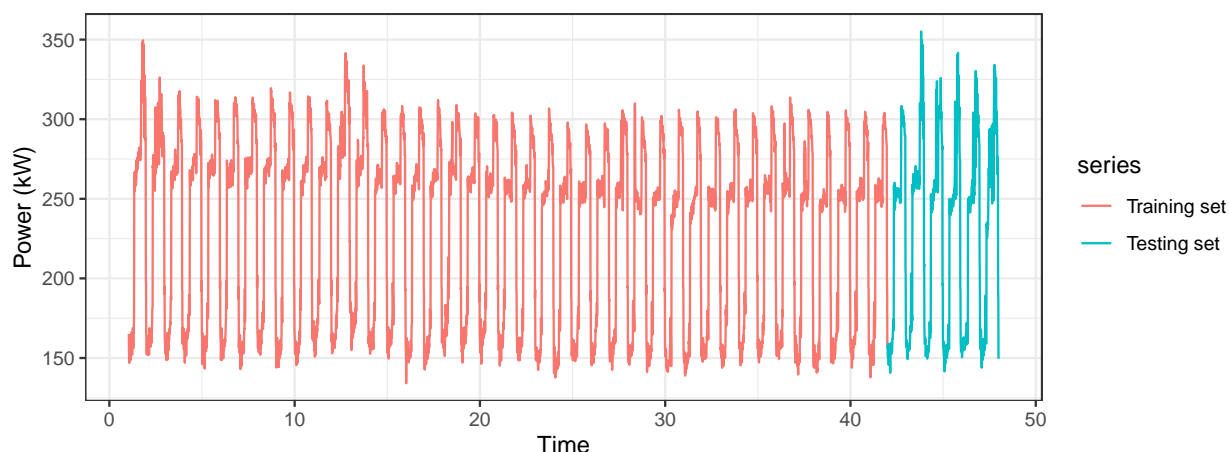
This report explains the different models that have been considered in order to forecast the electricity consumption of a building. An additional variable, outdoor air temperature, will also be considered to eventually improve the forecast. The dataset contains time series measured every 15 minutes and starting from 2010-01-01 to 2010-02-16. Our main purpose is to forecast electricity consumption for 2010-02-17. This forecast will be performed using both univariate and multivariate time series models. In the first section, we discuss the forecast in the univariate context and in the second one we will consider multivariate time series models; In the former case, we will consider exponential smoothing, NNAR, and ARIMA models, and in the latter case, we will model the series using dynamic regression, NNAR, and VAR models.

To train the model, we will split the data into two sets: the training set which starts from 2010-01-01 to 2010-02-10, and the testing set which starts from 2010-02-11 to 2010-02-16. The testing set contains 576 observations and represents about 12.78% of the total number of observations.

Forecast without using outdoor temperature

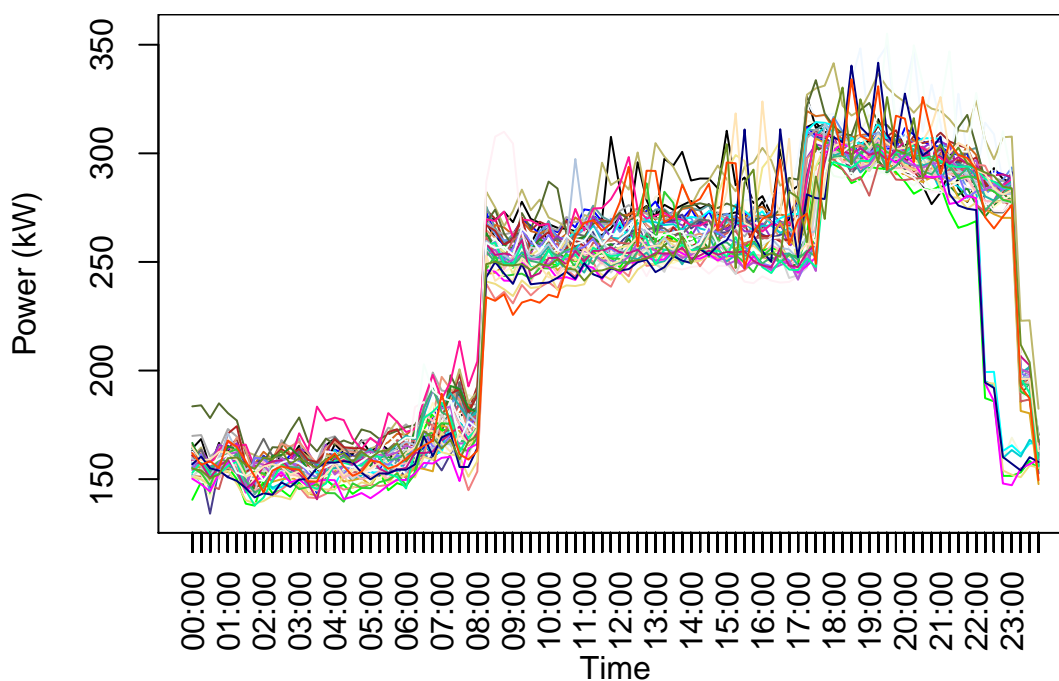
Let's start by plotting the electricity consumption series in order to have an insight about its pattern; As shown in Figure 1, electricity consumption is strongly seasonal although it is difficult to identify the periodicity visually.

Figure 1: Electricity consumption



To have a more detailed idea about the pattern of the cycles of this series, Figure 2 plots its seasonal pattern.

Figure 2: Seasonal plot of electricity consumption

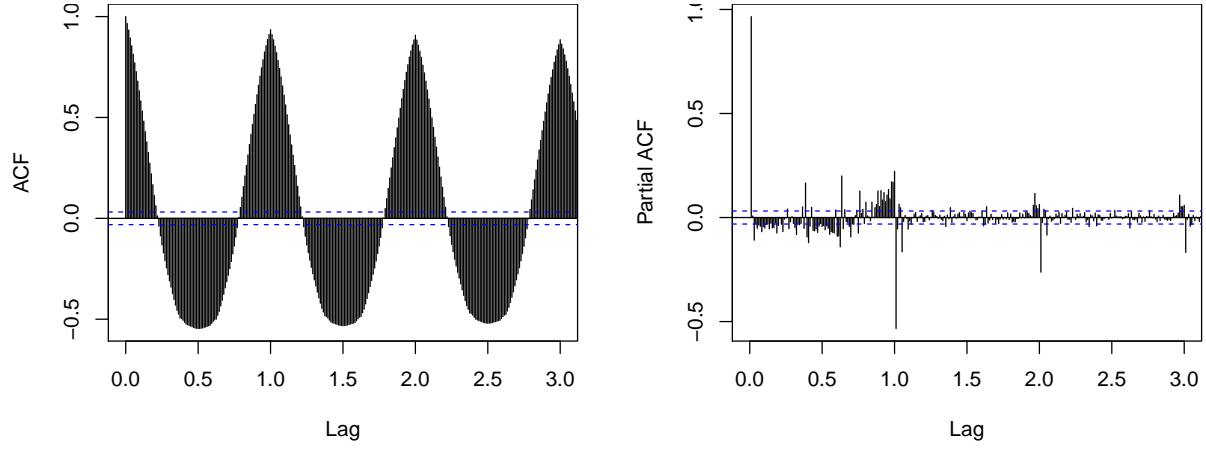


As the plot clearly shows, there is a strong seasonal pattern. A cycle can be roughly split into three main phases each of which has its own characteristics. From 00:00 to 08:15, we observe a electricity consumption (low power(KW)); this is likely due to the fact that at night people are sleeping. After, we observe a sudden jump from 08:15/08:30; this could be justified, for example, by the fact during the day people are working (or almost all companies are fonctionning). These two first phases also shows a slightly increasing trend. The last phase starts at around 17:15 till 23:00 where the electricity consumption increases further. After it sharply falls down around its starting point.

The features of this plot suggests the use of a model that takes into account a seasonal pattern. That is, models like seasonal-exponential smoothing, seasonal ARIMA, seasonal NNAR, or seasonal VAR.

To determine more precisely the models that could be used to forecast this time series, let's look at the ACF and PACF of the series.

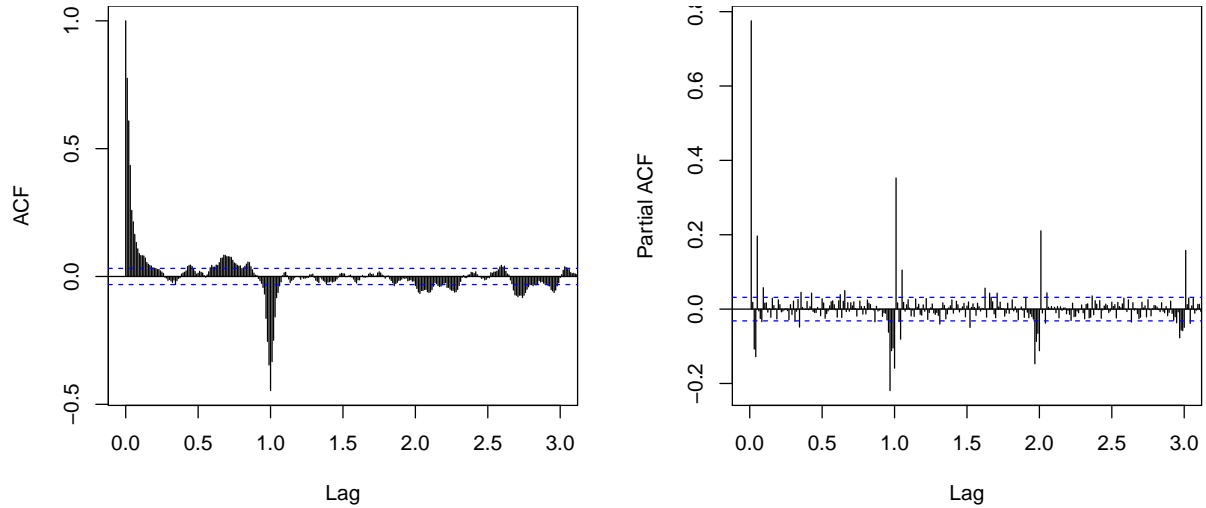
Figure 3: ACF and PACF of electricity consumption (in KW)



In Figure 3, the ACF plot (left panel) suggests that the series has a trend and is correlated; the pacf plot (right panel) is very persistent and shows a large correlation at 1, 2 etc. which, in our data, correspond to lag 96×1 , 96×2 , etc.¹. Further it tends to increase and then decrease repeatedly indicating that there is a seasonal pattern in the data. The ACF looks like a pure seasonal pattern series. Based on these characteristics, we can deduce that seasonal exponential smoothing (with and without dumping effect) are already two potential candidates to modelize our series.

In order to remove the linear trend and seasonal component in the data, we compute the first difference at lag 96 (length of a periodicity) of the series and then plot the ACF and PACF again.

Figure 4: ACF and PACF of the first difference of electricity consumption (in KW) at lag 96

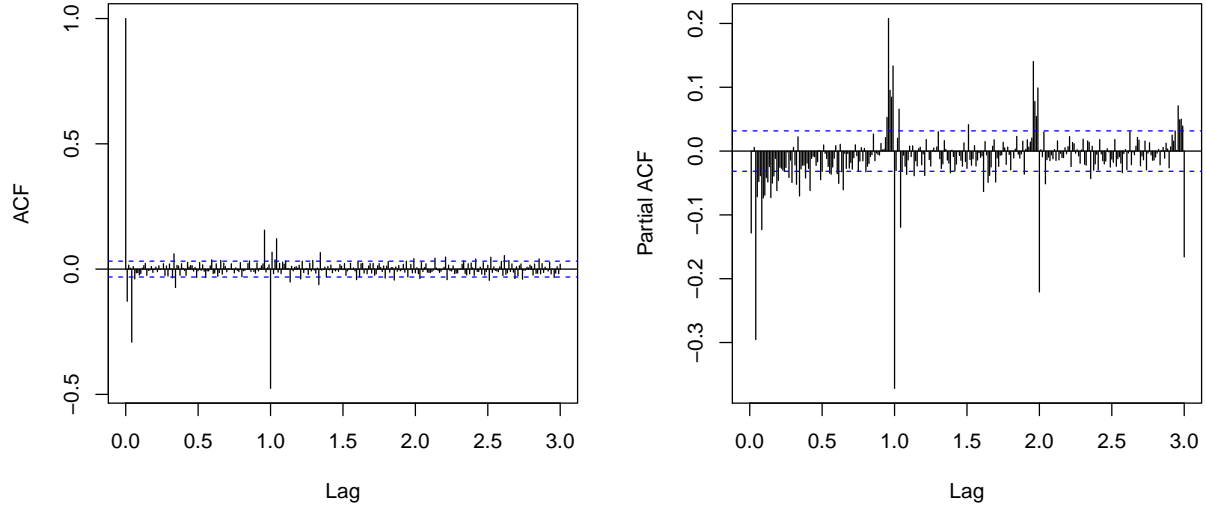


The plot, Figure 4, still shows that a strong exponential decaying autocorrelation and a persistent but exponentially decaying pacf at lag 96×1 , lag 96×2 , etc.

Let's difference the data once more.

¹Note that the periodicity in the series normally contains 96 observations (the number of data points observed in a day)

Figure 5: ACF and PACF of the 2nd difference of electricity consumption (in KW)



The significant ACF at lag 1 and 2 suggests an AM_2 for the seasonal part of SARIMA; and the significant at lag 96 (1 on the graph) suggests an MA_1 for the seasonal component; this leads to a $SARIMA_{(0,1,2)(0,1,1)}$. Similar interpretation of the PACF suggests an AR_1 or AR_2 or even more for the seasonal part and to do the same for the non-seasonal component. As a consequence, we have to consider $SARIMA_{(p,1,0)(1,1,1)}$ or $SARIMA_{(p,1,0)(2,1,1)}$ for several reasonable values of p .

When we continue to difference this series (even more than 10 times), the ACF and PACF do not show any improvement; that is, we continue to observe a persistent correlation on both graphs. This is a clear sign that ARIMA model (and SARIMA) is likely not appropriate for this data.

The last model that we will consider for this time series is a seasonal neural network auto-regressive model.

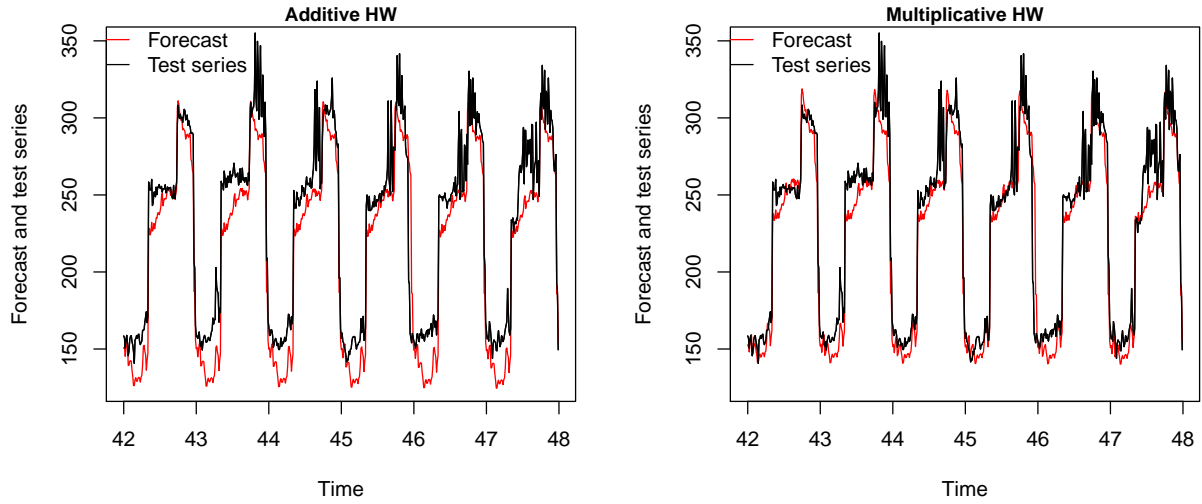
To summarize, the potentials models that we selected are: seasonal HW smoothing (with and without damping effect), seasonal $ARIMA$ (SARIMA), and seasonal neural network.

Let estimates these model in turn and check their performance in forecasting using the test dataset.

Exponential smoothing: seasonal HW smoothing

Figure 6 plots the true values of the test series and the forecast values obtained using seasonal HW smoothing models where the seasonal component of the series enter the model both additively and multiplicatively.

Figure 6: Forecast vs Test series using seasonal HW smoothing



In Figure 6, the forecast with multiplicative seasonal component is relatively closer to the true values of the series compared to that with additive seasonal component; that is, when looking at the graph, the model seasonal HW smoothing performs better. To quantify the forecast performance, let compute the root mean square error in both cases.

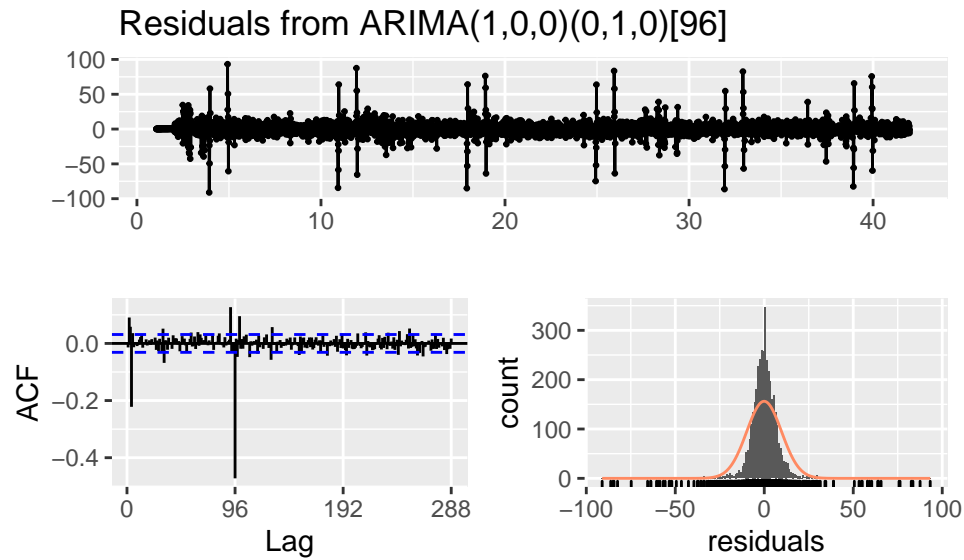
The RMSE using Multiplicative seasonal Holt-Winters is lower than that obtained with Additive seasonal Holt-Winters (17.27 vs 22.79)

A noticeable remark on this plot is that forecast tend to be systematically lower than the true value of the series. This effect is more pronounced when considering additive seasonality, making the Holt-Winters model with multiplicative seasonality compact relatively better.

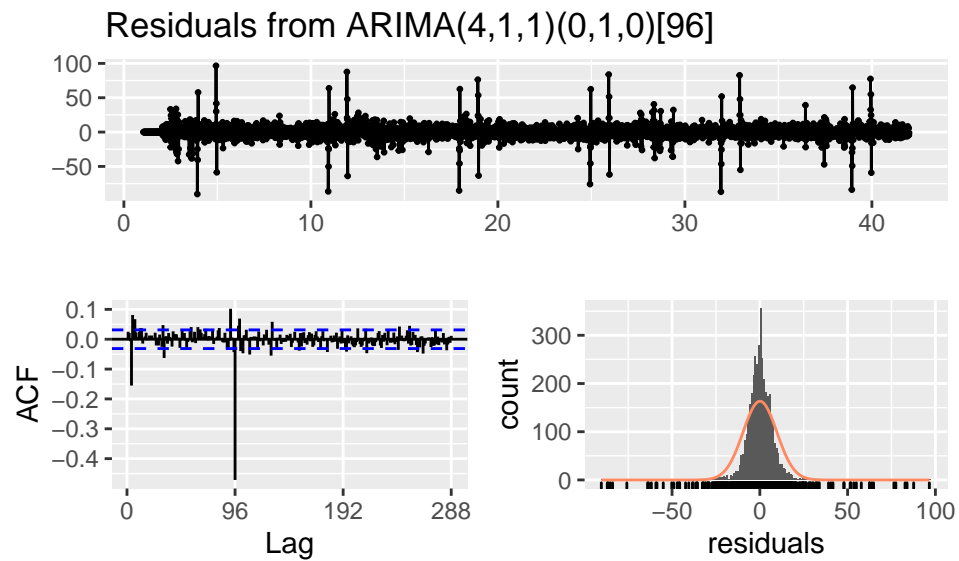
Unfortunately, The seasonal Holt-Winters model with damped effect cannot be performed on our series due to the fact that the frequency per period (96 observations per period) is higher than the maximum value allowed (24 observations per period) in the `hw` function of the `forecast` package. Therefore, this model will not be considered.

Seasonal ARIMA

Because of the large size and frequency of the series, almost all seasonal SARIMA require a more powerful machine for its execution. Nevertheless, two cases have been implemented but the analysis of the residuals is far from being satisfactory. The normality test does not hold at all (regardless of the number of times we compute the series difference)



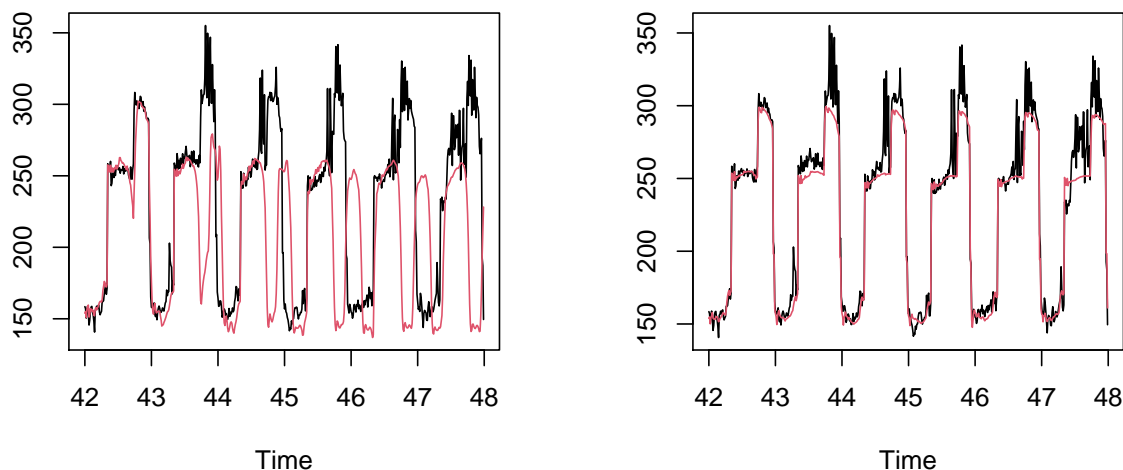
```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(1,0,0)(0,1,0) [96]
## Q* = 1531.9, df = 191, p-value < 2.2e-16
##
## Model df: 1.   Total lags used: 192
```



```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(4,1,1)(0,1,0) [96]
## Q* = 1425, df = 187, p-value < 2.2e-16
##
## Model df: 5.   Total lags used: 192
```

Seasonal Neural Network Auto-Regression (NNAR)

In this section, we will consider two cases of seasonal NNAR: one where the parameters values are automatically chosen by the function and another where the lag orders for seasonal and non-seasonal components are set to 7. The value 7 has been chosen in order to potentially capture days of weeks effect.

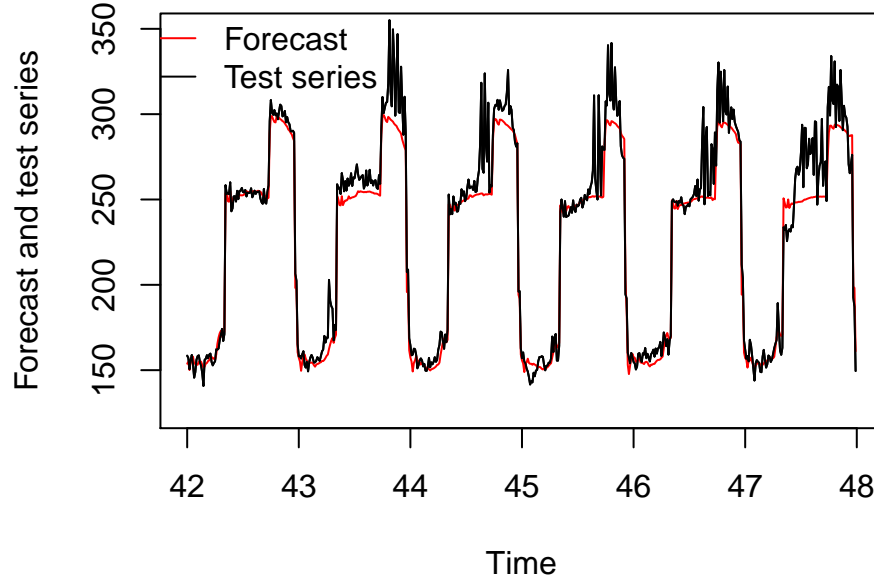


The RMSE using auto-selection of parameter values is actually much more poor than that obtained by considering the order of autocorrelation of seasonal and non-seasonal components of the series to be 7.

Further, the RMSE using Neural Network Auto-Regression (NNAR) is largely lower than that obtained using Exponential smoothing models before (14.05 vs 17.27). This happens only when we modify the parameters of the NNAR compared to automatic NNAR which has a large RMSE. The new value of RMSE is about 20.78% the value obtained with automatic choice of the values of the parameters

The forecast and test series are both plotted in Figure 7.

Figure 7: Forecast vs Test series using seasonal NNAR



As shown in Figure 7, the forecast is clearly better than that obtained using exponential smoothing model. This forecast solves (partially) the problem observed with smoothing model, which systematically underestimates the series; but even in this case, the series tends to be underestimated around the peak of the seasonality pattern.

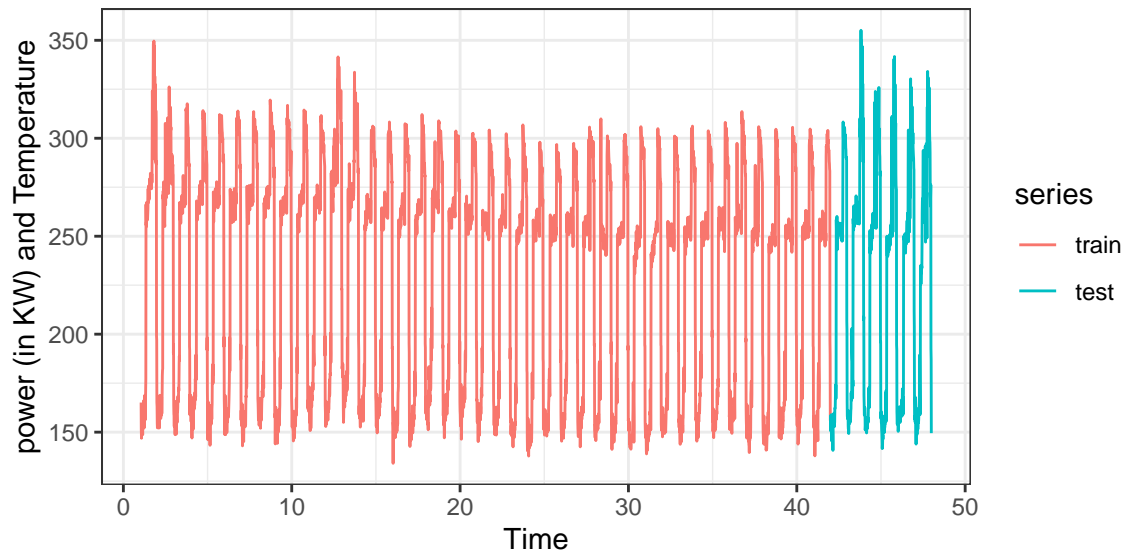
To summarize, the model that performs better among all those used so far is the seasonal NNAR.

Forecast using outdoor temperature

In this section, we will move further by considering the temperature together with past values of the electricity consumption to make the forecast; the aim being to use temperature to potentially improve the forecast performance.

The data used for this section is plotted in Figure 8

Figure 8: Electricity consumption and Temperature

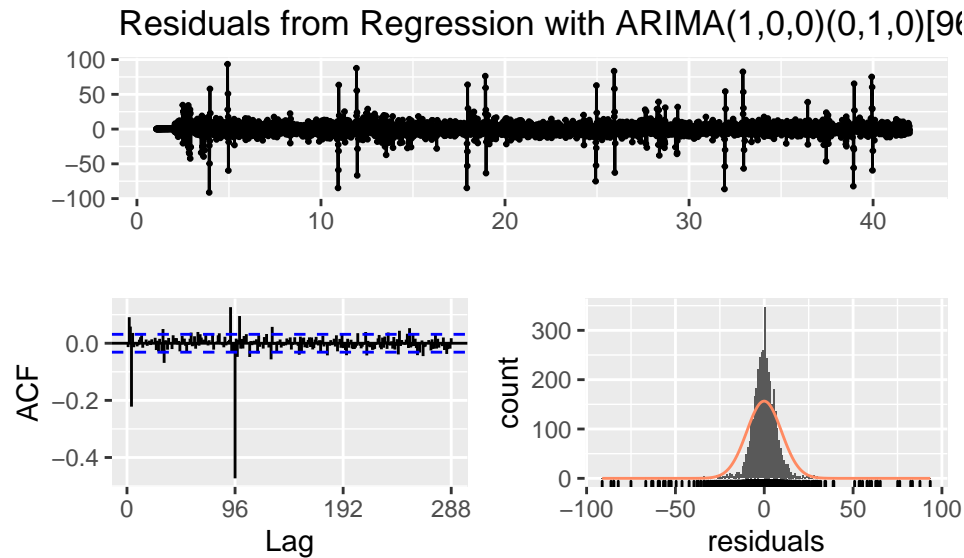


Multivariate seasonal ARIMA (Dynamic regression model)

Let estimate a dynamic regression using the `auto.arima` function. I also tried the model used before, an $ARIMA_{(2,1,1)(2,1,1)}$, and $ARIMA_{(2,1,1)(1,1,1)}$ but in both cases the algorithm used in `Arima` complained and did not allow to estimate the model. For this reason, only the automatic arima will be considered. `D` is set to 1 to enforce seasonality.

When looking at the seasonality plot, it is obvious that including temperature in the model did not allow to obtain residuals that are normally distributed in practice. Adding lag values did not allow to obtain satisfactory residuals.

```
## Series: train[, "Power (kW)"]
## Regression with ARIMA(1,0,0)(0,1,0)[96] errors
##
## Coefficients:
##      ar1      xreg
##      0.7733  0.3313
## s.e.  0.0103  0.2388
##
## sigma^2 estimated as 98.26:  log likelihood=-14237.77
## AIC=28481.54  AICc=28481.55  BIC=28500.3
##
## Training set error measures:
##              ME      RMSE      MAE      MPE      MAPE      MASE
## Training set -0.06930023  9.788178  5.726581 -0.1197672  2.640756  0.7190028
##              ACF1
## Training set -0.01352257
```



```
##
##  Ljung-Box test
##
## data:  Residuals from Regression with ARIMA(1,0,0)(0,1,0)[96] errors
## Q* = 1534.6, df = 190, p-value < 2.2e-16
##
## Model df: 2.    Total lags used: 192
```

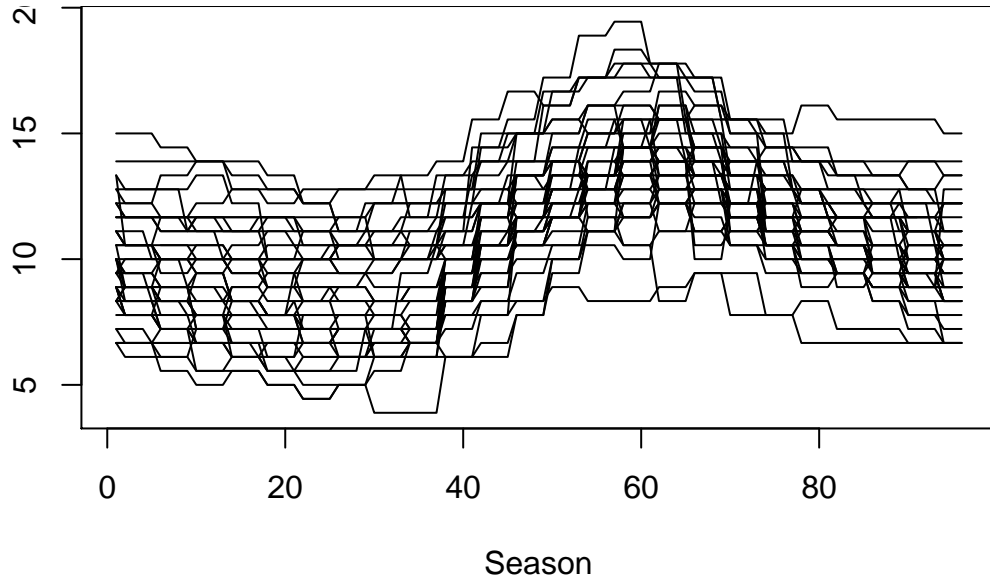
VAR model

Let select the order of the autoregressive part of the model.

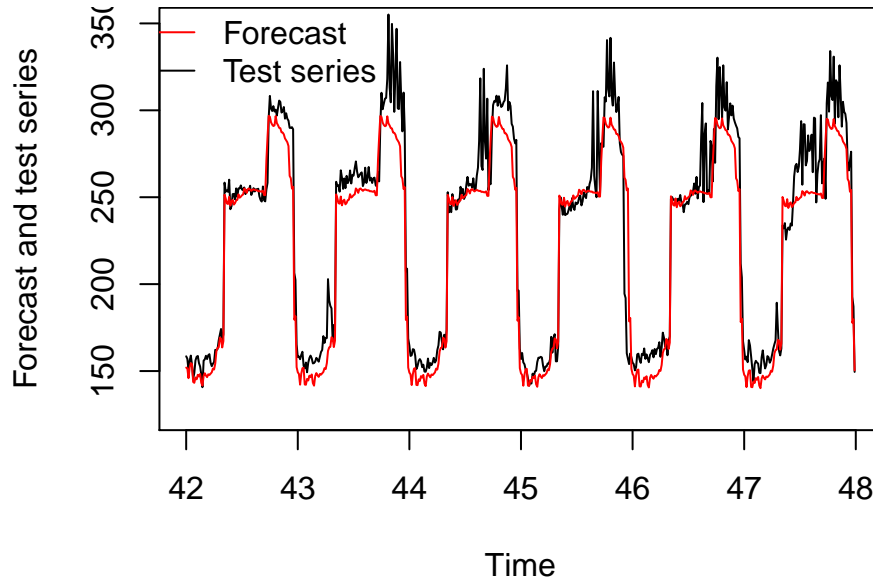
```
## AIC(n)  HQ(n)  SC(n) FPE(n)
##      5      5      5      5
```

Because both series (electricity and temperature) are strongly seasonal (see Figure 2 about electricity consumption and Figure 9 about temperature), we add seasonal dummy in the VAR to control the seasonal effect. From Figure 2, we deduce that the electricity consumption is slightly trended and not centered around 0. As a consequence, we need to add both a trend and constant to the model in addition to the seasonal dummy variable, which account for the seasonal nature of the series.

Figure 9: Seasonal plot for temperature



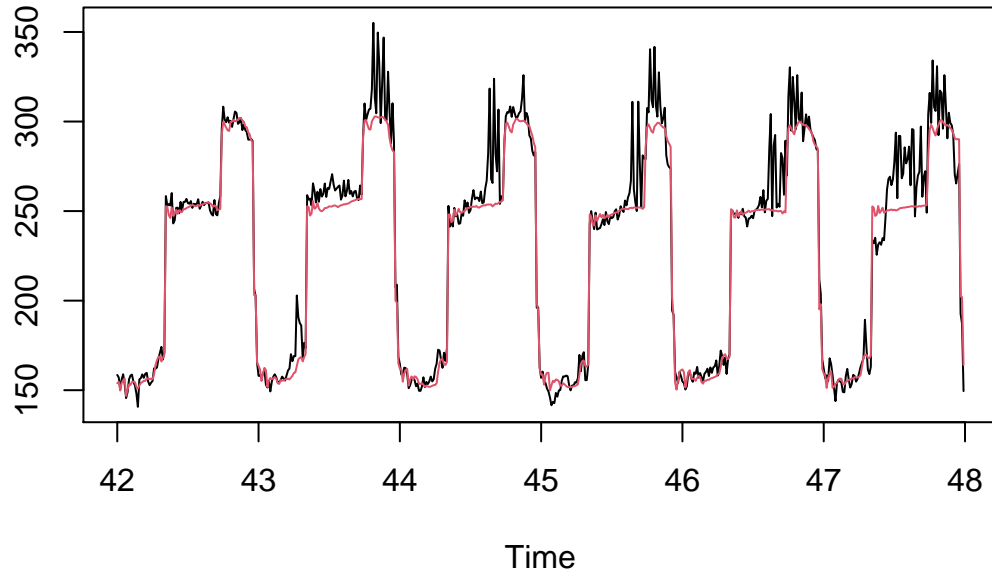
By comparing the plot of forecast and test series, we can conclude that the seasonal NNAR, which is the best model so far, also outperforms the VAR model, which has the same drawback found in the context of exponential smoothing model (systematic underestimation of the series).



The RMSE for VAR model is estimated to be 18.22, which is lower than its counterpart 14.05 obtained using NNAR model (in the univariate case).

Now, we consider an NNAR with temperature embedded in the model.

Figure 10: Forecast and test series using NNAR and temperature



The resulting RMSE is equal to 13.24 is slightly lower than that obtained using VAR model. Therefore, NNAR performs better than all the other models considered, either in the univariate case or in the multivariate one (when we account for temperature).

Forecasting electricity consumption on 2010-02-17.

Since the best among all the models tried is NNAR, let re-estimate the models using all the dataset and then forecast electricity consumption on 2010-02-17 assuming that the temperature is known. The output data is exported to an excel file named *ChristianKamgang.xlsx*.