MÁSTER UNIVERSITARIO EN INGENIERÍA INDUSTRIAL + MÁSTER EN INDUSTRIA CONECTADA

Forecasting Assignment

Authors:

Manuel Trabado de la Cruz

Alfonso Sanz Giner

Ignacio Ortiz de Zúñiga

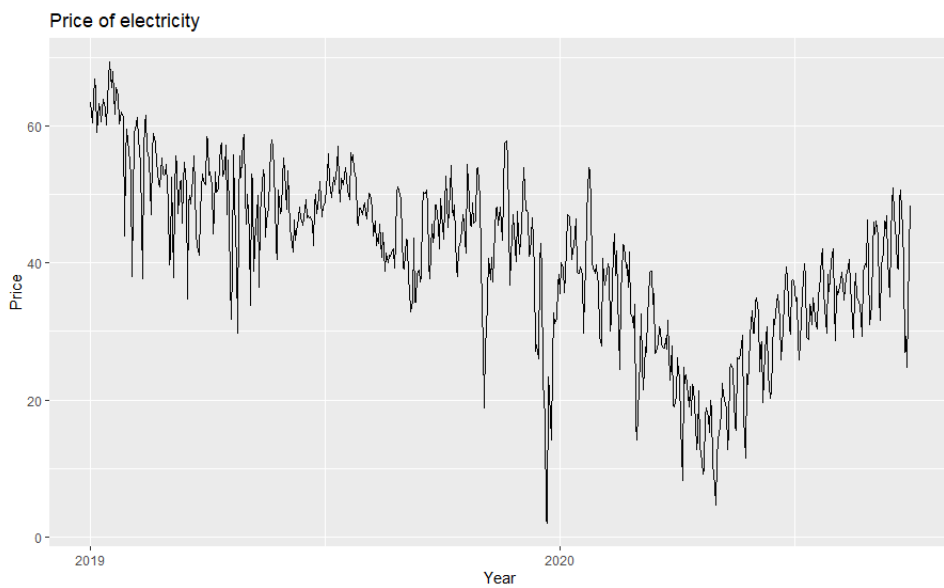
Madrid

22/04/2021

Contents

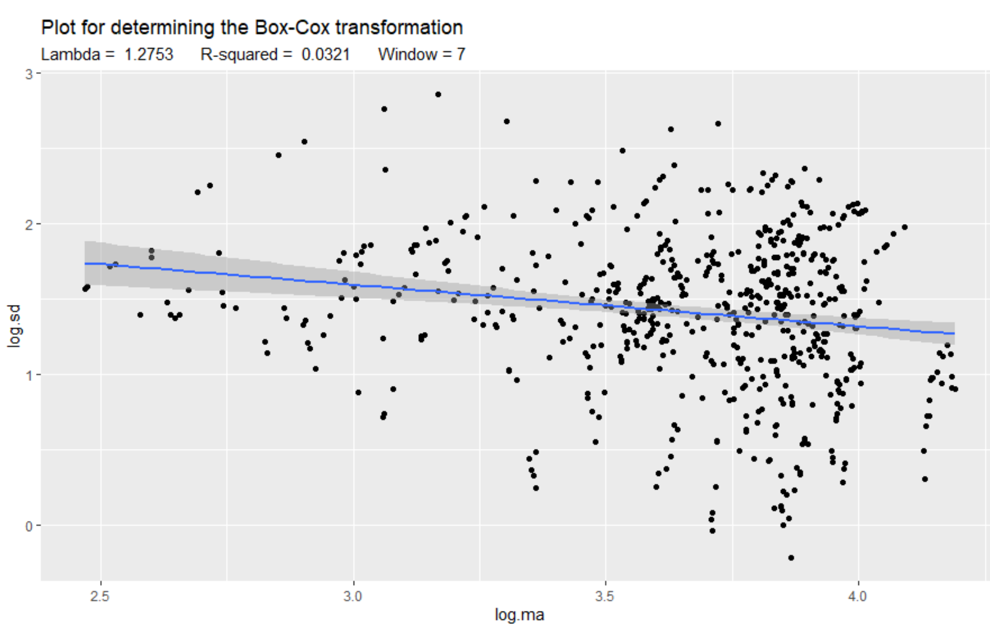
Introduction

This assignment focuses on a forecasting problem of the daily price of electricity calculated from the demand and wind between the years 2019 and the 30th of September of 2020 (included):

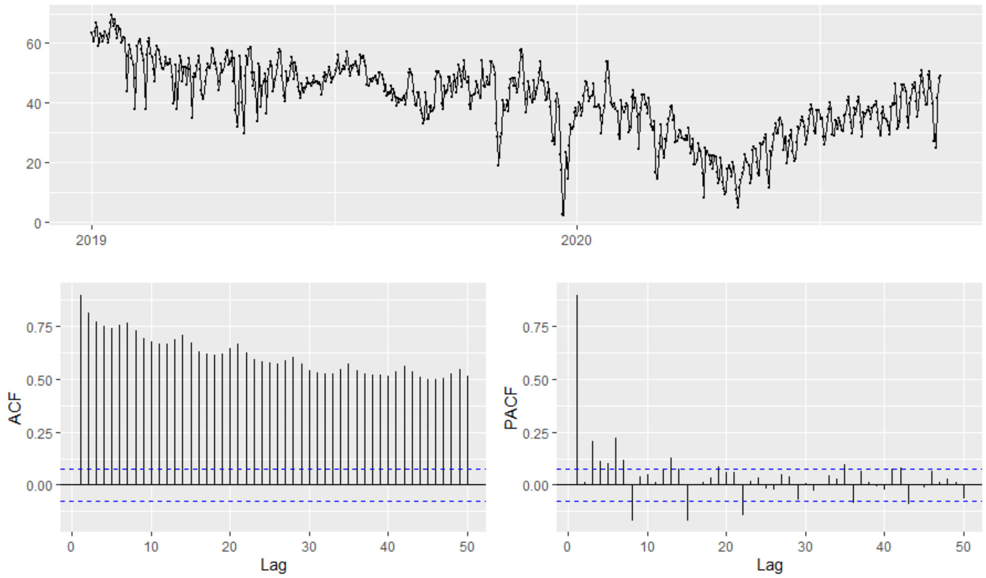


SARIMA model

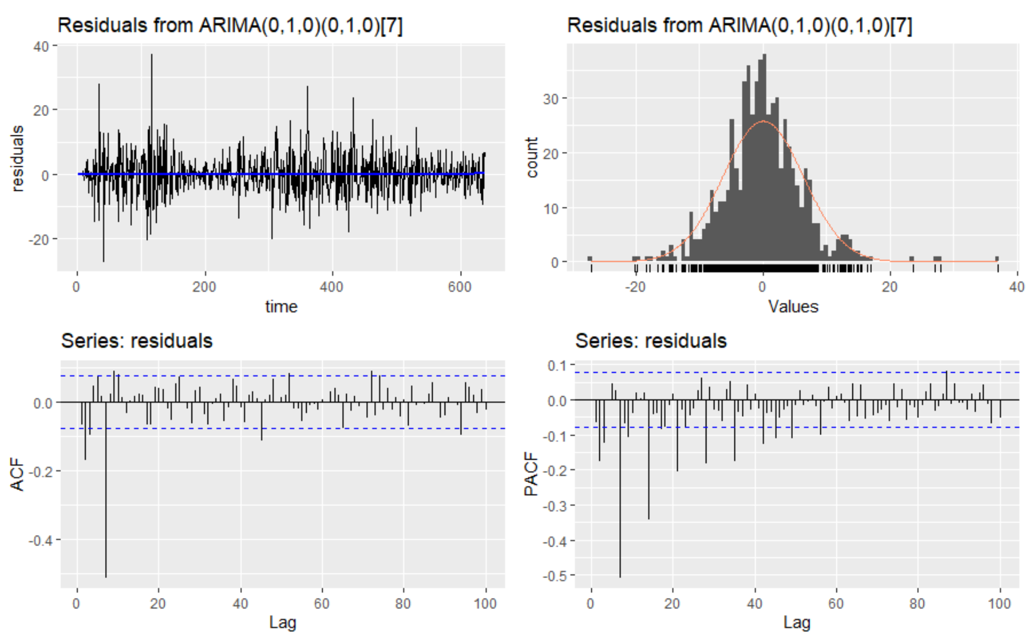
First, we make an analysis the price curve of the dataset without a introducing any model to determine if it is a stationary function. As we can observe the mean is not constant while the variance is. Even though, we decided to make a Box-Cox analysis to determine that the variance is constant. As we can observe there is not noticeable pattern:



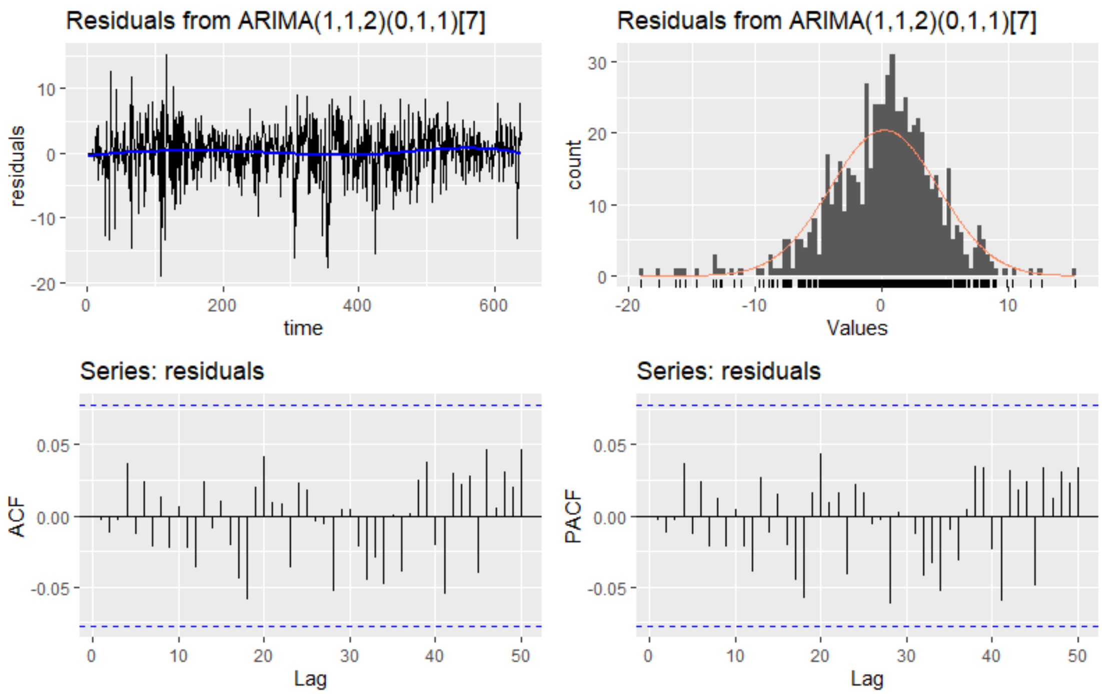
After analysing the variation of the mean and the continuous decrease of the ACF, we have determined the need of using a differentiation and seasonal differentiation:



With an ARIMA model of (0,1,0) both in order and seasonal we have stabilised the mean and there is a very high correlation between y(t) and y(t-7) and an exponential response in the PACF due to the seasonality of the data, this implies the existence of a seasonal MA1.

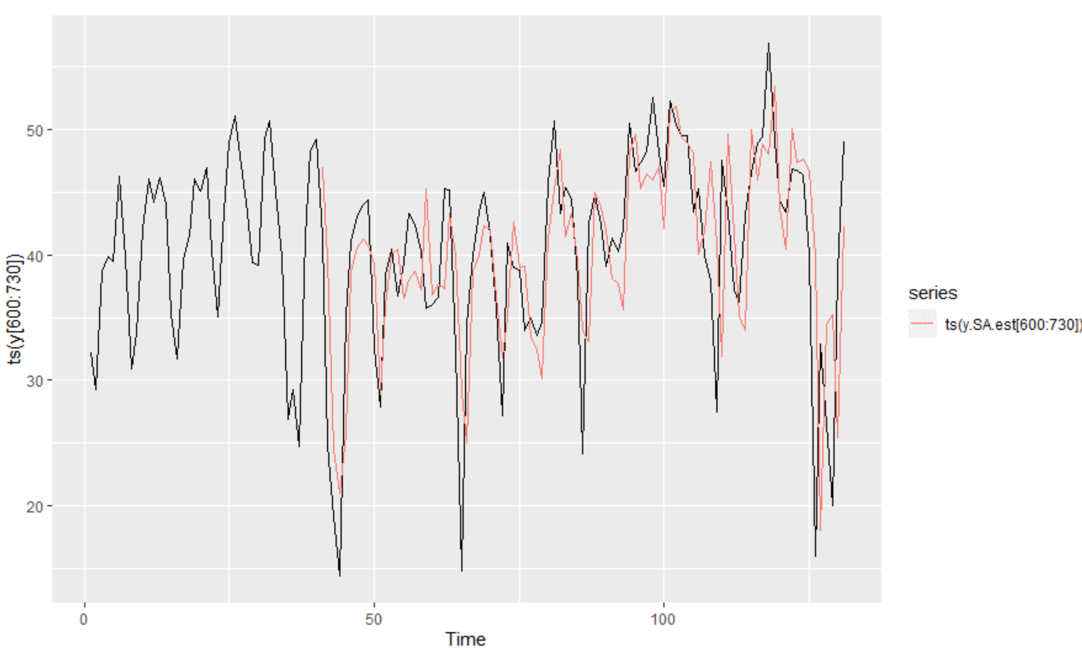


The existence of a smaller correlation in y(t-1) than in y(t-2) indicates there are a regular AR and MA. From now on we start an analysis of the significance of the AR and MA parameters for different orders based on the p-value.



The final model fitted to the data is an ARIMA (1,1,2)(0,1,1)[7]. The residuals can be assumed as white noise because the p-value of the Ljung-Box test is 0.986 (bigger than 0.05).

The forecast based on this model, compared to the real data, is the following:



Transfer function time series

For the next analysis, a transfer function is introduced for the input variables so the model can predict the price considering the changes in demand or wind.

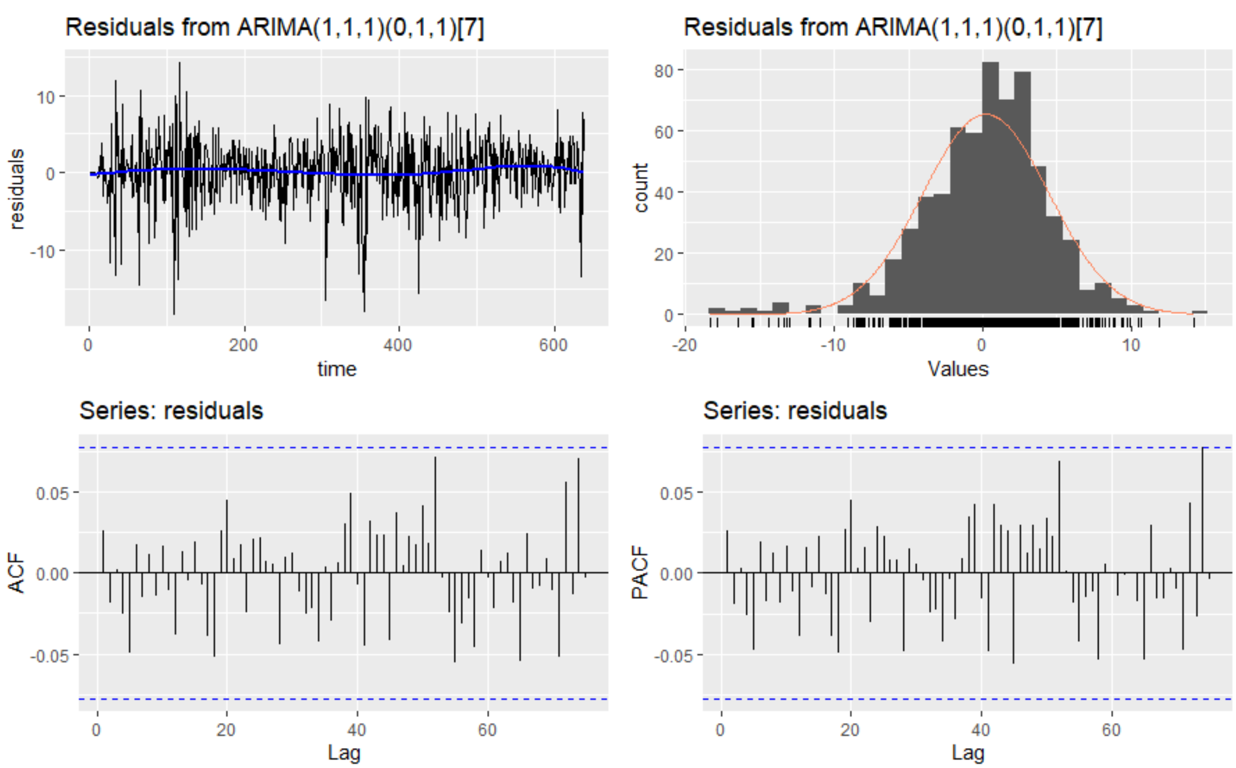
There have been considered two models based on the demand and wind respectively, but the model that takes into account the wind has proven more precise. Therefore, only the wind function will be used in the final model.

Starting from the previous SARIMA model it has been added to this model a transfer function for the wind. And, after some testing, we have concluded that the best transfer for the model is of order (0,1) with no intercept, meaning the transfer function is:

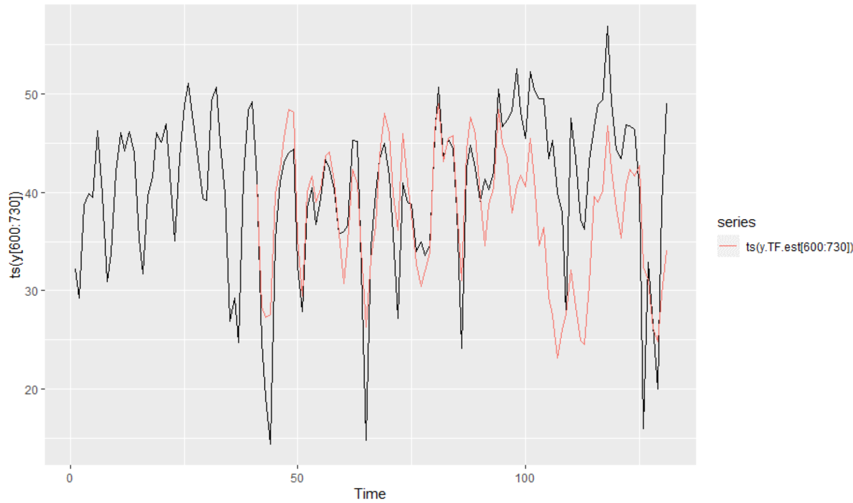
$$y(t) = \ 0.011x(t) + 0.007x(t-1) + v(t)$$

```
Coefficients:
         ar1      ma1     sma1   T1-MA0  T1-MA1
      0.7618  -0.9460  -0.9706   0.0108  0.0071
s.e.  0.0396   0.0193   0.0181   0.0033  0.0032
```

Due to the changes introduced by this transfer function our SARIMA model has been affected and to improve its performance we have reduced the autoregressive order:



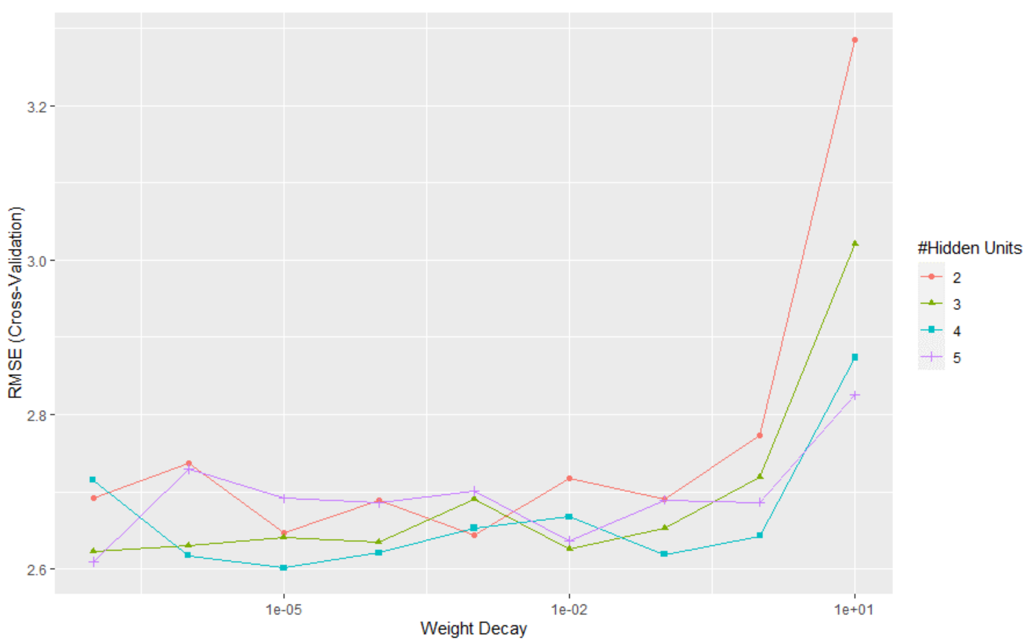
The resulting forecast is shown in the next graph:



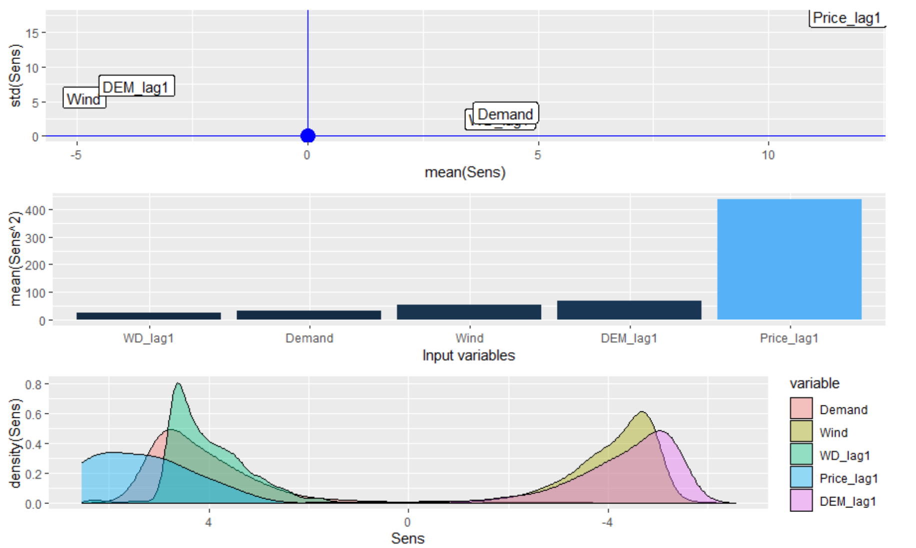
Non-linear regression model

The last model we are going to analyze is an MLP with both inputs, wind and demand.

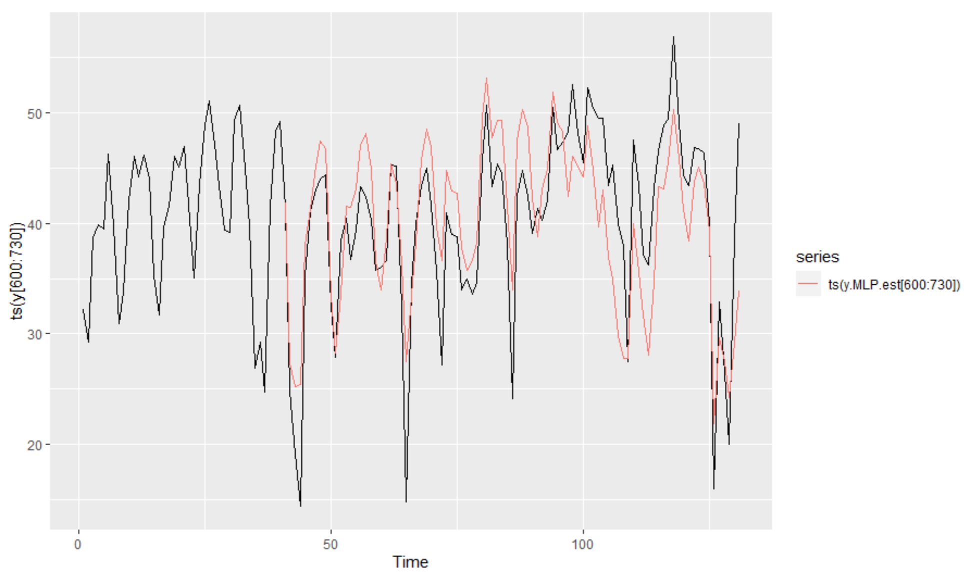
After testing for different combinations of decay and number of neurons, the best model we have found based on the RMSE is the one with size 4, a decay 10^{-5} and using the lag1 of demand, wind and price. With a training RMSE of 2.60 and R-squared of 0.95:



The sensibility analysis proves that the lag of one day of price is the most relevant followed by the lag1 of the demand and the demand.

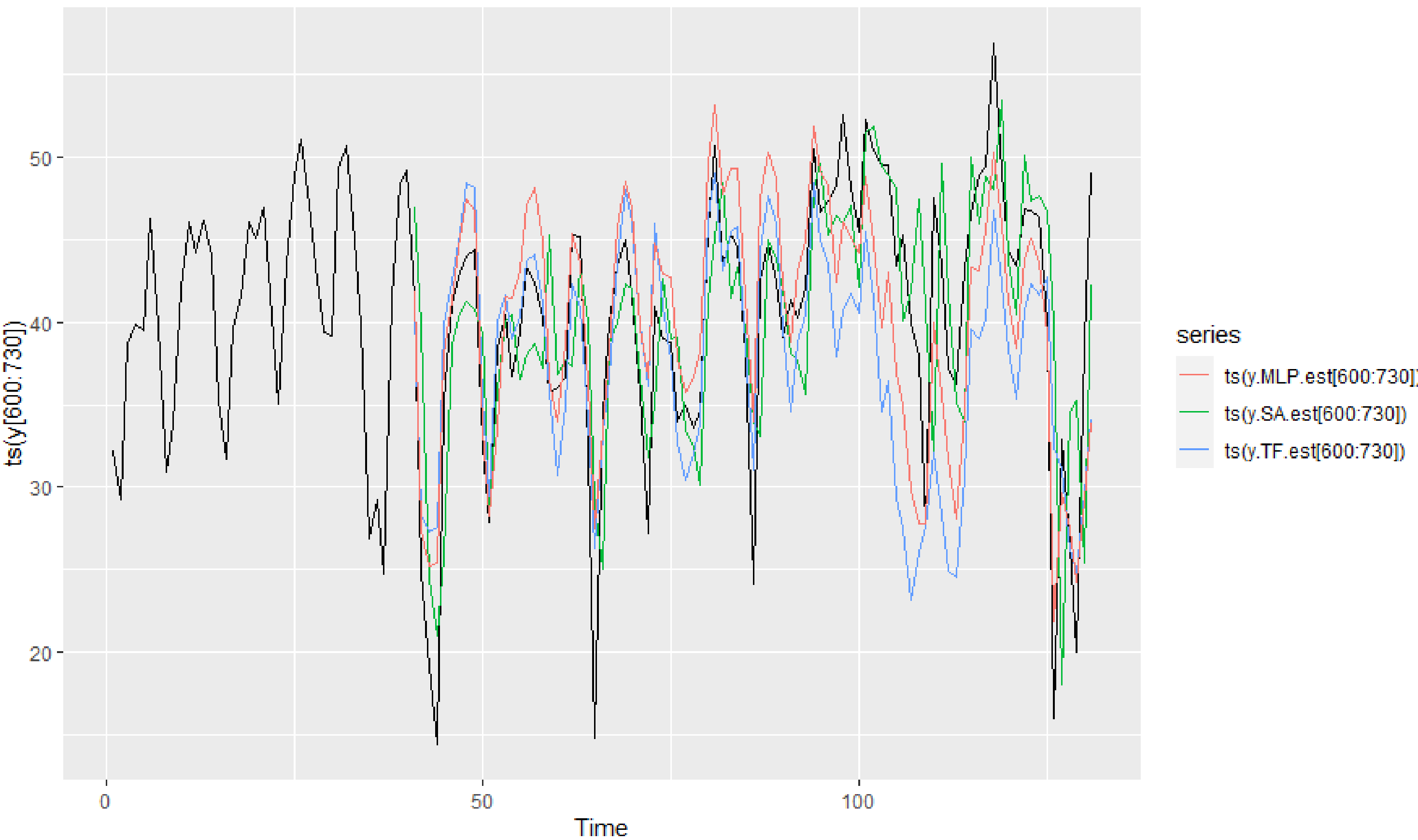


The prediction of the test dataset proves to be accurate as shown in the following graph:



Comparative analysis

Although all models are quite accurate, the SARIMA model with a transfer function seems to fall behind predicting the test data, while the MLP proves to be able to predict to the closest values to the real data.



Finally, we can compare the RMSE and the R-squared of the three models:

Model	**RMSE**
SARIMA	6.46
Transfer function	7.71
MLP	5.54

Normally the transfer function model is more accurate than the SARIMA model thanks to having more information, especially if that information is relevant as it has been proven by the MLP model. But in this case the SARIMA model has a better fit.