RAF Predictive Analytics

Acciona agua www.acciona-agua.com	Alejandro Beivide (alejandro.beivide.garcia@acciona.com) Frank Zamora (frank.zamora.avendano@acciona.com) Clara Sanroma (clara.sanroma.flores@acciona.com) Elena Crespo (elena.crespo.olazabal@acciona.com) Jessica Ruiz (jessica.ruiz.lopez@acciona.com) Cosmin Koch (ckoch@acciona.com) Jose Ivorra (jivorra.bec@acciona.com)
Artelnics www.artelnics.com	Roberto Lopez (robertolopez@artelnics.com) Fernando Gomez (fernandogomez@artelnics.com) Emilio Lucas (emiliolucas@artelnics.com)

Index

1. Summary

From 02/01/2017 to 26/10/2017	299 days	3 variables

Plant description

We have two entry points in the plant.

The coordinates of the Point 1 are (25.2124585, 51.640231).

The coordinates of Point 2 are (25.210405, 51.637468). T

The distance between both points is 0.36km.

2. Temperature model

Temperature model 1-day

Variable selection temperature 1-day

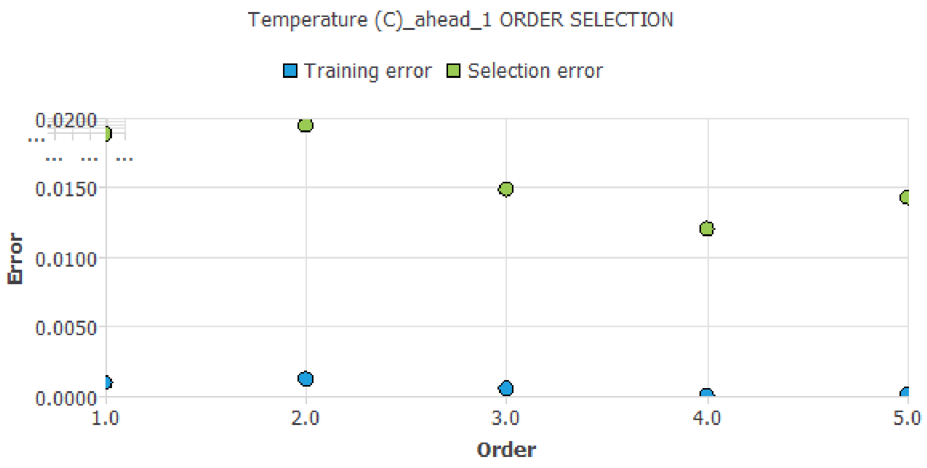
Variable	Region
Temperature	Plant
Sea floor potential temperature	Plant
Surface temperature	Southern Shallows
Sea floor potential temperature	Southern Shallows

The most influent variables for the temperature ahead 1 are the temperature lags at the plant.

Then, the most influential region is the Southern Shallows.

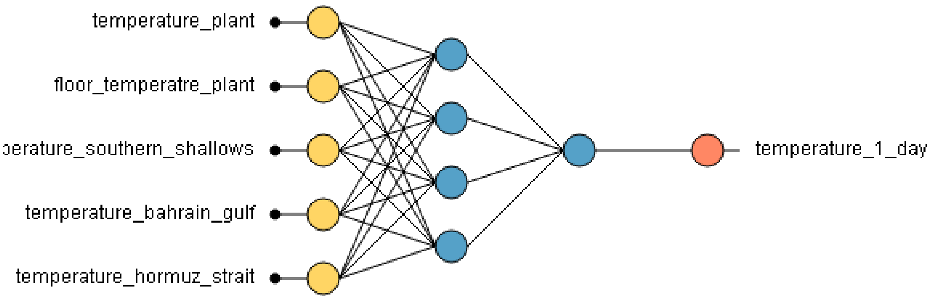
Order selection temperature 1-day

The optimal network architecture is that for which the selection error is minimum.



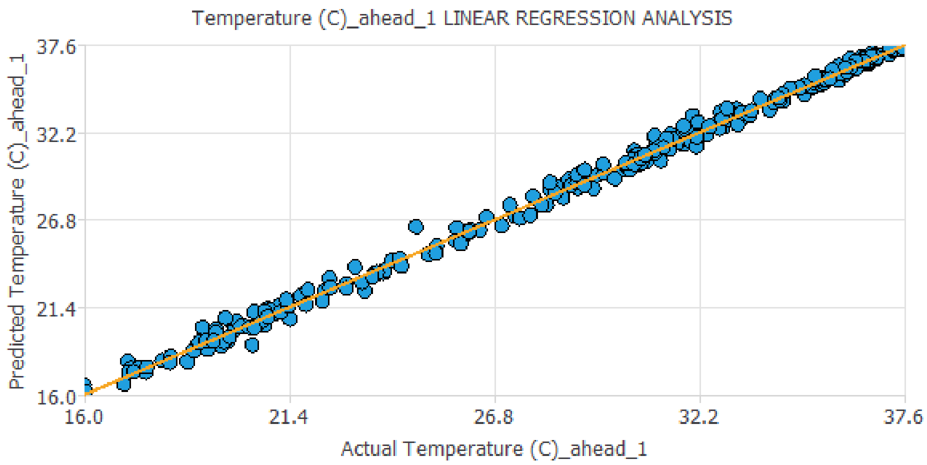
Training 0.003 Selection 0.012

Neural network architecture temperature 1-day



The predictive model takes the temperature of different regions as inputs.

Linear regression analysis temperature 1-day



The 1-day temperature forecast is very accurate.

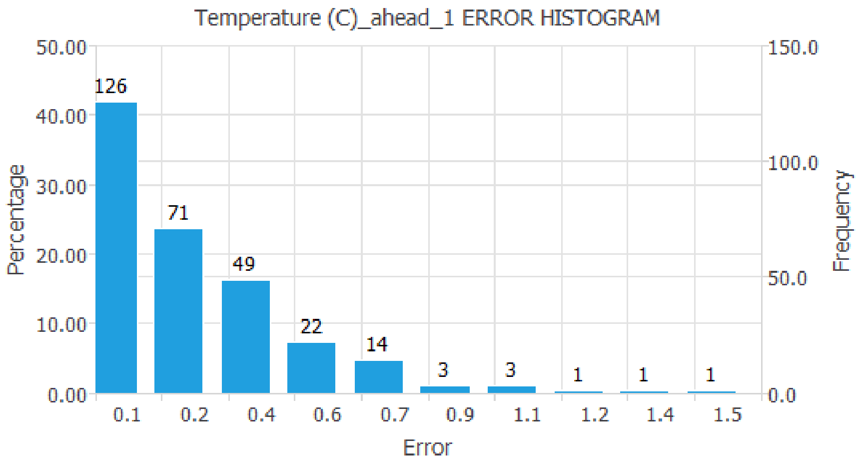
The correlation coefficient between the predicted and the real temperatures is 0.99.

Error analysis temperature 1-day

	Absolute error	Relative error
Minimum	0.00	0.00%
Maximum	1.41	5.55%
Mean	0.27	1.27%
Standard deviation	0.24	1.14%

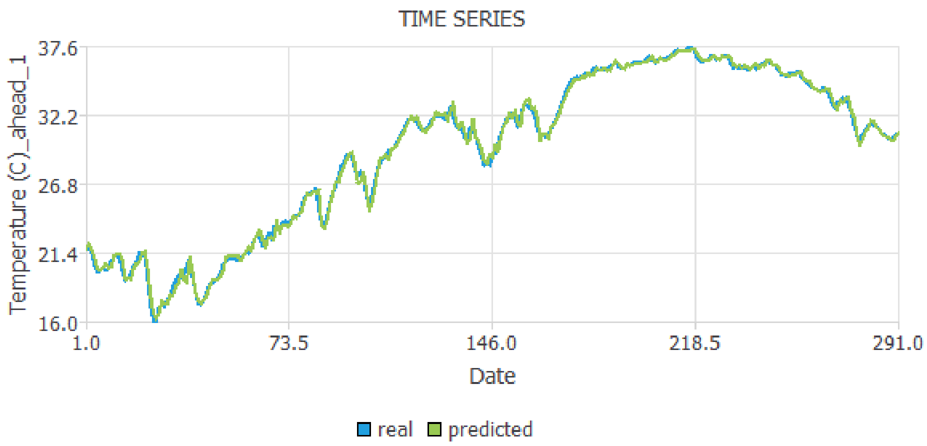
The errors of the 1-day prediction for the temperature are very low.

The mean error is 0.27 C and the maximum error is 1.41 C.



Most prediction error are around 0.1 C. Only one error is aroun 1.5 C.

Prediction time series temperature 1-day



The model has nearly a perfect prediction for the temperature for the next day.

Temperature model 2-days

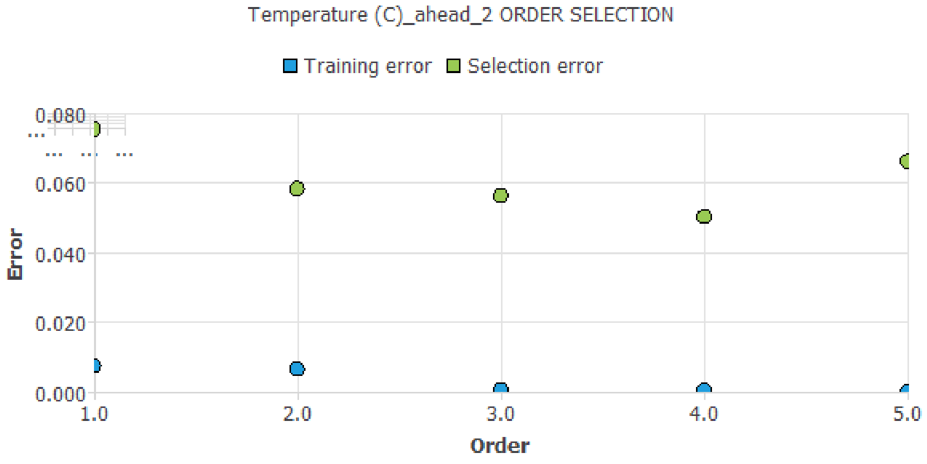
Variable selection temperature 2-days

Variable	Region
Temperature	Plant
Surface temperature	Southern Shallows
Sea floor potential temperatura	Southern Shallows
Surface temperature	Bahrain Gulf
Surface temperature	Hormuz strait

For 2-day forecasting, the most influential variables are the past temperatures at the plant, and the Southern Shallows.

Order selection temperature 2-days

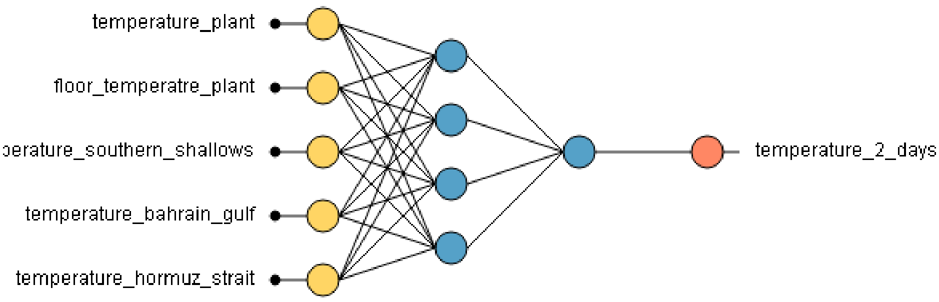
The optimal network architecture is that for which the selection error is minimum.



The smallest selection error is 0.050, and the corresponding training error is 0.008.

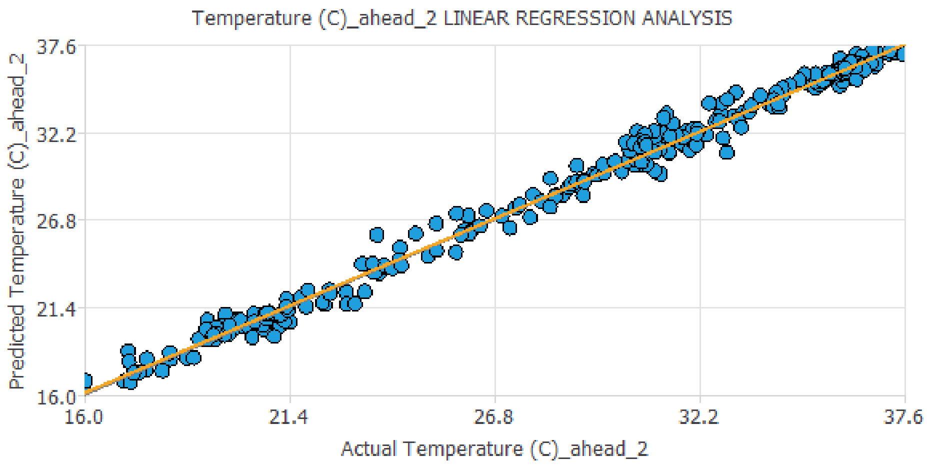
The smallest selection error is produced for a neural network with 4 hidden neurons.

Neural network architecture temperature 2-days



The 2-day neural network for the temperature is very similar to the 1-day one.

Linear regression analysis temperature 2-days

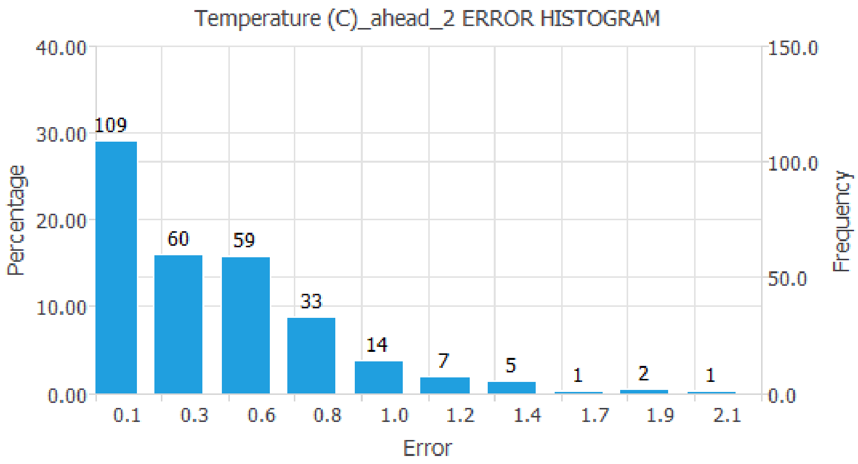


The 2-days predictions of the water temperature are also very accurate.

The correlation coefficient here is 0.99.

Error analysis temperature 2-days

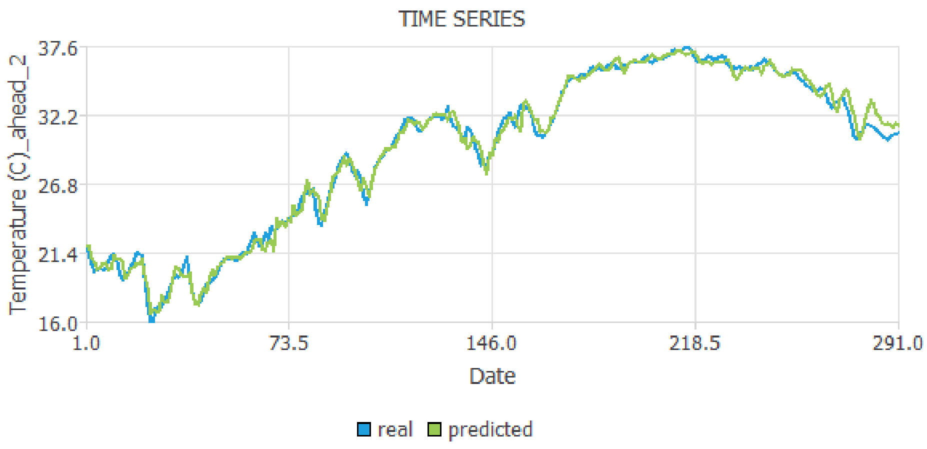
	Absolute error	Relative error
Minimum	0.00	0.00%
Maximum	2.14	9.92%
Mean	0.44	2.05%
Standard deviation	0.42	1.96%



The mean error for 2-day forecasting is 0.44 C, and the maximum error is 2.14 C.

Most errors are around 0.1 C.

Prediction time series temperature 2-days



The 2-day forecasting of the water temperature is still very accurate.

Temperature model 3-days

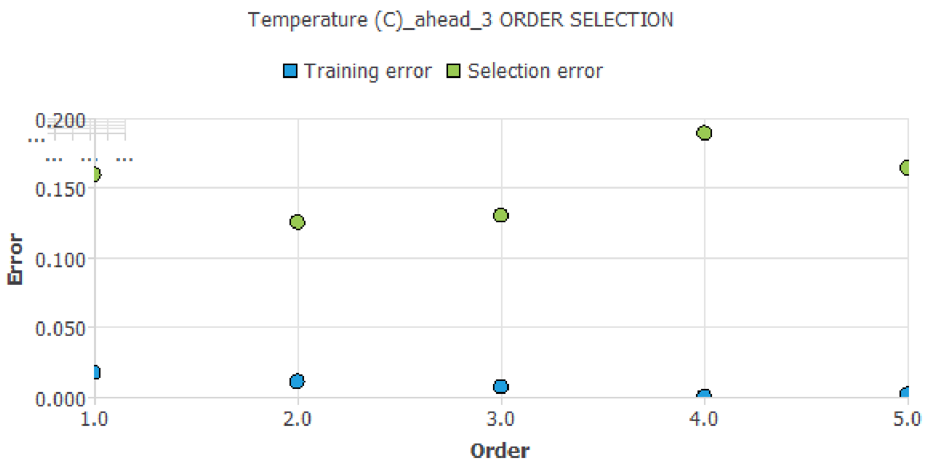
Variable selection temperature 3-days

Variable	Region
Temperature	Plant
Surface temperatura	Southern Shallows
Sea floor potential temperature	Southern Shallows
Surface temperature	Bahrain Gulf
Surface temperature	Bahrain Gulf
Surface temperatura	Hormuz Strait

Southern Shallows, Bahrain Gulf and Hormuz Strait temperatures are important for the 3-day forecasting.

Order selection temperature 3-days

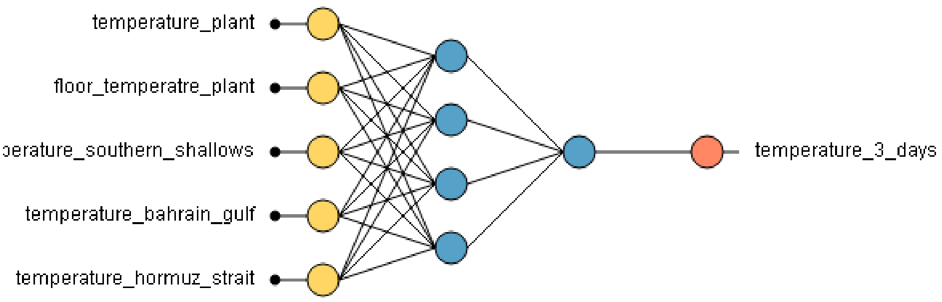
The optimal network architecture is that for which the selection error is minimum.



The optimal order here is 3.

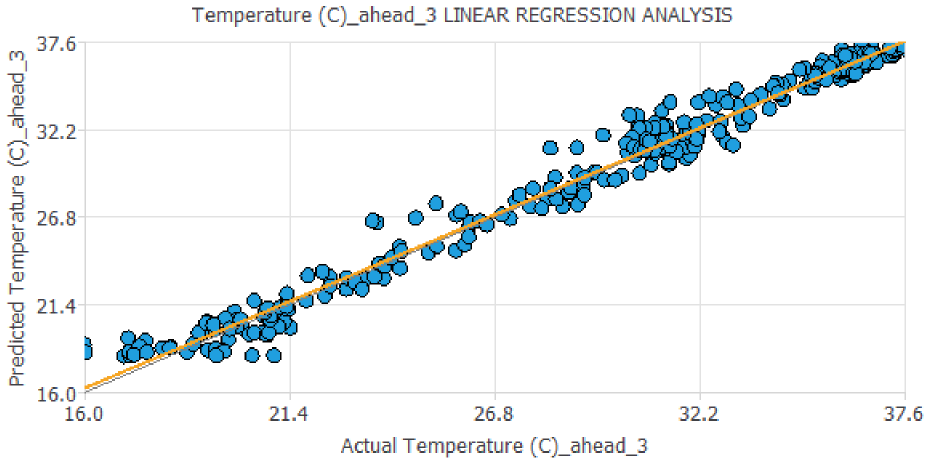
For this neural network complexity, the selection error is 0.02 and the corresponding training error is 0.12.

Neural network architecture temperature 3-days



The 3-day temperature is predicted as a function of the temperature at the different regions and for different lags.

Linear regression analysis temperature 3-days

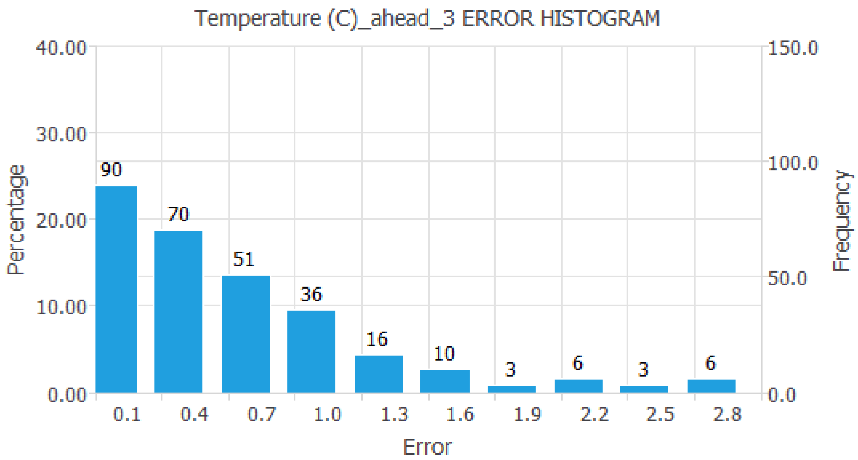


The 3-day forecasting values are very similar to the actual values.

The correlation coefficient here is 0.98.

Error analysis temperature 3-days

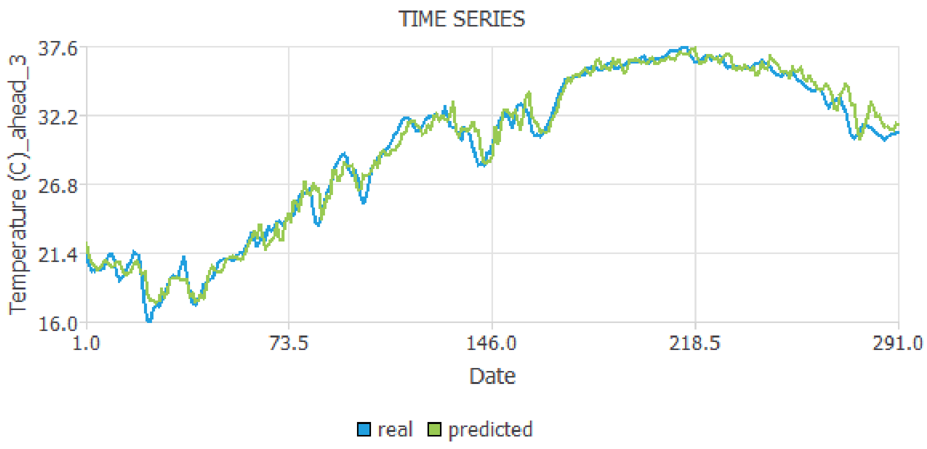
	Absolute error	Relative error
Minimum	0.00	0.00%
Maximum	2.92	13.56%
Mean	0.66	3.09%
Standard deviation	0.59	2.76%



The errors of the model are higher than the other days, but the average error is still very low.

The mean error is 0.66 C and the maximum error is 2.92 C for 3-day forecasting.

Prediction time series temperature 3-days



The model still predicts the trend of the temperature very well.

Here we can see bigger deviations between the real and the prediction signals.

3. Conductivity model

Conductivity model 1-day

Variable selection conductivity 1-day

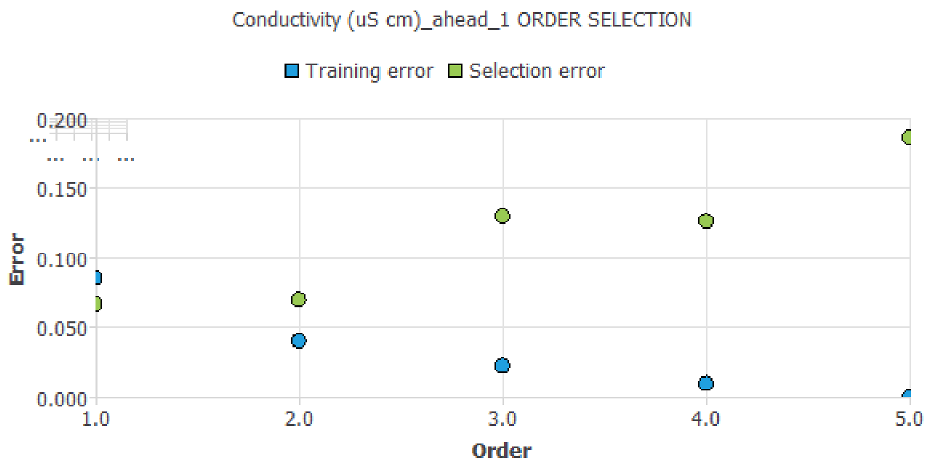
Variable	Region
Conductivity	Plant
Normalized fluorescence line height	Iran Coast
Mass concentration of chlorophyll	Southern Shallows
Particle organic carbon	Southern Shallows
Salinity	Southern Shallows
Salinity	Iran Coast

The most correlated variables for the conductivity are the previous values of this variable.

Some satellite measurements at the Iran Coast and the Southern Shallows have also an impact.

Order selection conductivity 1-day

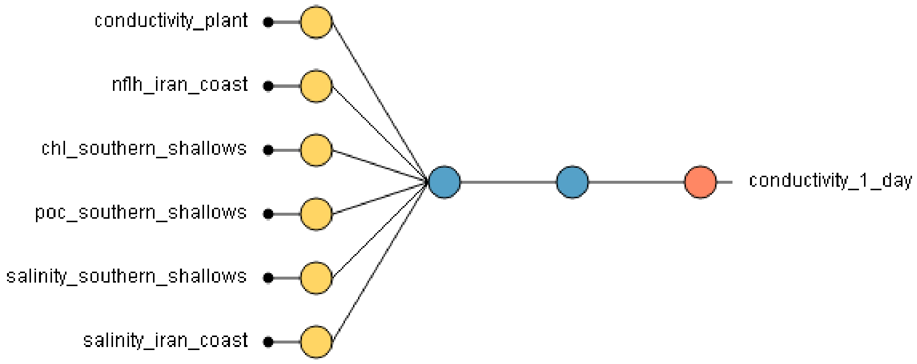
The optimal network architecture is that for which the selection error is minimum.



The neural network that best predicts the 1-day conductivity has low complexity (1 hidden neuron).

The selection error for that neural network is 0.07, and the corresponding training error is 0.11.

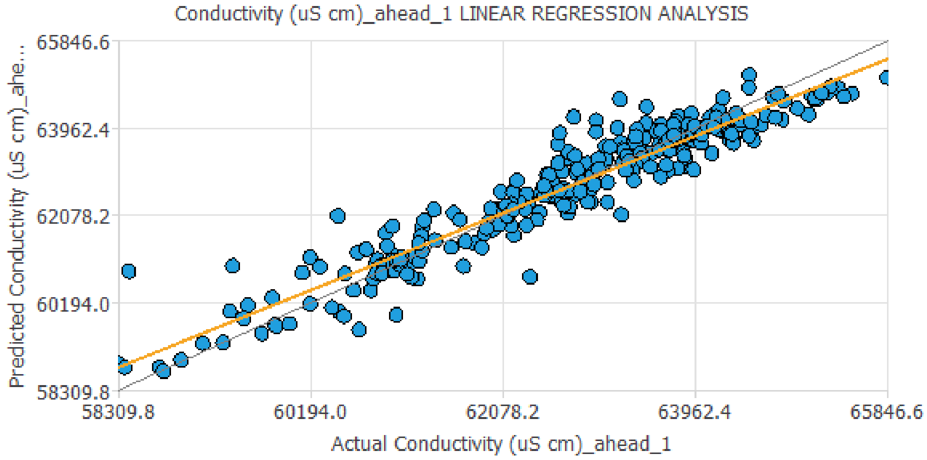
Neural network architecture conductivity 1-day



The neural network takes both endogenous and exogenous variables to predict the conductivity.

The real number of inputs is much greater, since this figure does not represent the 5 lag values for each variable.

Linear regression analysis conductivity 1-day

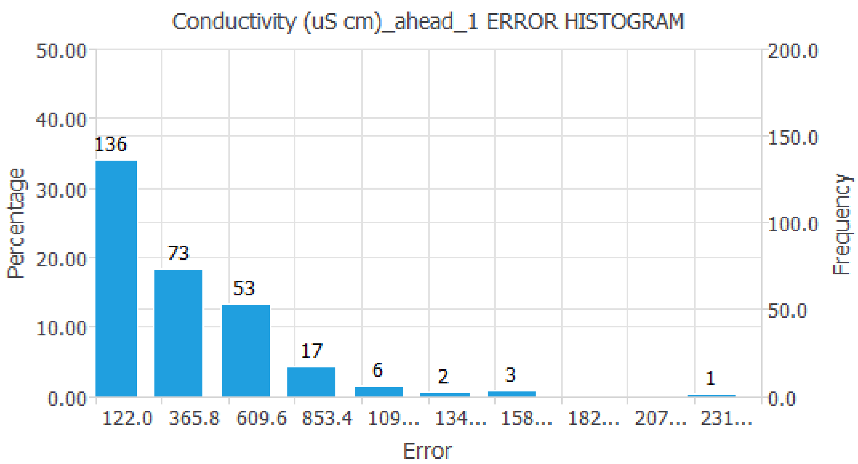


The conductivity predictions for 1-day are very similar to their actual values.

The correlation between predictions and measurements here is 0.91.

Error analysis conductivity 1-day

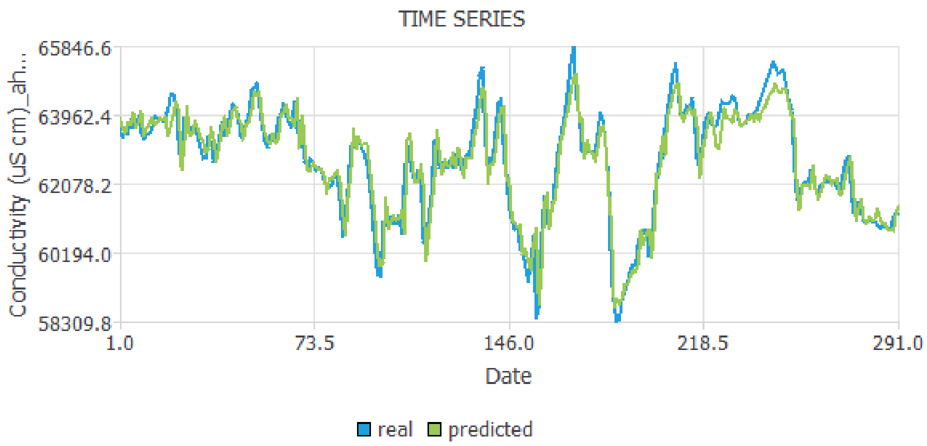
	Absolute error	Relative error
Minimum	0	0.00%
Maximum	2437	32.14%
Mean	350	4.65%
Standard deviation	323	4.29%



The mean error in the conductivity forecast is about 350 µS/cm.

Most of the times, the prediction errors are smaller.

Prediction time series conductivity 1-day



The prediction signal fits the actual values for the conductivity.

Conductivity model 2-days

Variable selection conductivity 2-days

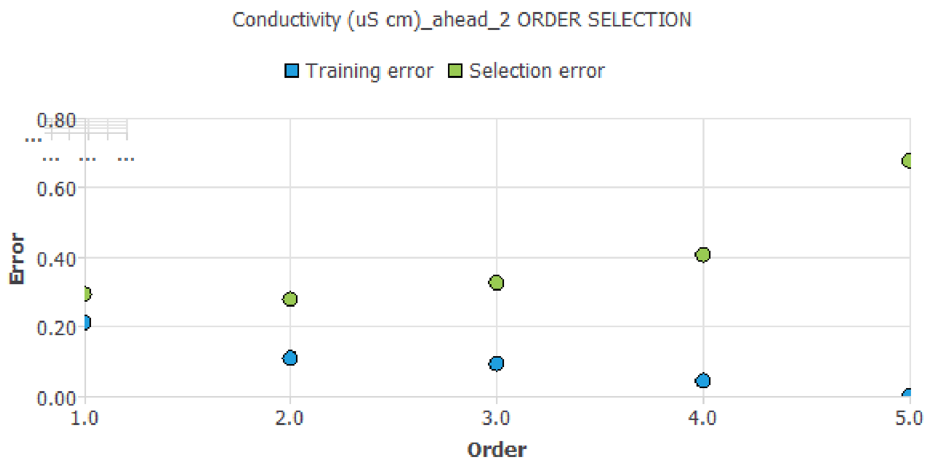
Variable	Region
Conductivity	Plant
Particle organic carbon	Southern Shallows
Normalized fluorescence line height	Iran Coast
Mass concentration of chlorophyll	Southern Shallows
Salinity	Southern Shallows
Salinity	Iran Coast

For 2-day forecasting of the conductivity, the past conductivity, POC, NFLH, CLH and salinity are the most influential variables.

The most important regions are the plant, the Southern Shallows and the Iran Coast.

Order selection conductivity 2-days

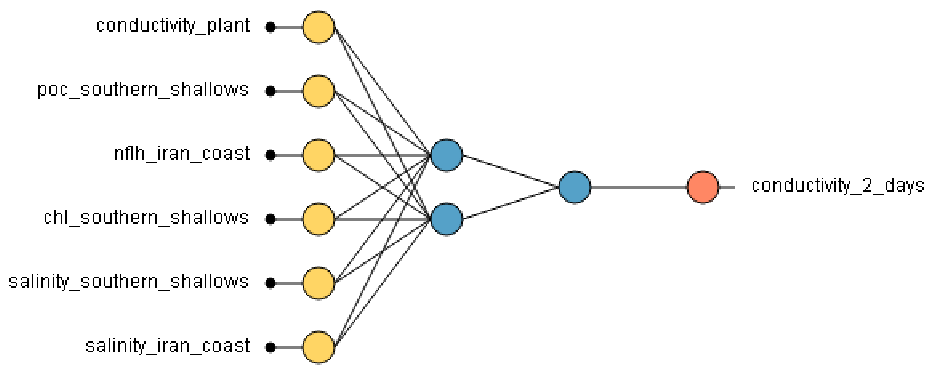
The optimal network architecture is that for which the selection error is minimum.



For this model, the optimal order is 2.

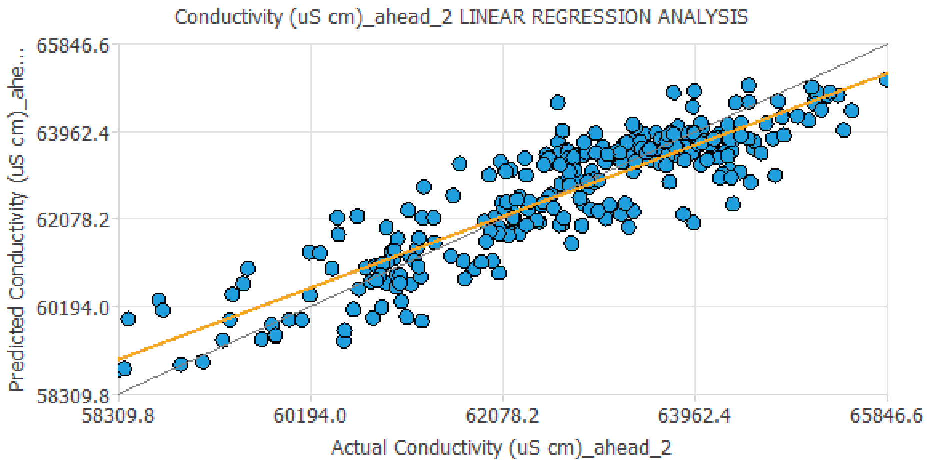
The selection error is 0.27, and the corresponding training error is 0.15.

Neural network architecture conductivity 2-days



We have created a different neural network for each variable and each step ahead.

Linear regression analysis conductivity 2-days

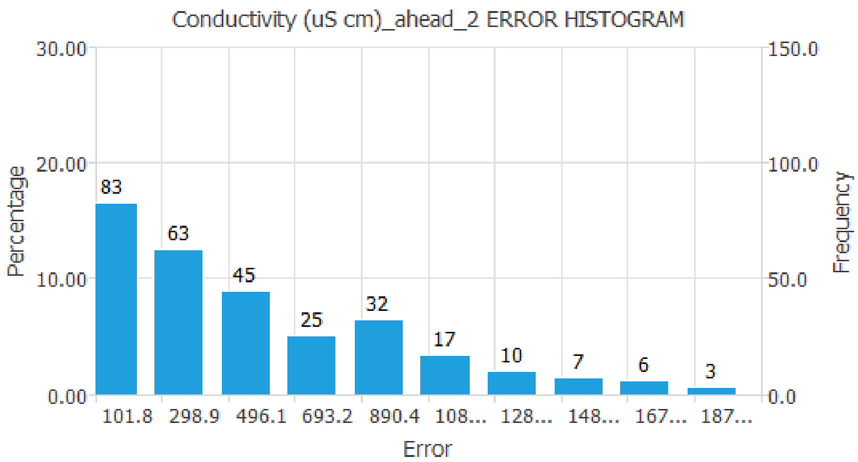


The 2-days predictions are more scattered than the 1-day ones, but the correlation is still 0.89.

The more steps ahead, the less accuracy in the predictions.

Error analysis conductivity 2-days

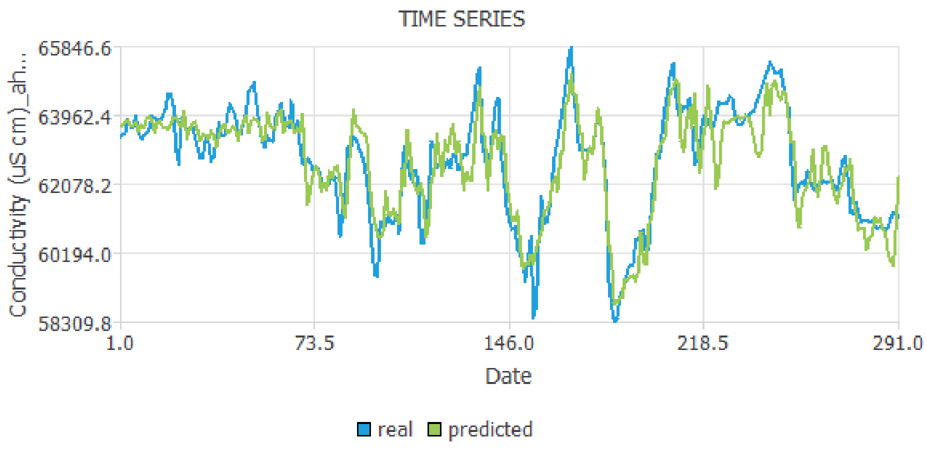
	Absolute error	Relative error
Minimum	3	0.04%
Maximum	1974	26.20%
Mean	523	6.94%
Standard deviation	433	5.75%



The mean error has passed from 4.65% (1-day) to 6.94 (2-days).

The most frequent errors are still the low ones.

Prediction time series conductivity 2-days



For 2 days ahead, the model is still predicting well the conductivity, but the error here is bigger.

Conductivity model 3-days

Variable selection conductivity 3-days

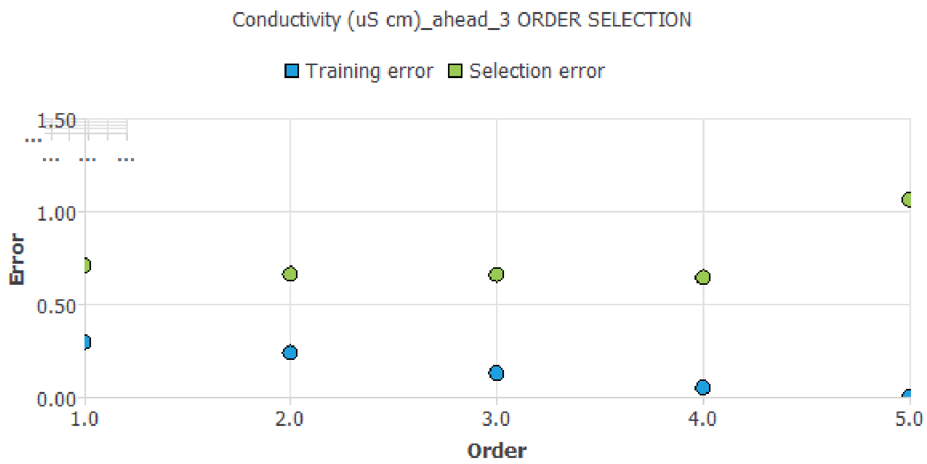
Variable	Region
Conductivity	Plant
Particle organic carbon	Southern Shallows
Normalized fluorescence line height	Iran Coast
Salinity	Southern Shallows
Mass concentration of chlorophyll	Southern Shallows
Salinity	Iran Coast

The most influential variables for the conductivity in 3 days are the conductivity lags, the particle organic carbon, the normalized fluorescence line height, the salinity and the mass concentration of chlorophyll.

The most influential regions are the plant itself, the Southern Shallows and the Iran Coast.

Order selection conductivity 3-days

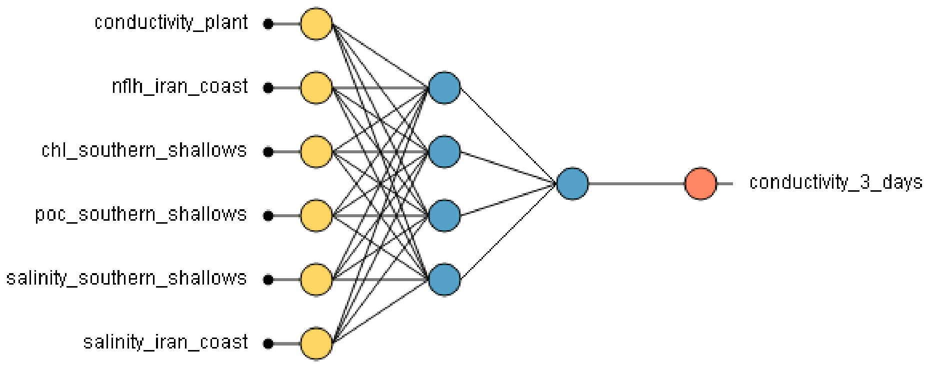
The optimal network architecture is that for which the selection error is minimum.



The neural network that betters predict the conductivity in 3 days has 4 hidden neurons.

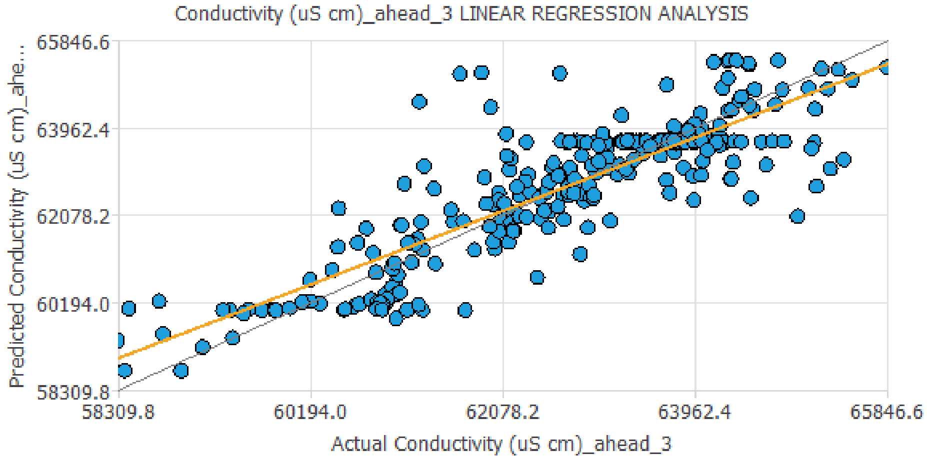
For that architecture the training error is 0.13, and the corresponding selection error is 0.64.

Neural network architecture conductivity 3-days



This neural network has 30 inputs (6 variables x 5 lags).

Linear regression analysis conductivity 3-days



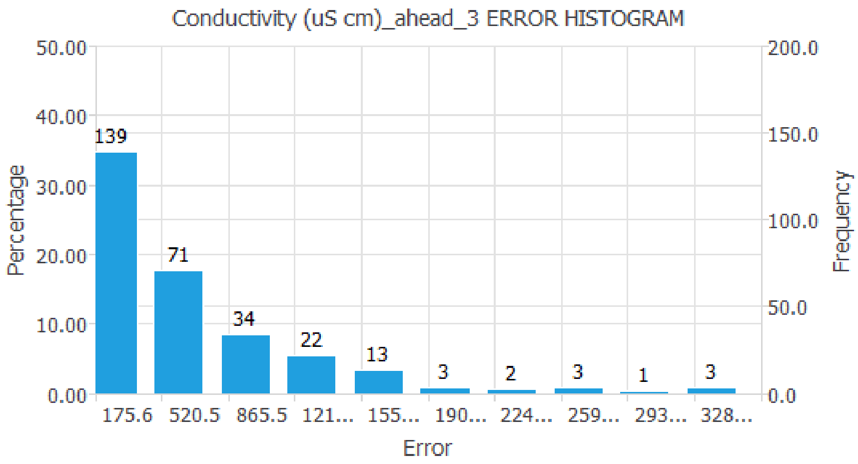
The 3-day predictions are still similar to the actual values.

However, there are some points with big deviations.

The correlation between the real and the forecasted values is 0.85.

Error analysis conductivity 3-days

	Absolute error	Relative error
Minimum	3	0.04%
Maximum	3320	33.55%
Mean	588	7.80%
Standard deviation	553	7.34%

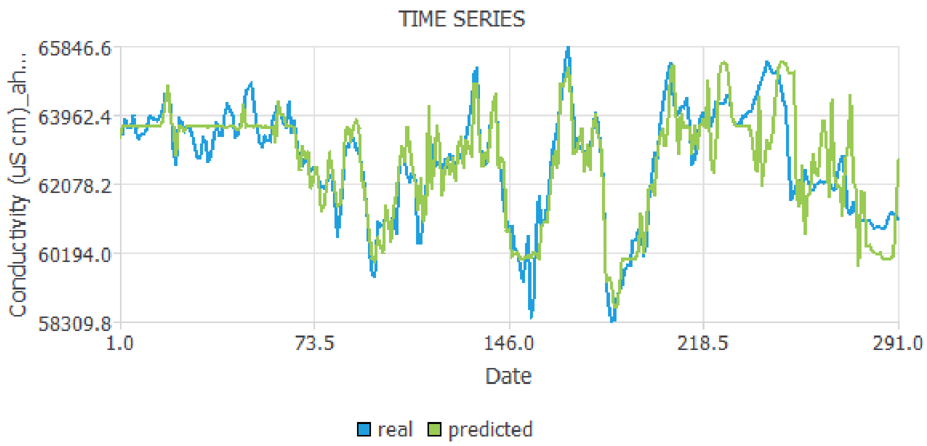


The mean error is 7.80%, and the maximum error is 33.55%.

Most of the predictions have a low error.

However, there are greater errors than for 1 and 2 days predictions.

Prediction time series conductivity 3-days



The 3-day prediction signal matches the actual values, but the error is bigger than for 1 and 2 days.

4. Turbidity model

Turbidity model 1-day

Variable selection turbidity 1-day

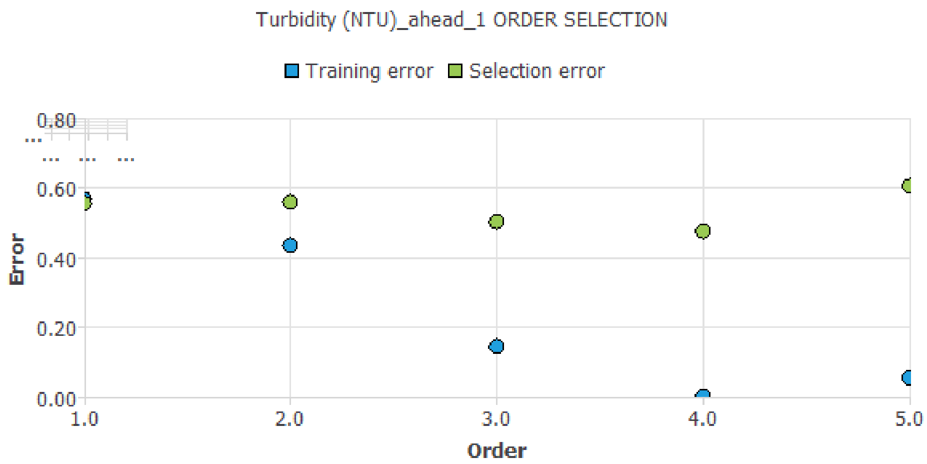
Variable	Region
Turbidity	Plant
Surface temperature	Plant
Particle organic carbon	Southern Shallows
Salinity	Southern Shallows
Photosynthetically available radiation	Bahrain Gulf
Salinity	Bahrain Gulf
Northward velocity	Hormuz Strait

The turbidity of the previous days has a high correlation with the turbidity.

The most influential regions for 1-day forecasting are the Southern Shallows and the Bahrain Gulf.

Order selection turbidity 1-day

The optimal network architecture is that for which the selection error is minimum.

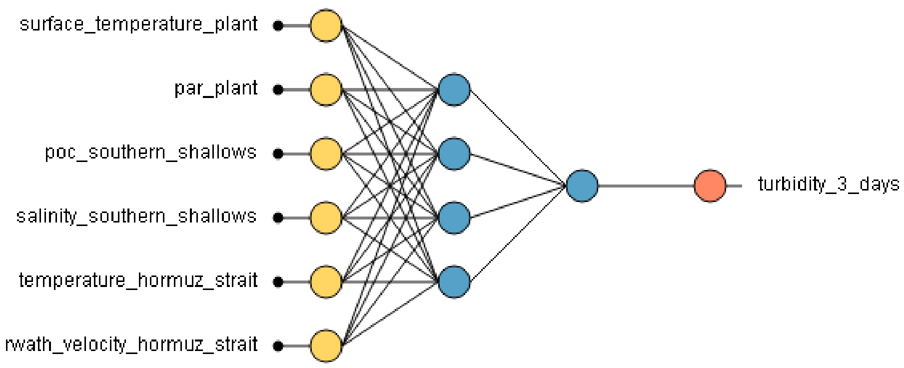


An order selection algorithm has been applied to find the architecture with best generalization properties.

The optimal number of hidden neurons found by that algorithm is 4.

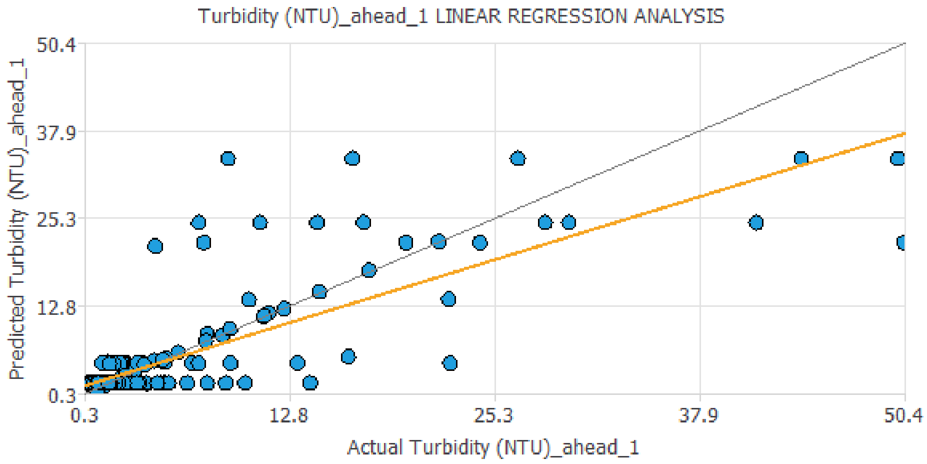
The selection error is 0.47, and the corresponding training error is 0.08.

Neural network architecture turbidity 1-day



The above figure represents the neural network that has been used to forecast the turbidity.

Linear regression analysis turbidity 1-day

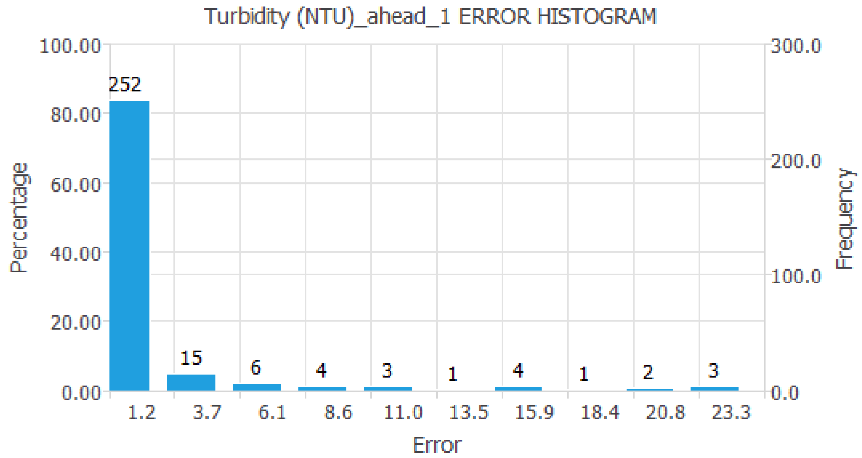


Since the turbidity has big peak values, this is the most difficult variable to predict.

The correlation coefficient between the predicted and the actual values here is 0.83.

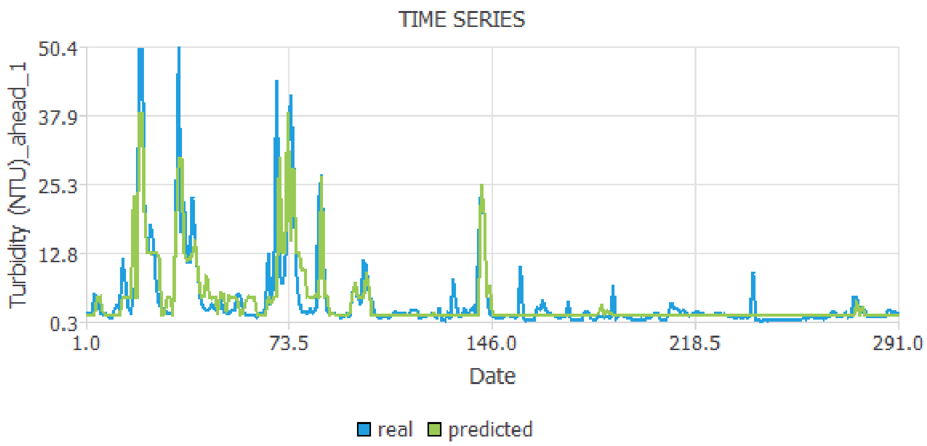
Error analysis turbidity 1-day

	Absolute error	Relative error
Minimum	0.00	0.00%
Maximum	24.51	48.12%
Mean	1.8	3.69%
Standard deviation	3.88	7.76%



The mean error for the turbidity forecasting is 3.69%.

The maximum error (48.12%) must be caused because of the peak values.



Prediction time series turbidity 1-day

The peaks are predicted but they usually have lower values than the real ones.

Most of the high turbidity scenarios are caught, but there are some ones that the neural network does not predict.

Turbidity model 2-days

Variable selection turbidity 2-days

Variable	Region
Surface temperatura	Plant
Sea floor potential temperatura	Plant
Surface temperature	Hormuz Strait
Surface temperatura	Southern Shallows
Photosynthetically avilable radiation	Bahrain Gulf
Turbidity	Plant
Salinity	Bahrain Gulf
Northward velocity	Hormuz Strait

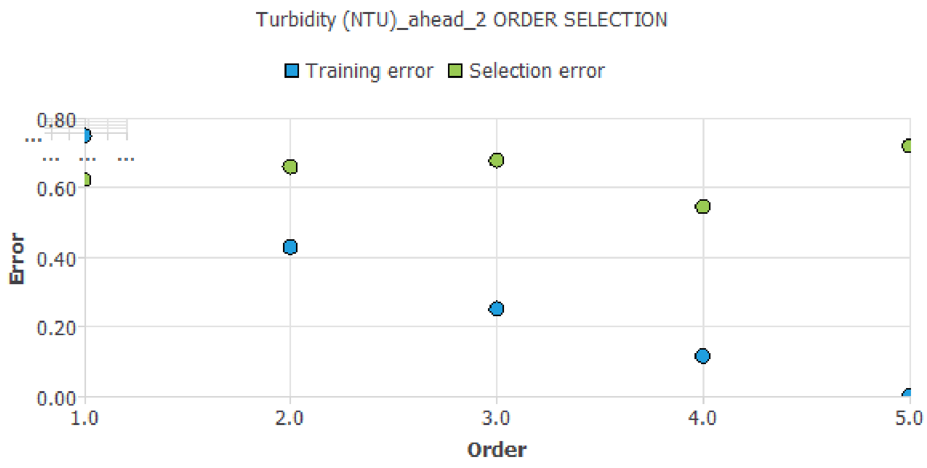
The turbidity with 2 days ahead are not well correlated with the internal data.

The region of the strait of Hormuz in the previous week has high influence in the turbidity.

There are many variables which influence the turbidity.

Order selection turbidity 2-days

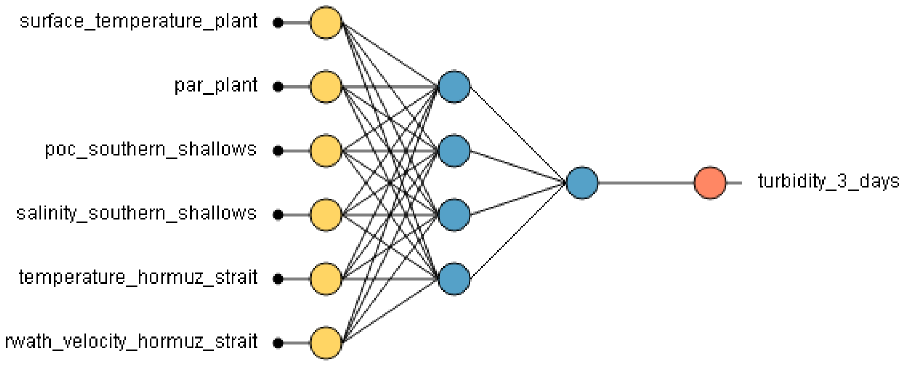
The optimal network architecture is that for which the selection error is minimum.



The optimal neural network has 4 neurons in the hidden layer.

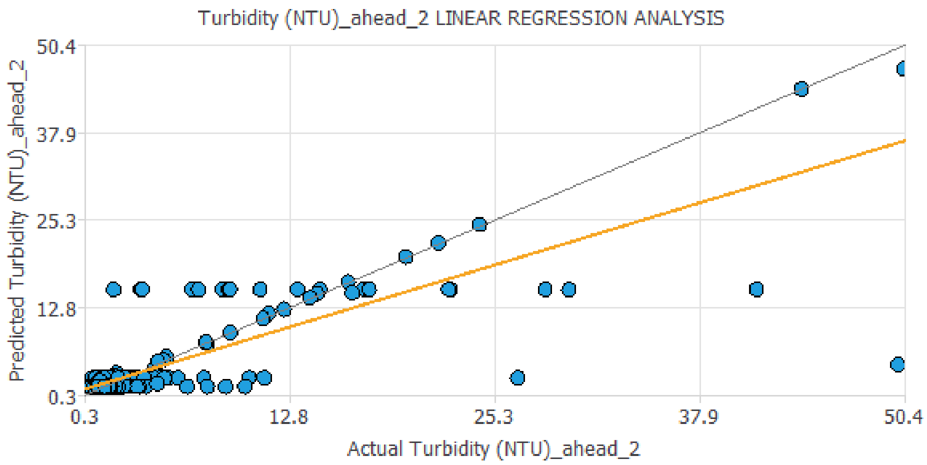
The selection error is 0.60, and the corresponding training error is 0.06.

Neural network architecture turbidity 2-days



An model selection analysis has been applied to find the inputs and the complexity of the neural network.

Linear regression analysis turbidity 2-days

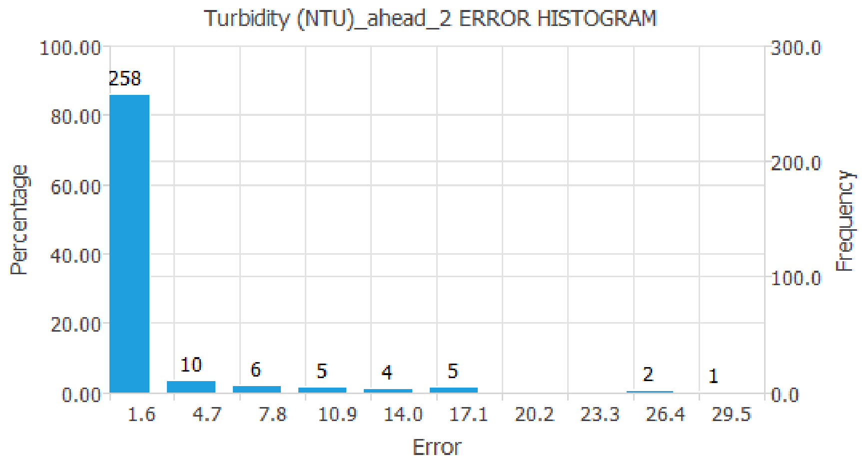


Most of the 2-day turbidity values have a small error, but there are some ones with big errors.

The correlation coefficient between the predicted and the actual values is 0.82.

Error analysis turbidity 2-days

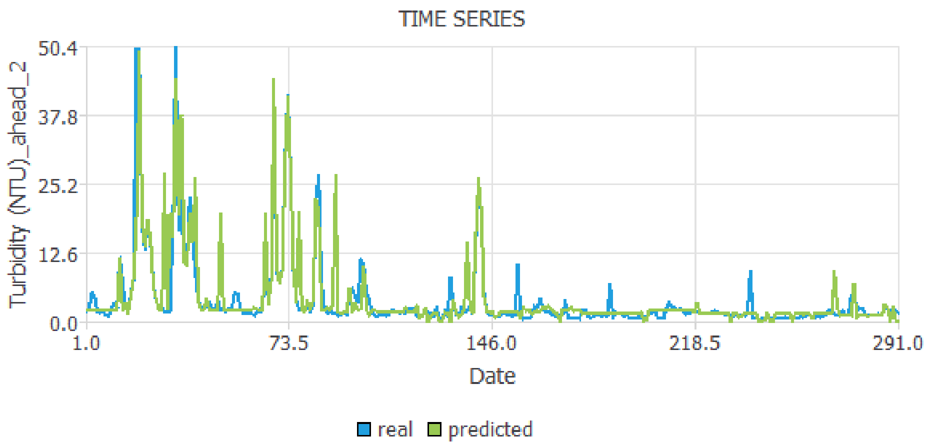
	Absolute error	Relative error
Minimum	0.00	0.00%
Maximum	31.07	59.93%
Mean	3.75	3.75%
Standard deviation	4.71	8.34%



The average error has increased from the 1 day ahead prediction.

The higher errors are not so usual in this model.

Prediction time series turbidity 2-days



The model detects most of the turbidity peaks.

However, there are some peaks not detected, and some normal scenarios with high predicted values.

Turbidity model 3-days

Variable selection turbidity 3-days

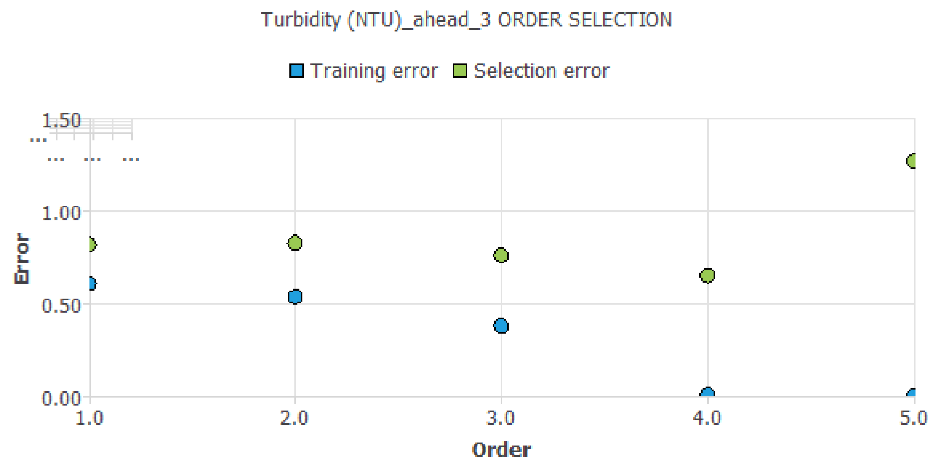
Variable	Region
Photosynthetically available radiation	Plant
Surface temperatura	Plant
Sea floor potential temperatura	Plant
Surface temperature	Hormuz Strait
Surface temperature	Southern Shallows
Turbidity	Plant
Eastward velocity	Hormuz Strait

The strait of Hormuz is the most influential region for the 3-day turbidity.

The temperatures of the previous days are very important for the turbidity.

Order selection turbidity 3-days

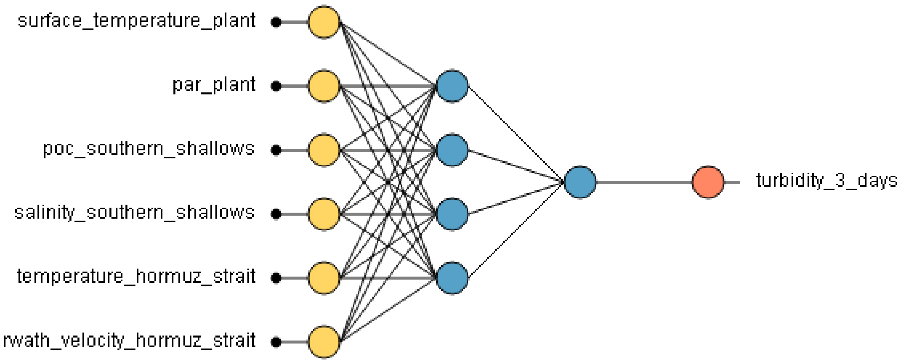
The optimal network architecture is that for which the selection error is minimum.



The optimal order is 4.

The selection error is 0.65.

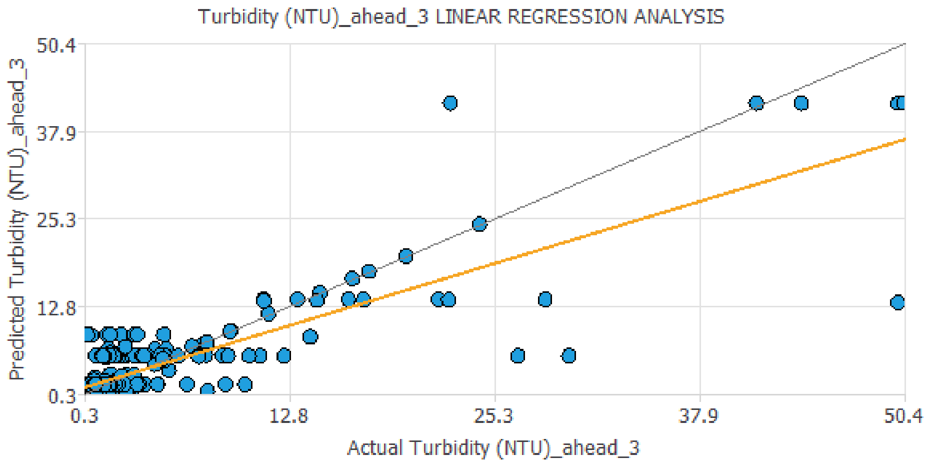
Neural network architecture turbidity 3-days



To predict the 3-day turbidity, the neural network uses endogenous and exogenous variables.

We use 5 lag values for each variable.

Linear regression analysis turbidity 3-days



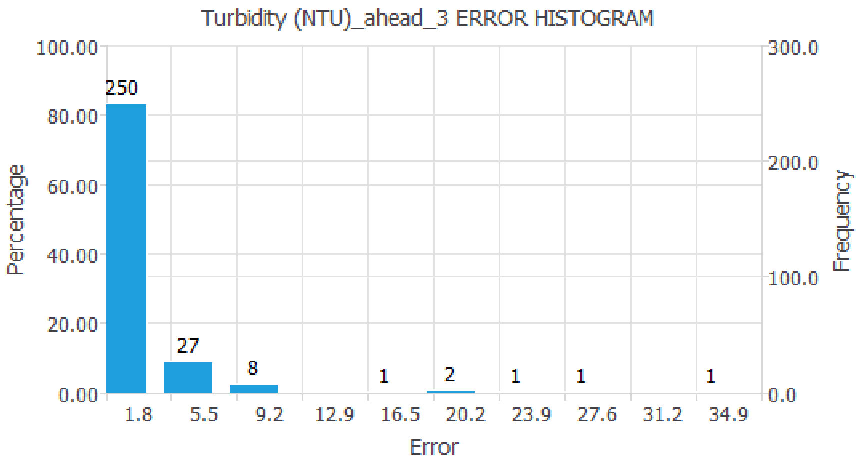
In general, the predictions are good.

However, there are some underestimated and overestimated values.

The correlation value is 0.81.

Error analysis turbidity 3-days

	Absolute error	Relative error
Minimum	0.01	0.02%
Maximum	31.06	61.05%
Mean	1.75	5.74%
Standard deviation	4.09	9.77%

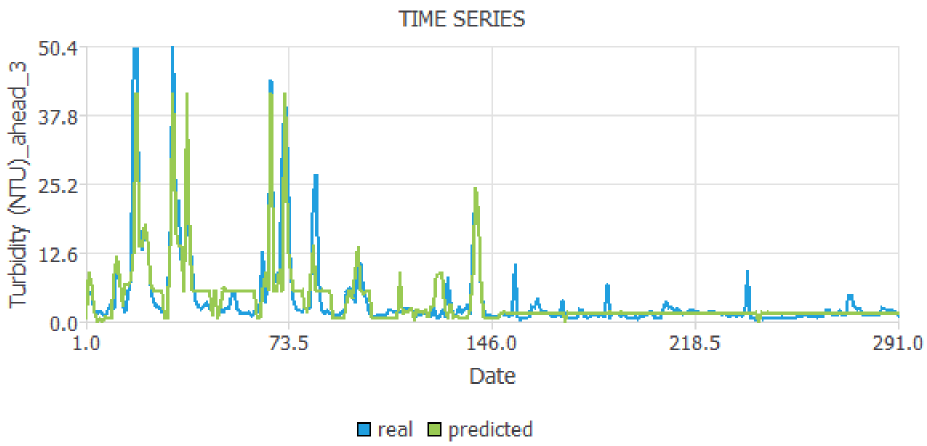


The mean error is 5.74%, which is a small value.

The maximum error is 61.05%, which corresponds to a high turbidity scenario which has not been predicted.

In more than 80% of the cases, the errors are very small.

Prediction time series turbidity 3-days



The turbidity prediction for 3-days ahead can be considered good.

The model still detects the peaks of turbidity, although there are some scenarios not detected.