Wave Power Prediction - Report

Ido Zohar

Introduction

The purpose of this project is to predict the wave power given the time span for the prediction.

This report documents wave power prediction in a specific location in the Mediterranean Sea using different learning methods.

The previous report used KNN regressor in order to predict the wave power.
That work focused on finding the best number of features for the KNN regressor in order to minimize the prediction error.
The first part of this report presents different regressors that try to predict the wave power for various numbers of hours ahead (will be referred below as the '*hours-ahead'* value) more accurately.
The second part focuses on prediction with Neural Networks, starting with basic network and trying to improve it with different methods.

Methods

This report based on data from the Marine Copernicus 006_017 database[1], this database contains measurements from several locations in the Mediterranean Sea.
The report predicts the wave power of the coordinate at latitude=34.64583, longitude=18.75.
In order to predict the wave power I based on the following parameters from the Copernicus database[2]:

- VHMO: spectral significant wave height (Hm0).
- VTM10: spectral moments (-1,0) wave period (Tm-10).
- VTM02: spectral moments (0,2) wave period (Tm02).
- VTPK: wave period at spectral peak / peak period (Tp).
- VMDR: mean wave direction from (Mdir).
- VPED: wave principal direction at spectral peak.
- VSDX: stokes drift U.
- VSDY: stokes drift V.
- VHM0_WW: spectral significant wind wave height.
- VTM01_WW: spectral moments (0,1) wind wave period.
- VMDR_SW1: mean wind wave direction from.
- VHM0_SW1: spectral significant primary swell wave height.
- VTM01_SW1: spectral moments (0,1) primary swell wave period.
- VMDR_SW1: mean primary swell wave direction from.

[1]https://resources.marine.copernicus.eu/?option=com_csw&view=details&product_id=MEDSEA_ANALYSIS_FORECAST_WAV_006_017
[2] https://resources.marine.copernicus.eu/documents/QUID/CMEMS-MED-QUID-006-017.pdf#page=5

- VHM0_SW2: spectral significant secondary swell wave height.
- VTM01_SW2: spectral moments (0,1) secondary swell wave.

When the wave power is calculated by the formula[3]:

$$Wave\ power = 0.49 \cdot VHMO^2 \cdot VTM10$$

The input features for the models in this report are these parameters for the 48 hours before the prediction time, i.e. in total there are $18 \cdot 48 = 864$ features.

The training data for the models is the measurements of the year 2018 (and in the second part the measurements of 2017-2018) from the Copernicus database.
The test data for the models is the corresponding measurements from the first five month of 2019 (January-May).

Each model presented in the report will be evaluated by MSE metric, i.e. the Mean Square Error of the prediction compared to the real wave power measurement.

Part I – Additional Regressors

In this part, several regressors were used in order to check if they can improve the accuracy of the prediction. The models that were checked:

- KNN Regressor
- SGD Regressor
- Linear Regressor
- Decision Tree
- Elastic Net

In addition, the first three regressors were also preprocessed using K-means algorithm with two clusters. Therefore, 8 regressors were used in total.

Each regressor was checked with 5, 10, 20, 50 best features chosen by the Information Gain of each feature on the training set. For each combination of regressor and number of features the prediction for 1, 2, 5, 10, 20, 24, 30 hours-ahead was checked.

Results

The following tables present the results of the regressors' prediction. Each table presents the MSE with different number of features. In each row the regressor that minimizes the MSE for marked with dark shade. For each hours-ahead, the result of the best combination (regressor and number of features) is marked red.

[3] https://en.wikipedia.org/wiki/Wave_power#Wave_power_formula

5 best features - MSE:

hours-ahead	enet-5	lin-5	sgd-5	dtree-5	knn-5	kmeans-knn-5	kmeans-sgd-5	kmeans-lin-5
1	6.739	2.459	2.507	9.877	5.695	4.851	2.694	2.583
2	19.103	8.874	9.310	70.313	13.626	13.634	9.819	9.665
5	91.091	65.352	69.479	153.823	81.392	83.238	68.936	68.758
10	248.933	222.263	258.207	464.502	236.519	236.345	232.720	223.185
20	454.998	435.671	497.278	798.416	460.719	460.105	467.573	434.825
24	503.786	480.556	558.175	897.438	561.667	570.254	561.425	490.523
30	622.302	775.671	717.610	821.145	641.775	678.222	714.352	654.510

10 best features - MSE:

hours-ahead	enet-10	lin-10	sgd-10	dtree-10	knn-10	kmeans-knn-10	kmeans-sgd-10	kmeans-lin-10
1	5.148	2.514	2.386	9.855	10.456	10.468	2.577	2.552
2	14.907	9.091	**8.562**	43.549	22.185	18.485	9.773	9.494
5	82.906	66.642	70.558	186.531	81.470	93.994	70.175	67.484
10	241.010	224.579	255.424	602.244	238.811	240.171	257.089	224.396
20	455.162	439.259	504.216	1087.462	479.831	480.321	467.505	437.403
24	503.848	480.642	558.990	820.260	586.223	591.792	565.518	487.773
30	622.302	775.430	717.027	844.445	635.734	636.793	712.938	618.700

20 best features - MSE:

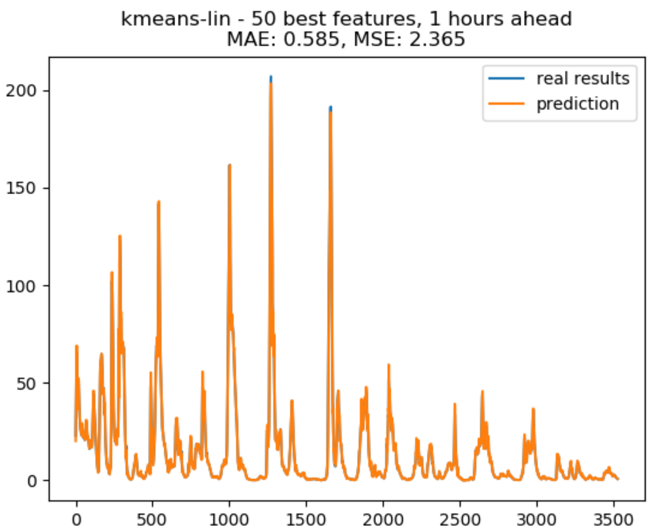
hours-ahead	enet-20	lin-20	sgd-20	dtree-20	knn-20	kmeans-knn-20	kmeans-sgd-20	kmeans-lin-20
1	5.221	2.529	2.459	12.832	18.948	19.059	2.602	2.565
2	15.130	9.127	9.123	50.744	31.675	31.883	9.421	9.554
5	84.170	67.032	69.037	158.060	103.015	103.832	72.502	69.539
10	240.721	219.480	252.664	423.964	274.443	275.138	258.804	215.463
20	442.568	432.429	503.377	823.019	512.030	517.244	492.889	445.989
24	494.412	**480.080**	556.713	873.201	571.358	587.406	555.049	483.228
30	622.302	775.814	716.759	796.787	632.892	634.862	714.555	639.493

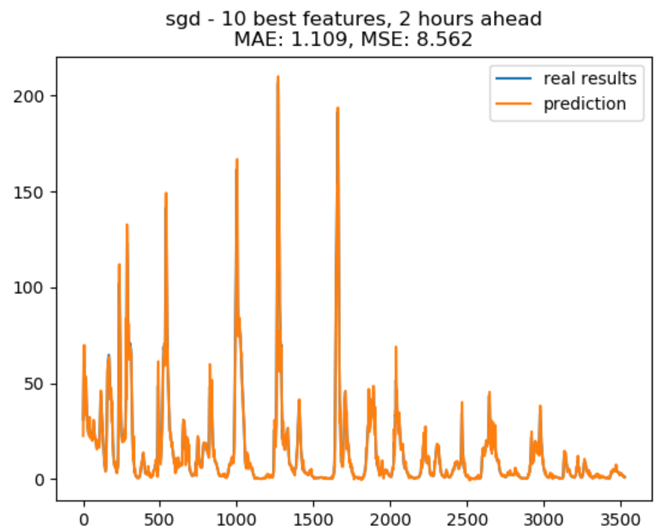
50 best features - MSE:

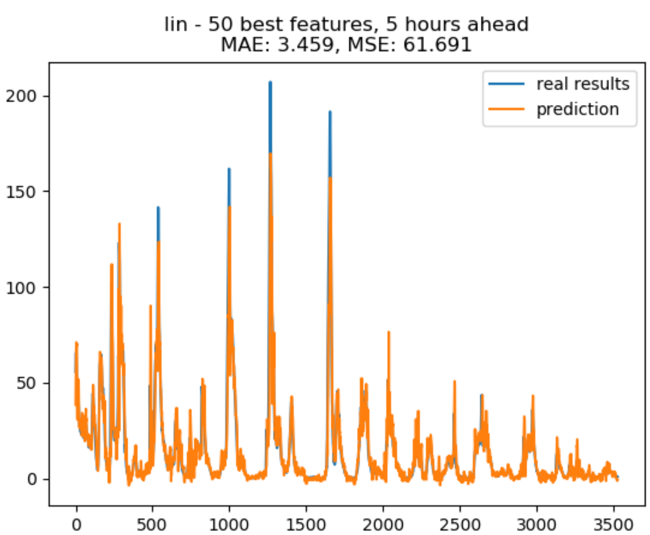
hours-ahead	enet-50	lin-50	sgd-50	dtree-50	knn-50	kmeans-knn-50	kmeans-sgd-50	kmeans-lin-50
1	5.045	2.412	2.379	23.506	47.798	46.651	2.476	**2.365**
2	14.306	8.692	8.605	58.609	63.406	68.520	9.010	8.905
5	76.803	**61.691**	71.207	275.456	152.798	153.646	66.382	66.966
10	230.484	**214.203**	249.255	712.373	305.070	305.211	247.372	228.707
20	448.820	433.198	501.767	1225.166	542.029	545.997	502.487	**431.681**
24	490.450	484.545	555.012	843.706	591.177	591.964	555.630	514.659
30	**554.194**	555.034	630.405	915.934	623.887	625.308	631.898	558.545

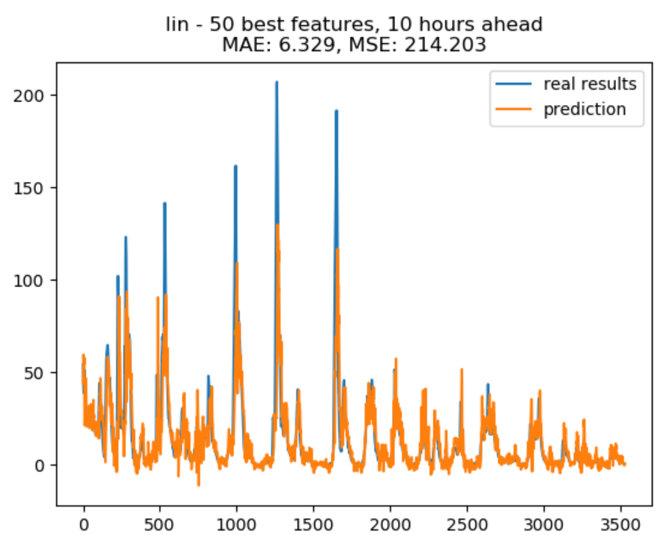
Due to the enormous number of combinations, the following graphs refer to the best combination for each hours-ahead prediction (marked in red above). The rest of the graphs are attached to the report separately.

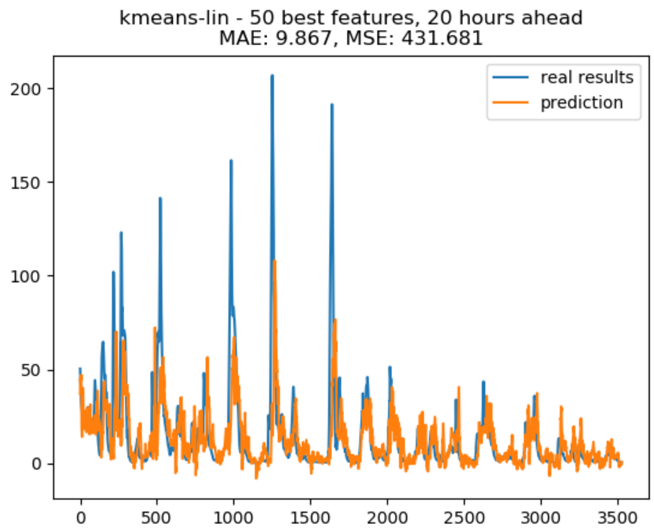
Each graph shows the real wave power compared to the wave power prediction. The y-axis represents the wave power and the x-axis represents the timeline (the number of hours from the beginning of 2019).

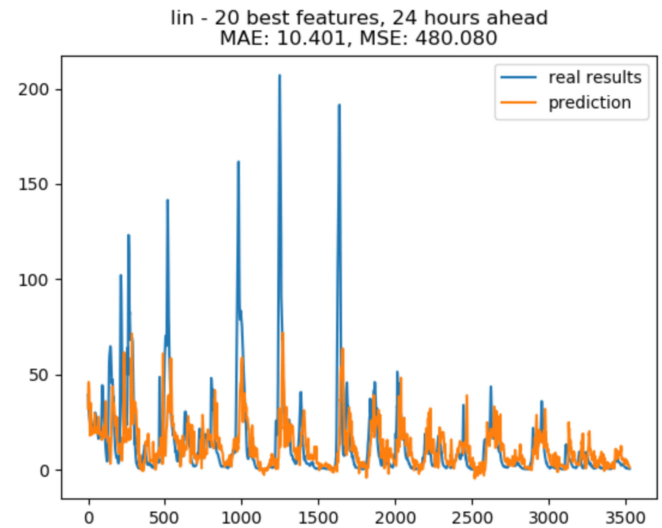


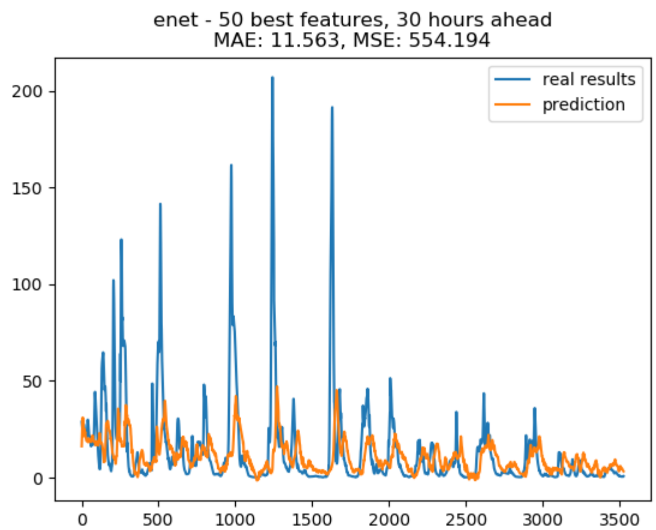










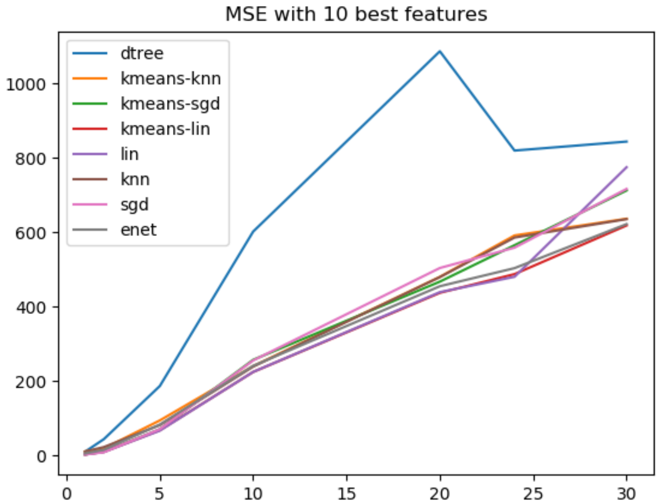


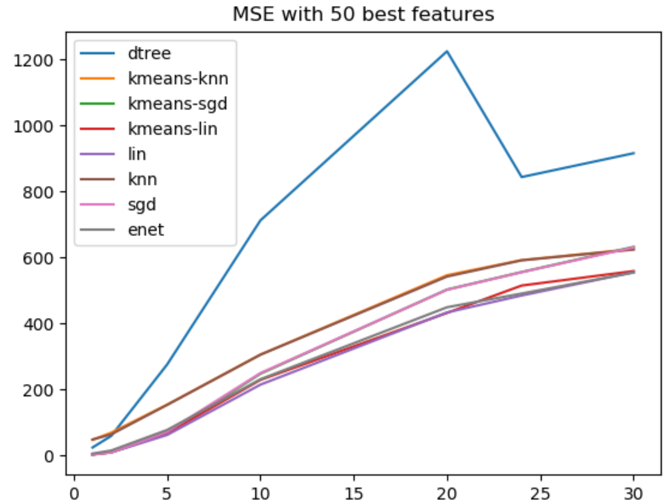
Conclusions

As expected, the accuracy of the prediction decreases as the hours-ahead increases. Actually, the error increases in accordance with the hours-ahead value, regardless the number of features and chosen regressor. This behavior makes sense as it is likely that the wave power's dependence in measurements decreases when those measurements were taken from longer hours ago.

The above tables and the graphs clearly show that the prediction of 1, 2 or 5 hours-ahead resolve similar results to the real wave power, while the prediction of 10 or 20 hours-ahead starts to fake. However, predicting for 24 or 30 hours-ahead leads to result that are far worst. Moreover, you can notice from the graphs above that when the MSE is higher there are noticeable misses in the prediction of 'storms', i.e. sudden increases in the wave power.

The following graphs shows the error value of each regressor in every hours-ahead option with best 10 or 50 features (similar graphs for 5 and 20 features attached separately):





These graphs illustrate the conclusions mentioned above, the MSE (y-axis) increases as the hours-ahead increase (x-axis). We can also see that there are hours that most of the regressors have similar MSE, and in general, most of the regressors are not far better from each other.

We can see from the tables above that in most of the cases the optimal number of features, which gives the lowest MSE value, is 50. This goes against the conclusion of the previous report, which find that the number of the features for KNN regressor is increasing when the hours-ahead increase. For example, the previous report finds that for 1 to 10 hours-ahead the optimal number of features is 5, while for 20 hours-ahead the optimal is 10 features, and for 24 hours-ahead the optimal is 20 features. In this report we can see that at most cases the optimal number of features is 50 regardless to the number of hours-ahead (in the other cases the error of 50 features is not so far from the optimal error).

Lastly, in general look it seems that the linear regressor obtains the best results. As seems from the tables above regardless to the number of features or the number of hours-ahead, the linear regressor gives the optimal prediction or close enough to the optimal (with or without K-means preprocessing).

Part II – Neural Networks

In this part, several types of neural networks were implemented in order to see if they can improve the wave power predictions.

Basic Network

First basic network was built in order to see the trivial predictions with neural networks.

The network in this section contains five layers, three Dense layers and two Dropout layers, alternately. In each Dense layer 64 units.

You can see in the figure below the basic model summary:

```
Layer (type)                 Output Shape              Param #
=================================================================
dense_1 (Dense)              (None, 64)                55360
_________________________________________________________________
dropout_1 (Dropout)          (None, 64)                0
_________________________________________________________________
dense_2 (Dense)              (None, 64)                4160
_________________________________________________________________
dropout_2 (Dropout)          (None, 64)                0
_________________________________________________________________
dense_3 (Dense)              (None, 1)                 65
=================================================================
```

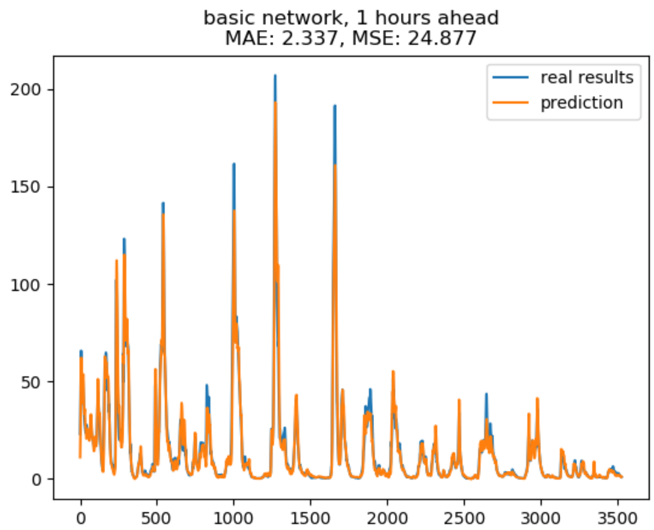
Similar to the previous part, the training set for the network is the data of the year 2018, and the testing set is from January-May 2019, while 20% of the training set used as validation set.

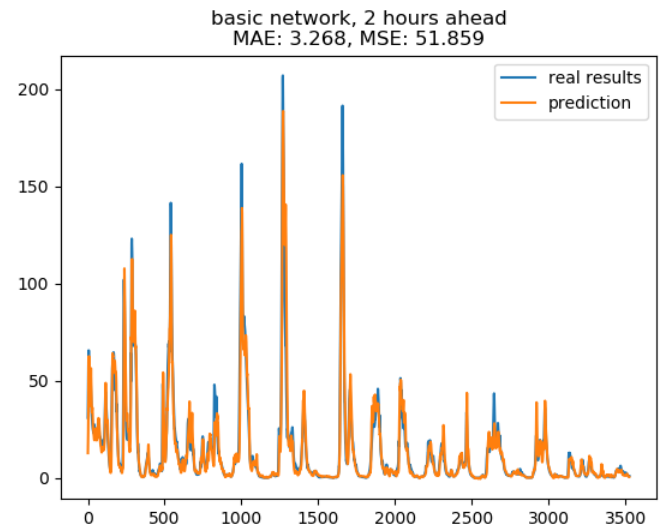
The network was trained and tested for prediction of 1, 2, 5, 10, 20, 24, 30 hours-ahead each time. As before the network was evaluated by MSE metric.

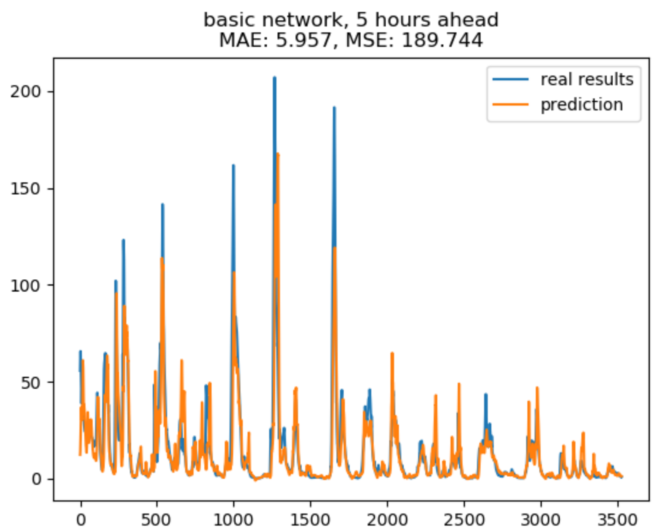
The MSE for each hours-ahead value is presented in the table below:

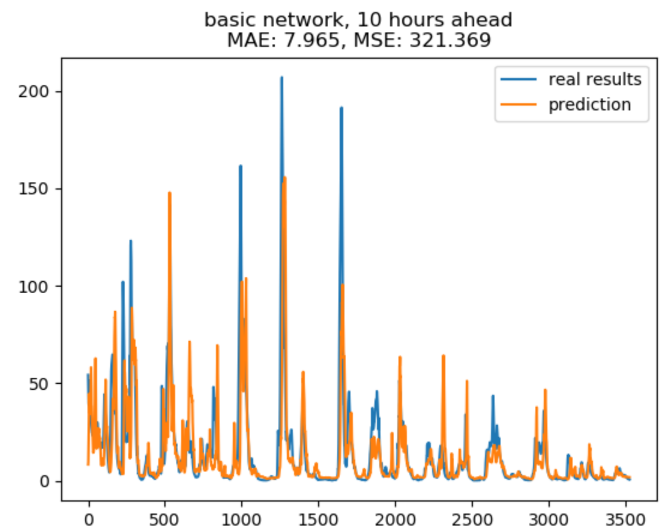
hours ahead	basic-net 18
1	24.877
2	51.859
5	189.745
10	321.369
20	573.334
24	634.349
30	681.574

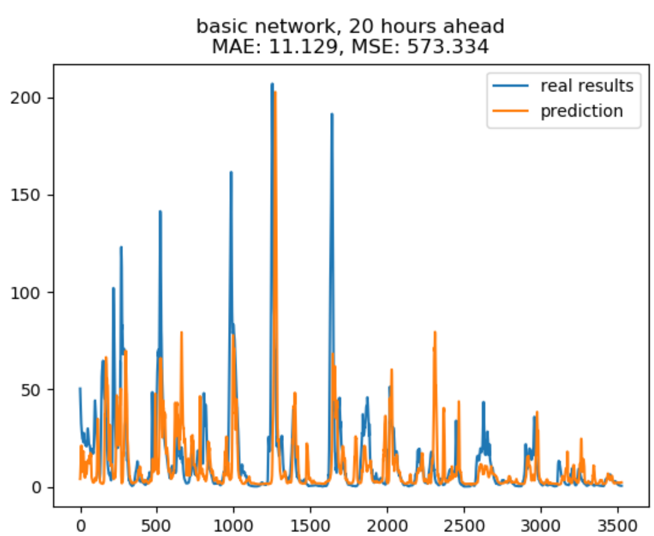
The following graphs show the real wave power compared to the predicted wave power for each hour-ahead value. The y-axis represents the wave power and the x-axis represents the timeline:

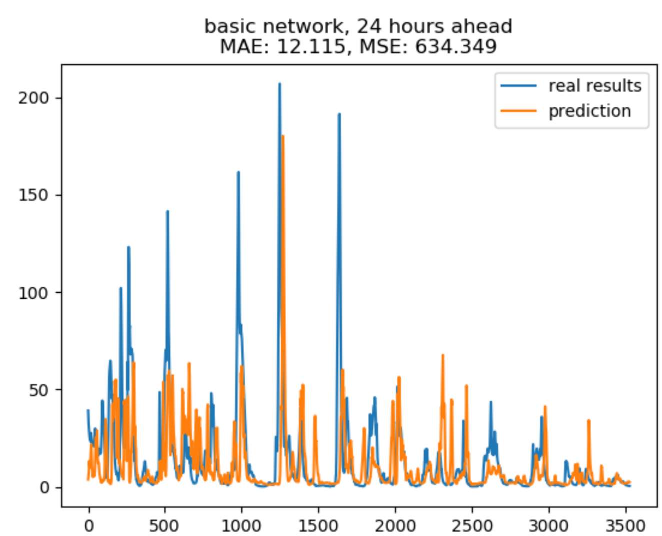

basic network, 1 hours ahead
MAE: 2.337, MSE: 24.877
real results
prediction

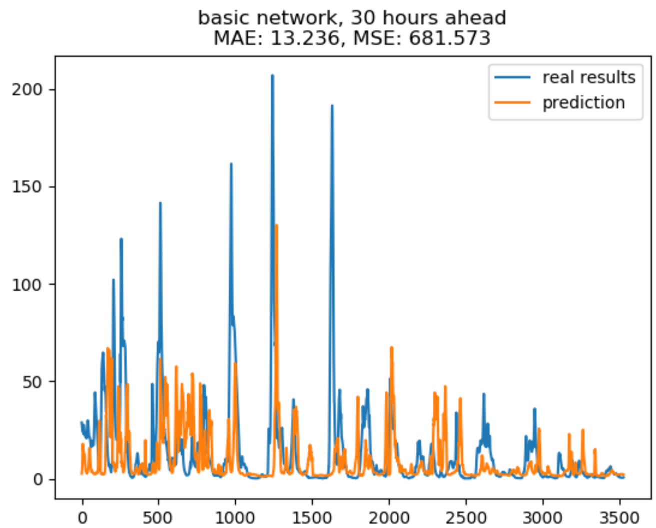

basic network, 2 hours ahead
MAE: 3.268, MSE: 51.859
real results
prediction


basic network, 5 hours ahead
MAE: 5.957, MSE: 189.744
real results
prediction


basic network, 10 hours ahead
MAE: 7.965, MSE: 321.369
real results
prediction


basic network, 20 hours ahead
MAE: 11.129, MSE: 573.334
real results
prediction


basic network, 24 hours ahead
MAE: 12.115, MSE: 634.349
real results
prediction



As can be seen from the graphs and the table above, the accuracy of the prediction decreases as the hours-ahead increases (as we saw in part 1).

The MSE of the network is far higher than the best MSE of the results obtained by the classic regression methods shown on part 1.

Due to that, the following sections suggest few improvements to the basic network in order to get better results in our prediction problem.

Enlarging the Training Set

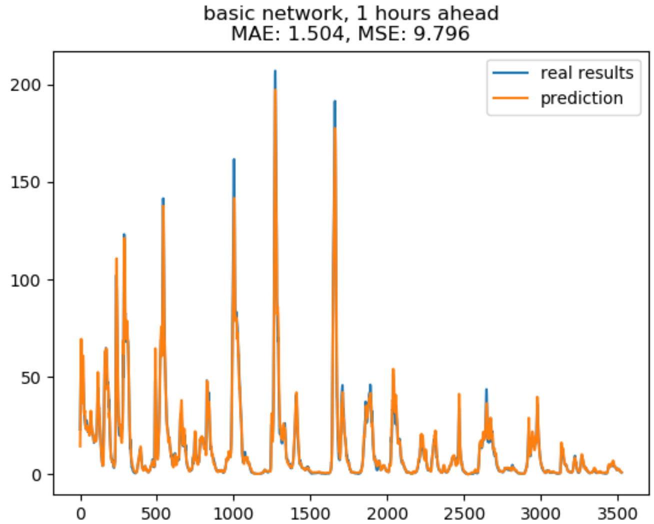
It is reasonable that the network requires bigger training data than the classic regressors, in order to converge to the right prediction. Therefore, the first try to improve the network results was to enlarge the training data. While the training data for the basic network was the measurements of the year 2018, in this network the measurements of the year 2017 (from the same coordinates) were added as well.

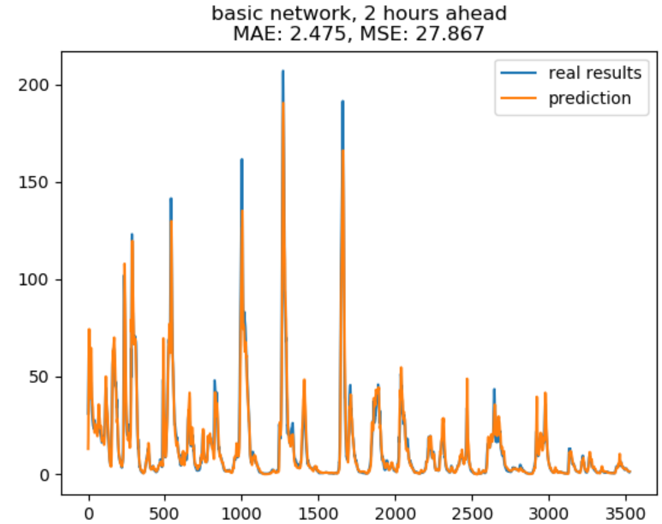
The table below compares the MSE values of the basic network trained only with 2018 measurements and of the network that was trained with 2017-2018 measurements:

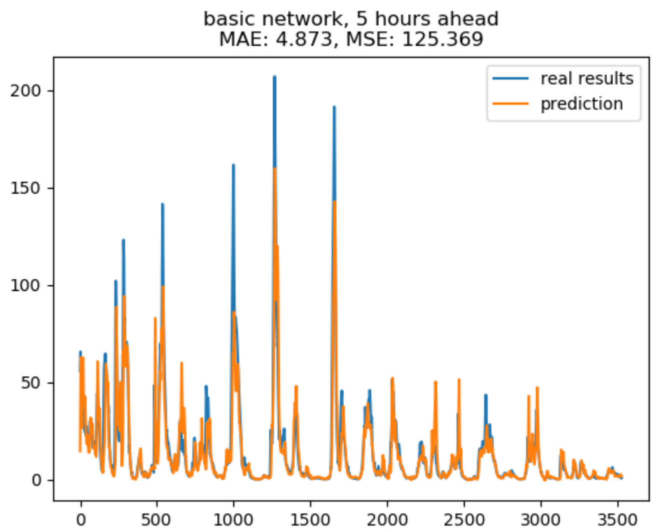
hours ahead	basic-net 2018	basic-net 2017-18	% Improvement
1	24.877	9.796	60.622 %
2	51.859	27.867	46.264 %
5	189.745	125.369	33.928 %
10	321.369	271.761	15.436 %
20	573.334	499.426	12.891 %
24	634.349	566.541	10.689 %
30	681.574	620.095	9.020 %

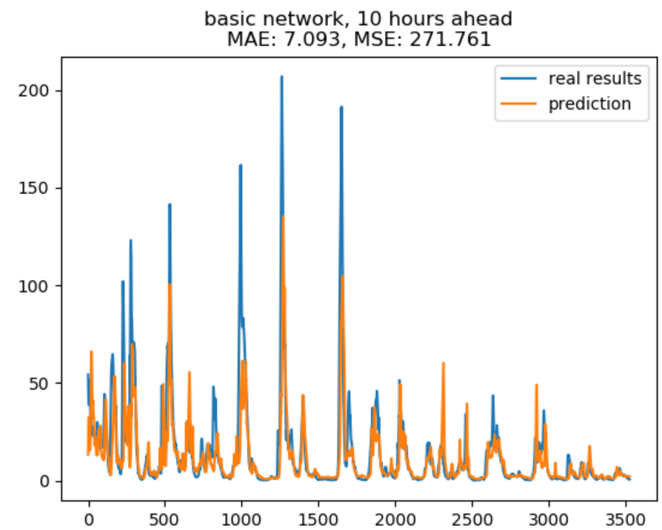
The table above clearly shows that the new network obtained more accurate predictions than the previous network, as assumed.

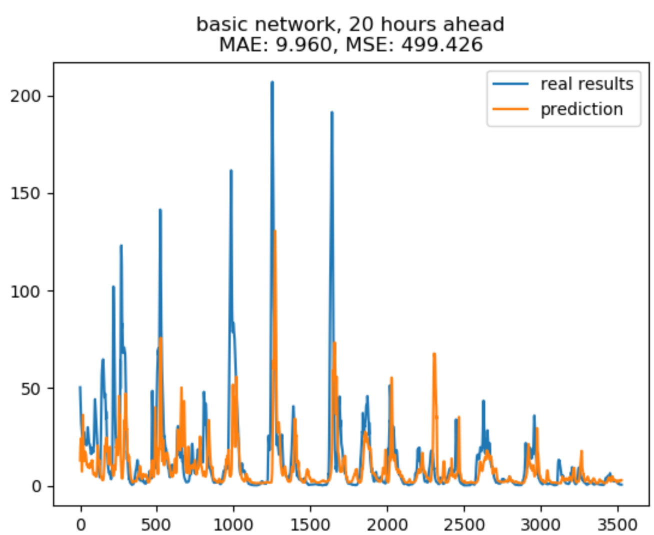
The following graphs show the real wave power compared to the predicted wave power for each hour-ahead for the network trained with 2017-2018 measurements:

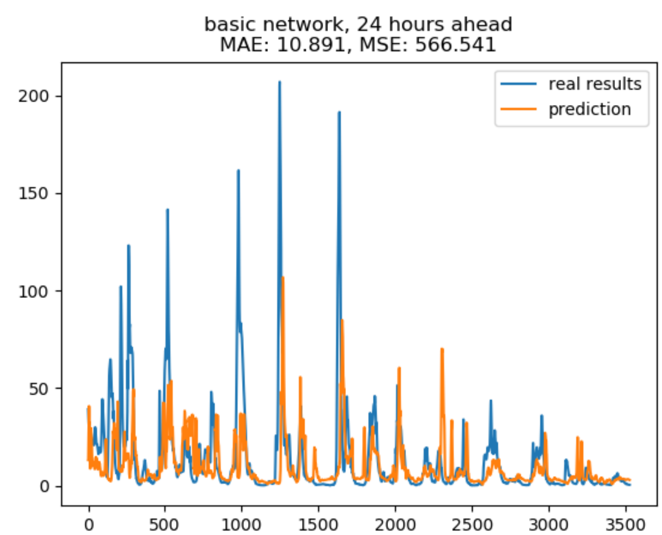


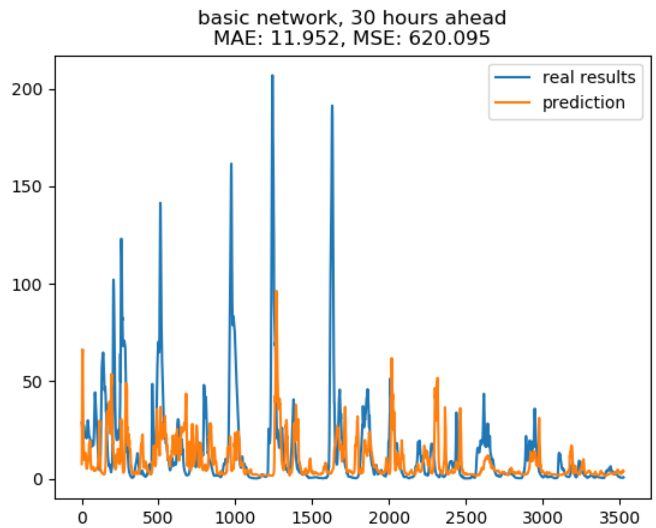




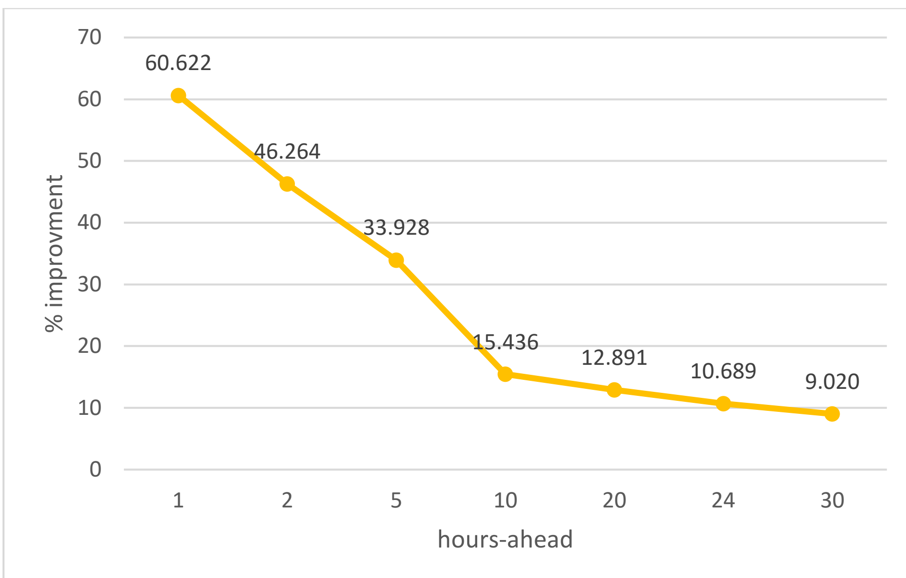








Notice that the improvement is inversely proportional to the number of hours-ahead prediction. The graph below demonstrates that:



In conclusion enlarging the training data improved the results significantly, but still the MSE is higher than the best MSE of the classical regressors used in the first part.

Using LSTM Layers:

The second try to improve network results was to use LSTM Layers. The idea behind this is to add a memory aspect to the network by organizing the input data differently:
In the basic network the input was a vector of $18 \cdot 48 = 864$ values. In this network the input is an 18-length vector, when each entry of the vector is a sequence of 48 hours measurements of a specific parameter.

This new network contains seven layers, LSTM layer and three Dense layers, with Dropout layers between them. In each Dense layer 128 units, and the LSTM layer with 360 units.

You can see in the figure below the LSTM model summary:

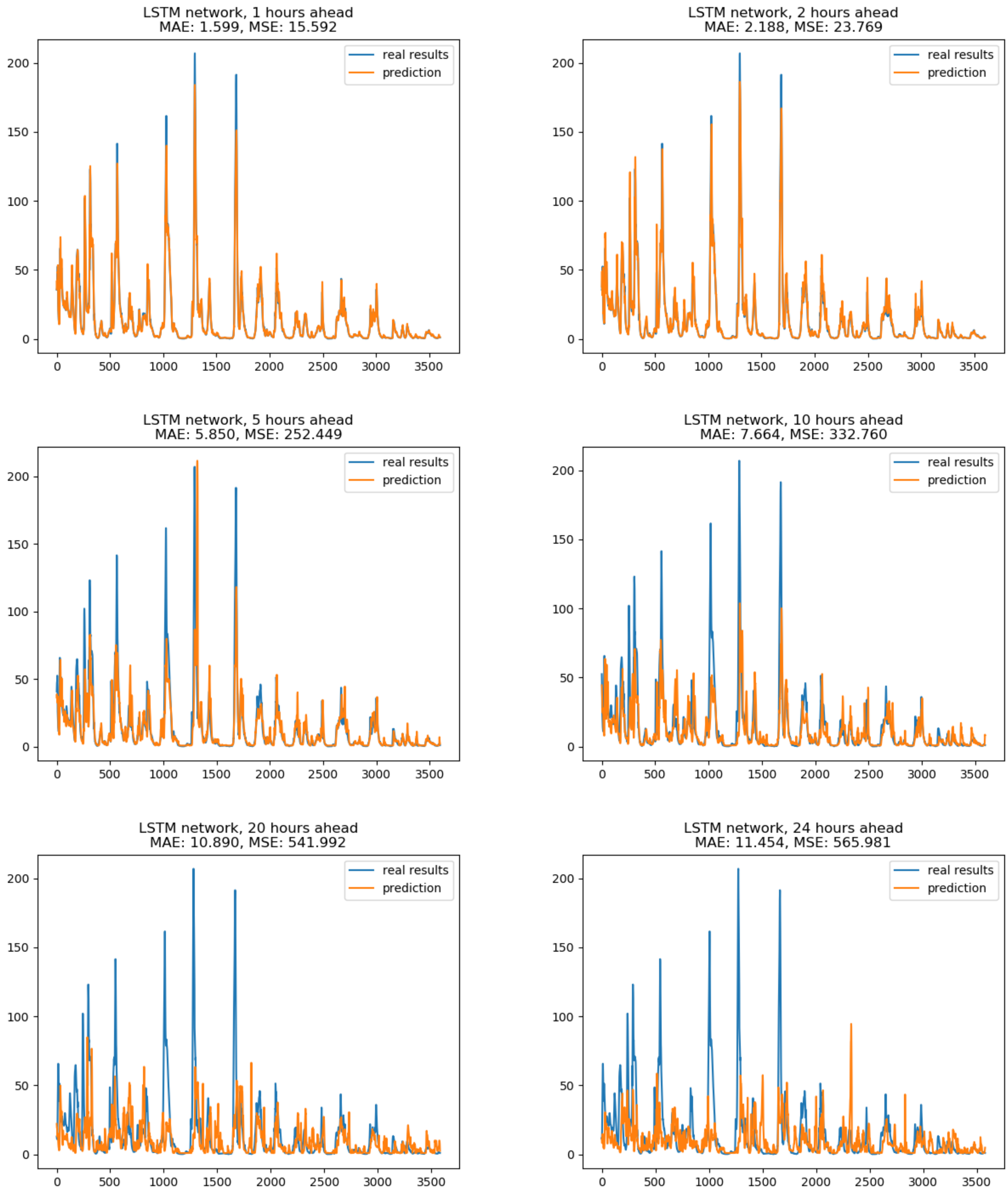
```
Layer (type)                 Output Shape              Param #
=================================================================
lstm_1 (LSTM)                (None, 360)               545760
_________________________________________________________________
dropout_1 (Dropout)          (None, 360)               0
_________________________________________________________________
dense_1 (Dense)              (None, 128)               46208
_________________________________________________________________
dropout_2 (Dropout)          (None, 128)               0
_________________________________________________________________
dense_2 (Dense)              (None, 128)               16512
_________________________________________________________________
dropout_3 (Dropout)          (None, 128)               0
_________________________________________________________________
dense_3 (Dense)              (None, 1)                 129
=================================================================
```

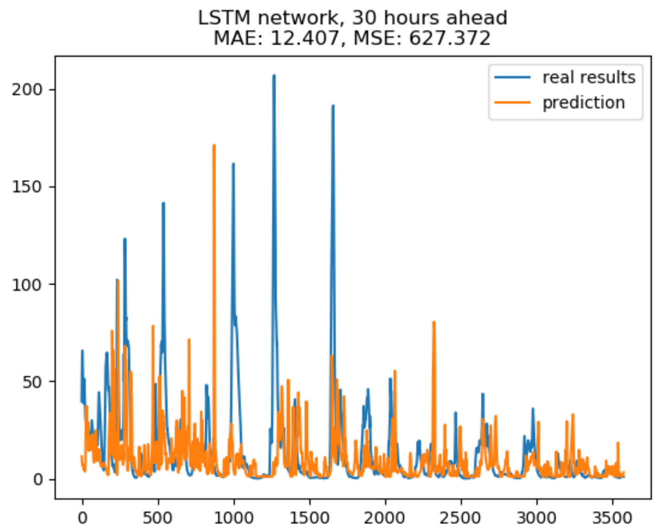
The table below compares the MSE values of the basic network and of the LSTM-network. Both networks were trained with 2017-2018 measurements:

hours ahead	basic-net 2017-18	LSTM-net	% Improvement
1	9.796	15.59231091	-59.170 %
2	27.867	23.76920891	14.705 %
5	125.369	252.4490814	-101.365 %
10	271.761	332.7595215	-22.446 %
20	499.426	541.9915161	-8.523 %
24	566.541	565.9812012	0.099 %
30	620.095	627.3722534	-1.174 %

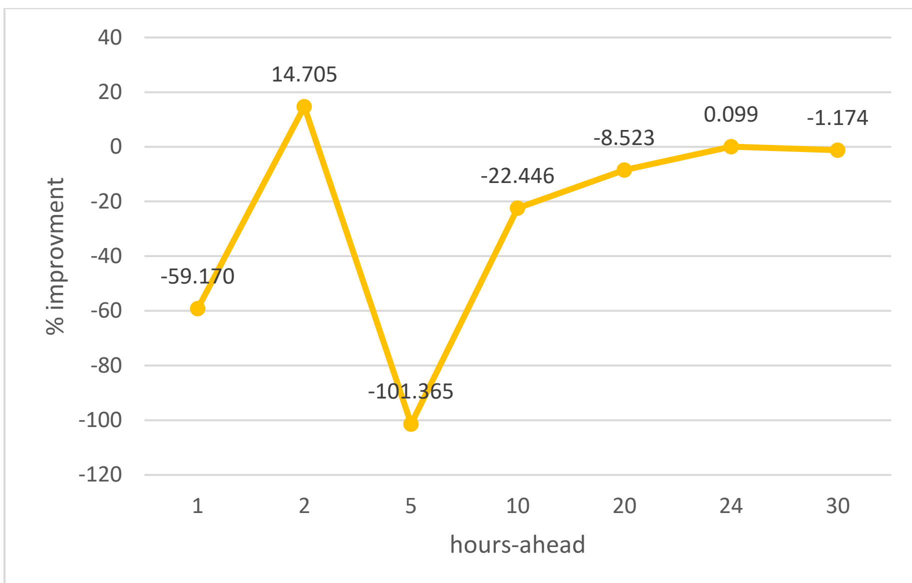
The table above shows that in most cases the new network got worse results than the basic network. Although, the new network got better results at predicting 2 hours-ahead, and the prediction for 24 hours-ahead showed a minor improvement.

The following graphs shows the real wave power compared to the predicted wave power for each hours-ahead value for the LSTM-network:


LSTM network, 1 hours ahead
MAE: 1.599, MSE: 15.592
real results
prediction
LSTM network, 2 hours ahead
MAE: 2.188, MSE: 23.769
real results
prediction
LSTM network, 5 hours ahead
MAE: 5.850, MSE: 252.449
real results
prediction
LSTM network, 10 hours ahead
MAE: 7.664, MSE: 332.760
real results
prediction
LSTM network, 20 hours ahead
MAE: 10.890, MSE: 541.992
real results
prediction
LSTM network, 24 hours ahead
MAE: 11.454, MSE: 565.981
real results
prediction



The graph below demonstrates the improvement ratio:

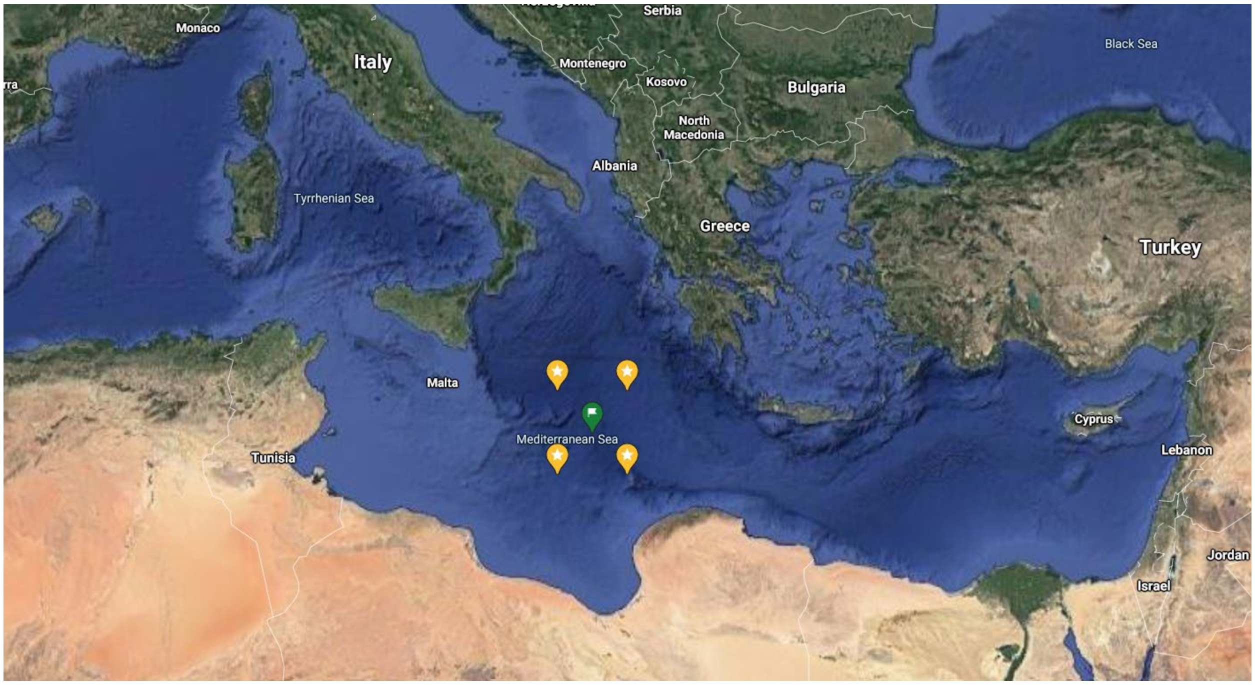


Adding the Spatial Aspect

The third try to improve network results was adding spatial aspect to the problem.

While until now the training data for the networks was the measurements at the location of the prediction. In this network the wave power of 4 other locations (will be referred below as the '*nearby locations*') were added as features to the input vector. For each location, the wave power measurements of the previous 48 hours were used, therefore, totally $48 \cdot 4 = 192$ new features were added.

The 4 chosen locations were (33.64583,17.75), (33.64583,19.75), (35.64583,17.75), (35.64583,19.75). These coordinates create $144km \times 144km$ square around the coordinate of the prediction - (34.64583,18.75). Shown in the map below:

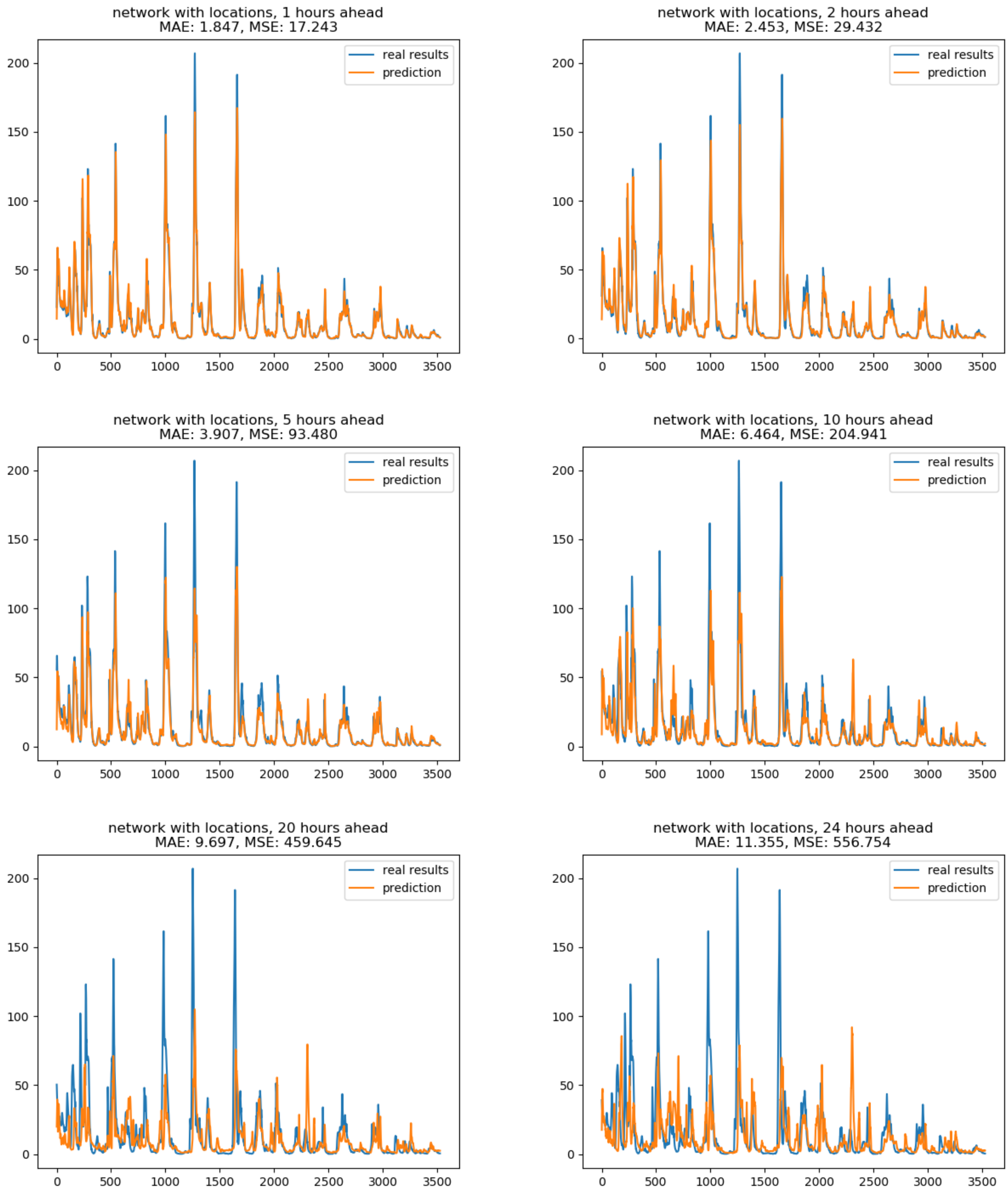


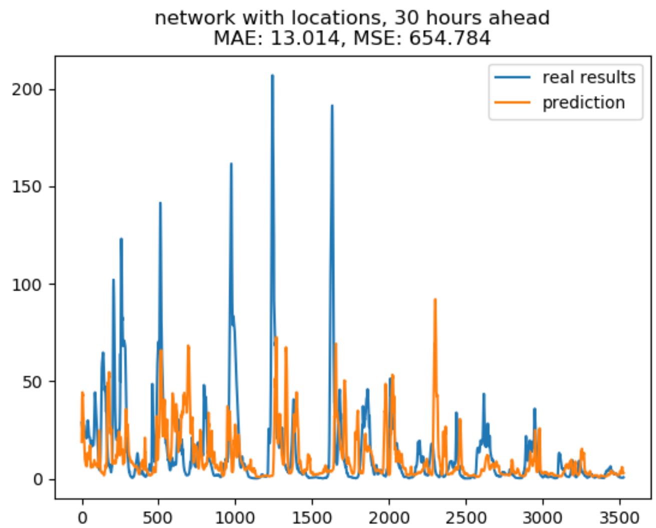
The idea for adding these features is the spatial aspect of our problem. It makes sense that knowing the nearby wave power, would help to predict trends in the waves' behavior.

The table below compares the MSE values of the basic network and of the network with the spatial data. Both networks were trained with 2017-2018 measurements:

hours ahead	basic-net 2017-18	loc-net 17-18	% Improvement
1	9.796	17.243	-76.021 %
2	27.867	29.432	-5.616 %
5	125.369	93.480	25.436 %
10	271.761	204.941	24.588 %
20	499.426	459.645	7.965 %
24	566.541	556.754	1.728 %
30	620.095	654.784	-5.594 %

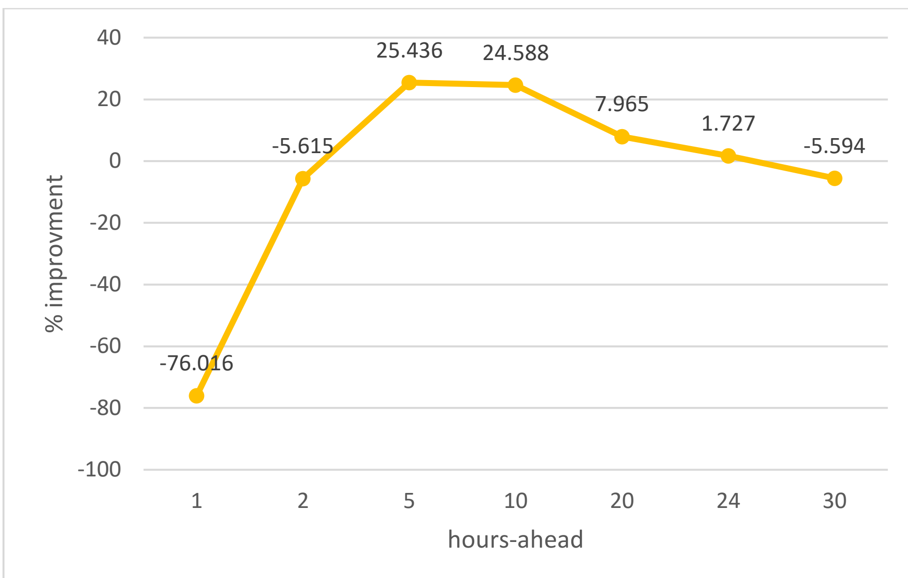
The following graphs shows the real wave power compared to the predicted wave power for each hours-ahead value for the network with the spatial data:


network with locations, 1 hours ahead
MAE: 1.847, MSE: 17.243
network with locations, 2 hours ahead
MAE: 2.453, MSE: 29.432
network with locations, 5 hours ahead
MAE: 3.907, MSE: 93.480
network with locations, 10 hours ahead
MAE: 6.464, MSE: 204.941
network with locations, 20 hours ahead
MAE: 9.697, MSE: 459.645
network with locations, 24 hours ahead
MAE: 11.355, MSE: 556.754
real results
prediction



From table above it seems that this effort got ambiguous improvements. For the 5-24 hours-ahead predictions it seems that adding the spatial data was beneficial, while for 1 hour-ahead this change caused a significant decrease in the accuracy.

The graph below demonstrates the improvement ratio:



Adding the Spatial Aspect – Advanced Approach

The last try to improve network results was to refine the previous idea of spatial aspect.

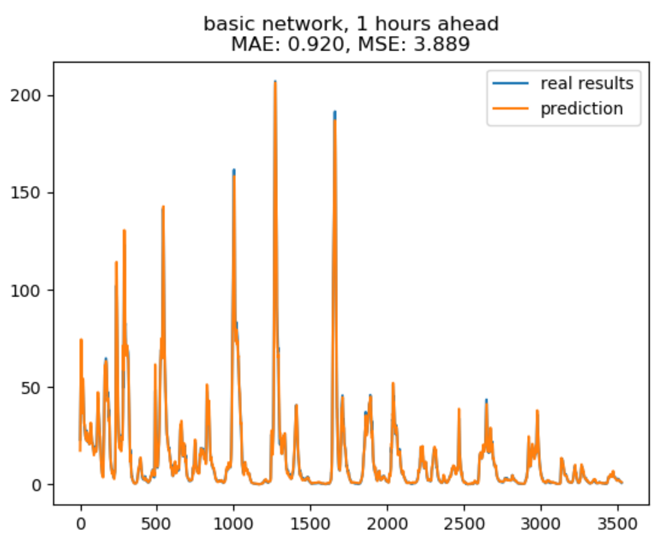
While in the previous section the original input vector was extended with the wave power at 4 nearby locations as new features, in this section the measurements of the 4 nearby locations will be added to the training set as new input vectors. The outcome of this change is multiplying the training data size by 5. The thought behind this approach is to enrich and diversify the training set in order to avoid overfitting, and at the same time trying not to contaminate the training set with data that behaves differently. Reasonable assumption is that waves at close enough coordinates will behave similarly, while too close coordinates will not cause the desired diversity.

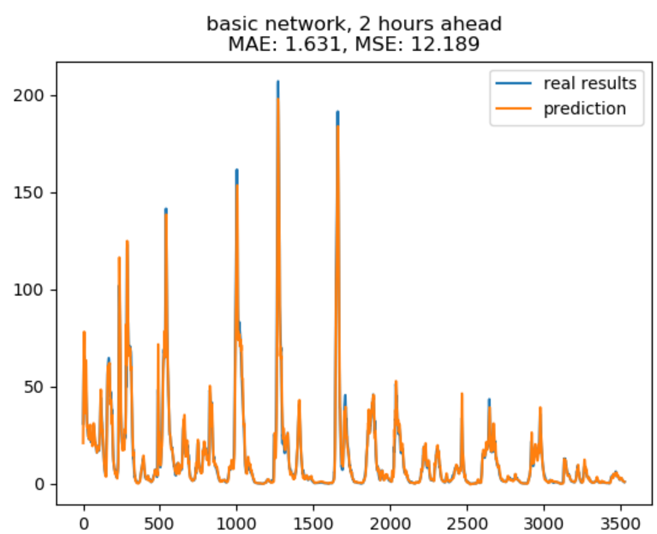
The table below compares the MSE values of the network trained with 2017-2018 measurements and of the network trained with 2017-2018 measurements of all five coordinates:

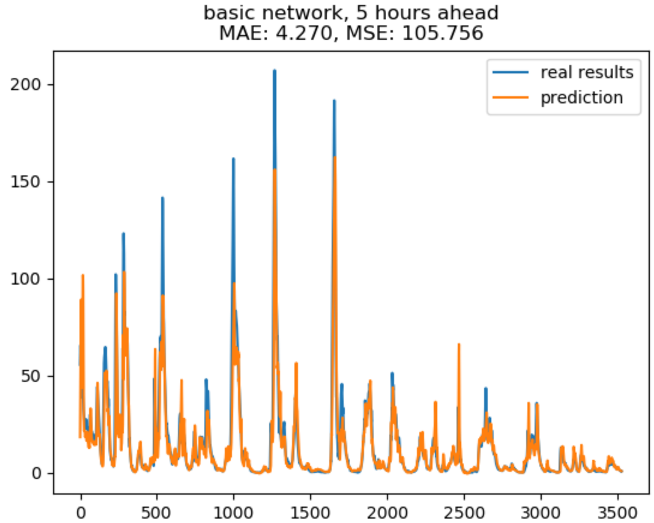
hours ahead	basic-net 2017-18	basic-net 5 locations	% Improvement
1	9.796	3.888982058	60.300 %
2	27.867	12.18929291	56.259 %
5	125.369	105.7561264	15.644 %
10	271.761	267.7641602	1.471 %
20	499.426	486.296814	2.629 %
24	566.541	543.3109131	4.100 %
30	620.095	653.3363647	-5.361 %

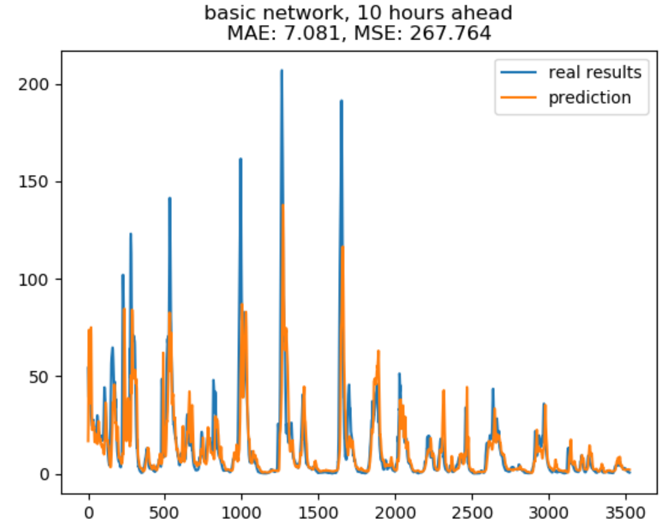
From the table above it seems that this effort got significant improvement on predicting 1-5 hours ahead, while predicting more hours ahead got minor (if any) improvements.

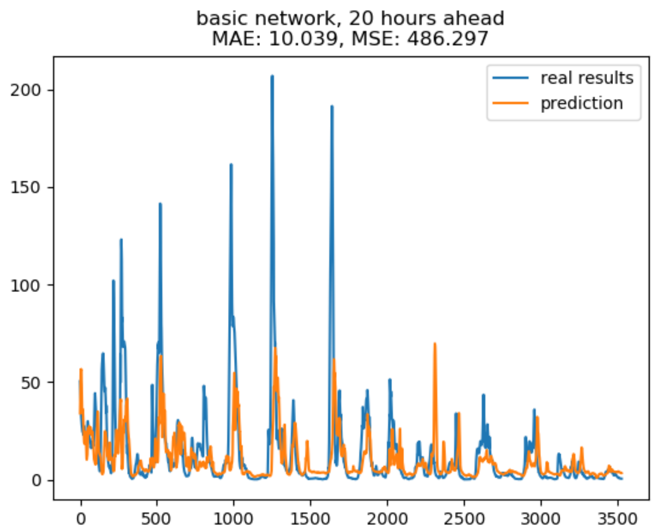
The following graphs shows the real wave power compared to the predicted wave power for each hour-ahead value for the network trained with the 2017-2018 measurements of these five coordinates:

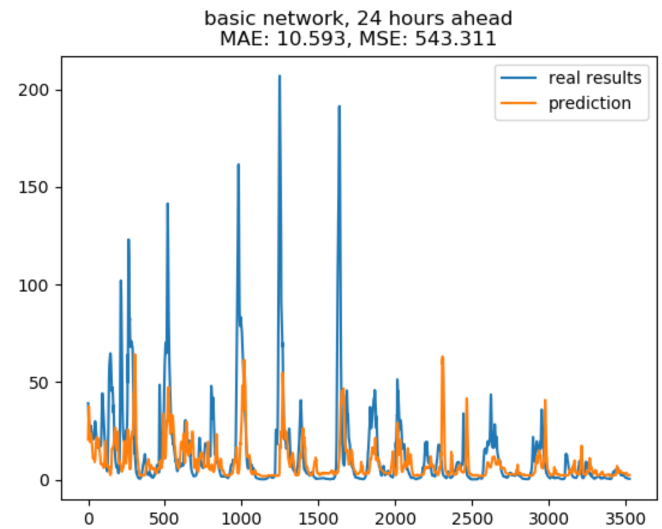


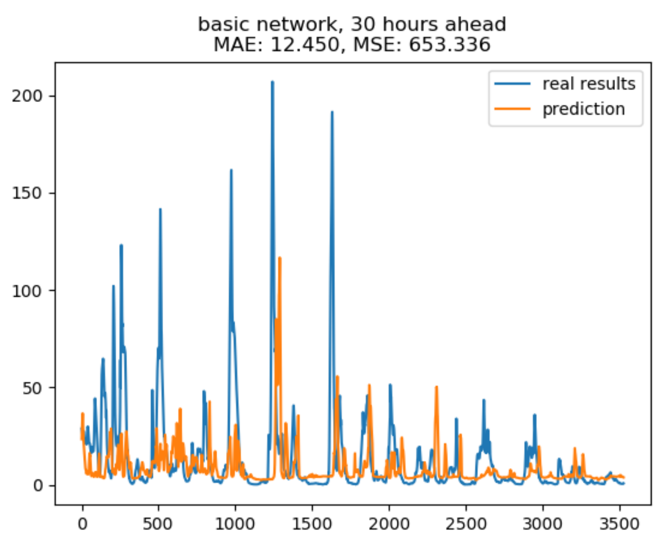



basic network, 5 hours ahead
MAE: 4.270, MSE: 105.756
real results
prediction
0
50
100
150
200
0
500
1000
1500
2000
2500
3000
3500

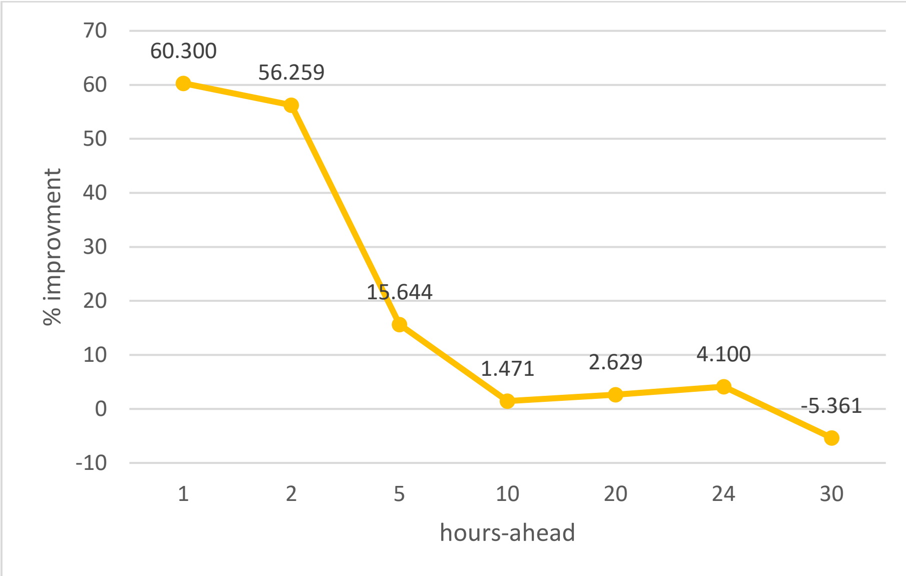

basic network, 10 hours ahead
MAE: 7.081, MSE: 267.764
real results
prediction
0
50
100
150
200
0
500
1000
1500
2000
2500
3000
3500


basic network, 20 hours ahead
MAE: 10.039, MSE: 486.297
real results
prediction
0
50
100
150
200
0
500
1000
1500
2000
2500
3000
3500


basic network, 24 hours ahead
MAE: 10.593, MSE: 543.311
real results
prediction
0
50
100
150
200
0
500
1000
1500
2000
2500
3000
3500


basic network, 30 hours ahead
MAE: 12.450, MSE: 653.336
real results
prediction
0
50
100
150
200
0
500
1000
1500
2000
2500
3000
3500

The graph below demonstrates the improvement ratio:



Conclusions

In this part serval neural network used in order to predict the wave power.
Similarly to the first part, the accuracy of the predictions decreases as the hours-ahead increases.
While the MSE results of the basic network were very high, enlarging the training set with 2017 measurements leads to much better results. Adding the spatial data improves some of the results, but not all of them.
The LSTM network did not obtain better results than the basic network's in most of the cases.

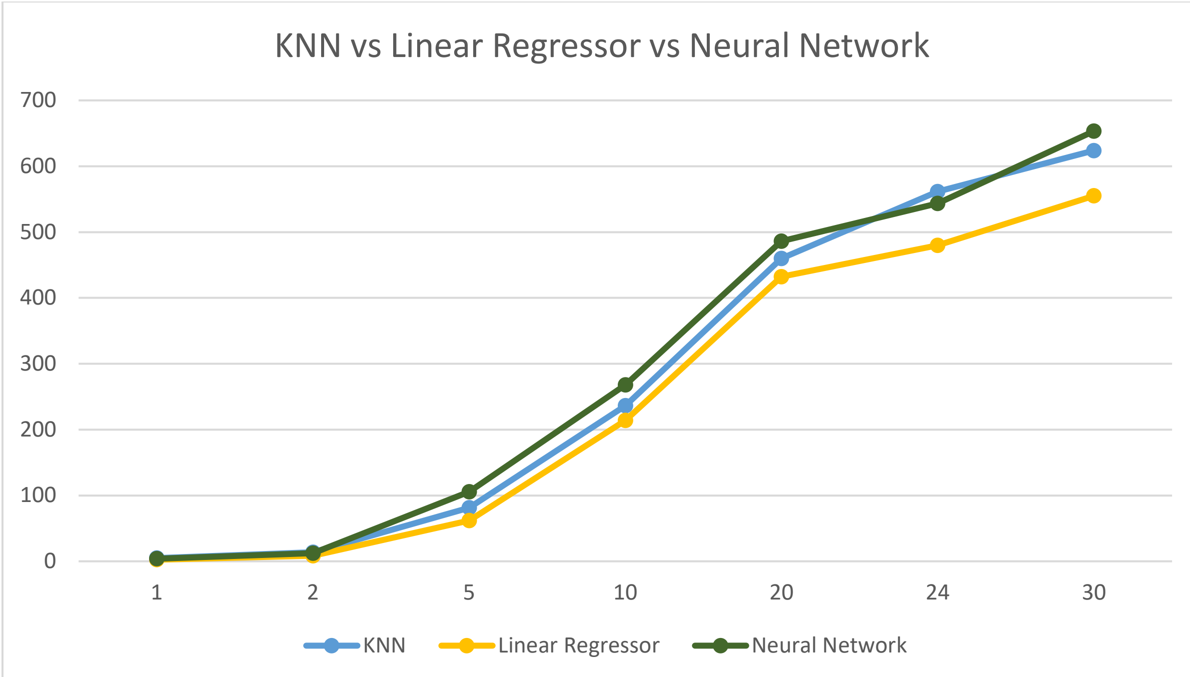
Summary

The purpose of this work was to improve the accuracy the wave power prediction obtained in the previous work, two methods were presented in this report:

The first method was using classic regressors for the prediction. Using this method led to improvement in the results.

The second approach was using Neural Networks for the prediction. In this part several versions of networks were built and used. Although some of the versions obtained better results than the basic network, those result were worse than the results obtained using the classic regressors.

The graph below compares the MSE values of the KNN, Linear Regressor (the best results in each hour-ahead value) and the Neural Network (trained with input of 5 locations):



As you can see, the Linear Regressor obtained the best results in any hours-ahead prediction. On the other hand, the Neural Network failed to improve the KNN results (of the previous report).

Despite the disappointing results of the networks, they are not that far from the KNN results, and it seems that with additional work the network could obtain better.

Future work could be trying other types of network and see how they influence the results.

Since enlarging the training set improved the results, it reasonable to try enlarging the training data with other methods.

Additional direction could be trying to adjust the number and coordinates of the nearby locations.

Moreover, it could be interesting to test the performance of these models on other locations in order to see if these models could be generalized and still maintain the quality of the prediction.