9

Experiments and Results

This chapter will present the experiments done for assessing the performance of the different forecasting models and their performances. The RF model and the SVR model were both used with the direct and the recursive forecasting methods as described in chapter 4. Due to limitations in computational capacity and time constraints related to the submission deadline, the LSTM model was used only for one of the wind farm locations and for three different time horizons. According to the performance of the persistence model for 1-hour ahead predictions on all wind farm locations it was shown that the model had the worst performance on Havøygavlen. The persistence model can in some cases be used as measure of the complexity of the dataset. Because Havøygavlen seemed to be the wind farm location that was the most difficult to predict, the LSTM model was tested on this location.

For every location and every split of the dataset, predictions were made for all the forecasting horizons. The MAE, RMSE, and the NRMSE for every forecasting horizon presented in tables 9.1-9.5 are the average performance for each split of the datasets. This way, some of the uncertainty in using non-deterministic models is eliminated, and the models are evaluated on different seasons of the year rather than for one particular time of the year.

The experiment results are first presented for each of the wind farm locations in sections 9.1 - 9.5, and in section 9.6 the overall results of the different models are shown.

9.1 Raggovidda Wind Farm

The results from Raggovidda wind farm show that the best performing model across all time horizons is the recursive SVR model, while the worst performing model is the direct RF model, especially at large forecasting horizons. All of the implemented models, except the ARIMAX and the direct RF model, outperform the persistence model at forecasting horizons $h > 12$. This can be seen in figure 9.2 and figure 9.3. The average results across all the dataset splits, from all models and all forecasting horizons are shown in table 9.1.

Results From Raggovidda Wind Farm											
Method	Error Metric	Forecasting Horizon									
		h=1	h=2	h=3	h=4	h=5	h=6	h=9	h=12	h=18	h=24
Persistence	MAE	0.64	0.85	1.66	2.53	3.19	2.97	2.40	2.94	6.70	7.62
	RMSE	1.54	3.00	4.18	4.93	5.53	5.85	4.98	4.68	7.41	7.13
	NRMSE	0.03	0.05	0.10	0.15	0.19	0.18	0.16	0.22	0.46	0.50
RF(Recursive)	MAE	6.44	4.99	3.80	3.17	3.75	4.10	3.93	4.23	4.10	4.37
	RMSE	6.44	5.22	4.39	3.89	4.78	5.40	4.97	5.16	5.14	5.56
	NRMSE	0.30	0.24	0.20	0.18	0.22	0.25	0.22	0.24	0.24	0.26
RF(Direct)	MAE	4.07	3.96	3.11	3.39	4.69	5.64	6.73	8.04	10.36	12.06
	RMSE	4.07	3.97	3.38	3.64	5.51	6.82	8.08	9.47	12.18	13.84
	NRMSE	0.19	0.18	0.15	0.17	0.25	0.31	0.37	0.43	0.56	0.64
SVR(recursive)	MAE	2.60	3.37	2.92	3.00	3.73	4.18	3.65	3.56	3.69	3.85
	RMSE	2.60	3.54	3.31	3.30	4.95	5.68	5.05	4.76	5.05	5.48
	NRMSE	0.12	0.16	0.15	0.15	0.23	0.26	0.23	0.22	0.23	0.25
SVR(Direct)	MAE	2.69	1.86	2.43	2.99	4.02	4.21	3.86	4.07	5.80	6.79
	RMSE	2.69	2.08	2.85	3.80	5.10	5.38	5.17	5.27	7.75	9.04
	NRMSE	0.12	0.10	0.13	0.17	0.23	0.25	0.24	0.24	0.36	0.42
ARIMAX	MAE	3.63	3.43	3.45	3.62	4.30	5.04	6.40	7.62	7.78	9.54
	RMSE	3.63	3.78	4.17	4.37	5.02	5.85	7.54	8.71	9.11	11.03
	NRMSE	0.17	0.17	0.19	0.20	0.23	0.27	0.35	0.40	0.42	0.51

Table 9.1: Results from Raggovidda wind farm. The results show that the persistence model is the best performing model for short forecasting horizons, but when $h > 12$ the persistence model is outperformed by the recursive RF and SVR and the direct SVR.

From the table it can be seen that the direct approaches of the RF and the SVR models have a low error across all error measures for short forecasting horizons, and the error increases as the forecasting horizon gets bigger. Figure 9.1 shows the 24 hour predictions for one of the test sets. It can be seen that both the recursive forecasting approaches do a better job in approximating the actual power output of the wind farm, whereas the direct approaches seem to estimate some trend or overall average of the power output. This indicates why the recursive forecasting method outperforms the direct method for large forecasting horizons. The reason might lie in the different way the models are built when using the recursive vs. the direct forecasting approach, which will be further discussed in chapter 10. It is also noticeable that the ARIMAX model follows the pattern of the power output for short time horizons, but when the forecasting horizon increases the model accuracy decreases.

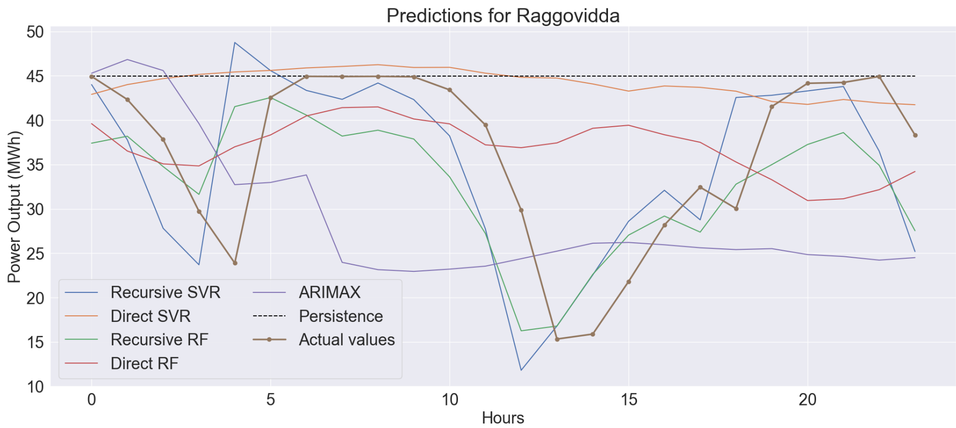


Figure 9.1: 24 hours predictions for Raggovidda wind farm from the different forecasting models on one of the test datasets.

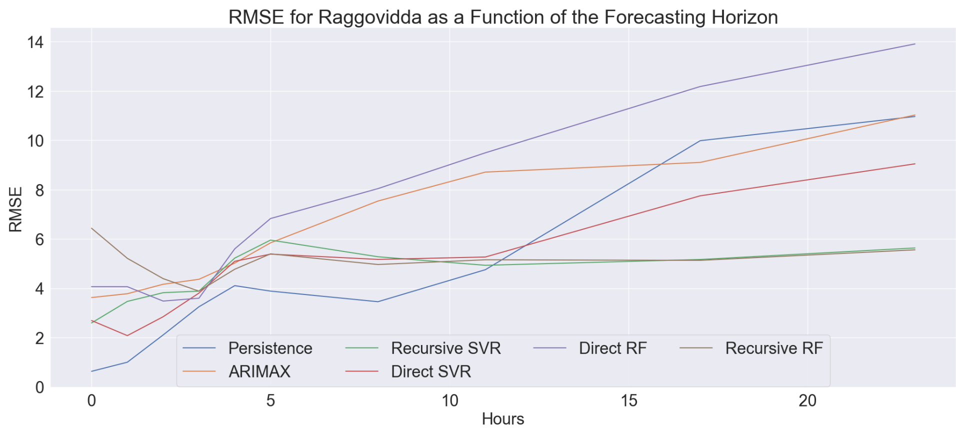


Figure 9.2: RMSE as a function of the forecasting horizon for Raggovidda

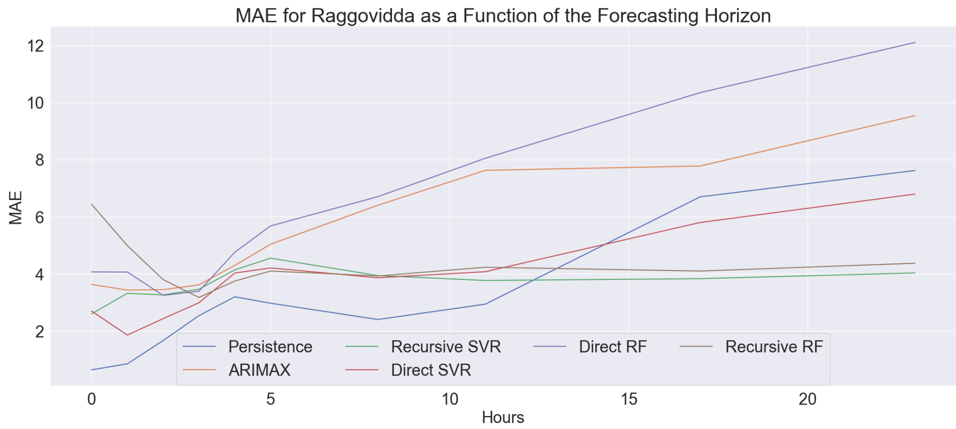


Figure 9.3: MAE as a function of the forecasting horizon for Raggovidda

9.2 Kjøllefjord Wind Farm

The results from Kjøllefjord wind farm show that the recursive SVR model yields the best results for time horizons $h > 3$ in terms of all the error measures, which can be seen in figures 9.5 and 9.6. The results for all forecasting horizons and all the different forecasting models are shown in table 9.2.

Results From Kjøllefjord Wind Farm											
Method	Error Metric	Forecasting Horizon									
		h=1	h=2	h=3	h=4	h=5	h=6	h=9	h=12	h=18	h=24
Persistence(Recursive)	MAE	1.21	4.17	6.41	7.16	6.73	6.04	5.56	6.38	8.47	8.38
	RMSE	1.98	6.38	8.63	9.22	9.18	8.53	8.19	9.20	11.59	11.34
	NRMSE	0.10	0.45	0.63	0.67	0.63	0.58	0.55	0.65	0.82	0.79
RF(Recursive)	MAE	5.82	4.62	5.01	5.21	4.64	4.47	4.74	4.49	3.99	4.09
	RMSE	5.82	4.82	5.23	5.38	4.99	4.87	5.25	5.11	4.73	4.99
	NRMSE	0.47	0.39	0.42	0.42	0.44	0.39	0.43	0.41	0.38	0.40
RF(Direct)	MAE	6.21	5.12	6.05	6.31	5.88	6.02	7.09	7.23	7.46	7.85
	RMSE	6.21	5.32	6.45	6.68	6.66	6.72	8.35	8.54	8.67	9.01
	NRMSE	0.50	0.43	0.52	0.54	0.54	0.54	0.68	0.69	0.70	0.73
SVR(recursive)	MAE	6.78	5.31	4.09	4.06	3.79	3.88	3.91	4.01	3.62	3.69
	RMSE	6.78	5.73	4.77	4.62	4.42	4.60	4.58	4.75	4.41	4.48
	NRMSE	0.55	0.46	0.38	0.37	0.36	0.37	0.37	0.38	0.36	0.36
SVR(Direct)	MAE	7.22	4.37	4.76	4.69	4.41	4.72	5.61	5.79	5.56	6.56
	RMSE	7.22	5.22	5.55	5.38	5.12	5.31	6.43	6.68	6.63	7.86
	NRMSE	0.58	0.42	0.45	0.44	0.41	0.43	0.52	0.54	0.54	0.64
ARIMAX	MAE	9.78	12.05	13.77	14.23	13.50	12.73	12.39	11.36	10.67	10.02
	RMSE	9.78	13.17	14.96	15.19	14.60	14.00	13.72	12.88	12.34	11.75
	NRMSE	0.79	1.07	1.21	1.23	1.18	1.13	1.11	1.04	0.99	0.95

Table 9.2: Results from Kjøllefjord wind farm. The results show that the persistence model is the best performing model for forecasting horizons $h < 3$. When $h > 3$ the persistence model is outperformed by both the recursive and direct implementations of the RF and SVR model.

For one-hour-ahead predictions the persistence model is the best performing model, but when $h = 2$, the persistence model is outperformed by the direct SVR and both the RF models in terms of the NRMSE. The worst performing model for Kjøllefjord wind farm is the ARIMAX model that is not competitive with the persistence model for any time horizons. In figure 9.4 it can be seen that the ARIMAX model indeed does a bad job in approximating the power output at Kjøllefjord wind farm for any time horizons. However, the persistence model is acting very poorly on this particular test dataset as well. The results in table 9.2 is the average results across all splits of the dataset, and possibly the persistence model performs better on the rest of the year. The recursive SVR model is the best at estimating the pattern of the power output, closely followed by the recursive RF model. However, it seems as if the RF underestimates the power output in the peak values, and over estimates in the minimum values.

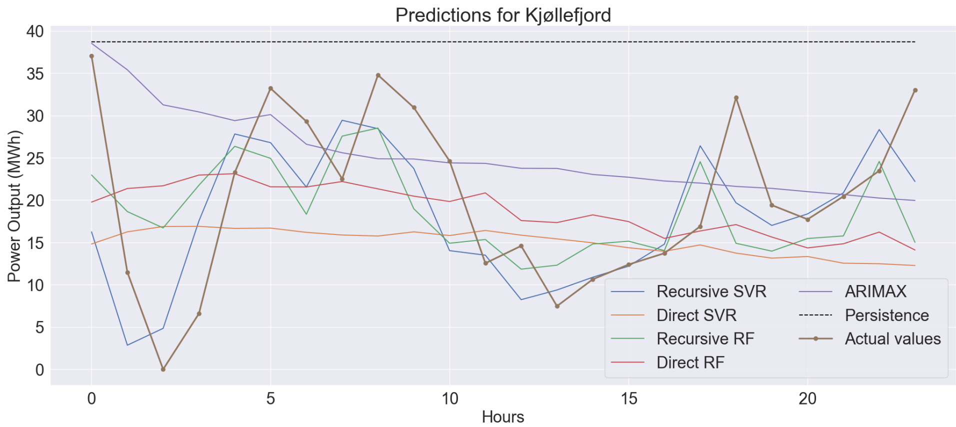


Figure 9.4: 24 hours predictions for Kjøllefjord wind farm from all the different forecasting models on one of the test datasets.

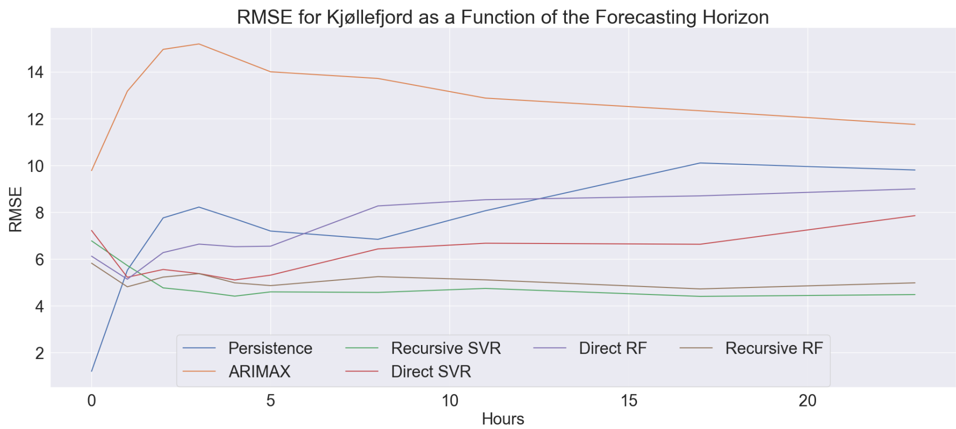


Figure 9.5: RMSE as a function of the forecasting horizon for Kjøllefjord

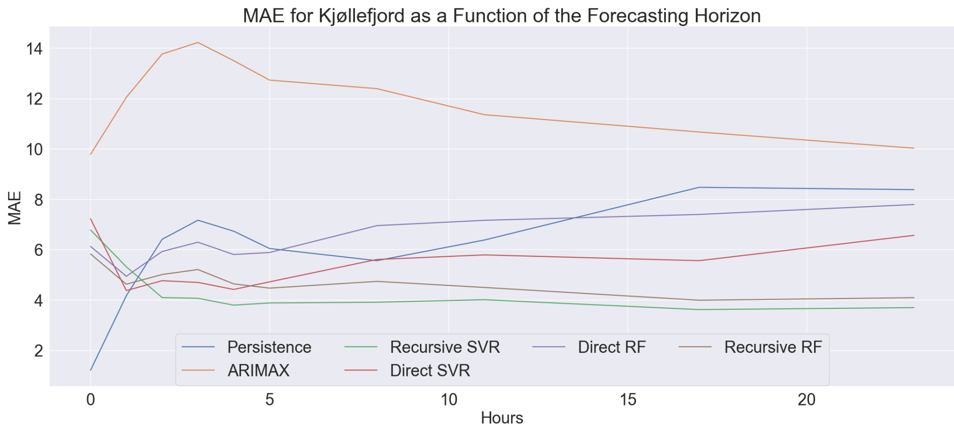


Figure 9.6: MAE as a function of the forecasting horizon for Kjøllefjord

9.3 Havøygavlen Wind Farm

At Havøygavlen wind farm the LSTM model was also tested for three of the time horizons. In this thesis, it seems as if the LSTM model is not competitive with the simple machine learning algorithms, namely the RF and the SVR models. The reason for this is most likely that the hyperparameters that were found for the LSTM model were not optimal. It was very time consuming to tune the hyperparameters of the LSTM model and several assumptions had to be made in order to limit the computational power necessary for reaching the submission deadline of the thesis.

The overall best model for Havøygavlen wind farm is the recursive SVR model, which outperforms the persistence model for all forecasting horizons as shown in figures 9.8 and 9.9. At $h > 1$ the recursive RF model also shows a good performance. The worst performing model is the ARIMAX model, closely followed by the LSTM model, which was only tested for three forecasting horizons. In figure 9.7 the 24 hours ahead predictions from the different forecasting models are shown. Similarly to the rest of the locations it is seen in figure 9.7 that the recursive implementation of the SVR and the RF best follows the pattern of the power output. The RF model struggles to predict the maximum and minimum values, and the direct implementation of the RF and SVR models seems to be estimating the overall trend of the power output.

Results From Havøygavlen Wind Farm											
Method	Error Metric	Forecasting Horizon									
		h=1	h=2	h=3	h=4	h=5	h=6	h=9	h=12	h=18	h=24
Persistence	MAE	4.78	4.45	3.56	3.27	2.84	2.64	3.36	4.64	6.46	7.46
	RMSE	4.58	4.07	3.41	4.26	4.85	5.20	6.09	8.42	11.42	13.38
	NRMSE	0.46	0.44	0.39	0.38	0.36	0.35	0.44	0.59	0.77	0.86
RF(Recursive)	MAE	5.13	3.92	4.03	3.66	3.36	3.56	3.55	3.49	3.32	3.47
	RMSE	5.13	4.16	4.23	4.06	3.83	4.08	4.04	4.02	4.15	4.40
	NRMSE	0.50	0.40	0.41	0.39	0.37	0.40	0.39	0.39	0.40	0.43
RF(Direct)	MAE	3.95	3.97	4.33	4.81	4.89	4.59	4.62	4.97	4.47	6.06
	RMSE	3.95	4.14	4.59	5.18	5.30	5.07	5.45	5.99	6.73	7.18
	NRMSE	0.38	0.40	0.45	0.50	0.51	0.49	0.53	0.58	0.65	0.70
SVR(recursive)	MAE	2.08	2.51	2.30	2.78	2.71	2.96	3.23	3.59	3.48	3.44
	RMSE	2.08	2.65	2.54	3.34	3.22	3.65	3.77	4.14	4.20	4.44
	NRMSE	0.20	0.26	0.25	0.32	0.31	0.35	0.37	0.40	0.41	0.43
SVR(Direct)	MAE	1.94	1.45	2.35	2.79	3.35	3.93	4.57	5.06	5.54	6.18
	RMSE	1.94	1.67	2.78	3.23	3.86	4.54	5.38	6.19	6.92	6.18
	NRMSE	0.19	0.16	0.27	0.31	0.37	0.44	0.52	0.60	0.67	0.73
ARIMAX	MAE	11.43	10.89	10.62	10.35	10.08	9.80	9.41	8.96	8.82	8.44
	RMSE	11.43	11.07	10.83	10.55	10.40	10.31	10.15	10.04	10.24	9.87
	NRMSE	1.11	1.07	1.05	1.02	1.01	1.00	0.98	0.97	1.00	0.96
LSTM	MAE	5.15							8.56		8.95
	RMSE	7.56							10.61		11.14
	NRMSE	0.73							1.03		1.08

Table 9.3: Results from Havøygavlen wind farm. The results show that the persistence model was outperformed by the recursive SVR model for all forecasting horizons closely followed by the recursive RF model that is better than the persistence for $h > 2$.

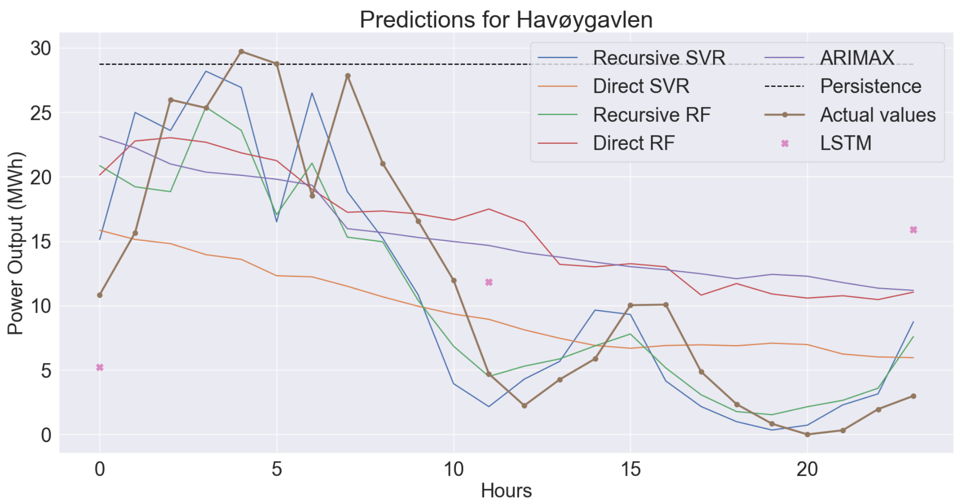


Figure 9.7: 24 hours predictions for Havøygavlen wind farm from all the different forecasting models on one of the test datasets.

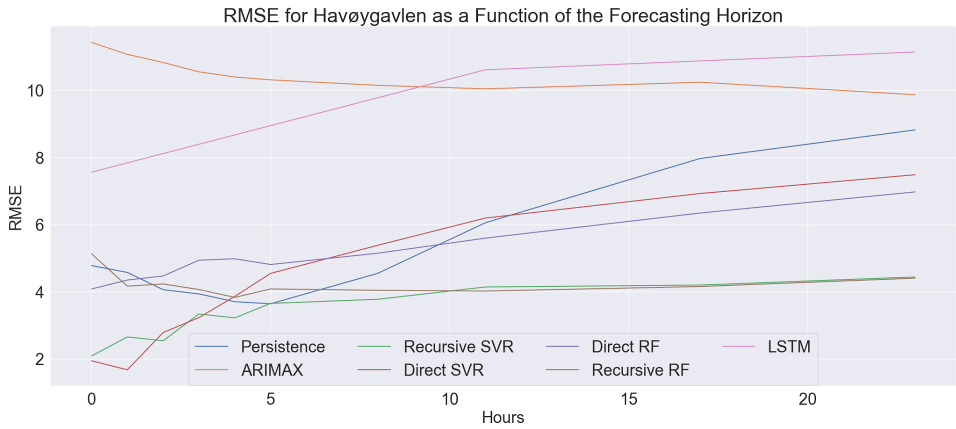


Figure 9.8: RMSE as a function of the forecasting horizon for Havøygavlen

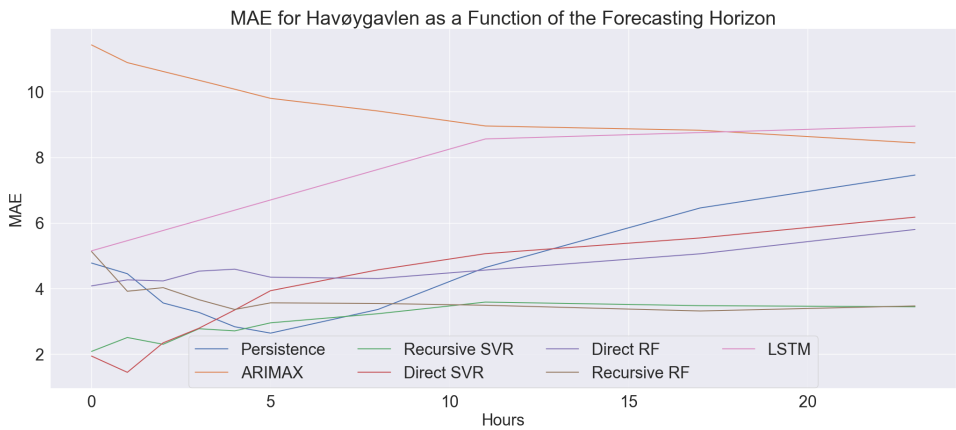


Figure 9.9: MAE as a function of the forecasting horizon for Havøygavlen

9.4 Fakken Wind Farm

At Fakken wind farm the best performing model is again the recursive SVR model. For short forecasting horizons the direct SVR also outperforms the persistence model, but the recursive SVR is better for large forecasting horizons. The recursive RF model is outperformed by the persistence model for $h < 3$, but it is better than the persistence model for large forecasting horizons. The direct RF model is better than the persistence model for large forecasting horizons, but it is worse than the recursive implementations of both the RF and SVR model. The ARIMAX model shows poor results across all timesteps. Figures 9.11 and 9.12 gives an overview of the results. The average results from all splits of the dataset for all the forecasting horizons and all the models are presented in table 9.4.

Results From Fakken Wind Farm											
Method	Error Metric	Forecasting Horizon									
		h=1	h=2	h=3	h=4	h=5	h=6	h=9	h=12	h=18	h=24
Persistence	MAE	4.35	6.05	7.02	7.08	6.97	7.61	10.56	12.51	14.00	16.36
	RMSE	4.58	5.82	6.31	6.42	6.44	7.07	8.35	10.42	14.21	17.26
	NRMSE	0.29	0.42	0.51	0.51	0.51	0.56	0.81	0.94	1.09	1.24
RF(Recursive)	MAE	8.35	7.67	7.12	7.55	7.41	8.18	7.15	6.55	6.70	6.79
	RMSE	8.35	7.97	7.50	7.91	7.78	8.76	7.81	7.31	8.02	8.37
	NRMSE	0.55	0.52	0.49	0.52	0.51	0.57	0.51	0.48	0.53	0.55
RF(direct)	MAE	5.84	5.98	5.96	7.28	8.08	8.73	9.86	11.22	12.45	14.90
	RMSE	5.84	6.44	6.32	7.97	8.79	9.44	10.82	12.32	14.49	17.48
	NRMSE	0.38	0.42	0.41	0.52	0.58	0.62	0.71	0.81	0.95	1.15
SVR(recursive)	MAE	3.94	3.33	3.16	3.30	3.17	4.22	3.84	3.76	4.80	5.23
	RMSE	3.94	3.92	3.83	3.88	3.75	5.31	4.85	4.97	6.39	6.97
	NRMSE	0.26	0.26	0.25	0.26	0.25	0.35	0.32	0.33	0.42	0.46
SVR(Direct).	MAE	3.45	2.65	2.85	4.42	5.38	6.19	6.44	7.34	9.92	11.32
	RMSE	3.45	3.10	3.32	5.48	6.45	7.22	7.62	8.42	12.20	13.73
	NRMSE	0.23	0.20	0.22	0.22	0.42	0.47	0.50	0.55	0.80	0.90
ARIMAX	MAE	14.69	15.98	15.93	15.86	15.92	15.79	16.24	16.67	16.84	16.75
	RMSE	14.69	16.21	16.36	16.50	16.54	16.44	17.10	17.64	18.89	19.22
	NRMSE	0.96	1.06	1.07	1.08	1.09	1.08	1.12	1.16	1.24	1.26

Table 9.4: Results from Fakken. The results show that the best performing model is the recursive SVR model. The direct SVR model also performs well for short forecasting horizons.

Figure 9.10 shows the 24 hours ahead predictions for Fakken wind farm. It is seen that the recursive implementations of the RF and the SVR model capture large variations in the timeseries quite well, while the ARIMAX and the direct SVR models are almost unaffected by the peak in power output at 15 hours. The direct RF model shows some adjustments to this event, but the overall performance of the model for large forecasting horizons is not comparable to the recursive implementation.

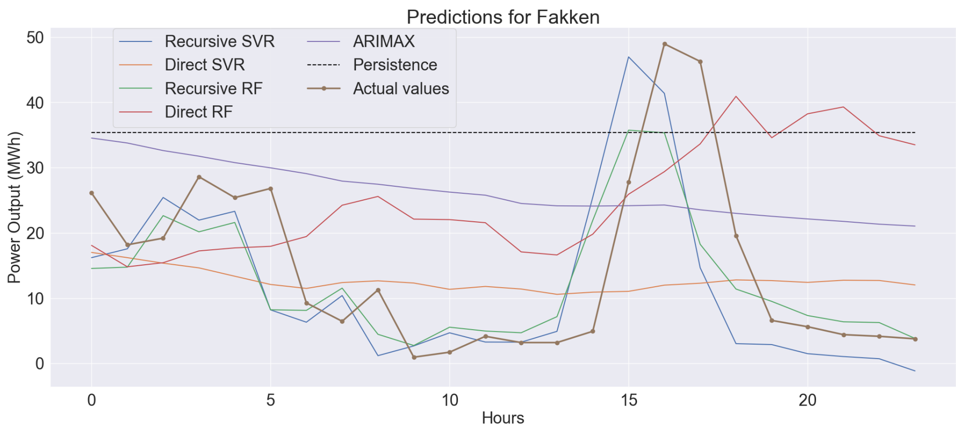


Figure 9.10: 24 hours predictions for Fakken wind farm for all forecasting models.

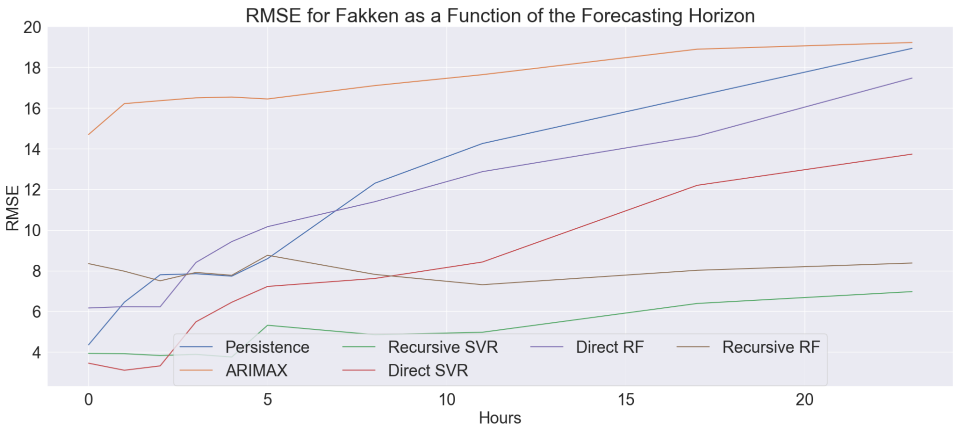


Figure 9.11: RMSE as a function of the forecasting horizon for Fakken

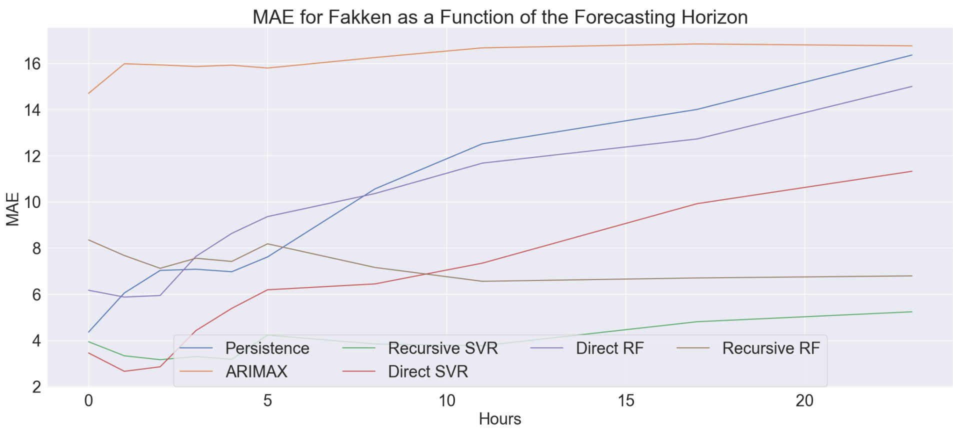


Figure 9.12: MAE as a function of the forecasting horizon for Fakken

9.5 Nygårdsfjellet Wind Farm

At Nygårdsfjellet wind farm the results show that both the recursive RF and the recursive SVR model outperform the persistence model model across all forecasting horizons in terms of all error measures. The direct SVR model also does a good job on this dataset for all forecasting horizons, but it is seen that in terms of the RMSE the persistence model is better for $h > 6$. The persistence model also outperforms the ARIMAX model for all timesteps. The observations are illustrated in figures 9.14 and 9.15. The average results from all splits of the dataset for all forecasting horizons and all models are presented in table 9.5.

The predictions from each of the models for one of the test dataset for Nygårdsfjellet wind farm are shown in figure 9.13. It can be seen that the recursive forecasting strategy is, in addition to detecting the pattern of the power output, also better at prediction the zero level in the time series. In the wind power dataset the zero values are often consecutively repeated values, because the wind turbines are shut down for a period of time, most likely due to high winds or maintenance work. Since the high winds is a determining weather factor for a turbine shut down, but the maintenance events are unaccounted for in the dataset it may be harder for some models to interpret the zero level. For example the recursive models will know from the recursive predictions that the last value is low, but the direct methods may be unaffected by the events that led to zero wind power output. This will be further discussed in chapter 10.

Results From Nygårdsfjellet Wind Farm											
Method	Error Metric	Forecasting Horizon									
		h=1	h=2	h=3	h=4	h=5	h=6	h=9	h=12	h=18	h=24
Persistence	MAE	2.35	3.51	4.44	5.01	5.63	5.67	6.81	7.01	7.02	7.69
	RMSE	1.10	1.50	2.78	4.23	5.16	5.26	5.14	4.62	4.00	4.94
	NRMSE	0.21	0.33	0.43	0.48	0.54	0.55	0.65	0.67	0.68	0.81
RF(Recursive)	MAE	2.18	3.16	3.38	3.05	2.96	3.10	2.83	2.58	2.59	2.62
	RMSE	2.18	3.37	3.60	3.31	3.22	3.51	3.23	2.99	3.07	3.31
	NRMSE	0.20	0.30	0.32	0.30	0.29	0.32	0.29	0.27	0.28	0.30
RF(Direct)	MAE	1.38	1.89	1.71	1.95	2.61	3.11	3.97	6.15	9.30	10.46
	RMSE	1.38	2.00	1.88	2.27	3.34	3.91	4.98	7.85	11.16	11.99
	NRMSE	0.12	0.18	0.17	0.21	0.30	0.35	0.45	0.70	1.00	1.08
SVR(recursive)	MAE	1.18	1.66	2.11	2.56	2.30	2.53	2.40	2.03	1.74	1.91
	RMSE	1.18	1.75	2.52	3.04	2.82	3.13	2.97	2.66	2.41	2.80
	NRMSE	0.11	0.16	0.23	0.27	0.25	0.28	0.27	0.24	0.22	0.25
SVR(Direct)	MAE	0.89	0.84	1.11	1.62	2.78	3.67	4.27.	4.62	5.89	6.82
	RMSE	0.89	0.99	1.30	2.14	3.87	4.84	5.63	6.07	7.39	8.46
	NRMSE	0.08	0.09	0.12	0.19	0.35	0.43	0.51	0.55	0.66	0.76
ARIMAX	MAE	6.81	7.25	7.39	6.94	6.46	6.14	7.27	8.08	8.88	9.14
	RMSE	6.81	7.33	7.55	7.22	7.24	7.22	8.28	9.07	9.89	10.08
	NRMSE	0.61	0.66	0.68	0.65	0.65	0.65	0.74	0.81	0.89	0.911

Table 9.5: Results from Nygårdsfjellet. The results show that the recursive RF and SVR outperforms the persistence model for all forecasting horizons in terms of all error measures. The worst performing model is the ARIMAX model.

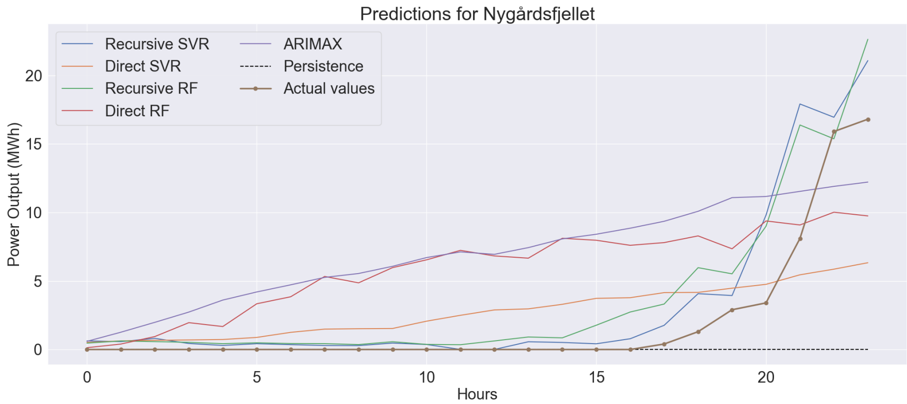


Figure 9.13: 24 hours predictions for Nygårdsfjellet wind farm from all the different forecasting models.

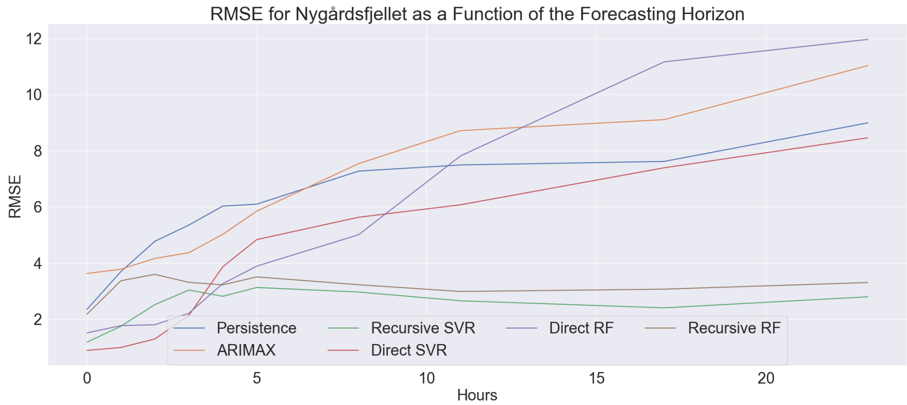


Figure 9.14: RMSE as a function of the forecasting horizon for Nygårdsfjellet

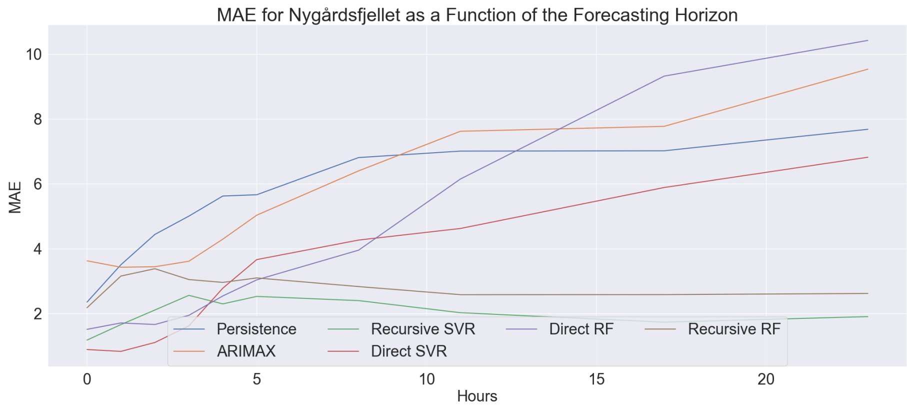


Figure 9.15: MAE as a function of the forecasting horizon for Nygårdsfjellet

9.6 Overall Results

When considering the overall performance of the models across all the five different wind farm locations the NRMSE is used for comparison. The comparison of the modes are done according to the overall NRMSE across all forecasting horizons, and the performance for different forecasting horizons. In figure 9.16 the average NRMSE across all the different wind farm locations as a function of time is shown.

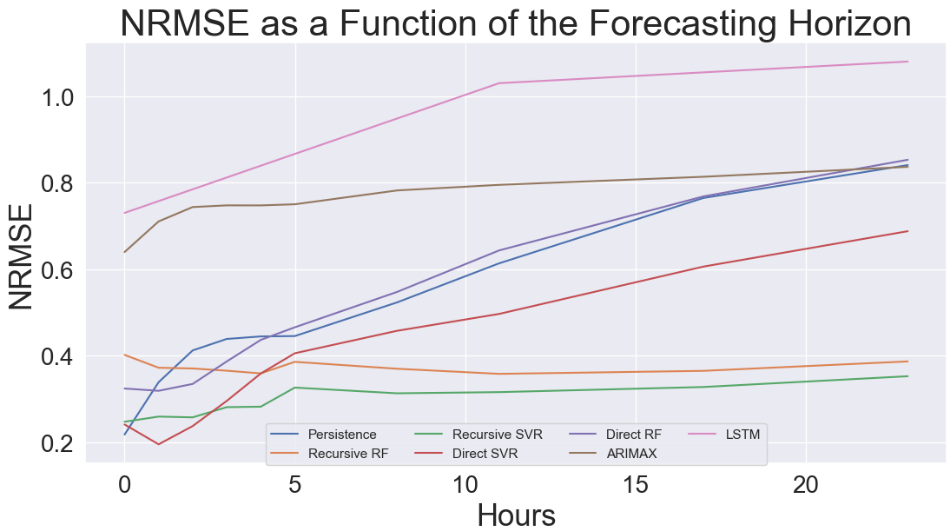


Figure 9.16: Average NRMSE as a function of forecasting horizons

As seen in sections 9.1-9.5 in figures 9.1-9.13 it is evident that the recursive implementations of the models does a better job in detecting the pattern of the power output from all wind farms while the direct approaches usually has a good performance on short timescales, but their accuracy decreases as the forecasting horizon increases, as expected. This is confirmed by looking at figure 9.16. It is seen that the recursive forecasting approaches, quite unexpectedly, have a relatively constant error as the forecasting horizon increases, whereas the error of the direct approaches increases as the forecasting horizon increases. The ARIMAX model has a more constant error for long forecasting horizons, but it is seen that the error is slightly lower for shorter timescales than for longer timescales. For one hour predictions the persistence model is the best performing model, but the error of the persistence model rapidly increases along with the forecasting horizon. The NRMSE of the LSTM model is also included in the plot, but the error is only recorded for Havøygavlen, so the overall error is not comparable with the other averaged errors, particularly since Havøygavlen is thought to be the most difficult location to forecast based on the results of the persistence model. However, it is seen that also for the

LSTM model the error increases with the forecasting horizon, but it seems as if he growth rate of the error is lower than for the persistence model.

In figure 9.17 it is seen that it is the recursive SVR model that has the overall best performance across all datasets, closely followed by the recursive RF model. The worst performing model is the ARIMAX model, and the second worst is the persistence model.

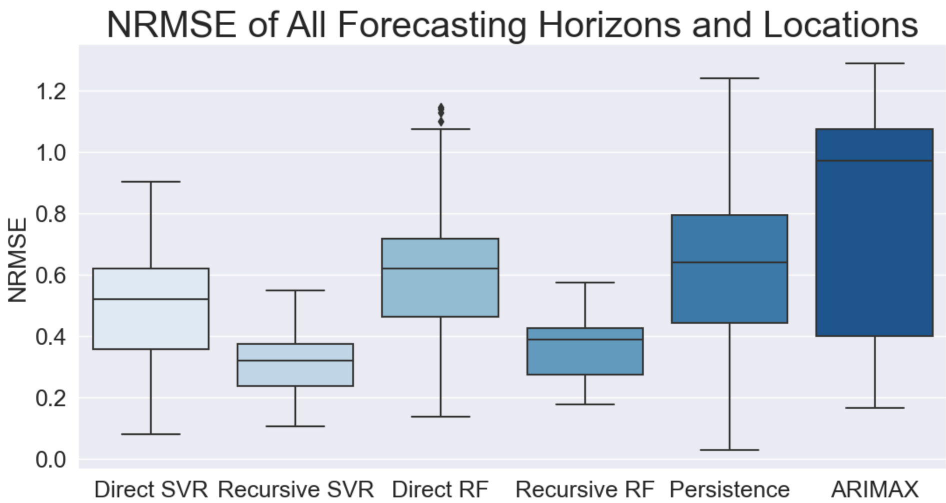
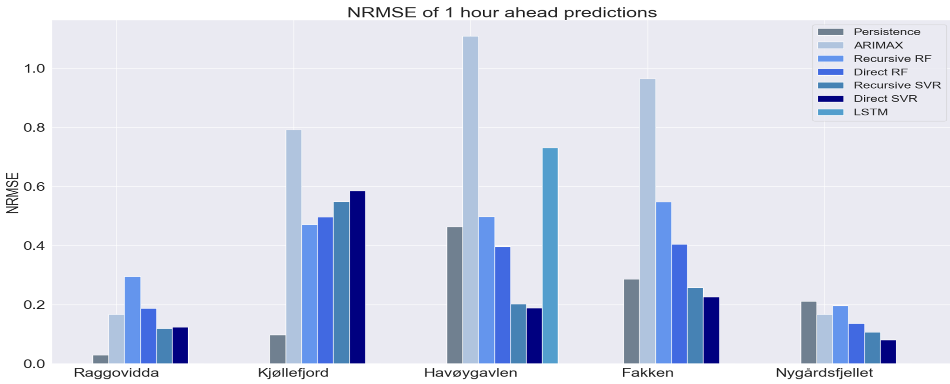
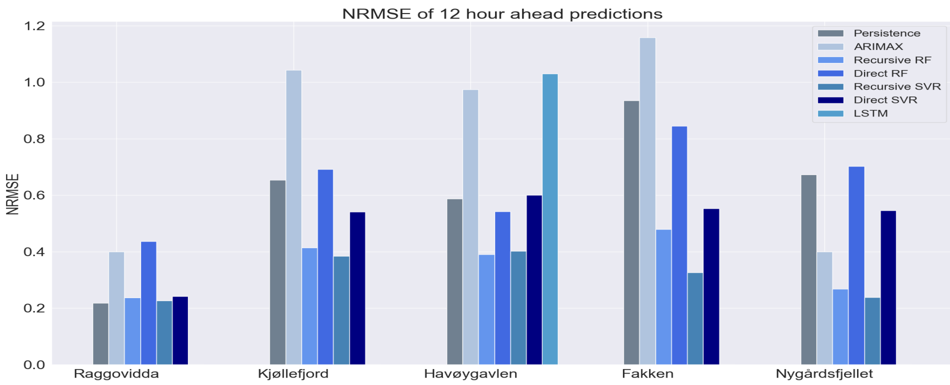


Figure 9.17: NRMSE of all models across all forecasting horizons and locations shown as a box plot. The plot shows that the ARIMAX model is the worst performing model across all datasets with the largest spread of the NRMSE and the highest median NRMSE. The persistence model is the second worst. It has a much lower median NRMSE than the ARIMAX, but with larger spread than the rest of the models. The best performing model is the recursive SVR model that has the lowest spread and median NRMSE.

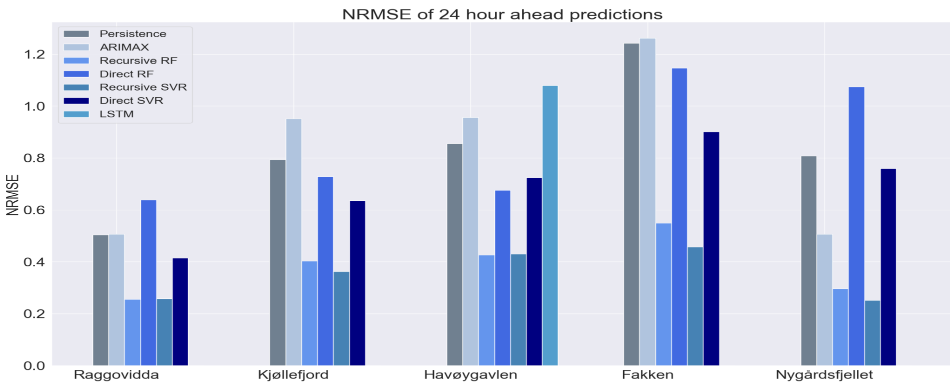
Another observation that is made through section 9.1 - 9.5 is the there is a big difference across locations in which models that performs well on different datasets. This is illustrated in figure 9.18.



(a) The NRMSE of different locations for one-hour-ahead predictions



(b) The NRMSE of different locations for 12 hours ahead predictions



(c) The NRMSE of different locations for 24 hours ahead predictions

Figure 9.18: The NRMSE for all wind farm locations and different forecasting horizons.

It is seen in figure 9.18 that the ARIMAX model, for example, shows a good performance for one-hour-ahead predictions at Raggovidda and Nygårdsfjellet, but at Havøygavlen it is by far the worst performing model. For one-hour-ahead predictions at Nygårdsfjellet wind farm the direct SVR is the best performing model, while for Kjøllefjord wind farm the direct SVR model is the second worst performing model.

To summarize, it is noted in this chapter that the best model for wind power predictions, as implemented in this thesis, is the recursive SVR model. The worst model was the ARIMAX model. The persistence model often outperformed the more complicated models for short forecasting horizons, but when the forecasting horizon increased the error rapidly increased. The recursive RF model also showed promising results for large forecasting horizons, and the direct SVR and RF models showed good results for short timescales, but were outperformed by the recursive implementations for longer forecasting horizons. It was also seen that the models' performance is highly dependent on the forecasting horizons they are evaluated by and the locations the predictions are made for.