Ship power prediction

Martin Alexandersson

June 16, 2021

ABSTRACT

Full-scale measurement data of ship's propulsion power is analyzed. Various statistics and machine learning models are fitted to predict the propulsion power based on features about the ship operational condition and data about the encountered metocean environments.

DATA

The data used in this study is described in Tab.1.

Class	Category	Description	Attributes
Input features	Operation	Ship speed through water	V [knots]
		Ship draft	T_{fwd} & T_{aft} [m]
		Heading	HDG [°]
	Metocean	Significant wave height	H_s [m]
		Mean wave period	T_z [s]
		Mean wave direction	D_{wave} [°]
		Wind speed	U_{wind} & V_{wind} [m/s]
Output target	Operation	Propulsion power	P_b [kW]

Figure 1: Spring-mass-damper system

Exploratory data analysis

The ship speed V, ship draughts T_{aft} and T_{fwd} were all negative in the raw data file. This was imidiatelly corrected, to be more in line with what would be expected from a more general sign convention. The data seems to have been collected in time cronological order, giving a time series of data. For a time series, measurements close to each other in time have a high correlation, as they are experiencing similar envinronmental conditions etc. This is confirmed by looking at the autocorrelation plot in Fig.2. Dead reckoning (using ship speed and heading) has been used to atempt to describe motion of the ship as seen in Fig.3. The positions are given in an unknown longintude and latitude scale, as the time step between measurements is unknown. The speed of the ship is also indicated as a color gradient in this figure.

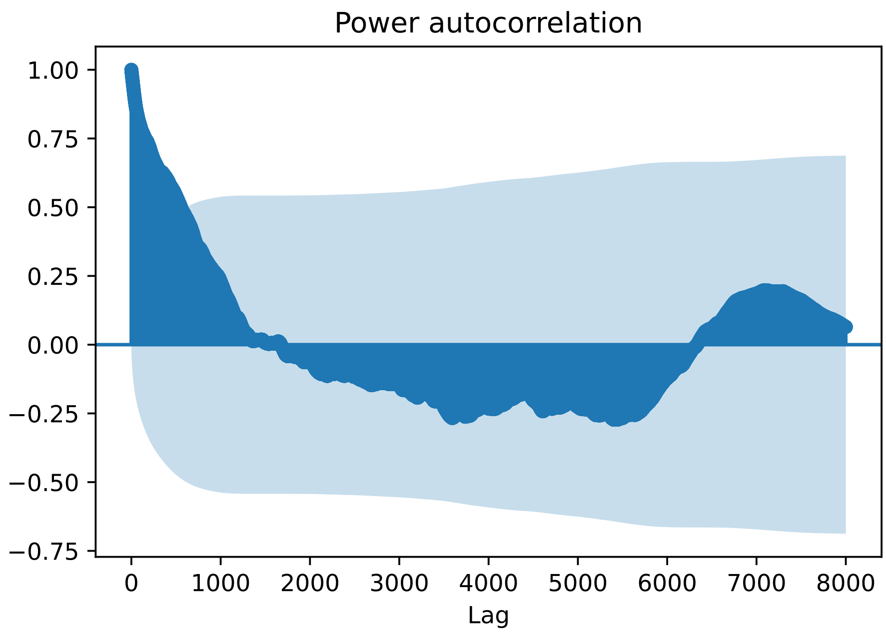

Figure 2: Autocorrelation plot of the Power data

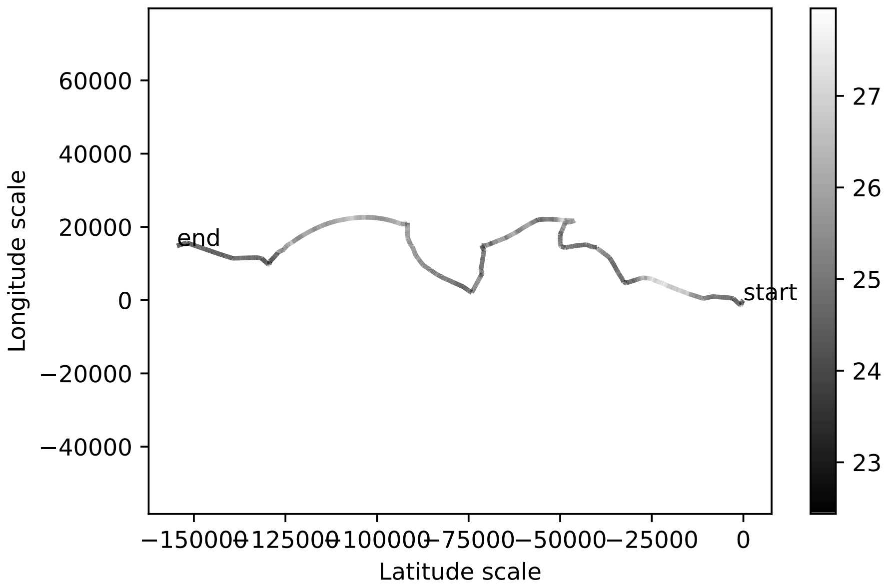

Figure 3: Dead reckoning of the position of the ship

Fig.4 shows a heat map of the absolute linear correlation coefficient between all of the features in the raw data. It can be seen that T_{aft} and T_{fwd} have the highest correlation with the $Power$. It can also be seen that the correlation between T_{aft} and T_{fwd} is also very high (approximately 1) implying a very high multicollinearity which is generally something that should be avoided in regression problems. These two features are instead replaced with the two features: mean draught T and $trim$. The correspondig heat map with the new features is shown in Fig.5. The mean draught T now seems to be a very important feature in this regression as it has the highest linear correlation with the $Power$. This can also be seen in Fig.6, where the $Power$ has been plotted together with the corresponding negative draught.

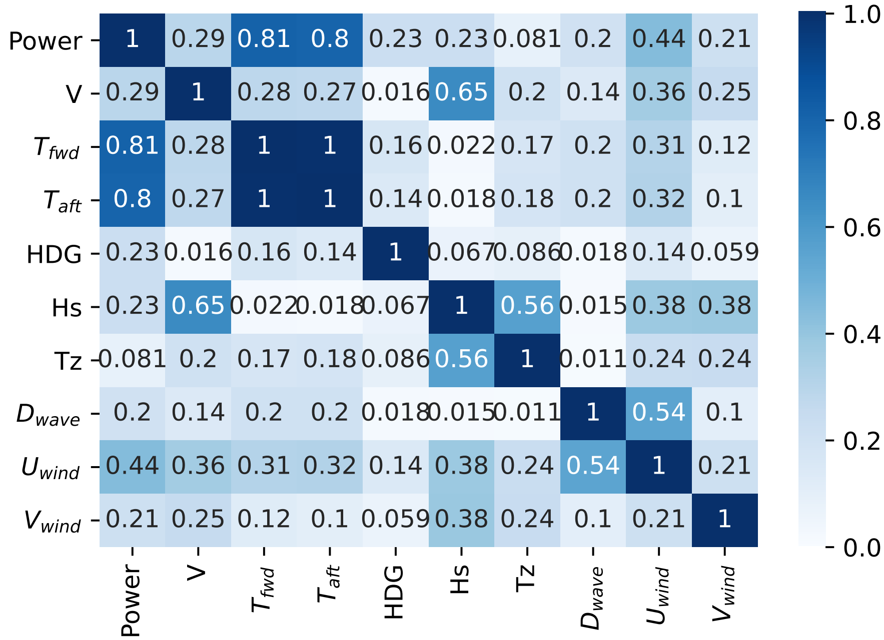

Figure 4: Heat map showing absolute value of correlation coefficient between features in raw data

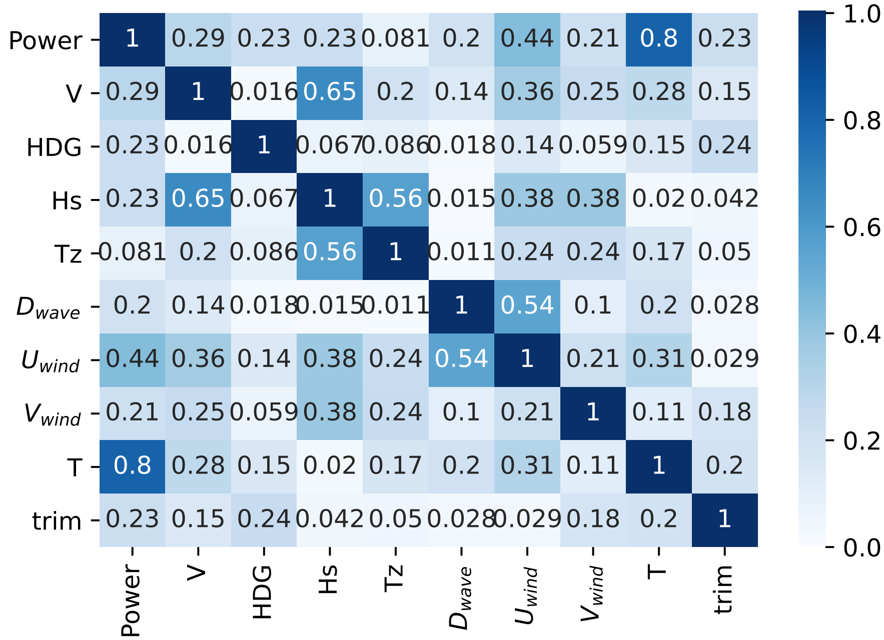

Figure 5: Heat map showing absolute value of correlation coefficient between features in transformed data

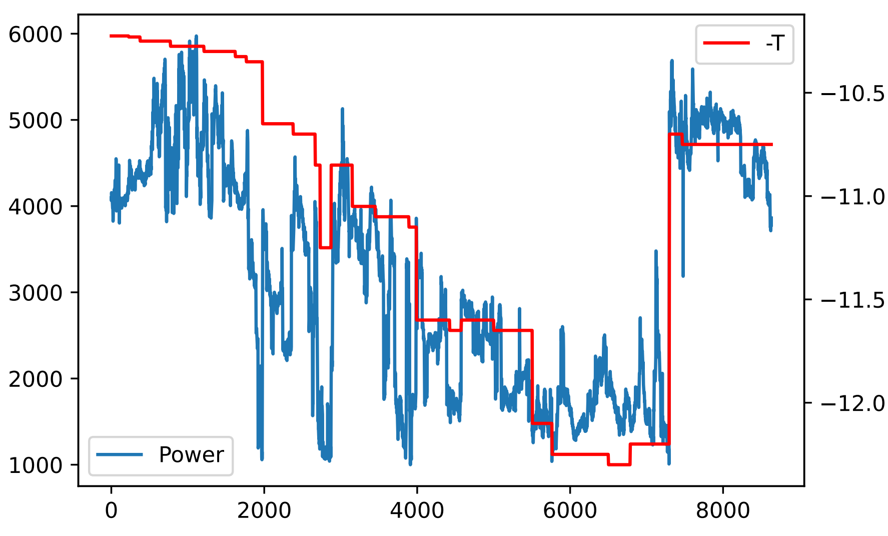

Figure 6: The power is highly correlated with the draught

Extend data

The data has also been extended by adding some additional features, describing the wind speed and wave direction in a ship fixed coordinate system:u_{wind}, v_{wind}, $wave_{direction}$. The new features have very low linear correlation with *Power* as seen in Fig.7. They are however kept in the data, to let the regression model decide if they are useable or not.

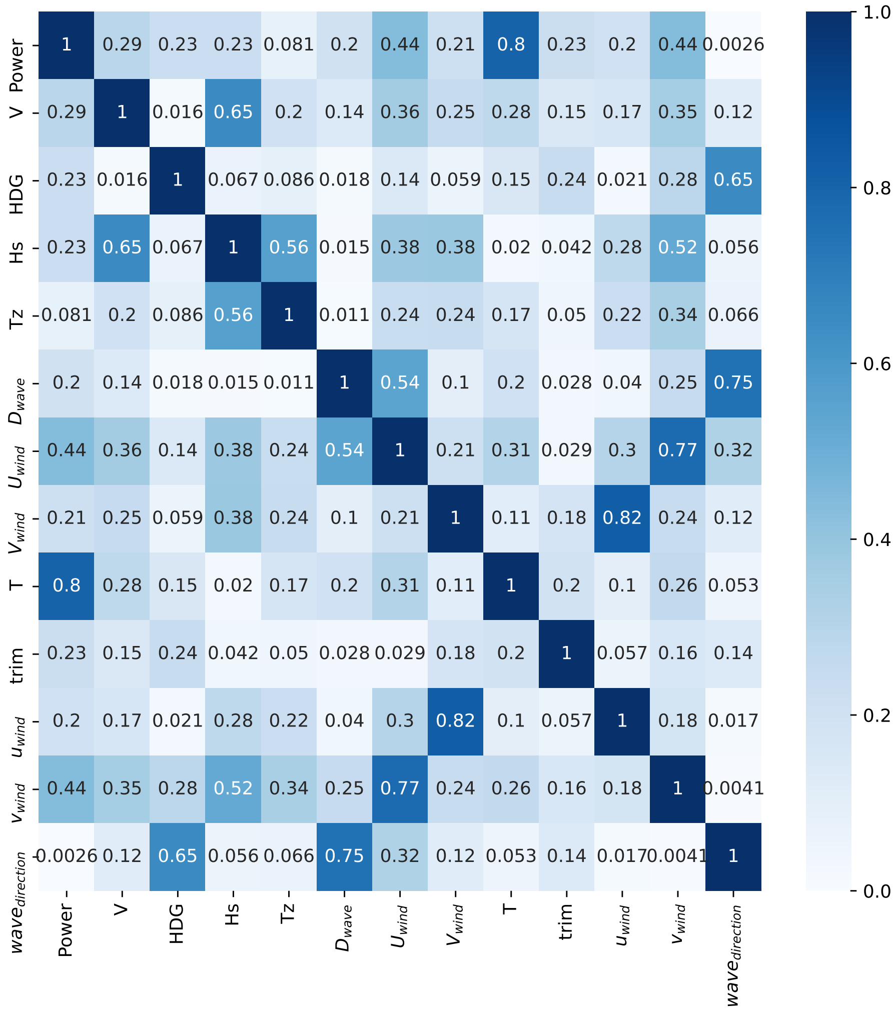

Figure 7: Heat map of data with additional features.

REGRESSION

The ship power data is split into a train and a test part. In the time series the samples are not independent, which means that splitting the data into a train and test set should not be done randomly! With a time series, training should be done on historic data to predict the future. This means that the training set must happen in time prior to the test set. Fig.8 shows how the train-test-split should been done for the present data, where the last 20% of the data is used for testing. In Fig.9 this train/test approach (time split) has been compared with the corresponding random train/test split (random split). A decision tree model has also been trained and tested to show that the random split overestimates the accuracy by far. To ensure that this is not just a coincidence for a 20% test, the test size has also been varied between 20% and 70% as seen in Fig.10, where the random split gets very low accuracy for the entire range, while the random split gets very good accuracy for all test sizes. Due to the same reasoning, cross-validation on a rolling basis is used for parameter tuning as seen in Fig.11.

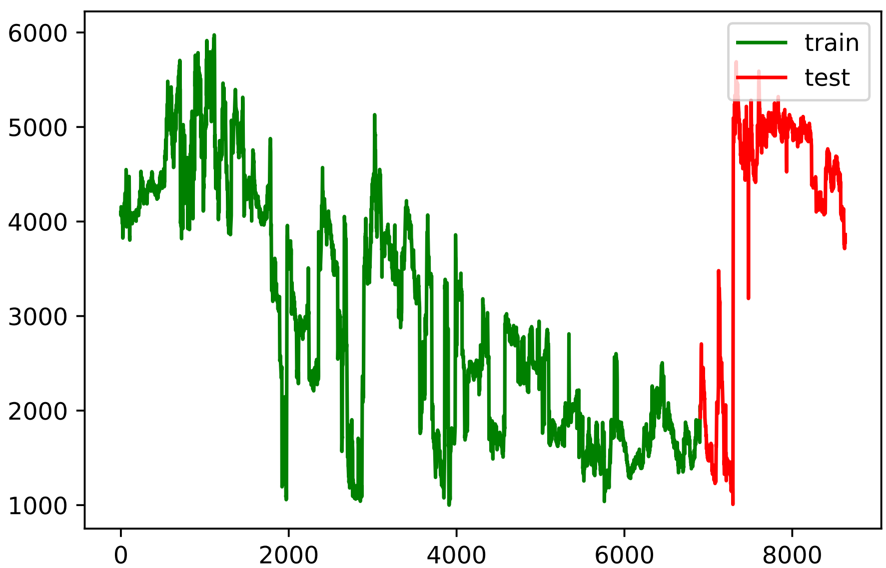

Figure 8: Data is split into a train and test part

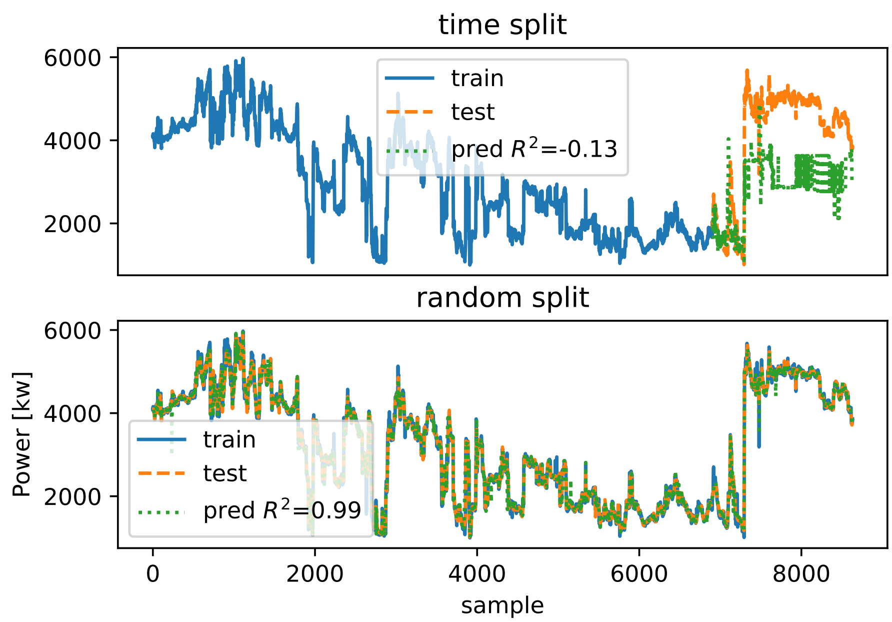

Figure 9: Comparison of using random or rolling base test set with a decision tree model

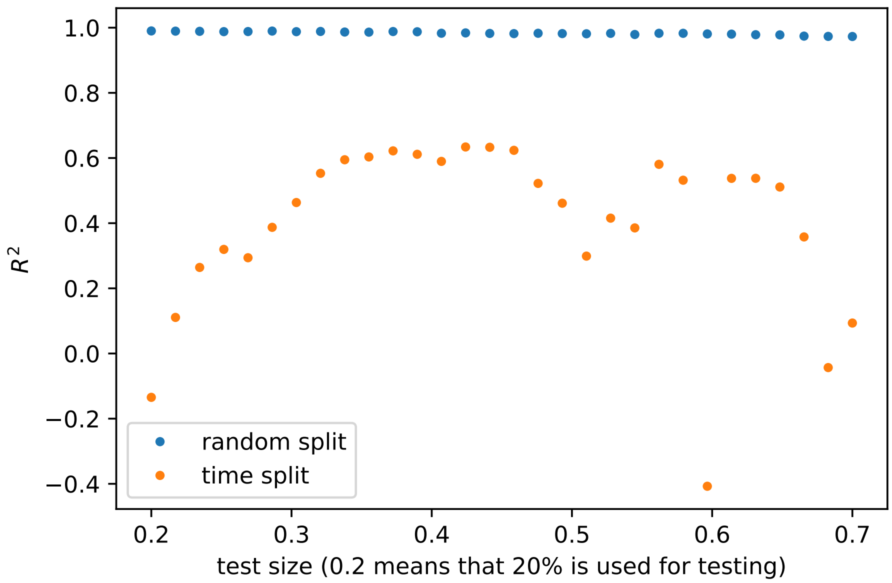

Figure 10: Accuracy with a decision tree model for varying test sizes, using random test or rolling base test

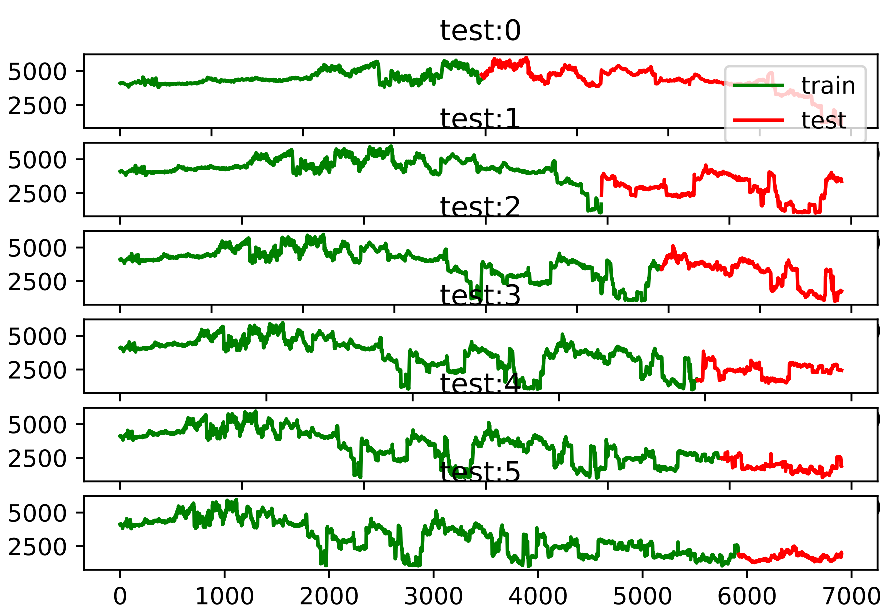

Figure 11: Cross-validation on a rolling basis

VALIDATION

The accuracy of the fitted models have been estimated using two test data sets, containing 20% and 30% of the data. The results from this evaluation in seen in Tab.1 and Fig.12. The actual predictions on the test dataset is also plotted in Fig.13. The Support Vector Regressor seems to be the model that performs the best. However, none of the models seem to perform very well on this dataset and the accuracy changes very much with the test size, which means that the ranking between the models is very unreliable. So one might as well use the simple polynomial regression. The explicit formula from this regression is shown in Eq.1. It can be seen from this expression that the *Power* depends on the ship draught T and one of the wind components U_{wind}. There is however also a very high interception term, which means that most of the *Power* is not explained by this model, if we assume that *Power* is zero when the ship is at rest in the harbour.

Table 1: Models evaluated with the test sets

model name	test size	r2 score	mean absolute error
	0.2	0.64	687.07
XGBoost	0.2	0.27	976.96
lasso	0.2	0.29	999.13
polynomial	0.2	0.3	997.36
ridge	0.2	0.3	997.54
SVR	0.3	0.83	510.07
XGBoost	0.3	0.55	856.03
lasso	0.3	0.45	1063.0
polynomial	0.3	0.45	1060.51
ridge	0.3	0.45	1061.05

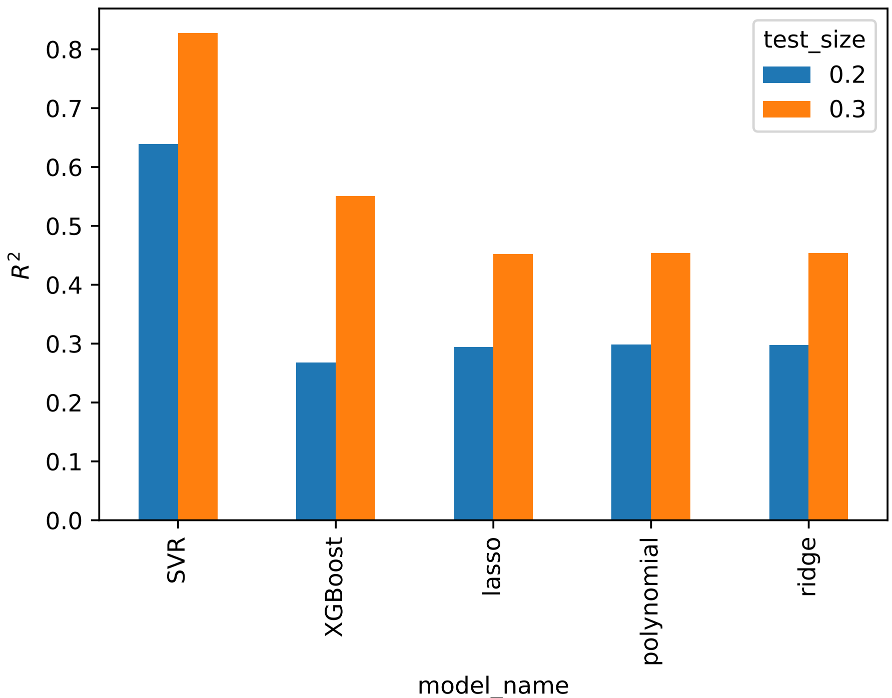

Figure 12: Evaluation of the fitted models accuracy

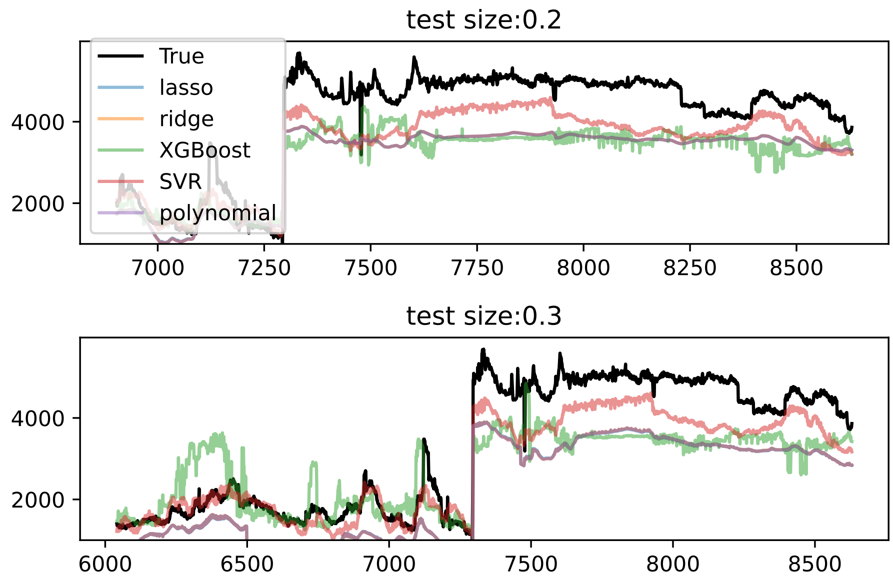

Figure 13: Predictions on the test data sets

$$y = -944.923049T + 244.319815U_{wind} + 2964.727832 \quad (1)$$

CONCLUSIONS

This dataset is a time series with high autocorrelation, so that the samples cannot be assumed as independent. This means that randomly splitting the data into a train and test dataset will overpredict the accuracy by far. With a time series, training should be done on historic data to predict the future. This means that the training set must happen in time prior to the test set, which has been done in this report for the test evaluation as well as in the cross validation for parameter tuning. The Support Vector Regressor seems to be the model that performs the best. However, none of the models seem to perform very well on this dataset and the accuracy changes very much with the test size, which means that the ranking between the models is very unreliable. The dataset should be further examined and extended with additional features to enable for a better model fit.