Forecasting Wind Speed using Support Vector Regression and Feature Selection

Nicolene Botha
Modelling and Digital Science (MDS)
Council for Scientific and Industrial Research (CSIR)
Pretoria, South Africa
nbotha@csir.co.za

Christiaan van der Walt
Modelling and Digital Science (MDS)
Council for Scientific and Industrial Research (CSIR)
Pretoria, South Africa
email address

***Abstract*—Since the wind power generated by a wind farm is entirely dependent on meteorological conditions (such as wind speed, wind direction, humidity etc.), accurately forecasting wind speed based on these conditions over a 1 to 24 hour time horizon is crucial to predict the potential short term energy supply of a wind farm. These short term predictions in turn are crucial in assisting with wind farm planning so that the required base load provided to the electricity grid is always guaranteed (even in the presence of highly variable wind power outputs). In this work we show that the relative prediction performance of a short-term Support Vector Regression (SVR) wind forecasting system can be improved by up to 11.12% by systematically selecting and combining relevant input features that influence short term wind speed. We illustrate our results on meteorological data collected in Alexander Bay, South Africa over a three year period from 2011-2013.**

***Keywords*—Wind Speed Forecasting, Support Vector Regression, Feature Selection**

I. Introduction

Generating renewable energy is significantly cleaner than burning fossil fuels and with the overall abundance of wind, renewable wind energy generation will play a large part in renewable energy portfolios. The prediction of wind speed plays a key role in wind power generation, with a large correlation between wind speed, wind direction and wind turbine power generation.

Support Vector Regression (SVR) has successfully been used for wind speed prediction in the past [1] and the use of various optimisation algorithms for hyper parameter selection followed [2–4].

Hybrid methods employed, where some evolutionary methods were combined with SVR, with the evolutionary method determining the hyper parameters used in the SVR. Santanmaria-Bonfil *et. al.* [2] used a genetic algorithm for parameter tuning while SanchoSalcedo-Sanz *et. al.* [3] compared the performance of two evolutionary methods, evolutionary programming and particle swarm optimisation.

Wang *et. al.* [3] compared a genetic algorithm, particle swarm optimisation and cuckoo optimisation algorithm (COA) and found that the hybrid COA-SVR method outperforms the other methods. In a previous study of wind speed forecasting from 1 to 24 hours ahead it was found that Support Vector Regression with an RBF (Gaussian) kernel outperforms other regression methods [5].

In that study a comparison was done between Ordinary Least Squares, Bayesian Ridge Regression and Support Vector Regression with an RBF kernel using a moving time window, with a persistence forecast functioning as the benchmark. It has also been shown in various studies that SVR can outperform over other types of algorithms such as variations on artificial neural networks (ANN) [6, 7].

Due to the relative good prediction accuracy of SVRs, they are also utilised in conjunction with other methods to create hybrid algorithms [8–11]. Lui *et. al.* [8] used a wavelet transform to decompose the wind speed into two components to function as inputs in a SVM after mathematical manipulation and employing Granger causality, with a genetic algorithm used to optimise the hyper parameters. Chen *et. al.* [9] adopted SVR with a Kalman filter that updates the states under stochastic uncertainty. Hu *et. al.* [10] decomposed the wind speed data into a number of independent intrinsic mode functions (IMFs) and one residual series using ensemble empirical mode decomposition. The highest frequency band (IMF1) and the residual series are discarded and after re-scaling and normalisation the remaining bands are used as the input for the SVM. Wang *et. al.* [11] created two hybrid models by integrating SVR with seasonal index adjustment and Elman recurrent neural network methods. The SVR is used to detect and eliminate outliers during preprocessing and the Kruskal-Wallis test is done to confirm that the dataset's distribution is similar. Next the seasonal index adjustment method is used to extract the seasonal effects and the Elman recurrent neural network was used to predict the trends. The seasonal index is used to adjust the predicted trends to obtain the forecasted wind speed.

Combining a SVR with feature tailoring was explored by Liu *et. al.* [12] for forecasting wind power ramps. An orthogonal test was used to determine which additional features, such as wind speed, wind direction, temperature, relative humidity and pressure, would be the optimal input for a SVR. Niu *et. al.* [13] used ant colony optimisation to optimise the feature selection mechanism for forecasting the power load. Short term, 1 hour ahead, wind speed forecasting was done by Liu *et. al.* [14], using a Support Vector Machine and compared the results with that of an ANN. A principal component analysis was employed to calculate the contribution rates of various additional features to the prediction. The hyper parameters were optimised using particle swarm optimisation. Four models were created where different features were incorporated and it was found that a combination of the air temperature and air pressure has the largest influence on the wind speed prediction accuracy 1 hour ahead from the features that were investigated. It was found in a previous study [5] that by adding additional features such as wind speed acceleration, change in wind speed acceleration and the wind speed at other heights that the prediction accuracy can be improved.

While feature selection has been explored [12–14], this paper will explore the influence of feature selection on a 1 to 24 hour ahead forecast using an extensive list of atmospheric features. In this paper we fully explore the effect of feature selection on the forecasting accuracy of a 1 to 24 hour ahead wind speed prediction at a 60 m hub height, using Support Vector Regression. Atmospheric data from Alexander Bay, South Africa [15], including wind speed at the same location but at five metmast hub heights, wind speed acceleration, change in wind speed acceleration, wind direction, air pressure, air temperature and relative humidity, are combined and incorporated in several models.

II. Support Vector Regression

Support Vector Regression (SVR) is a regression algorithm where the cost function ignores any training data inside a tunable margin (determined by the hyper parameter epsilon) around the model prediction, [16]. The margin thus controls the bias-variance trade-off of the SVR. The RBF kernel creates a new feature vector by mapping training examples to the Euclidian norm of the example and some chosen landmark in the data set with a bias function determining an offset. The SVR forecast is benchmarked against a persistence forecast. A moving time window is used in the SVR where, for each prediction step forward in time, the training data window is also shifted. The initial training of the SVR algorithm is done using the training data set, with the predicting done on the validation set. The training data set consists of the past data of the parameter to be predicted, additional features such as the wind speed at different hub heights, wind velocity and air temperature can be added to provide more information to the algorithm for a better prediction. Adding different features to the training data to improve the forecast accuracy is called feature selection or feature tailoring.

A. Theory

The non-linear epsilon-insensitive SVR [16] attempts to fit a function,

$$f(\overline{x}) = \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)k(\overline{x}_i, \overline{x}) + b, \tag{1}$$

through the training samples, where all training data points are within epsilon (ϵ) from the function. ϵ is the hyperparameter that represents the band where there would be no penalty to the function. $k(\overline{x}_i, \overline{x})$ is the kernel function, in this case the RBF kernel, defined by

$$k(\overline{x}_i, \overline{x}) = exp(-\gamma||\overline{x}_i - \overline{x}||^2). \tag{2}$$

α_i and α_i^* are the Lagrange multipliers that solve the objective function

$$\text{Maximize} = \begin{cases} -1/2\sum_{i,j=1}^{l}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)k(\overline{x}_i, \overline{x}) \\ \epsilon\sum_{i=1}^{l}(\alpha_i - \alpha_i^*) + \sum_{i,j=1}^{l}(\alpha_i - \alpha_i^*), \end{cases}$$

$$\text{subject to} \sum_{i,j=1}^{l}(\alpha_i - \alpha_i^*) = 0 \text{ and } \alpha_i, \alpha_i^* \in [0, C]. \tag{3}$$

l is the number of support vectors (training points with non-zero Lagrange multipliers), and b is the bias that can be solved by making use of the Karush-Kuhn-Tucker conditions

$$\begin{aligned} \max\{-\epsilon + \gamma_i - \langle\omega, \overline{x}_i\rangle | \alpha_i < C \text{ or } \alpha_i^* > 0\} \leqslant b \leqslant \\ \min\{-\epsilon + \gamma_i - \langle\omega, \overline{x}_i\rangle | \alpha_i > 0 \text{ or } \alpha_i^* < C\} \end{aligned} \tag{4}$$

In this formulation, the hyper parameter C determines the penalty assigned to target values outside of the epsilon-band and thus controls the degree of regularisation, while the hyper parameter gamma (γ) is inversely proportional to the width of the RBF kernel which is placed over each support vector.

TABLE I
Features with identifier and optimised hyper parameter for SVR.

Feature set	Features
A	Wind speed at 60 m.
B	Wind speed at 60 m and wind acceleration (delta).
C	Wind speed at 60 m, wind acceleration and change in acceleration (double delta).
D	Wind speed and wind direction at 60 m.
E	Wind speed at 60 m and air temperature.
F	Wind speed at 60 m and barometric pressure.
G	Wind speed at 60 m and relative humidity.
H	Wind speed at all heights (10, 20, 40, 60 and 62 m).
I	Wind speed at 60 m, air temperature and barometric pressure.
J	All heights, air temperature and barometric pressure.
K	All heights, air temperature, barometric pressure and relative humidity.
L	All heights, delta, air temperature and barometric pressure.
M	All heights, delta, double delta, wind direction, air temperature, barometric pressure and relative humidity.

B. Hyper parameter selection

These hyper parameters (ϵ, γ and C) used in the SVR are selected through a rough grid search using the validation set prediction, varying the hyper parameters to optimise the root mean square error (RMSE). The RMSE is calculated by

$$\text{RMSE} = \sqrt{\frac{1}{m}\sum_{i}^{m}(\overline{x}_i - x_i)^2}. \tag{5}$$

The ϵ, γ and C hyper parameter search grids were [1e^{-2}, 1e^{-1}, 1e^{0}, 1e^{1}, 1e^{2}], [1e^{-5}, 1e^{-4}, 1e^{-3}, 1e^{-2}, 1e^{-1}, 1e^{0}, 1e^{1}, 1e^{2}] and [1e^{-2}, 1e^{-1}, 1e^{0}, 1e^{1}, 1e^{2}, 1e^{3}, 1e^{4}] respectively. This process is repeated for each model with different training features. The final prediction is done on the testing data set. The training and validation data sets are combined to form a new training set to be used with the optimised hyper parameters. The hyper parameter selection is done using the validation set to lessen the chances of the SVR overfitting on the testing data set. The designated feature set identifier for the different features that are combined in the training data set are shown in Table I.

A prediction's (Feat. set Y av. RMSE) average improvement on another (Feat. set Z av. RMSE) is calculated by taking the difference between the two forecast average RMSES as a fraction of the one,

$$\frac{\text{Feat. set Z av. RMSE - Feat. set Y av. RMSE}}{\text{Feat. set Z av. RMSE}} \text{ X } 100 = x\%.$$

III. Results and Discussion

The RMSE of a SVR model with feature set A is compared to a persistence forecast for the purposes of benchmarking. The persistence forecast and SVR predicts the wind speed at a 60 m hub height, as shown in Figure 1. The persistence forecast is improved on at every step by the SVR, as expected from previous work [5]. The RMSE from the SVR shows a 8.930% improvement on the RMSE of the persistence forecast at 1 hour, 42.851% improvement at 12 hours and 17.246% improvement at 24 hours ahead.

Due to the 24 hour cyclic behavior of wind, the persistence forecast RMSE reaches a maximum at around 12 hours and a local minimum around 24 hours, in Figure 1. If the forecast window is increased the same cyclic behavior for the persistence forecast will be observed throughout the whole window. The SVR prediction of the wind speed at 60 m, using only the 60 m wind speed time series in the training data set (Feature set A), functions as the base case SVR prediction.

In the Figures 2-5 additional features are used in the training SVR data sets to predict the wind speed at 60 m. The different feature set combinations (defined in Table I) are compared with the base case.

Basic feature tailoring, such as calculating the wind speed acceleration (delta) and change in wind speed acceleration (double delta) from the 60 m wind speed data, are added to the training data set for predicting the wind speed at a 60 m hub

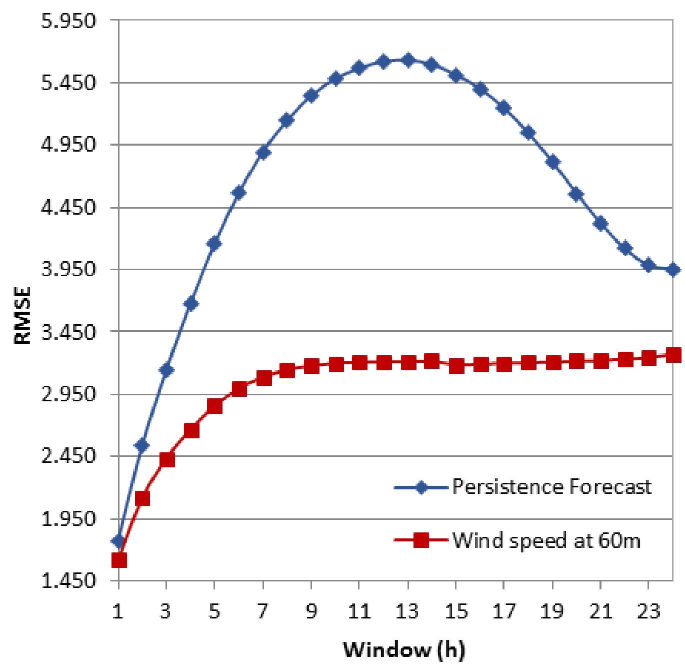

Fig. 1. Persistence and SVR forecast of the wind speed at a 60 m hub height (base case) [5].

height in Figure 2. The predictions, using tailored features in the training set (Feature sets B and C), outperforms the base case from 3 to 16 hours ahead. At 1 hour and from 19 to 24 hours the base case performs the best by a very small margin. From 2 to 18 hours ahead the prediction using Feature set B performs the best. On average the prediction using Feature set B performs the best with a 1.11% improvement on the base case RMSE.

Additional atmospheric measurements such as the wind direction measured at 60 m (Feature set D), the air temperature measured at 60 m (Feature set E), the barometric (atmospheric) pressure measured at 6 m (Feature set F) and relative humidity measured at 60 m (Feature set G), are added to the training data respectively, to improve the SVRs prediction of the wind speed at 60 m. These forecasts are compared with the base case prediction in Figure 3.

The prediction using wind direction (Feature set D) does not show any improvement on the base case, and at certain times, such 14 hours ahead the forecast is significantly worse. The forecasts using air temperature (Feature set E), barometric pressure (Feature set F) and relative humidity (Feature set G) show a clear improvement on the base case forecast. The RMSE of Feature set E is the lowest from 1-15 hours ahead and from 16-24 hours Feature set F produces the lowest RMSE. On average Feature set E has the lowest RMSE which is a 6.80% improvement on the base case prediction (Feature set A).

Different features are combined in Figure 4 to further investigate the effect of feature selection on the forecast accuracy. The forecast combining wind speed data at different

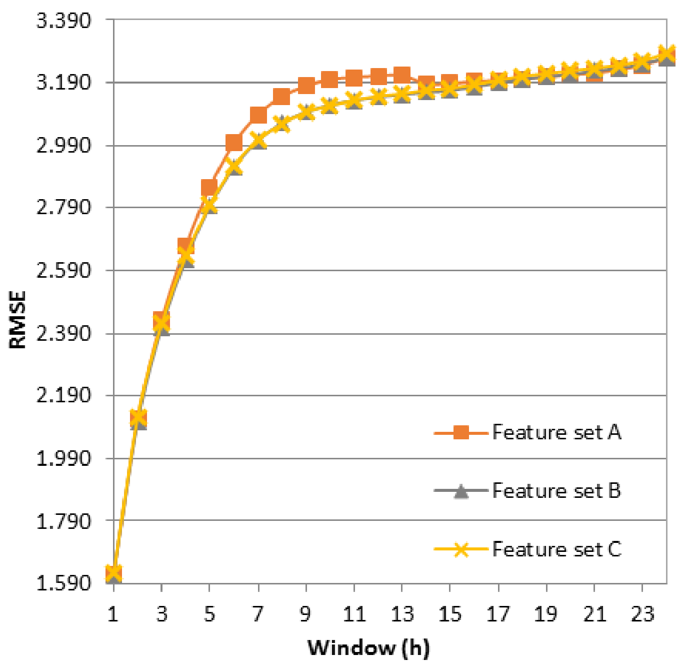

Fig. 2. SVR forecast of the base case forecast compared with the wind speed at a 60 m hub height with additional training features.

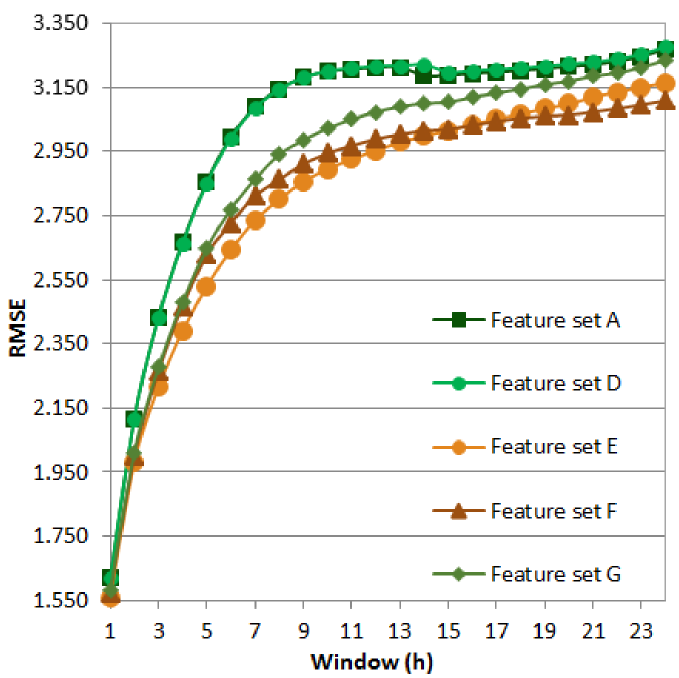

Fig. 3. SVR forecast of the wind speed at a 60 m hub height with additional training features.

hub heights 10 m, 20 m, 40 m, 60 m and 62 m (Feature set H) is displayed in Figure 4 and compared with the wind base case speed prediction (Feature set A). It is found that the prediction using the wind speed at various heights (Feature set H) improves on the base case average RMSE by 4.48%. To further improve the prediction accuracy of the base case forecast and forecast using Feature set H, the additional training features from the two best performing forecasts from Figure 3 (the air temperature and barometric pressure), are merged with the two sets of training data to create Feature sets I and J, respectively.

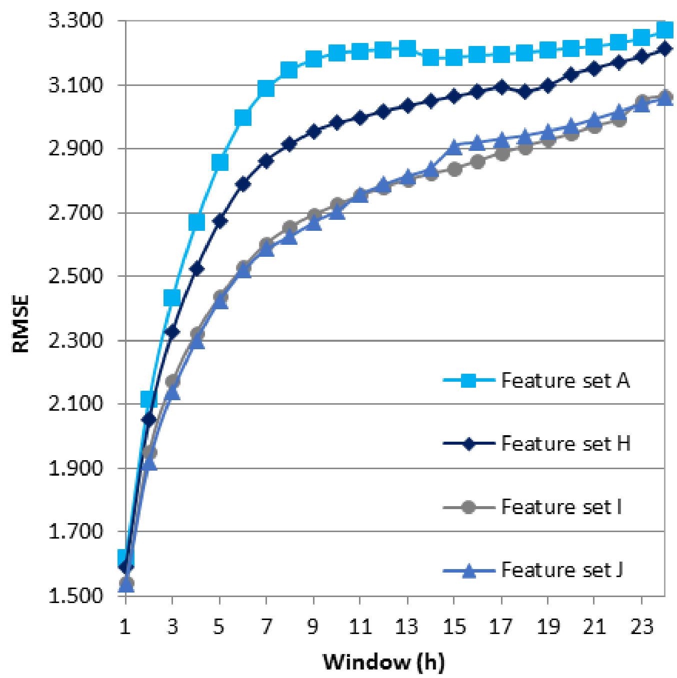

Fig. 4. SVR forecast of the wind speed at a 60 m hub height with additional training features combinations.

It is expected that the prediction using Feature set J will outperform the prediction using Feature set I and that both will outperform both the base case and the prediction using Feature set H. But even though more height data (Feature set H) improved the base case forecast, combining the various heights' wind speed data with the air temperature and barometric pressure (Feature set J) does not improve the RMSE further from the forecast using Feature set I and even performs worse from 15 hours ahead. Feature set I improves on average, the base case by 11.12% while Feature set J improves base case by 10.97%.

In Figure 5 several features are added to the combination that makes up Feature set I, to create Feature sets K and L. Feature set K is a combination of the wind speed at various heights, air temperature, barometric pressure and relative humidity. Feature set L is a combination of the wind speed at all the heights, air temperature, barometric pressure and wind speed acceleration (delta) and finally Feature set M is a combination of all the available features, see Table I.

The prediction using Feature set L shows the best overall

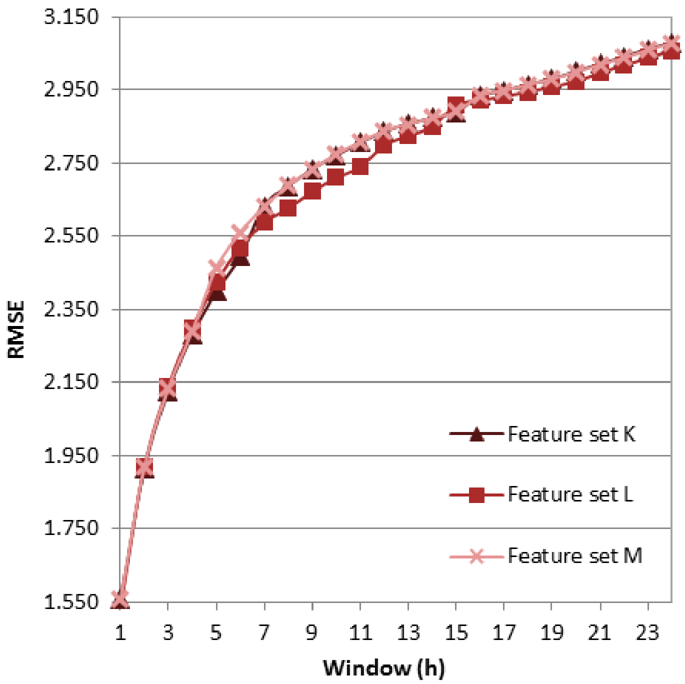

Fig. 5. SVR forecast of the wind speed at a 60 m hub height combining additional training features that performed well in Figures 2-4.

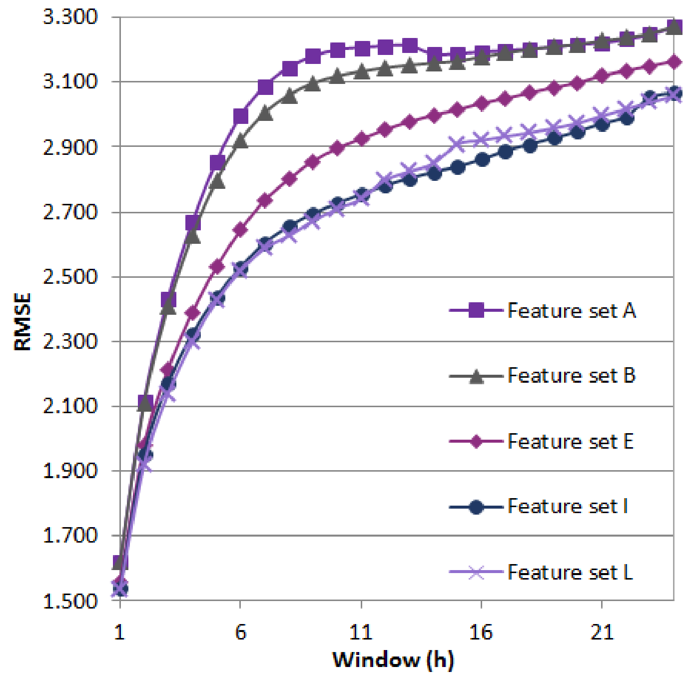

Fig. 6. Comparison of the best forecasts from Figures 2-5.

result in Figure 5 by a small margin, with the lowest RMSE between 7 to 14 hours ahead and 16 to 24 hours ahead. This is once again counter intuitive since the prediction using the feature relative humidity (Feature set G, Figure 3) performs better than the prediction using the delta (Feature set B, Figure 2). When all the features are combined in the training data (Feature set M), the RMSE is not improved when compared to Feature sets K and L and even performs visibly worse at 5 to 6 hours ahead. This highlights the importance of careful feature selection. The average improvement of the prediction using Feature set L on the base case is 10.95% while the prediction using Feature set M improves by 10.06%. The RMSE values for the SVR forecasts using Feature sets A, I, J, L and M respectively, are shown in Table III, Appendix A.

The predictions that show the most noticeable improvement on the base case (Feature set A) in each preceding Figure are shown in Figure 6, to provide an overall clear visual comparison of the most noted results.

The prediction using the air temperature (Feature set E) shows visible improvement on the base case for the whole 1 to 24 hour window. The predictions using Feature sets I and L, adds more features and further improves the prediction with Feature set E's RMSE. Feature sets I and L show similar RMSE and trends, with Feature set L performing the best from 1 to 11 hours and 23 to 24 hours ahead. On average, the prediction using the air temperature and barometric pressure (Feature set I) shows the most noticeable improvement, with an 11.12% improvement on the base case forecast.

The percentage of average improvement for each prediction on the base case prediction is summarised in Table II. From the results summary in Table II, every combination of features shown in Figures 2-5, improves on average the forecast with Feature set A to some degree, except the prediction using the wind direction (Feature set D) that performs marginally worse with a higher average RMSE.

By carefully tailoring the use of additional features in training a SVR algorithm, the RMSE of the wind speed predicted at 60 m hub height at Alexander Bay can be improved upon by up to 11.12%. The importance of feature tailoring is highlighted with the case where some features may actually decrease the prediction accuracy, such as the wind direction. Predictions that improve on the base case does not necessarily improve further on the forecast when combined with other features, such as the case of relative humidity and delta in Figure 5.

IV. Conclusion

In this paper we explored the effect of feature selection on the forecasting accuracy of a 1 to 24 hour ahead wind speed prediction at a 60 m hub height, by combining and incorporating wind speed at five metmast hub heights, wind speed acceleration, change in wind speed acceleration, wind direction, air pressure, air temperature and relative humidity, in several Support Vector Regression models.

We have show that the relative prediction performance of a short-term SVR wind forecasting system can be improved by up to 11.12% (compared to a base case of only wind speed data) by systematically adding and combining relevant input features that influence short term wind speed.

We also found that the addition of air temperature and barometric temperature to the 60m wind speed measurments

TABLE II
PERCENTAGE VAERAGE IMPROVEMENT ON THE RMSE OF EACH PREDICTION IN FIGURE 6 OVER THE BASE CASE (FEATURE SET A).

Feature set	**Improvement on the base case**
B	1.11%
C	0.90%
D	-0.15%
E	6.80%
F	6.23%
G	3.81%
H	4.48%
I	11.12%
J	10.97%
K	10.26%
L	10.95%
M	10.06%

(used in the base case) gave the most significant performance improvement and also observed that some feature combinations provided optimal performance at shorter time forecast horizons, while other feature combinations conversely provided optimal performance at longer time forecast horizons.

For future work we will investigate various feature engineering strategies to capture longer term behaviour in the input features.

REFERENCES

[1] G. Ji, P. Han and Y. Zhai, "Wind speed forecasting based on support vector machine with forecasting error estimation," Proceedings of the Sixth international Conference on Machine Learning and Cybernetics, Hong Kong, 19-22 August, 2007, pp. 2735-2739.

[2] G. Santanmaria-Bonfil, A. Reyes-Ballesteros and C. Gershenson, "Wind speed forecasting for wind farms: A method based on support vector regression," Renewable Energy, 85, 2016, pp. 790-809.

[3] S. Salcedo-Sanz, E.G. Ortiz-Garcia, A.M. Perez-Bellido, A. Portilla-Figueras and L. Prieto, "Short term wind speed prediction based on evolutionary support vector regression algorithms," Expert Systems with Applications, Vol 38, 2011, pp.4052-4057.

[4] J. Wang, Q. Zhou, H. Jiang and R. Hou, "Short-term wind speed forecasting using support vector regression optimized by Cuckoo optimization algorithm," Mathematical Problems in Engineering, Vol. 2015, 2015.

[5] G. N. Botha and C. M. van der Walt, "A comparison of regression algorithms for wind speed forecasting at Alexander Bay," 27th annual Symposium of the Pattern Recognition Association of South Africa and Robotics and Mechatronics (PRASA-RobMech) International Conference Proceedings, Stellenbosch, South Africa, 2016.

[6] M.A. Mohandes, T.O. Halawani, S. Rehman and A.A. Hussain, "Support vector machines for wind speed prediction," Renewable Energy, Vol 29, 2004, pp. 939-947.

[7] P. Zhao, J. Xia, Y. Dai and J. He, "Wind speed prediction using support vector regression," 5th IEEE Conference on Industrial Electronics and Applications, 2010, pp. 882-886.

[8] D. Liu, D. Niu, H. Wang and L. Fan, "Short-term wind speed forecasting using wavelet transform and support vector machines optimized by genetic algorithm," Renewable Energy, Vol 62, 2014, pp. 592-597.

[9] K. Chen and J. Yu, "Short-term wind speed prediction using an unscented Kalman filter based state-space support vector regression approach," Applied Energy, Vol 113, 2014, pp. 690-705.

[10] J. Hu, J. Wang and G. Zeng, "A hybrid forecasting approach applied to wind speed time series," Renewable Energy, Vol. 60, 2013, pp. 185-194.

[11] J. Wang, S. Qin, Q Zhou and H. Jiang, "Medium-term wind speeds forecasting utilizing hybrid models for three different sites in Xinjiang, China," Renewable Energy, Vol 76, 2015, pp. 91-101.

[12] Y. Lui, Y. Sun, D. Infield, Y. Zhao, S. Han and J. Yan, "A Hybrid Forecasting Method for Wind Power Ramp Based on Orthogonal Test and Support Vector Machine (OT-SVM)," IEEE Transactions on Sustainable energy, Vol 8, No 2, 2019, pp. 451-457.

[13] D. Niu, Y. Wang and D.D. Wu, "Power load forecasting using support vector machine and ant colony optimization," Expert Systems with Applications, Vol. 37, 2010, pp. 2531-2539.

[14] X. Liu, X. Kong and K.Y. Lee, "Wind Speed Prediction with high efficiency convex optimization Support Vector Machine," Proceeding of the 11th World Congress on Intelligent Control and Automation, Shenyang, China, June 29-July4, 2014, pp. 908-915.

[15] Wind Atlas for South Africa (WASA), http://www.wasaproject.info/, Last accessed on 12/08/2016.

[16] A.J. Smola and B. Schlkopf, "A tutorial on support vector regression." Statistics and computing, Vol 14, 2004, pp. 199-222.

APPENDIX

The RMSE of several forecasts, combining the 60 m wind speed with features such as the wind speed at different heights, air temperature and barometric pressure are shown in Table III. The forecasted values in Table III are also displayed in Figures 4 and 6.

TABLE III
ROOT MEAN SQUATE ERRORS DIPICTED IN FIGURES 4 AND 6.

Window	**SVR (A)**	**60m,Tair, Pbaro (I)**	**All heights, Tair, Pbaro (J)**	**All heights,Delta, Tair,Pbaro (L)**	**All features (M)**
1	1.619	1.541	1.536	1.536	1.556
2	2.116	1.953	1.920	1.920	1.920
3	2.434	2.172	2.138	2.139	2.134
4	2.668	2.323	2.299	2.300	2.291
5	2.856	2.438	2.425	2.425	2.463
6	2.996	2.529	2.519	2.519	2.559
7	3.088	2.604	2.586	2.587	2.631
8	3.146	2.655	2.625	2.626	2.688
9	3.180	2.694	2.671	2.671	2.732
10	3.200	2.726	2.706	2.709	2.774
11	3.207	2.755	2.758	2.740	2.807
12	3.212	2.781	2.788	2.798	2.838
13	3.213	2.804	2.815	2.824	2.856
14	3.186	2.823	2.841	2.849	2.874
15	3.188	2.840	2.908	2.908	2.890
16	3.193	2.862	2.920	2.921	2.935
17	3.198	2.889	2.932	2.933	2.948
18	3.203	2.908	2.942	2.944	2.964
19	3.208	2.929	2.957	2.958	2.980
20	3.216	2.948	2.973	2.973	2.998
21	3.221	2.973	2.994	2.996	3.019
22	3.232	2.994	3.017	3.017	3.038
23	3.246	3.053	3.040	3.037	3.058
24	3.270	3.068	3.060	3.057	3.078