



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Wind direction forecasting with artificial neural networks and support vector machines

Citation for published version:

Tagliaferri, F, Viola, IM & Flay, RGJ 2015, 'Wind direction forecasting with artificial neural networks and support vector machines' *Ocean Engineering*, vol. 97, no. 15, pp. 65-73.

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Early version, also known as pre-print

Published In:

Ocean Engineering

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Wind direction forecasting with artificial neural networks and support vector machines

F. Tagliaferri

Institute for Energy Systems, School of Engineering, The University of Edinburgh, United Kingdom

I.M. Viola

Institute for Energy Systems, School of Engineering, The University of Edinburgh, United Kingdom

R.G.J. Flay

Yacht Research Unit, Department of Mechanical Engineering, The University of Auckland, New Zealand

Abstract

We propose two methods for short term forecasting of wind direction with the aim to provide input for tactic decisions during yacht races. The wind direction measured in the past minutes is used as input and the wind direction for the next two minutes constitutes the output. The two methods are based on artificial neural networks (ANN) and support vector machines (SVM), respectively. For both methods we optimise the length of the moving average that we use to pre-process the input data, the length of the input vector and, for the ANN only, the number of neurons of each layer. The forecast is evaluated by looking at the mean absolute error and at a mean effectiveness index, which assesses the percentage of times that the forecast is accurate enough to predict the correct tactical choice in a sailing yacht race. The ANN forecast based on the ensemble average of ten networks shows a larger mean absolute error and a similar mean effectiveness index than the SVM forecast. However, we showed that the ANN

*Corresponding author

Email addresses: f.tagliaferri@ed.ac.uk (F. Tagliaferri), i.m.viola@ed.ac.uk (I.M. Viola), r.flay@auckland.ac.nz (R.G.J. Flay)

forecast accuracy increases significantly with the size of the ensemble. Therefore increasing the computational power, it can lead to a better forecast.

Keywords:

wind forecast, support vector machines, artificial neural networks, sailing yacht, race, tactics

1. Introduction

The speed of a sailing yacht depends on the wind speed and the course wind angle (the supplementary of the angle between the wind direction and the boat velocity). The boat speed can be represented as a polar diagram, such as the one shown in Fig. ??, where the radial coordinate is the boat speed of an AC72-class boat, for a fixed wind speed, while the angular coordinate is the boat's heading with respect to the true wind direction (data from ?). The dependency of the boat speed on the wind velocity is a key element when deciding the tactics during a yacht race. For instance, when navigating around the world, experienced sailors take advantage of large-scale weather changes, and for inshore races lasting less than one hour, minimal wind shifts can be used to gain an advantage on the competitors. In the latter case, the ability to forecast very-short-term wind changes can make the difference between a win and a loss. A prepared sailor consults wind forecasts before starting the race, and this is usually used to build a strategy aiming at sailing in the racing area where a higher wind speed and a more favourable wind direction are expected. During the race, it is possible to use only the information which is collected on board, including wind speed and direction, which are measured with a cup and vane anemometer on the top of the mast. In the present paper, we present two methods to forecast very-short-term wind shifts based only on the wind direction measured on board during the race. The proposed wind forecast is aimed at complementing the longer-term weather forecast available to the sailors up to the beginning of the race.

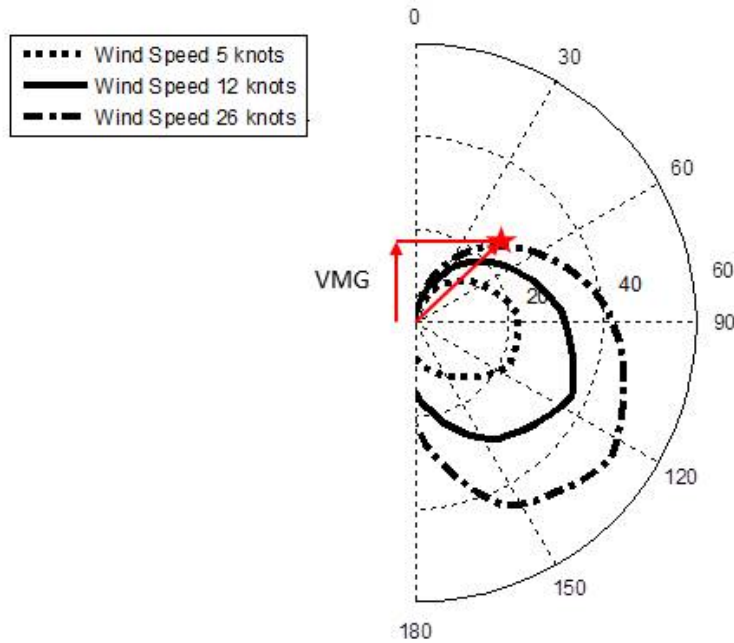


Figure 1: Polar diagram of an AC72-class yacht

1.1. Racing Tactics

A typical America's Cup race is made of several turns around an upwind and a downwind mark, where the marks are aligned with the wind. For instance, in an upwind leg, the boats start from the downwind mark and have to reach the upwind mark, located upstream. As shown in the polar plot in Fig. ??, it is not possible to sail directly upwind. The fastest route consists in keeping an optimum course wind angle that maximises the boat velocity in the upwind direction, i.e. the velocity for which the projection on the vertical axis of the plot in Fig. ?? is a maximum. This velocity component in the wind direction is known as VMG. As an example, the optimum course wind angle in an upwind leg for a wind speed of 26 knots is shown in Fig. ??, together with its projection. In most wind conditions this optimum angle is roughly 45° . Because the boat is not sailing straight towards the upwind mark, at a certain point a change of direction of about 90° will be needed. These changes of course are called

tacks. While tacking, the course wind angle decreases from the optimum value, leading to a speed loss. Therefore the total number of tacks should be minimised. Typically, an AC72-class yacht takes about 20 seconds to complete a tack, and during the manoeuvre the average boat speed is approx. 70% of the optimum speed. Therefore, a tack leads to a loss of about 6 s.

If the wind direction is constant, then the optimum route includes only one tack. Conversely, when the wind direction varies such as in real conditions, then the sailor can tack between two consecutive wind shifts in order to gain an advantage. However, this advantage must be greater than what is lost due to undertaking the tack. There is a high number of possible scenarios that can arise during a race and the problem of taking the best decision is not trivial (?). There are however some situations in which the best decision can be easily taken when a sailor is able to foresee how the wind is going to behave. In the following the two most likely scenarios are presented.

Figure ?? shows the route followed by two boats in order to reach an upwind mark in two different wind conditions. In the first case (left) the wind alternately shifts by 3° to the right and to the left. The first shift is towards the right and, while the two boats sail at their optimum course wind angles, they chose to sail in different directions. Both boats tack at every wind shift. The black boat is always sailing towards the left hand side of the race course when the wind shift is towards the right, and vice versa when the wind shift is towards the left. Conversely, the grey boat has the opposite strategy. Therefore, the black boat is always sailing to a closer angle to the mark than the grey boat, she sails a shorter course and she arrives first to the mark. The winning tactic of the black boat is that she always maximises her velocity towards the mark. However, this is not always a winning tactic. In fact, in the second case (right), the wind constantly shifts towards the right. The two boats follow the same strategy as in the previous case: experiencing a wind shift to the right, the black boat sails towards the left and the grey boat sails towards the right. Being the wind constantly shifting to the right, the two boats never tack until they reach a lay line, i.e. where a tack allows the mark to be reached without any further tacks.

The black boat needs to sail most of the race course before reaching the lay line and being able to tack to the mark, while the grey boat reaches the lay line before the black boat. In this case, even if the two boats have pursued the same strategy based on the wind observed at the time, the resulting course sailed by the black boat is longer than the course sailed by the grey boat, because the wind shifted regularly in the same direction instead of alternating to opposite directions. This shows that the tactical decision cannot be based on the wind direction observed at the time, but that the future wind shifts must be foreseen in order to develop a winning strategy.

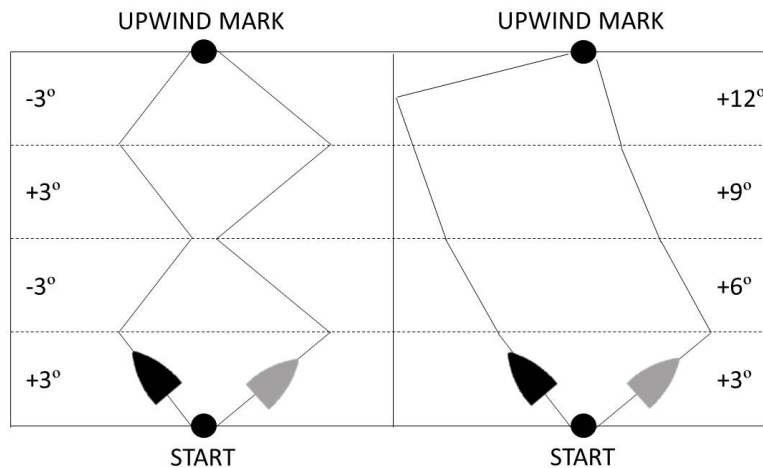


Figure 2: Yacht routes in oscillating wind (left) and in a permanent wind shift (right).

Figure ?? shows that the optimum strategy depends on if the wind shift has a wavelength shorter or longer than the distance to the lay line. If the race course is confined by boundaries, such as shore lines or forbidden areas, then wind shifts with wavelength up to the distance to the boundary should be considered.

In the 34th America's Cup, the racecourse was bounded by imaginary lines in order to allow the spectators to be closer to the racing boats. The boats took about two minutes to sail from one boundary to the opposite one, thus the

maximum wind shift period of interest was about two minutes.

Consider the black boat in Fig. ???. The boat starts from the centre of the course, she sails for say one minute, then tacks and comes back to the centre of the course. Say she took two minutes plus 6 seconds for the tack. If the optimum course wind angle is 45° , in two minutes a shift of 3° would lead to an advantage of about 6 seconds on the grey boat. On the contrary, if the grey boat had tacked at the start and followed the black boat from the beginning, she would be 6 seconds behind the black boat due to the time spent to tack at the start. Therefore a wind shift of 3° is the threshold at which the decision of tacking or sailing in the same direction has to be made.

1.2. Artificial Neural Networks

We use artificial neural networks (ANN) to accurately forecast wind shifts larger than 3° for the two minutes ahead. ANNs are computational models that emulate the ability of the human brain to learn from experience, similarly, for instance, to the capability of a human sailor to make predictions based on his lifelong experience. In the literature it is possible to find a vast number of applications where ANNs have been successfully used, especially to solve problems which are peculiar of humans, such as speech recognition (?), image classification (?), control of moving robots (?).

The constitutive unit of a neural network is a neuron, which is a singular processing unit that takes several inputs originating from other neurons, and produces an output that is then transmitted to other neurons. A representation of the structure of a neuron is shown in Figure ??. A neuron can be broken down into the following components:

1. A set of connecting links, called synapses, where the i -th synapse is characterised by a weight w_i (synaptic weight);
2. An adder within the neuron that sums each i^{th} input multiplied by weight w_i ;
3. An activation function, ψ , which transforms the sum computed by the adder into the neuron output y . If the activation function is linear, a

neuron results in a linear combination of the input values, while non-linear activation functions allow for the modelling of non-linear problems.

Therefore, a neuron can mathematically be described by Equation (??):

$$y = \psi(\zeta); \zeta = \sum_{i=1}^u w_i x_i \quad (1)$$

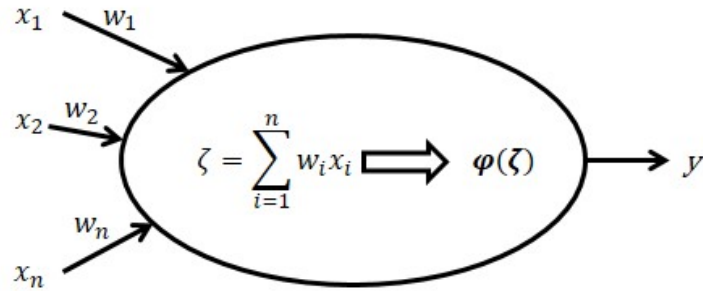


Figure 3: Schematic representation of a neuron

Neurons are assembled together into an integrated structure, the actual ANN, that depends on the kind of problem that the network has to solve. The input vector is processed by an input layer and the information moves through the structure of the network until the output layer.

The learning process, aimed at making the network able to model a specific problem, involves the continuous modification of the synaptic weights. A commonly used training algorithm is based on the principle of iterative error-correction. The synaptic weights of the various neurons are initialized to random values, then a training set of input and output data is presented to the network. For

each input vector, the initially generated output vector is compared with the known true output vector. The synaptic weights of the output layer are then modified by adding a factor that is proportional to the error and to a learning rate, and those corrections are extended to all of the weights in the network through a back-propagation process up to the input layer. This operation is iterated until successive changes in the synaptic weights are smaller than a given value, or when the errors begin to increase. For further details on ANNs, including training algorithms and validation processes, see ?.

1.3. Support Vector Machines

In this paper we compare the wind direction forecast performed with ANN with the forecast performed with support vector machines (SVM), which constitute another class of supervised learning models. SVM were originally developed for solving classification problems, where data need to be classified in two or more different categories, and have successively been used also to tackle regression problems (?). Regression using SVM is also called in the literature support vector regression (SVR).

The fundamental idea behind SVM is to map the data into a higher dimensional space where the problem is linearly separable, and then solve the problem on the new space. In the case of binary classification this process can be easily visualised as in Fig. ???. In this simple example, points on a plane belonging to two different categories are mapped into a three-dimensional space. The problem is not linearly separable on a plane, but it is linearly separable in three dimensions.

This concept is generalised for SVR, where the aim is to model a function $y = f(\bar{x})$. In this case we have to look for a function Φ which maps the problem to a higher dimensional space where it is possible to perform a linear regression, as shown in Equation (??):

$$y = \hat{f}(\bar{x}) = a \cdot \Phi(\bar{x}) + b \text{ where } \Phi : \mathbb{R}^n \rightarrow \mathcal{F} \subseteq \mathbb{R}^{n+m}, a, b \in \mathcal{F} \quad (2)$$

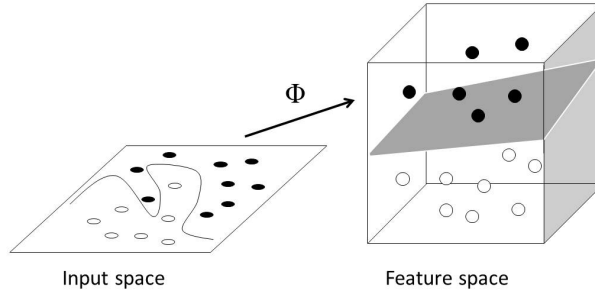


Figure 4: Mapping of 2D non linearly separable data to a 3D space where the problem becomes linearly separable.

A solution to this problem is defined as a minimum for the error function shown in Equation (??):

$$R[f] = \sum_{i=1}^n C(f(\bar{x}_i) - \hat{f}(\bar{x}_i)) + \lambda|b| \quad (3)$$

where n is the sample size, C is a cost function and the term $\lambda|b|$ is added to enforce flatness in the higher-dimension space. The function Φ can be found as the unique solution of a quadratic programming problem (??). Equation (??) can be rewritten in terms of this solution as

$$\hat{f}(\bar{x}) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) k(\bar{x}_i, \bar{x}) + b \quad (4)$$

where k is a symmetric kernel function (see ? for details) and α_i, α_i^* constitute the solution to the quadratic programming problem. Details of this method can be found in ?.

An important difference between SVR and ANN approaches is that SVR lead to a unique deterministic model for each data set, while ANNs depend on a random initial choice of synaptic weights. Therefore, in order to minimise the effects of this randomness in an ANN model, we train different ANNs and then average their outputs. This is not necessary for SVR, which constitutes an advantage in terms of computational time.

1.4. Wind speed and wind direction

The vast majority of the literature is focused on wind speed forecasting, as opposite to wind direction forecasting (??). ANNs have been used for wind speed forecasting, ranging from hours (?) to days (??), and have been mostly motivated by application in renewable energy. The wind direction has been used as an input for wind speed forecasting together with other quantities such as humidity and pressure to improve the forecast of the wind speed (?), which is the key parameter to predict the amount of energy available in the wind stream. Conversely, in the present application we are focused on nowcasting wind direction, i.e. on wind direction forecasting of few minutes ahead. In fact, an increase in the wind speed would lead to an increase in the boat speed, but the wind speed distribution over the racecourse is forecast more easily than the wind direction by visual observation of the surface of the sea. On the contrary, wind direction is forecast with difficulty by visual observation. Using wind speed, wind direction and other time series together can improve the accuracy of forecasts (?), but for nowcasting, wind speed and direction may be uncorrelated and therefore should not be used together. In particular, as discussed in the next section, the input data set that we used was particularly uncorrelated and using the wind speed as additional input would have not improved the forecast of the wind direction.

2. Method

2.1. Input and Output Data

The data set used for this work consists of registrations collected during the 34th America's Cup in San Francisco (?). Wind speed and direction are recorded from different moving and fixed sources at a frequency of 5 Hz, and averaged by the Media Data Server System with a proprietary algorithm. Because the data is the average of several measurements taken in different locations across the the race course, records for the same day show a correlation coefficient spanning from -0.4 to 0.6, while the global correlation coefficient was

-0.38. Also the correlation coefficient is highly volatile with respect to the subset used. Therefore only the wind direction is used as input. As far as known by the present authors, this is the first wind direction forecast based on ANN or SVR using only wind direction as input.

The data include registrations collected during 34 days. We use the last 100 minutes to test the performance of the ANN and SVR forecasts, and the rest of the data to train both models. Due to the high sampling frequency (5 Hz) and the limit of precision to 0.1 degrees, several consecutive values of the data set were identical. Therefore we average the data over 30 seconds using 150 consecutive values, leading to a re-sampled dataset with one value every 30 seconds (0.03 Hz).

As an example, the last 35 minutes of the re-sampled data set, corresponding to the last race of the 34th America’s Cup finals, are shown in Figure ???. The beginning of the race is characterised by small wind shifts, such as those on the left of Figure ??, while in the second half of the race there is a consistent wind shift towards the right such as on the right of Fig. ?. During the 35 minute period shown, the wind shifts by more than 3° for periods in excess of 30 seconds 16 times.

The time history of the re-sampled data set shows significant fluctuations with continuous change of direction every 30 seconds. In order to further smooth the data set, a moving average is used. The length of the moving average is optimised based on the performance indexes, as discussed in Section 3. Increasing the length of the moving average, the time series results increasingly smoother, while the sampling frequency remains constant. As an example, in Fig. ?? the data set smoothed with a moving average of six minutes is also shown.

A vector of consecutive past data was used as input and its length was optimised over the indexes described in Sec. ??.

The outputs of the models are the wind directions averaged over one minute ahead, and averaged between one and two minutes ahead. Therefore we use an ANN and SVR to approximate a function f that expresses future values as function of past ones as in Equation (??):

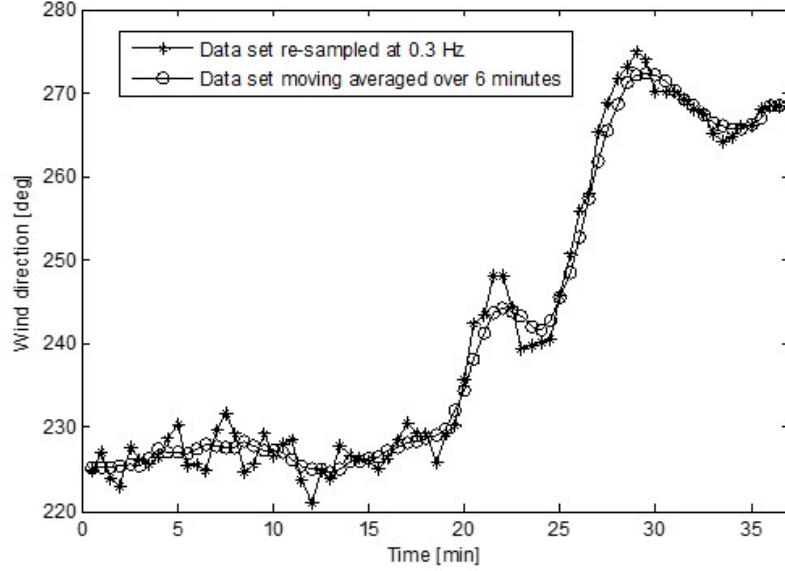


Figure 5: Example of data set averaged every 30 seconds

$$\left(\frac{x_{t+1} + x_{t+2}}{2}, \frac{x_{t+3} + x_{t+4}}{2} \right) = f(x_t, x_{t-1}, \dots, x_{t-m}) \quad (5)$$

where x_t is the wind speed at discrete time step t and m is an optimised integer.

2.2. ANN forecast

The chosen structure for the ANN is a feed-forward multi-layer perceptron which was implemented in Matlab. The hyperbolic tangent sigmoid function was used as activation function. Figure ?? shows one of the feed-forward structures tested. In a multi-layer structure, neurons are organised in layers, where neurons do not receive input from any other neuron in the same layer. In a feed-forward structure, the information flows in only one direction, i.e. the output of each layer is input for the successive layer. Both the number of layers and the number of neurons per layer can be increased at discretion, but in general any network with more than two layers can be reduced to a two-layers network with an

adequate number of neurons (?). The example shown in Fig. ?? shows a network with two hidden layers of five neurons each, taking past wind directions as input and giving as output the average wind direction in the following two minutes. In this study we present a comparison between two-layer perceptrons with different number of neurons. In Sec. ?? we present a parameter study for different lengths of the input vector, lengths of moving average and number of neurons. It will be shown that there is an optimum for each of these parameters that allows the maximum accuracy of the forecast.

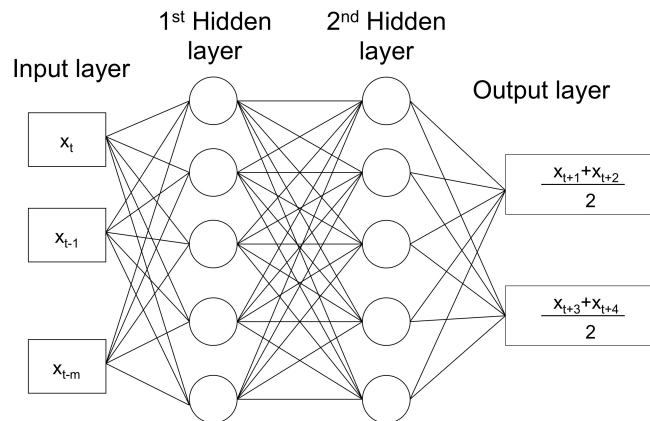


Figure 6: Feed forward multi layer ANN

Because the synaptic weights are randomly set and then iteratively adjusted with the training of the network, networks initialised with different sets of random weights lead to slightly different output. Therefore what it is called *ANN forecast* in the following is the average of the outputs of one ensemble made of ten identical networks trained independently using different sets of initial synaptic weights. This averaging process decreases the chances of large errors due to unfortunate initialization-training combinations (?). The size of the ensemble is chosen arbitrarily and in Section 3 it will be shown that a larger ensemble

can lead to a more accurate forecast but to higher computational time.

2.3. SVR forecast

SVR model is implemented using the software LIBSVM ?, together with its Matlab interface. This software has already been used successfully in time series forecasting (?). The kernel function is the radial basis function kernel shown in Equation (??):

$$k(\bar{x}, \bar{y}) = \exp\left(-\frac{\|\bar{x} - \bar{y}\|^2}{2\sigma^2}\right) \quad (6)$$

Similarly to the ANN case, a parameter study for different lengths of the input vector and lengths of moving average will be presented in Sec. ?. It will be shown that there is an optimum for each of these parameters that allows the maximum accuracy of the forecast.

2.4. Performance Indices

In order to assess the performance of the forecast we use two different evaluation indices, the mean absolute error and the mean effectiveness index.

The mean absolute error (MAE) is defined as in Equation (??):

$$MAE = \langle \overline{\hat{x}_t - x_t} \rangle \quad (7)$$

where \hat{x}_t is the 30-second-averaged wind direction forecast for the time t , while x_t is the 30-second-averaged wind direction measured at time t , the bar represents the average over the last 100 minutes of the data set and the $\langle \cdot \rangle$ operator represents the average of 100 ensembles. It should be reminded that, when using ANNs, each forecast is the average over an ensemble of ten ANNs, therefore MAE includes the results of 1000 ANNs. In Section ??, MAE is presented with the 95%-confidence-level error bars computed as two standard deviations of the absolute error over 100 ensembles.

The effectiveness index (EI) is defined to evaluate the performance of the forecast in relation to the specific application that it is intended for. As discussed in Section ??, we want to forecast wind shifts larger than 3° for two minutes ahead. In particular, we want to know if the average wind direction for one

minute ahead will change by more than 3° and in which direction; and also if and in which direction it will change in the second minute with respect to the first one in order to understand if it is a temporary or a permanent wind shift. Figure ?? shows a schematic diagram of the effectiveness index. Firstly we measure the average wind direction over the past minute that we use as a baseline. Then we forecast the average wind direction from now to one minute ahead, and from one to two minutes ahead. We compare the average of the first minute ahead with the baseline and we determine if the wind shift is smaller than 3° or, otherwise, the sign of the wind shift. Similarly we do for the average of the second minute ahead with respect to the one of the first minute ahead. Therefore the shift can be negligible, positive or negative for the first minute ahead, and negligible, positive or negative for the second minute ahead. If the forecast is correct for the first minute then the effectiveness index at one minute is one. If the forecast is correct for both the first and the second input, the effectiveness index at two minutes is one. In this case we will be able to recognise those wind shifts where tacking is worthwhile and we will distinguish between temporary and permanent shifts.

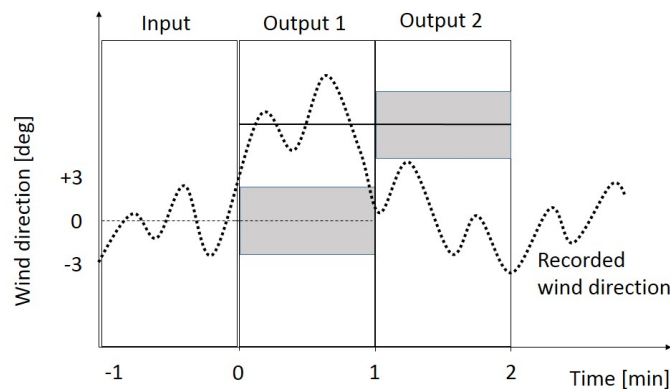


Figure 7: Effectiveness index

The mean effective index (MEI) is defined for the ANN as in Equation (??), where the bar represents the average over the test set and the $\langle \cdot \rangle$ operator represents the average of 100 ensembles. With the same notation, the MEI for the SVR forecast is defined in Equation (??):

$$MEI = \langle \overline{EI} \rangle \text{ for ANN} \quad (8a)$$

$$MEI = \langle EI \rangle \text{ for SVR} \quad (8b)$$

2.5. Test Matrix

In order to optimise the network, the MAE and the MEI are computed for different lengths of the moving average, lengths of the input vector and number of neurons in the hidden layers. In particular, moving averages from one to ten minutes, input vector lengths from 4 to 30 values (i.e. from 2 to 15 minutes) and from 5 to 100 neurons per hidden layer are tested.

The longer the moving average, the smoother the data and the easier to model the trend. However, increasing the length of the moving average leads to a loss of information in the lower frequencies that may result in the inability to forecast low frequency fluctuations. Therefore the optimum moving average length depends on the lowest frequency which needs to be forecast.

The longer the input vector the more information is fed to the model. However feeding unnecessary information makes the convergence of the synaptic weights more difficult to reach and, given a constant length of the available data set, the longer the input vector the fewer the vectors which can be built for training the network. Therefore, while a too-short vector length lacks of the necessary information for the forecast, a too-long vector length results in an undertrained network. It should also be noted that the computational time increases significantly with the length of the input vector.

When using ANN, increasing the number of neurons allows the network to model more complicated nonlinear trends. On the other hand, more neurons lead to more connections and thus to more synaptic weights which have to converge during the training (i.e. more degrees of freedom). Ultimately, for a constant

length of available data set, too many neurons leads to non converged synaptic weights.

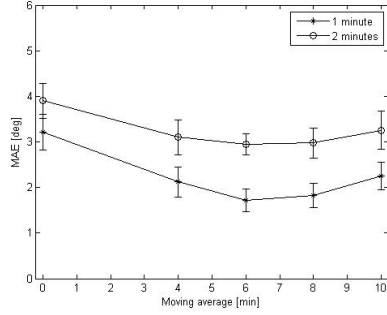
3. Results

In this section, first we present a parameter study of 5 different lengths of the moving average used to pre-process the input data, 14 lengths of the input vector and, for the ANN forecast, also 6 different sizes of layers. Then, using the optimum parameters selected from the parameter study, we compare the performance of the ANN and SVR forecasts.

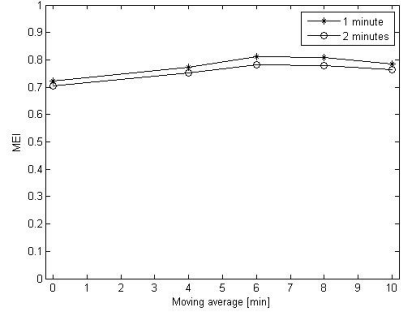
3.1. ANN optimisation

Figures ??-?? present MAEs and MEIs of one and two minutes ahead for different lengths of the moving average, different lengths of the input vector, and different number of neurons in the hidden layers. Due to the multidimensionality of the test matrix, the trend of the two performance indices versus each parameter is shown only for the best combination of the other two parameters. Figures ?? and ?? show the MAE and the MEI as functions of the length of the moving averages, where each hidden layer has 20 neurons and the size of the input vector is 18 (9 minutes). Figure ?? shows that for a moving average over 6 minutes, the MAE of one minute ahead is $1.7^\circ \pm 0.3^\circ$, while the MAE of the two-minute-ahead forecast is $3.0^\circ \pm 0.2^\circ$. As expected, the MAE increases both for smaller and larger lengths of the moving average. Importantly, the MAE of both one and two minutes ahead is minimum for a length of the moving average of six minutes. Figure ?? shows that the two-minute-ahead MEI is 0.78 when a six minute moving average is used, i.e. 78% of the time the forecast is able to predict the correct tactical decision.

Figures ?? and ?? show the MAE and the MEI, respectively, as a function of the length of the input vector, where a moving average of six minutes are used to smooth the data set and each hidden layer has 20 neurons. Figure ?? shows that the MAE for one and two minutes ahead is a minimum for an input



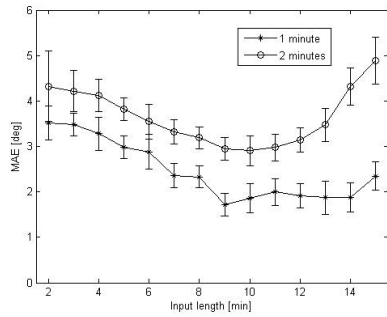
(a) Mean absolute error



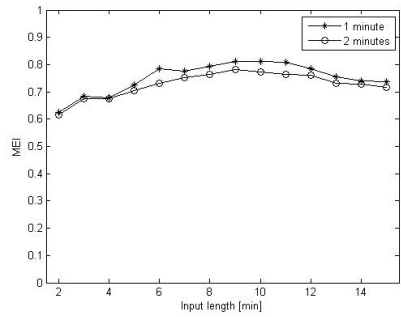
(b) Mean effectiveness index

Figure 8: ANN performance indices versus length of the moving average

vector size of 18 (9 minutes) and 20 (10 minutes), respectively. The MAE does not show a unique optimum input vector size. However, Fig. ?? shows that the two-minute-ahead MEI is a maximum for an input vector size of 18 (9 minutes). Therefore, for this specific application, the best forecast is performed using the past 9 minutes.



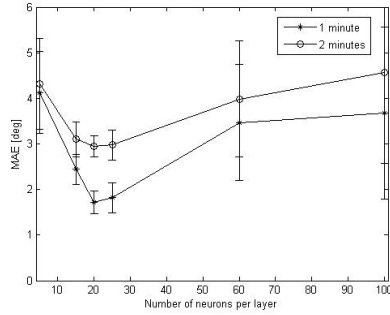
(a) Mean absolute error



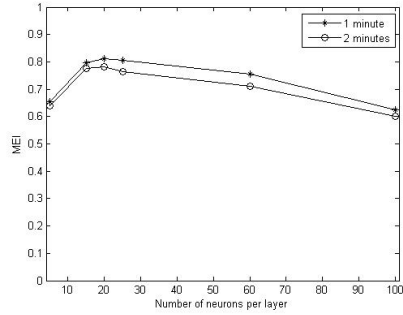
(b) Mean effectiveness index

Figure 9: ANN performance indices versus input length

Figures ?? and ?? show the MAE and the MEI, respectively, as a function of the number of neurons per each hidden layer, where a moving average of six minutes and an input vector size of 18 (9 minutes) are used. In this case, both indices show a clear optimum when 20 neurons per hidden layer are used.



(a) Mean absolute error

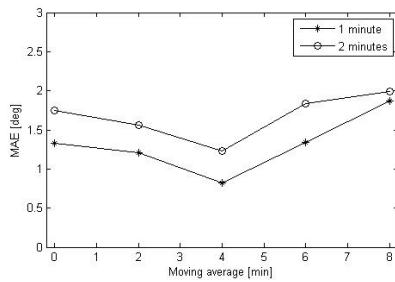


(b) Mean effectiveness index

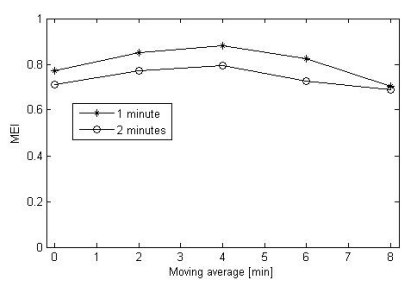
Figure 10: ANN performance indices versus number of neurons per layer

3.2. Support vector regression

Figures ?? and ?? show the MAE and MEI, respectively, as a function of the length of the moving average used to pre-process the data, using an input of 8 minutes. In this case the error bars are not needed, as there is only one possible model for each data set. Overall, the SVR shows a significantly lower MAE, both for one and two-seconds ahead predictions. The best performance is achieved for a moving average over 4 minutes, allowing a MAE of 0.8° and 1.2° for one and two minutes ahead, respectively, and a MEI of 0.79.



(a) Mean absolute error



(b) Mean effectiveness index

Figure 11: SVR performance indices versus length of the moving average

Figures ?? and ?? show the MAE and MEI, respectively, as a function of input vector for the SVR model, where a moving average of 4 minutes is used to

smooth the data. The lowest MAE is achieved for an input length of 8 minutes, and is of 0.8° and 1.2° for one and two minutes ahead, respectively. However, the maximum MEI achievable is of 0.79, differing only by 0.01 from the one achievable with the optimal ANN ensemble forecast.

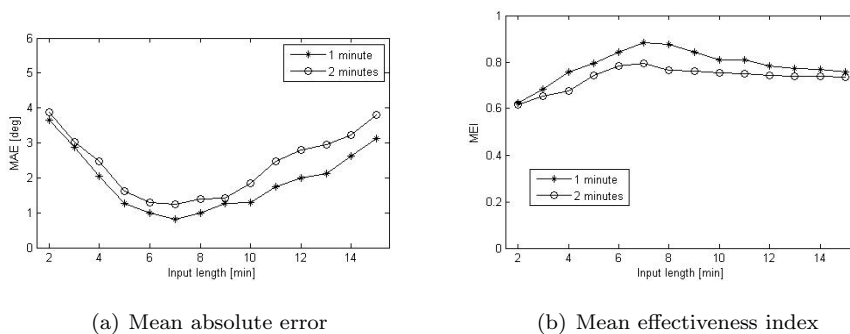


Figure 12: SVR performance indices versus input length

3.3. Performance of the optimised ANN and SVR

The results of the parameter study, which was based on a test set of 100 minutes, showed that the optimum ANN configuration allows a MAE of 1.7° and 3.0° and a MEI of 0.81 and 0.78 for one and two minutes ahead, respectively, while the optimum SVR allows a MAE of 0.8° and 1.2° and a MEI of 0.88 and 0.79 for one and two minutes ahead, respectively.

In Fig. ?? and ?? we test the optimum configuration for ANN and SVR on the last race of the 34th America’s Cup.

The input data set for this race is the one shown in Fig. ?. For clarity, we show only every second forecast, thus one forecast per minute. For each minute, the solid dots shows the mean recorded wind direction of that minute, while the stars and the circles show the forecast mean wind direction of that minute computed one and two minutes before, respectively. The plots are coloured with vertical bars showing the value of the effectiveness index. Each minute is coloured green (light grey when printed in black and white) if the effectiveness index is one, and is coloured red (dark grey) if zero. Therefore, for each green

minute the forecast would have led to the optimum tactical decision, while for each red minute it could have led to a mistake. In particular, every red minute underlines the cases in which the combination of one-minute-ahead and two-minute-ahead forecasts was not accurate enough to predict the correct tactical decision.

Fig. ?? shows that the ANN forecast leads to the correct tactical choice in all cases but four, corresponding to a MEI of 0.86. Figure ?? shows the SVR

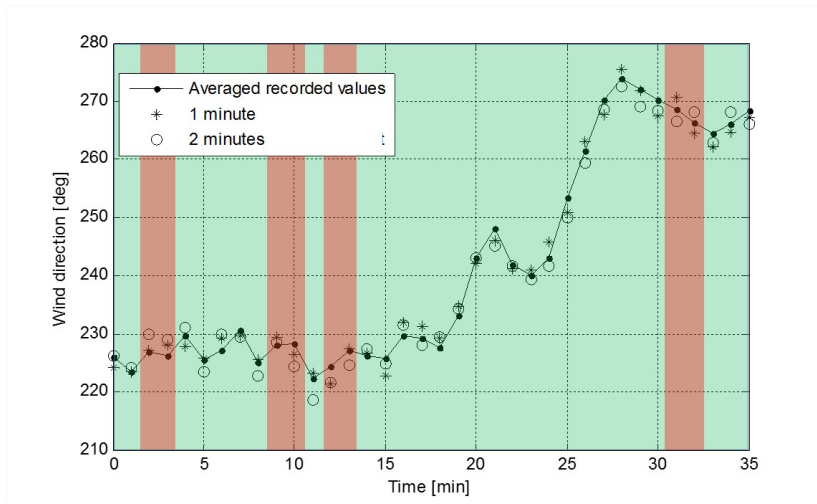


Figure 13: Example of forecast for the optimised ANN. Red stripes highlight wrong tactical decision, green stripes correct ones.

forecast on the same set. This forecast leads to three mistaken decisions, corresponding to a MEI, on this particular race, of 0.91.

The choice of obtaining the ANN forecast as an ensemble average of the outputs of ten networks constituted a compromise between training time and computational resources. With adequate hardware and lighter software, the variance of the error could be reduced further. As an example, we tested an ensemble average of 1000 networks (instead of 10) on the test set shown in Fig. ?? and the MEI reached 0.97, corresponding to just one potential mistaken decision.

In conclusion, SVR allows a better forecast in terms of accuracy and compu-

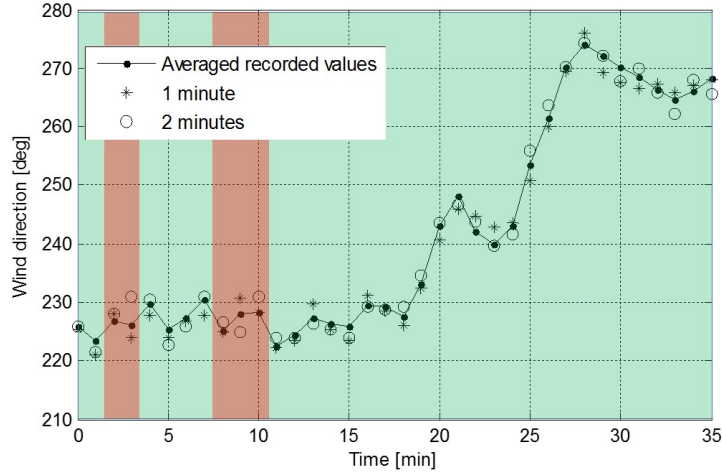


Figure 14: Example of forecast for the optimised SVR. Red stripes highlight wrong tactical decision, green stripes correct ones.

taton time, but increasing the computational power it is possible to obtain a better forecast from the combination of ANN models.

4. Conclusions

In this study we present two methods for short term wind direction forecasting based on ANN and SVR. Both methods use the knowledge acquired from previous recordings of wind direction to forecast the near future values. The reliability of the forecast is evaluated by computing the mean absolute error and the mean effectiveness index of the forecast, the latter being an index of the percentage of times in which the forecast is able to predict the correct tactical decision during a sailing yacht race.

The optimum ANN configuration allows a mean absolute error of 1.7° and 3.0° , and a mean effectiveness index of 0.81 and 0.78 for one and two minutes ahead, respectively. The optimum configuration based on the mean absolute error and the one based on the mean effectiveness index are almost identical but for a marginal difference in the optimum length of the input vector. Therefore these

results can be generalised to a certain extent to other applications. In particular it is expected that the optimum length of the moving average may decrease if wind shifts averaged over less than one minutes are desired; while longer input vectors and more neurons may be used if a longer data set is available to train the networks.

The optimum SVR forecast allows a mean absolute error of 0.8° and 1.2° . SVR outperforms ANN, both in terms of mean error and computational time. However, the performance in terms of MEI is similar to the ANN model, increasing only from 0.78 for the ANN to 0.79 for the SVR for the two-minutes ahead forecast.

In order to decrease the dependency of the forecast from the training of the ANN, the forecast is made of the ensemble average of ten ANNs subjected to different trainings. The mean effectiveness index was found to significantly increase with the size of the ensemble. For instance, a test performed on the wind recorded during the last race of the 34th America's Cup shows that, increasing the ensemble size from 10 to 1000, the mean effectiveness index increases from 0.79 to 0.97, corresponding to a reduction of the number of potential mistaken decisions from three to one during the entire race. Therefore, with adequate computational resources, the use of large ensembles for ANN forecasts can lead to better performance.

5. References

- America's Cup Event Authority, March 2013. AC34 Race Data Export Format Specification, 1.03.
- Barbounis, T., Theocharis, J., Alexiadis, M., Dokopoulos, P., 2006. Long-term wind speed and power forecasting using local recurrent neural network models. *IEEE Transactions on Energy Conversion* 21 (1), 273–284.
- Bitner-Gregersen, E., Bhattacharya, S., Chatjigeorgiou, I., Eames, I., Ellermann, K., Ewans, K., Hermanski, G., Johnson, M., Ma, N., Maisondieu, C.,

- Nilva, A., Rychlik, I., Waseda, T., 2014. Recent developments of ocean environmental description with focus on uncertainties. *Ocean Engineering* 86, 26–46.
- Chang, C.-C., Lin, C.-J., 2011. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology* 2, 27:1–27:27, software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm> visited on 01/09/2014.
- Chaouachi, A., Nagasaka, K., 2012. A novel ensemble neural network based short-term wind power generation in a microgrid. *ISESCO Journal of Science and Technology* 8 (14), 2–8.
- Chen, B.-J., Chang, M.-W., Lin, C.-J., 2004. Load forecasting using support vector machines: A study on EUNITE Competition 2001. *IEEE Transactions on Power Systems* 19 (4), 1821–1830.
- Costa, A., Crespo, A., Navarro, J., Lizcano, G., Madsen, H., Feitosa, E., 2008. A review on the young history of the wind power short-term prediction. *Renewable and Sustainable Energy Reviews* 12 (6), 1725–1744.
- Fierro, R., Lewis, F., 1998. Control of a nonholonomic mobile robot using neural networks. *IEEE Transactions on Neural Networks* 9 (4), 589–600.
- Haykin, S., 1994. *Neural networks: a comprehensive foundation*. Prentice Hall PTR.
- Li, G., Shi, J., 2010. On comparing three artificial neural networks for wind speed forecasting. *Applied Energy* 87 (7), 2313–2320.
- Mabel, M. C., Fernandez, E., 2008. Analysis of wind power generation and prediction using ann: A case study. *Renewable Energy* 33 (5), 986 – 992.
- Morgan, N., Bourlard, H. A., 1995. Neural networks for statistical recognition of continuous speech. *Proceedings of the IEEE* 83 (5), 742–770.

- Muller, K.-R., Smola, A., Rtsch, G., Schlkopf, B., Kohlmorgen, J., Vapnik, V., 1997. Predicting time series with support vector machines. *Artificial Neural Networks ICANN'97* 1327, 999–1004.
- Park, S., Lee, J., Kim, S., 2004. Content-based image classification using a neural network. *Pattern Recognition Letters* 25 (3), 287–300.
- Philpott, A., Mason, A., 26-27 Jan 2001. Optimising yacht routes under uncertainty. In: 15th Chesapeake Sailing Yacht Symposium. Society of Naval Architects and Marine Engineers, Annapolis, MD, pp. 89–98.
- Schölkopf, S. P., Vapnik, V., Smola, A., 1997. Improving the accuracy and speed of support vector machines. *Advances in neural information processing systems* 9, 375–381.
- Vapnik, V., 2000. *The nature of statistical learning theory*. Springer.