Improved Multi-beam Neural Network Scatterometer Forward Models NCRG/2000/008

Dan Cornford, Ian T. Nabney, Guillaume Ramage
Neural Computing Research Group, Aston University, Birmingham B4 7ET, UK

Abstract

Current methods for retrieving near surface winds from scatterometer observations over the ocean surface require a foward sensor model which maps the wind vector to the measured backscatter. This paper develops a hybrid neural network forward model, which retains the physical understanding embodied in CMOD4, but incorporates greater flexibility, allowing a better fit to the observations. By introducing a separate model for the mid-beam and using a common model for the fore- and aft-beams, we show a significant improvement in local wind vector retrieval. The hybrid model also fits the scatterometer observations more closely. The model is trained in a Bayesian framework, accounting for the noise on the wind vector inputs. We show that adding more high wind speed observations in the training set improves wind vector retrieval at high wind speeds without compromising performance at medium or low wind speeds.

1. Introduction

Obtaining wind vectors over the ocean is important to Numerical Weather Prediction (NWP) since the ability to produce a forecast of the future state of the atmosphere depends critically on knowing the current state accurately [*Haltiner and Williams,* 1980]. However, the observation network over the oceans (particularly in the southern hemisphere) is very limited [*Daley,* 1991]. Thus it is hoped that the global coverage of ocean wind vectors provided by satellite borne scatterometers [*Offiler,* 1994] will improve the accuracy of weather forecasts by providing better initial conditions for NWP models [*Lorenc et al.,* 1993]. The scatterometer data also offers the ability to improve wind climatologies over the oceans [*Levy,* 1994] and the possibility of studying, at high resolution, interesting meteorological features such as cyclones [*Dickinson and Brown,* 1996].

This study uses scatterometer data from the ERS-2 satellite; the on-board vertically polarised, microwave radar operates at 5.3 GHz and measures the backscatter from gravity-capillary waves on the ocean surface of around 5 cm wavelength. Measured backscatter from the ocean surface is given as the normalised radar cross section, generally denoted by σ^o, and has units of decibels[1]. A 500 km wide swathe is swept by the satellite to the right of the track of its polar orbit, with nineteen cells sampled across the swathe, each cell having dimensions of roughly 50 by 50 km. Thus there is some overlap between cells. Each cell is sampled from three different directions by the fore, mid, and aft beams respectively, giving a triplet, $\boldsymbol{\sigma^o} = (\sigma^o_{\rm f}, \sigma^o_{\rm m}, \sigma^o_{\rm a})$. This $\boldsymbol{\sigma^o}$ triplet, together with the incidence and azimuth angles of the beams (which vary across the swathe) is related to the average wind vector, (u, v), within the cell [*Offiler,* 1994]. We assume that the stability of the lower boundary layer and the effects of longer sea waves are largely related to wind speed and thus their impact is implicitly included in the empirical models. Other geophysical parameters such as rain, sea ice are believed to also have a small affect on the backscatter [*Stoffelen* 1998a]; however these are treated as additional noise sources in this paper since we have no independent measurements of them.

Section 2 reviews the current scatterometer forward models, while the neural network forward models are introduced in Section 3. The effect of training a non-linear model while accounting for noise on the inputs, (u, v), is discussed, as is data selection for training the model and the estimation method itself. Section 4 compares the performance of the neural network models with CMOD4 (the current operational model) using visualisation, distance to cone and wind retrieval. The results are summarised in Section 5 and conclusions are given in Section 6.

[1]We shall always assume we are working in decibel (or log) space where we write σ^o or $\sigma^o_{\rm dB}$ if the distinction is important. We use $\sigma^o_{\rm lin}$ to denote the raw measurement space, $\sigma^o = \sigma^o_{\rm dB} = 10\log_{10}(\sigma^o_{\rm lin})$.

2. Scatterometer Forward Models

Understanding the theoretical relation between σ^o and (u, v) is essential to retrieving wind vectors from scatterometer observations [*Offiler,* 1994]. The relation has been modelled based on both on studies of the physical processes that govern backscattering from water surfaces [*Ebuchi et al.,* 1993; *Janssen et al.,* 1998] and statistical analysis of the relation between wind vectors (both buoy observed and NWP derived) and scatterometer measurements [*Stoffelen and Anderson,* 1997a]. From these studies empirical forward models relating σ^o and (u, v) have been established of the general form:

$$\begin{aligned}\sigma^o_{\rm lin} \sim\ & b_0(s,\theta) + b_1(s,\theta)\cos(\chi) \\ & + b_2(s,\theta)\cos(2\chi)\end{aligned} \tag{1}$$

where the wind vector (relative to the satellite azimuth angle) is given in terms of wind speed, s, and relative wind direction, χ. θ denotes the beam incidence angle. Since there are three σ^o measurements for each observation, this functional form implies a double skinned cone-like response in $\boldsymbol{\sigma^o}$ space. Considering a point on the surface of the cone, the distance along the axis of the cone is largely related to wind speed, while the location around the cone is related to wind direction. The $\cos(2\chi)$ term dominates and, together with the presence of noise, is the source of the direction ambiguities in the solutions. The most widely used operational forward model is known as CMOD4 [*Stoffelen and Anderson,* 1997a] and has the form:

$$\begin{aligned}\sigma^o_{\rm lin} = B_0(s,\theta)[1 & + B_1(s,\theta)\cos(\chi) \\ & + B_3(s,\theta)\tanh(B_2(s,\theta))\cos(2\chi)]^{1.6}\end{aligned} \tag{2}$$

where the result is raised to the power 1.6 in order to make the dependence of $\sigma^o_{\rm lin}$ on χ a function of $\cos(\chi)$ and $\cos(2\chi)$ only.

This paper presents results which show improved performance of the neural network forward models

both in terms of wind vector retrieval and of their representation of the observation manifold in σ^o space. If the forward models are to be used directly in a variational data assimilation system (that is assimilate σ^o rather than retrieved (u, v)) both attributes will be important.

CMOD4 is the operational model and, as such, provides the benchmark by which other models may be measured. However, operationally CMOD4 is used together with some empirical corrections; for instance the UK Meteorological Office increase the retrieved wind speed by five percent. The VIERS-1 physically based theoretical ocean backscatter model [*Janssen et al.,* 1998] is shown to improve upon wind vector retrieval at high wind speeds when compared with CMOD4, although it does not fit the observed $\boldsymbol{\sigma^o}$ manifold as well as CMOD4. The VIERS-1 model was not available to us for comparison.

3. Neural Network Scatterometer Forward Models

One of the reasons that CMOD4 fits the $\boldsymbol{\sigma^o}$ observations poorly at high wind speeds is the restrictive functional form imposed by the use of up to second order Legendre polynomials in the parameterisation of $B.(s, \theta)$ in Equation 2. We relax the restrictions imposed by the functional form and produce an alternative model using a Multi-Layer Perceptron (MLP) with the CMOD4 functional form to produce a more flexible, hybrid model.

3.1. Neural Networks

Neural networks are universal, non-linear function approximators which can approximate any continuous mapping to arbitrary accuracy, given sufficient hidden units. The MLP used is a non-linear statistical model, which has the advantage of being efficient to train, due to the back-propagation method for determining derivatives of the outputs (and thus the cost function) with respect to the weights [*Bishop,* 1995]. The MLP parameters are generally referred to as weights can be determined using standard gradient based algorithms to minimise a cost function. More details of the implementation are given below. Further details on neural networks can be found in *Bishop* [1995].

3.2. Hybrid model

In order to make use of existing knowledge on the physics of backscattering we imposed constraints on the functional form of the neural network model. A obvious functional form for the hybrid model is:

$$\begin{aligned}\sigma^o_{\text{lin}} =& a_0(1 + 0.37\tanh(a_1)\cos(\chi) \\ &+ 0.62\tanh(a_2)\cos(2\chi))^p \,,\end{aligned} \tag{3}$$

where p, a_0, a_1, and a_2 are functions of the model inputs and $\tanh(\cdot)$ is used to ensure the expression remains real for all inputs. The values 0.37 and 0.62 are simply scaling parameters chosen so as to sum to a value less than one. Their relative values have little importance as the network weights allow rescaling. Now, σ^o_{dB} can be written in log space:

$$\begin{aligned}\sigma^o_{\text{dB}} = \tfrac{10}{\ln(10)}\Big(a_0 + p\ln\big[1 + 0.1\tanh(a_1)\cos(\chi) \\ + 0.8\tanh(a_2)\cos(2\chi)\big]\Big) \,,\end{aligned} \tag{4}$$

this model being referred to as NN2CMOD. The model is shown graphically in Figure 1. The MLP takes the wind speed and sine of the beam incidence angle as inputs. The inputs were chosen to keep the mapping as simple as possible. The outputs are $[a_0, a_1, a_2, p]$ which, together with the relative wind direction, χ, are then used with Equation (4) to yield the backscatter measurement in decibels. We have chosen to use σ^o in decibels at all times since this renders the multiplicative noise on σ^o_{lin} additive on σ^o_{dB}. There remains the question on the form of the noise distribution in σ^o_{dB} space, which is discussed later.

3.3. Multi-beam model

During model validation (see Section 4, Figures 4 and 5) it became apparent that the mid-beam antenna σ^o value was not being well modelled by either CMOD4 or NN2CMOD. This was not due to the inability of the models to represent the relation between σ^o and θ at low incidence angles, this being verified by the use of a more flexible neural network with 12 hidden units, which exhibited the same features. Although there is no physical basis for this, a three beam model was constructed, with two NN2CMOD models with four hidden units in the MLP, one for the mid-beam and one for both fore- and aft-beams. This model, denoted NN3CMOD, has the advantage that during the training process three σ^o measurements are used to infer the 'true' wind vector, as opposed to one during the training of NN2CMOD.

3.4. Bayesian Parameter Estimation in the Presence of Input Noise

NN2CMOD depends upon weights, $\boldsymbol{w}$, which are determined from the training data. We adopt a prag-

matic Bayesian approach for the estimation of the weight vector in the presence of input noise, details of which can be found in *Cornford et al.,* [1999b] and *Wright* [1998]. If the input noise is not properly accounted for then non-linear models will learn a biased estimate of the true underlying function.

Using Bayes' theorem the posterior distribution of the weights given the noisy training data, $p(\boldsymbol{w} \mid \boldsymbol{D})$, can be expanded as:

$$\int_{\tilde{\boldsymbol{x}}_{\boldsymbol{n}}} \prod_n \underbrace{p(t_n \mid \tilde{\boldsymbol{x}}_{\boldsymbol{n}}, \boldsymbol{w})}_{p_1} \underbrace{p(\boldsymbol{x}_{\boldsymbol{n}} \mid \tilde{\boldsymbol{x}}_{\boldsymbol{n}})}_{p_2} \underbrace{p(\tilde{\boldsymbol{x}}_{\boldsymbol{n}})}_{p_3} \underbrace{p(\boldsymbol{w})}_{p_4} \, d\tilde{\boldsymbol{x}}_{\boldsymbol{n}} \,, \tag{5}$$

where $\boldsymbol{D}$ is the noisy training set, $\boldsymbol{D} = \{t_n, \boldsymbol{x}_{\boldsymbol{n}}\}$, t_n are the (noisy) targets in the training data, $\boldsymbol{x}_{\boldsymbol{n}}$ are the corresponding noisy inputs, and $\tilde{\boldsymbol{x}}_{\boldsymbol{n}}$ are the associated noiseless inputs. Training the network consists of determining the maximum *a posteriori* probability (MAP) weight vector and noiseless inputs, by minimising the negative logarithm of Equation (5). Here we are making a sub-optimal choice since ideally we should sample $\boldsymbol{w}$ from the distribution $p(\boldsymbol{w} \mid \boldsymbol{D})$ and use the samples to approximate the predictive integral:

$$p(t^* \mid \boldsymbol{x}^*) = \int_{\boldsymbol{w}} p(t^* \mid \boldsymbol{x}^*, \boldsymbol{w}) p(\boldsymbol{w} \mid \boldsymbol{D}) \, d\boldsymbol{w} \,, \tag{6}$$

where $\boldsymbol{x}^*$ is a new noise-free input and t^* is the corresponding predicted target. However in operational use this fully Bayesian approach would be too time consuming.

In order to evaluate the maximum *a posteriori* probability value of $p(\boldsymbol{w} \mid \boldsymbol{D})$, we compute four errors $E_i = -\ln(p_i)$, derived from Equation (5). Writing $t_n = \{\sigma^o\}$, $\boldsymbol{x}_{\boldsymbol{n}} = \{s, \chi, \theta\}$ and $\tilde{\boldsymbol{x}}_{\boldsymbol{n}} = \{\tilde{s}, \tilde{\chi}, \theta\}$ these terms become:

$E_1 = -\ln(\prod_n p_1(\sigma^o | \tilde{s}, \tilde{\chi}, \theta, \boldsymbol{w}))$, the error of the model, calculated using the observed satellite measurements and modified wind vectors $(\tilde{s}, \tilde{\chi})$ which tend to the noise free ('true') values during training. The distribution of p_1 is assumed to be Gaussian in σ^o_{dB} space, thus:

$$\begin{aligned} E_1 = & \frac{1}{2} \left(\log(2\pi) + \log(\sigma_t^2) \right) \\ & + \sum \left(f(\tilde{s}, \tilde{\chi}, \theta; \boldsymbol{w}) - t_n \right)^2 / (2\sigma_t^2) \,, \end{aligned} \tag{7}$$

where the sum is over all patterns in the training set, σ_t^2 is the variance of the errors in the t_n (target) measurements and $f(\tilde{s}, \tilde{\chi}, \theta; \boldsymbol{w})$ is the output obtained by propagating the noise free inputs $(\tilde{s}, \tilde{\chi})$ and θ through the model. Note when training NN3CMOD this will be the sum of three such terms, one for each antenna.

$E_2 = -\ln(\prod_n p_2(s, \chi | \tilde{s}, \tilde{\chi}))$, the error due to the noise free wind vectors differing from the corresponding noisy wind vector. The distribution p_2 is assumed to be spherically Gaussian with variance σ_u^2:

$$\begin{aligned} E_2 = & \log(2\pi) + \log(\sigma_u^2) + \\ & \sum \left((\tilde{u} - u)^2 + (\tilde{v} - v)^2 \right) / (2\sigma_u^2) \,, \end{aligned} \tag{8}$$

the summation being over the patterns. Note $\tilde{u} = \tilde{s} \sin(\tilde{\chi})$, $\tilde{v} = -\tilde{s} \cos(\tilde{\chi})$ following the meteorological convention. This component of the cost function could also represent the discrepancy between the ECMWF 10 m wind vector and the local surface stress vector, which is what actually generates the ocean surface ripples [*Stoffelen* 1998a].

$E_3 = -\ln(p_3(\tilde{s}, \tilde{\chi}))$, the assumed prior distribution of noise free wind vectors in the training set. In practice we rarely know the true distribution of the wind vectors so in this case we assume a constant prior distribution (that is a uniform distribution in wind speed and direction). This is a reasonable assumption because of the data selection method used, but future work could investigate the effect of this assumption.

$E_4 = -\ln(p_4(\boldsymbol{w}))$ is the prior over the weights which controls the complexity of the MLP [*Bishop,* 1995]. The weight decay prior:

$$E_4 = \sum_{\boldsymbol{w}} \boldsymbol{w}^2 / (2\sigma_{\boldsymbol{w}}^2) \tag{9}$$

is used, where $\sigma_{\boldsymbol{w}}^2$ is the variance of the weights, which were fixed on the basis of experimentation to be 0.005 for the weights and 0.1 for the biases. The effect of this term is to produce smoother network mappings as the weight variance is decreased.

This is very similar to the cost function used to determine the parameters of CMOD4 in *Stoffelen and Anderson* [1997a], with the addition of a prior model for the distribution of the true wind vectors and a prior model for the weight vector. Finding the MAP solution is essentially the same as a variational determination of the weights. We used 20,000 iterations of scaled conjugate gradient optimisation to determine the MAP weight values to ensure convergence, particularly in the estimation of the noise free input values.

3.5. Data Selection

When using data driven models, the quality of the trained model is only as good as the data used to

train it. It is possible to bring additional information to model determination, such as using certain model classes, within the Bayesian framework adopted, but we still depend critically on careful data selection.

We have used ERS-2 data collected over the period March 1996 to January 1998 in the Northern Hemisphere to create our training sets. The ERS-2 data was collocated with European Centre for Medium Range Weather Forecasting (ECMWF) 10m wind vectors by the French Research Institute for the Exploitation of the Sea (IFREMER)[2]. The ECMWF wind vectors assimilated CMOD4 retrieved scatterometer winds. If the data assimilation system of the ECMWF model is working well then this will only improve the quality of the (u, v) data in the training set. To further improve the quality of the data set we carefully corrected the $\boldsymbol{\sigma^o}$ observations to account for calibration changes over the data acquisition period. We also insisted that the signal to noise ratio in the scatterometer observations was less than 7%.

We make the usual assumption that the observations in the training set are independent, thus we selected the observations so that they are separated in space by at least 300 km. This distance was chosen to achieve a compromise between independence and obtaining sufficient samples at high wind speeds. A filter computed the variance of the wind field within a circle of 1 degree of latitude or longitude. If the summed variance of the wind components was greater than 2.5 $m^2 s^{-2}$ the central wind vector was not selected, to reduce the impact of incorrectly positioned fronts and cyclones in the ECMWF model on the quality of the wind vectors in the training set. This variance was chosen on the basis of experimentation.

3.5.1. Outlier Removal. We further 'clean' the training data using an interactive, manual outlier removal procedure. We know that the noise on the $\boldsymbol{\sigma^o}$ observations is small thus visualisation in $\boldsymbol{\sigma^o}$ space can quickly identify outliers in terms of $\boldsymbol{\sigma^o}$, which we have found to be present in practice. As the assumed $\boldsymbol{\sigma^o}$ variance is very small these can have a very large effect on the trained models. By using three linked plots it was possible to eliminate the extreme $\boldsymbol{\sigma^o}$ outliers present in the dataset. A further two linked plots allowed us to examine outliers in wind speed and direction.

We considered each fixed the mid-beam incidence angle separately and plotted σ^o_{f} against σ^o_{a}, σ^o_{f} against σ^o_{m} and σ^o_{a} against σ^o_{m} in σ^o_{dB} space. We also plotted NWP wind speed against $(\sigma^o_{\mathrm{f}} + \sigma^o_{\mathrm{m}} + \sigma^o_{\mathrm{a}})/3$ and NWP wind direction against $\sigma^o_{\mathrm{f}} - \sigma^o_{\mathrm{a}}$. To help find directional errors we highlighted those points for which $(\sigma^o_{\mathrm{f}} + \sigma^o_{\mathrm{m}} + \sigma^o_{\mathrm{a}})/3 > \min(\sigma^o_{\mathrm{f}} + \sigma^o_{\mathrm{m}} + \sigma^o_{\mathrm{a}})/3 + 8$, since these are the points with higher wind speeds and thus should have a more consistent relation to wind direction. During outlier removal 1.9% of the training data was rejected.

3.6. Parameter Estimation

We used estimates in *Stoffelen and Anderson* [1997a] to set the error variances on the σ^o and the wind vectors. We used values of $\sigma_l^2 = 0.04\ dB^2$ and $\sigma_u^2 = 2.25\ m^2 s^{-2}$ in our training (see Equations (7) and (8)). To verify these assumptions we trained models using these values and looked at their performance on an independent validation set, which had also undergone the process of outlier removal.

4. Validation of Forward Models

There are several measures which one might use when determining the performance of the various forward models. The natural choice, related to the error function used during training, is the root mean square error (RMSE) of the $\boldsymbol{\sigma^o}$ observations, given the 'true' wind vector. However the requirement for accurate wind vector retrieval means that the vector RMSE of the retrieved (u, v) observations is more important. Other measures include biases in the models, accuracy of the first (most probable) solution or the so called Figure of Merit, as proposed by David Offiler of the UK Meteorological Office (UKMO) [*Cornford et al.,* 1999]. We present a wide range of performance indicators to allow a complete assessment of the models. In common with standard practice, since we are interested in the quality of the local models, when computing error measures in (u, v) space, we pick the wind vector (from the 2 to 4 returned) that is closest to the NWP winds [*Offiler,* 1994]. We employed three validation methods for our models: visualisation, quantitative measures in $\boldsymbol{\sigma^o}$ space and quality of locally retrieved winds.

4.1. Visualisation

This section qualitatively assesses the degree to which the models fit the $\boldsymbol{\sigma^o}$ observations. This has proved to be a powerful tool for the rapid examination of forward models, and can be instructive in suggesting where improvements may be necessary. The

[2]See `http://www.ifremer.fr/cersat/ACTIVITE/E_CERACT.HTM` for details.

results can be seen in Figures 3 and 4, where the outline of the model manifolds are plotted over the range 2–28 ms^{-1}, viewed from above (in the plane $\sigma^o_{\rm f}$, $\sigma^o_{\rm a}$) and the side (in the plane $(\sigma^o_{\rm f} + \sigma^o_{\rm a})/2$, $\sigma^o_{\rm m}$). Every $\boldsymbol{\sigma^o}$ observation from an independent validation set of ERS-2 observations is also plotted for the given incidence angle, which thus fill the 'centre' of the models. CMOD4 can be seen to fit the $\boldsymbol{\sigma^o}$ observations well at larger incidence angles, but is not sufficiently flexible at low incidence angles. The model extends beyond the regions within which $\boldsymbol{\sigma^o}$ is observed, mainly at higher winds speeds, which correspond to greater $\boldsymbol{\sigma^o}$ values. NN2CMOD fits the $\boldsymbol{\sigma^o}$ observation well at higher wind speeds, but does not fit well at lower wind speeds. This is related to the difference between the behaviour of the mid-beam compared with the fore- and aft-beams. NN3CMOD, which uses separate models for these beams, fits the $\boldsymbol{\sigma^o}$ observations very well.

4.2. Distance to Model Cone - Validation in $\boldsymbol{\sigma^o}$ Space

In order to obtain quantitative results on the fit of the models in $\boldsymbol{\sigma^o}$ space we have looked at the distance to the cone for a validation set which has the same distribution in wind speed and direction as the training set. The results can be seen in Table 1 where the Jacobians of the models were used to determine the exact minimum distance to the cone using a scaled conjugate gradient minimisation algorithm. The minimum distance to the cone, which corresponds to the minimum distance of the $\boldsymbol{\sigma^o}$ observation from the model manifold shows that NN3CMOD fits the cone more tightly than CMOD4 and NN2CMOD. The distance from the point on the cone corresponding to the retrieved wind vector closest to the NWP wind vector, denoted dist(best) also shows the improvement given by NN3CMOD. The variances of the $\boldsymbol{\sigma^o}$ errors on the individual beams were also calculated for the three models.

In Table 1 this shows that the fore- and aft-beams have smaller noise levels than the mid-beam, for all models. For CMOD4 and NN2CMOD, Figure 5 shows this is related to the poor fitting (bias) of the models to σ^o at small incidence angles (the mid-beam incidence angles are generally smaller). However for NN3CMOD the mid-beam σ^o still has a higher variance, despite an unbiased fit to the σ^o observations (Figure 5c). Figure 5c shows no evidence of a systematic dependence on θ, rather there seems to be a distinctly different variance for the mid-beam σ^o. This suggests that it would is preferable to have a separate model for the mid-beam and a joint model for the fore- and aft-beams, as done in NN3CMOD.

Figure 6 shows the misfit of models in $\boldsymbol{\sigma^o}$ space plotted as a function of retrieved wind speed. CMOD4 shows a great deal more scatter than the neural network models, particularly at wind speeds above 8 ms^{-1}. The results for NN3CMOD (Figure 6c) suggest that the variance of the σ^o observations decreases with increasing wind speed, however reference to Figure 3 suggests this may partly reflect the poor fit of all models at low σ^o values, which correspond to lower wind speeds. The misfit in $\boldsymbol{\sigma^o}$ space as a function of retrieved relative wind direction can be seen in Figure 7 and show that all models are relatively insensitive to wind direction, with NN3CMOD exhibiting the closest fit to the data.

4.3. Local Wind Retrieval - Validation in (u, v) Space

Since the forward models will ultimately be used for wind vector retrieval, it is this evaluation measure that is the most important from a user perspective. In this section we present the results of the local retrieval of wind vectors using the forward models. The models are inverted using their Jacobians as done on the validation set. As the ECMWF wind vectors used in training the models already have some influence from CMOD4, an independent test set of $\boldsymbol{\sigma^o}$ measurements was used. This test set used UK Meteorological Office (UKMO) first guess at appropriate time winds as targets. These are unified model zero to six hour forecast winds [*Andrews and Bell,* 1998], interpolated to the $\boldsymbol{\sigma^o}$ observation locations. Three days of global scatterometer observations from 10/6/98, 25/1/99 and 7/2/99 were randomly subsampled, to provide the test set of 60,000 measurements, with a distribution similar to that observed in the atmosphere.

Table 2 shows the results on the test set. There has been no selection of the data which is collected in both Northern and Southern hemispheres. The VRMSE of the CMOD4+5% retrieval is larger than that of the neural network models by some 0.5 ms^{-1} which is a large margin, and certainly statistically significant with over fifty thousand observations. This figure must be interpreted carefully since on average CMOD4+5% returned 2.24 solutions per $\boldsymbol{\sigma^o}$ observation, while NN2CMOD returned 2.36 and NN3CMOD 2.33. A negative bias in wind speed remains in CMOD4+5% despite the 5% correction applied to the wind speed. The bias of the neural network models

is small, suggesting that the parameterisations of the lower boundary layers in the ECMWF and UKMO numerical models are similar, and thus our models could be used consistently with the UKMO unified model or the ECMWF model.

The direction biases are similar and small for all models, but CMOD4+5% has a larger direction standard deviation. Both neural network models consistently have much better performance in terms of getting the first (most probable) solution within 20° of the NWP wind vector compared with CMOD4+5%. This is probably related to the lower RMSE in $\boldsymbol{\sigma^o}$ space, and illustrates that a better fitting in $\boldsymbol{\sigma^o}$ space is important for (u, v) retrieval, particularly in ambiguity removal. This measure will not be affected by the number of solution returned. The Figure of Merit, which can be used to assess different models on many criteria, shows that NN2CMOD and NN3CMOD are very similar in overall performance on wind vector retrieval and better by 20% than CMOD4+5%.

5. Discussion

When tuning a non-linear model, data selection and quality control is very important. Although interactive data manipulation demands anecessitated large amount of user time, it can greatly improve the modelling exercise. The more flexible the model, the more important is data integrity. Despite this reduction of errors in the (u, v) inputs, it was necessary to train the neural network models, using a Bayesian procedure to learn both the forward model parameters and the 'true' (u, v) values. If standard training was used on the neural network models (that is disregarding input noise) the results, both in terms of fit to the $\boldsymbol{\sigma^o}$ observations and (u, v) retrieval were very poor.

The fit of the models in $\boldsymbol{\sigma^o}$ space shows that NN3CMOD fits better than both NN2CMOD and CMOD4 which is attributed to the use of a different model of the mid-beam in NN3CMOD. This is confirmed by visualisation, where it can be seen that NN3CMOD fits the $\boldsymbol{\sigma^o}$ observations well, although there remains room for improvement at low wind speeds, which could be investigated in further work.

Local wind vector retrieval is improved using the neural network models. This is related to their ability to fit the $\boldsymbol{\sigma^o}$ observations better but can be largely attributed to the Bayesian training procedure used to minimise the impact of input noise on the model parameters. Although NN3CMOD fits much better in $\boldsymbol{\sigma^o}$ space compared with NN2CMOD, the retrieval in (u, v) space is only marginally better. However,NN3CMOD has greater skill in determining which of the ambiguous solutions is the 'true' solution. Data assimilation systems which assimilate $\boldsymbol{\sigma^o}$ rather than (u, v) will be more accurate if the $\boldsymbol{\sigma^o}$ fit of the model is more accurate. Even for data assimilation systems which assimilate retrieved (u, v), the accurate fit of NN3CMOD will improve the estimation of (u, v) directly but also improve the estimate of the probability of each ambiguous solution which can be fed to the data assimilation system (or ambiguity removal algorithm).

Figure 8 shows the effect of using different distributions of wind speed in the training set. Three NN3CMOD models where trained using different training sets. The results illustrate that models trained with a near uniform distribution in wind speed (that is with more cases in the higher wind speed range) perform better when retrieving winds at higher wind speeds, but slightly worse when retrieving lower speed winds. It is also clear that mixing the training sets allows the model to learn well at both high and low wind speeds without compromising performance elsewhere. This suggests that on-line learning strategies could be used to enhance these models performance at high wind speeds, which is currently limited by data availability. The error for CMOD4+5% is also shown, illustrating the improvement in performance of NN3CMOD especially at higher wind speeds.

6. Conclusions

This paper has discussed two novel neural network based scatterometer forward models. An interactive outlier removal method was used with careful data selection but this still the use of a training method that accounted for the input noise in the 'cleaned' NWP wind vectors. Visualisation was used in a preliminary assessment of model accuracy in $\boldsymbol{\sigma^o}$ space, and in interpreting later results.

Using the model Jacobians we have shown that the neural network models fit the $\boldsymbol{\sigma^o}$ observations better than CMOD4. We also show that the mid-beam antenna has a different response to the fore- and aft-beams. This strongly suggests that a different model is required for the mid-beam σ^o measurements, although a joint model can be used for the fore- and aft-beams, as implemented in NN3CMOD.

The neural network models are shown to be more accurate for wind vector retrieval. The neural network models are unbiased with respect to wind speed

and direction retrieval on the ECMWF and UKMO datasets and have lower direction standard deviations compare with CMOD4+5%. When using look up tables the models will take the same amount of time to invert, and thus on the basis of performance the hybrid neural network model, NN3CMOD, might be preferred for operational use. In terms of the cost of inverting the models using the Jacobians NN3CMOD requires about 30% more floating point operations per pattern, compared to CMOD4.

Future work could consider a better model for the $\sigma^o_{\rm dB}$ error which should improve the fit of the model. Better understanding of the impact of the input noise in (u, v) (and its distribution) will also improve the model training.

Acknowledgments. The authors wish to thank David Offiler of the UKMO for his input. Ad Stoffelen, R.N. Hoffman and S.M. Leidner have greatly improved this paper through their constructive reviews. This work is funded by the European Union under contract ENV4–CT96–0314, and forms part of the NEUROSAT program.

References

Andrews, P.L. and Bell, R.S., Optimizing the United Kingdom Meteorological Office Data Assimilation for ERS-1 Scatterometer Wind. *Monthly Weather Review, 26*, 736–746, 1998.

Bishop, C. M., *Neural Networks for Pattern Recognition,* 482 pp., Oxford University Press, Oxford, 1995.

Cornford, D., G. Ramage, and I. T. Nabney, A Scatterometer Neural Network Sensor Model with Input Noise. *Neurocomputing Letters, 30*, 13–21, 2000.

Cornford, D., I. T. Nabney, and C. M. Bishop, Neural Network Based Wind Vector Retrieval from Satellite Scatterometer Data. *Neural Computing and Applications, 8*, 206–217, 1999.

Daley, R., *Atmospheric Data Analysis,* 457 pp., Cambridge University Press, Cambridge, 1991.

Dickinson, S. and R. A. Brown, A Study of Near-Surface Winds in Marine Cyclones Using Multiple Satellite Sensors, *Journal of Applied Meteorology, 35,* 769–781, 1996.

Ebuchi, N., H. Kawamura, and Y. Toba, Physical Processes of Microwave Backscattering from Laboratory Wind Wave Measurements, *Journal of Geophysical Research, 98,* 14669–14681, 1993.

Haltiner, G. J. and R. T. Williams, *Numerical Prediction and Dynamic Meteorology,* 477 pp., John Wiley, Chichester, 1980.

Janssen, P. A. E. M., H. Wallbrink, C. J. Calkoen, D. van Halsema, W.A. Oost, and P. Snoeij, VIERS-1 scatterometer model, *Journal of Geophysical Research, 103,* 7807–7831, 1998.

Levy, G., Southern-Hemisphere Low-Level Wind Circulation Statistics from the SeaSat Scatterometer. *Annales Geophysicae - Atmospheres, Hydroshperes and Space Sciences, 12,* 65–79, 1994.

Lorenc, A. C., R. S. Bell, S. J. Foreman, C. D. Hall, D. L. Harrison, M. W. Holt, D. Offiler, and S. G. Smith, The Use of ERS-1 Products in Operational Meteorology, *Advances in Space Research, 13,* 19–27, 1993.

Offiler, D., ERS-1 wind retrieval algorithms, *U.K. Meteorological Office Memorandum No. 86,* Meteorological Office, Bracknell, UK, 1987.

Offiler, D., The Calibration of ERS-1 Satellite Scatterometer Winds, *Journal of Atmospheric and Oceanic Technology, 11,* 1002–1017, 1994.

Ramage, G., Neural Networks for Modelling Wind Vectors, MSc Thesis, Aston University, 1998. Available from: `http://www.ncrg.aston.ac.uk/Projects/NEUROSAT`.

Stoffelen, A., *Scatterometry*, PhD Thesis, Universiteit Utrecht, 1998a.

Stoffelen, A., Error Modelling and Calibration; Towards the True Near Surface Wind Speed. *Journal of Geophysical Research, 103,* 7755–7766, 1998b.

Stoffelen, A. and D. Anderson, Scatterometer Data Interpretation: Estimation and Validation of the Transfer Function CMOD4. *Journal of Geophysical Research, 102,* 5767–5780, 1997a.

Stoffelen, A. and D. Anderson, Scatterometer Data Interpretation: Measurement Space and Inversion, *Journal of Atmospheric and Oceanic Technology, 14,* 1298–1313, 1997n.

Wright, W. A., Neural Network Regression with Input Uncertainty, In *Neural Networks for Signal Processing*, 8, 284–293, 1998.

D. Cornford and Ian T. Nabney and Guillaume Ramage, Neural Computing Research Group, Electronic Engineering and Computer Science, Aston University, Aston Triangle, Birmingham B7 4ET, UK (e-mail: d.cornford@@aston.ac.uk, i.t.nabney@@aston.ac.uk)

This preprint was prepared with AGU's LaTeX macros v5.01. File ncrg'tech'report'version formatted November 14, 2000.

Table 1. Fit of the models in $\boldsymbol{\sigma^o}$ space on a validation set of 15000 observations between 4 and 24 ms^{-1} which have undergone the manual outlier removal procedure.

			σ^o variance for beam		
	dist(min)[a] dB	dist(best)[b] dB	fore dB^2	mid dB^2	aft dB^2
CMOD4	0.29	0.48	0.104	0.181	0.105
NN2CMOD	0.31	0.45	0.056	0.152	0.056
NN3CMOD	0.22	0.29	0.025	0.096	0.025

[a]Mean minimum distance to the cone for all ambiguous wind vectors retrieved.

[b]Mean distance to the cone for the wind vector closest to the NWP wind.

Table 2. Performance of the models on the UKMO test set with an atmospheric distribution in wind speed and direction and 50,720 observations in the range 4–24 ms^{-1}. Wind speed is in ms^{-1} and wind direction is in degrees.

	vector RMSE	s bias	s std[a]	χ bias	χ std[a]	ONET[b]	FoM[c]
CMOD4+5%	3.26	−0.44	1.75	−0.9	22.4	32.0	1.07
NN2CMOD	2.76	−0.09	1.73	0.6	16.7	44.7	1.27
NN3CMOD	2.71	−0.19	1.71	0.7	16.3	51.1	1.29

[a]Standard deviation.

[b]Percentage of the most likely solutions within 20° of the NWP wind vector.

[c]The Figure of Merit as proposed by David Offiler of the UK Meteorological Office. This is unit-less, a value of 1 indicating that the scatterometer meets its design specifications, larger values reflecting better performance.

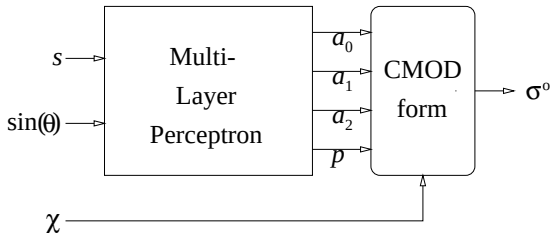


Figure 1. The hybrid neural network scatterometer model, symbols defined in the text.

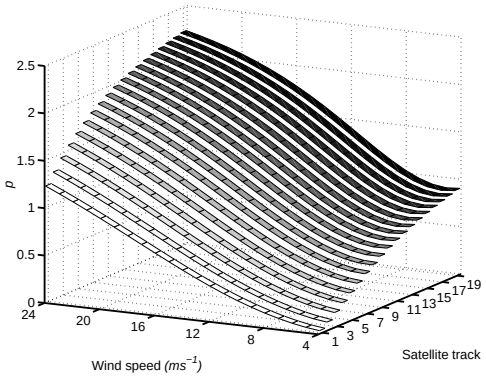


Figure 2. Graphical representation of the the power parameter p (Equation 3) of NN2CMOD as a function of wind speed and incidence angle.

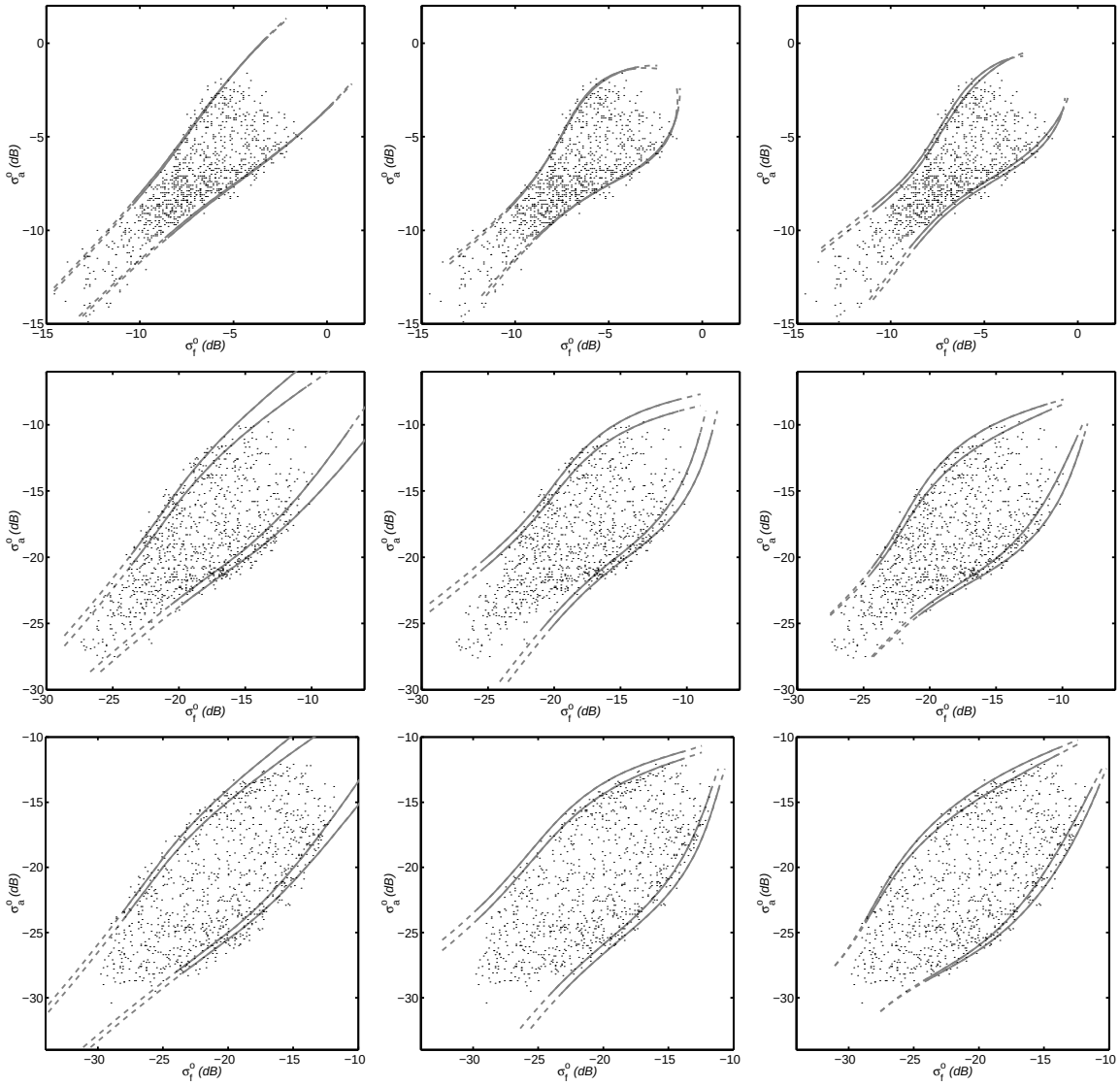


Figure 3. Visualising CMOD4 (left), NN2CMOD (middle) and NN3CMOD (right) using the 'top view' at mid-beam incidence angles of 18.0° (top), 33.4° (middle) and 45.4° (bottom). The solid line shows the model manifold for 4–24 ms^{-1}, the dotted line for 2–28 ms^{-1}. The small dots show every point in the validation set.

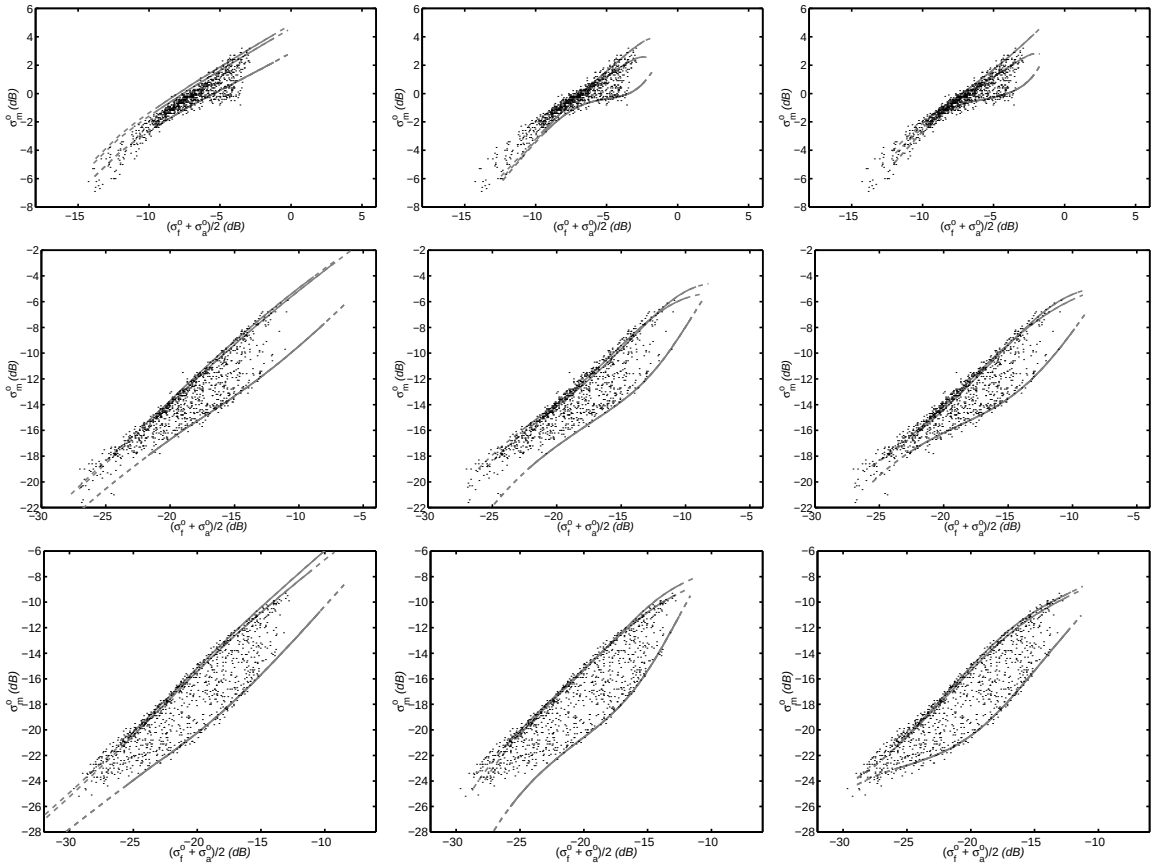


Figure 4. Visualising CMOD4 (left), NN2CMOD (middle) and NN3CMOD (right) using the 'side view' at mid-beam incidence angles of 18.0° (top), 33.4° (middle) and 45.4° (bottom). The solid line shows the model manifold for 4–24 ms^{-1}, the dotted line for 2–28 ms^{-1}. The small dots show every point in the validation set.

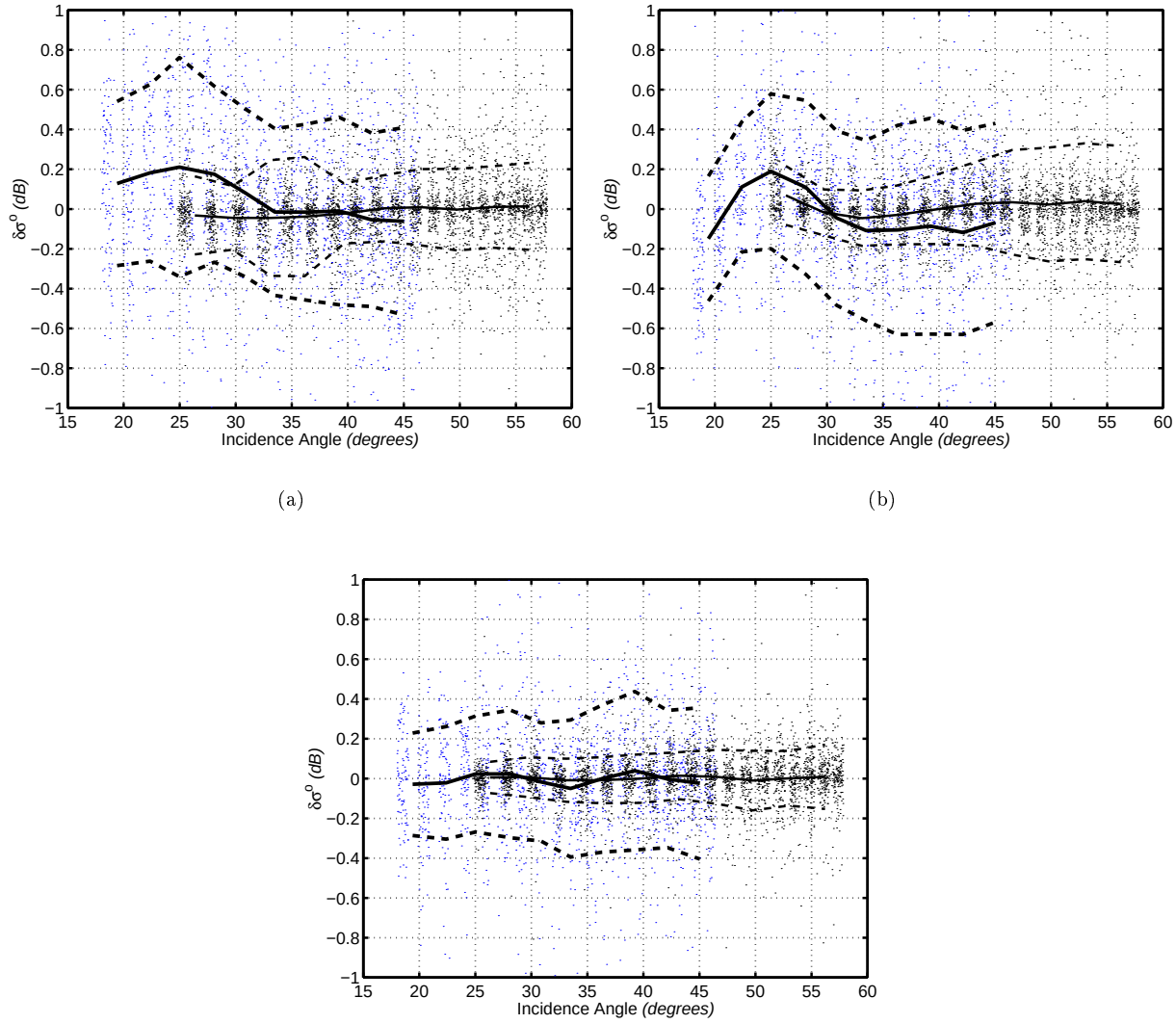
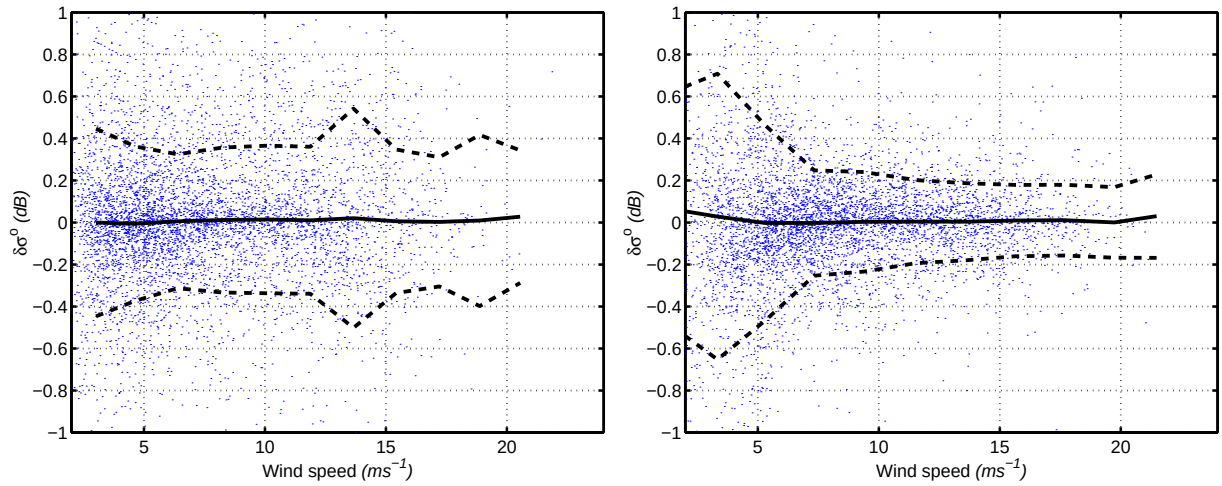
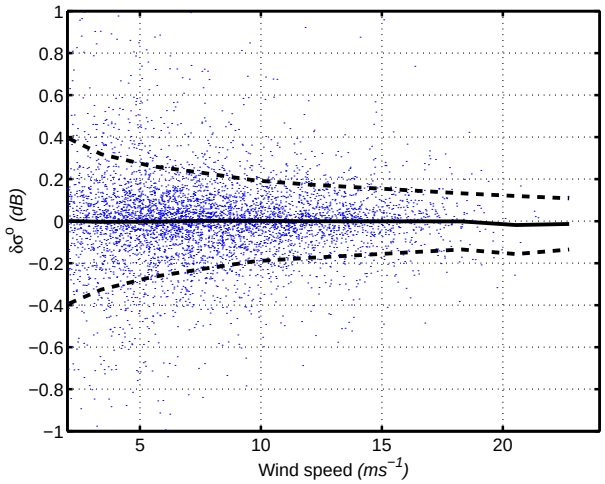


Figure 5. The residuals $\delta\sigma^o = \sigma^o_{predicted} - \sigma^o_{observed}$ plotted for every tenth point in the validation set as a function of incidence angle for a) CMOD4, b) NN2CMOD and c) NN3CMOD. The solid line gives the running mean, the dotted line ± one standard deviation. The thicker lines to the left are the mid-beam statistics, the thinner line to the right, the combined fore- and aft-beam statistics.



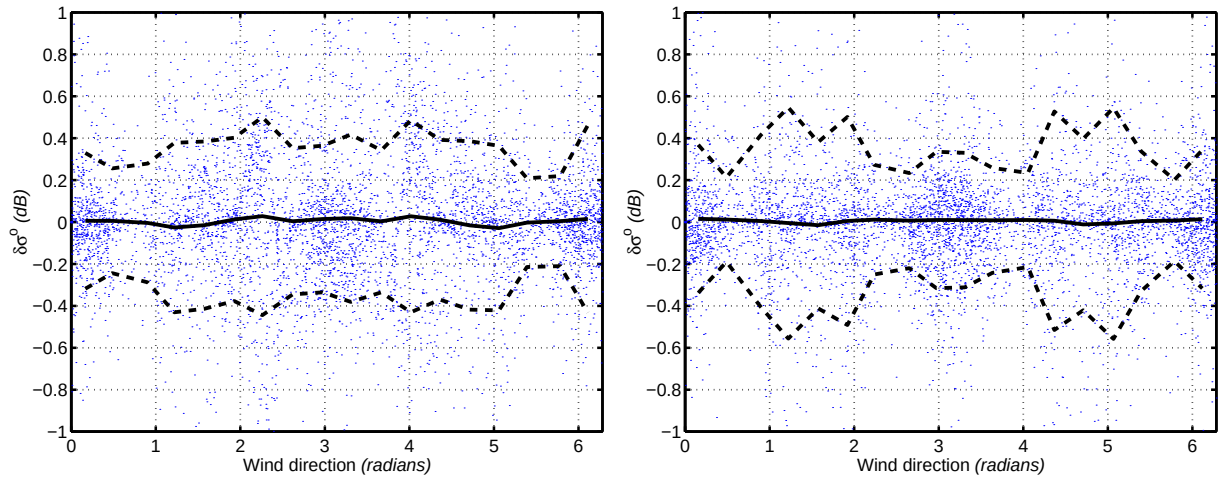
(a)

(b)



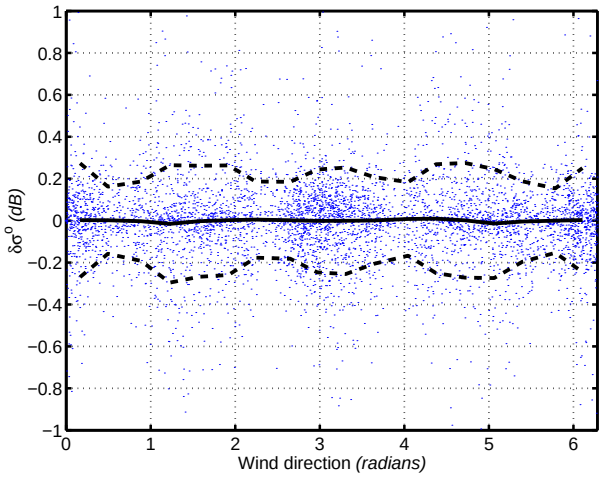
(c)

Figure 6. The residuals $\delta\sigma^o = \sigma^o_{predicted} - \sigma^o_{observed}$ plotted for every tenth point in the validation set as a function of retrieved wind speed for a) CMOD4, b)NN2CMOD and c) NN3CMOD. The solid line gives the running mean, the dotted line ± one standard deviation. All beams are considered together.



(a)

(b)



(c)

Figure 7. The residuals $\delta\sigma^o = \sigma^o_{predicted} - \sigma^o_{observed}$ plotted for every tenth point in the validation set as a function of retrieved wind direction (closest to NWP direction) for a) CMOD4, b) NN2CMOD and c) NN3CMOD. The solid line gives the running mean, the dotted line ± one standard deviation. All beams are considered together.

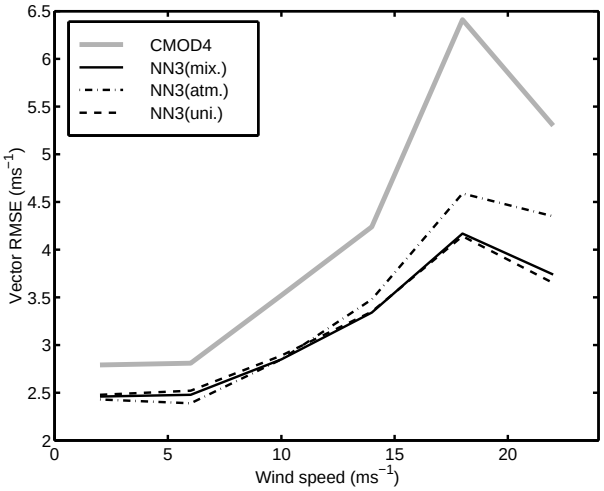


Figure 8. The vector RMSE as a function of retrieved wind speed, for CMOD4+5% and three versions of NN3CMOD trained using a mixed (mix.), atmospheric (atm.) and uniform (uni.) distribution of wind speed in the training set.