

# Improved neural network scatterometer forward models

Dan Cornford, Ian T. Nabney, and Guillaume Ramage

Neural Computing Research Group, Aston University, Birmingham, England, UK

**Abstract.** Current methods for retrieving near-surface winds from scatterometer observations over the ocean surface require a forward 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 midbeam 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 offer the potential of improved 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 onboard vertically polarized microwave radar operates at 5.3 GHz and measures the backscatter from gravity-capillary waves on the ocean surface of  $\sim 5$ -cm wavelength. Backscatter from the ocean surface is measured by the normalized radar cross section, generally denoted by  $\sigma^o$ , and has units of decibels. (We use  $\sigma_{\text{lin}}^o$  to denote the raw measurement space. We shall usually work in decibel (or log) space where measurements are denoted by  $\sigma^o$  or by  $\sigma_{\text{dB}}^o$  if the distinction is important. Note that  $\sigma_{\text{dB}}^o = 10 \log_{10}(\sigma_{\text{lin}}^o)$ .) A 500-km wide swathe is swept by the satellite to the right of the track of its polar orbit. There are 19 cells sam-

pled across the swathe, and each cell has dimensions of roughly 50 by 50 km, which implies that there is some overlap between cells. Each cell is sampled from three different directions by the fore, mid, and aft beams, giving a triplet,  $\sigma^\circ = (\sigma_f^\circ, \sigma_m^\circ, \sigma_a^\circ)$ . This  $\sigma^\circ$  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 and sea ice are also believed to have a small effect on the backscatter [Stoffelen, 1998]; 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 models are introduced in section 3. A method for training a nonlinear 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 visualization, distance to the model manifold, and wind retrieval. The results are summarized in section 5, and conclusions are given in section 6.

## 2. Scatterometer Forward Models

A theoretical understanding of the relation between  $\sigma^\circ$  and  $(u, v)$  is essential to retrieve wind vectors from scatterometer observations [Offiler, 1994]. The relation has been modeled based both on studies of the physical processes that govern backscattering from water surfaces [Ebuchi *et al.*, 1993; Janssen *et al.*, 1998] and on statistical analysis of the observed data [Stoffelen and Anderson, 1997a]. Statistical studies have established empirical forward models relating  $\sigma^\circ$  and  $(u, v)$  of the general form

$$\begin{aligned} \sigma_{\text{in}}^\circ \sim & b_0(s, \theta) + b_1(s, \theta) \cos(\chi) \\ & + b_2(s, \theta) \cos(2\chi), \end{aligned} \quad (1)$$

where the wind vector is expressed in terms of speed,  $s$ , direction relative to the satellite azimuth angle  $\chi$  and  $\theta$  denotes the beam incidence angle. The terms  $b_0$ ,  $b_1$ , and  $b_2$  are functions which encapsulate the dependency of  $\sigma^\circ$  on wind speed and incidence angle. Since there are three  $\sigma^\circ$  measurements for each satellite observation, this functional form implies a double-skinned cone-like

response in  $\sigma^\circ$  space [Thiria *et al.*, 1993]. 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 direction ambiguities in the retrieved wind vectors. The most widely used operational forward model is known as CMOD4 [Stoffelen and Anderson, 1997a] and has the form

$$\sigma_{\text{lin}}^\circ = B_0(s, \theta) \{1 + B_1(s, \theta) \cos(\chi) + B_3(s, \theta) \tanh[B_2(s, \theta)] \cos(2\chi)\}^{1.6}, \quad (2)$$

where the result is raised to the power 1.6 in order to make the dependence of  $\sigma_{\text{lin}}^\circ$  on  $\chi$  a function of  $\cos(\chi)$  and  $\cos(2\chi)$  only. The functions  $B_0$ ,  $B_1$ ,  $B_2$ , and  $B_3$  were estimated from ERS-1 data collocated with European Centre for Medium Range Weather Forecasting (ECMWF) wind vectors [Stoffelen and Anderson, 1997a]. In operation, empirical corrections are made to CMOD4; for instance, the UK Meteorological Office increases the retrieved wind speed by 5% to reduce bias (CMOD4+5%). The Verification and Interpretation of ERS-1 (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  $\sigma^\circ$  manifold as well as CMOD4. The VIERS-1 model was not available to us for comparison.

### 3. Neural Network Scatterometer Forward Models

Earlier work using neural networks modeled  $\sigma^\circ$  directly as a function of  $s$ ,  $\chi$ , and  $\theta$  [Mejia *et al.*, 1999] for the National Aeronautics and Space Administration scatterometer or to construct direct inverse models which infer  $s$  and  $\chi$  directly from  $\sigma^\circ$  and  $\theta$  [Richaume *et al.*, 2000] for ERS-1 observations. The advantage of using neural networks is that they do not require any knowledge of the functional form of the input–output relation; however, a disadvantage is that they cannot easily incorporate a priori knowledge about the input–output mapping, which often exists in geophysical applications. We employ a combination of neural network methods and knowledge of the functional form embodied in CMOD4 to produce an enhanced model. One of the reasons that CMOD4 fits the  $\sigma^\circ$  observations poorly at high wind speeds is the restricted functional form imposed by the use of second-order Legendre polynomials in the parameterisation of  $B_i(s, \theta)$  in (2). We

relax this restriction and create a model combining a multilayer perceptron (MLP) with the CMOD4 functional form to produce a more flexible, hybrid model.

### 3.1. Multilayer Perceptrons

MLPs are nonlinear statistical models which can approximate any continuous function to arbitrary accuracy, given sufficient hidden units. The MLP parameters, or weights, can be estimated from training data using standard gradient-based algorithms to minimize an appropriate cost function. The back-propagation method is used to determine the error derivatives with respect to the weights [Bishop, 1995].

### 3.2. Hybrid Model

In order to make use of existing knowledge about the physics of backscattering we imposed constraints on the functional form of the neural network model. The hybrid model is defined by

$$\sigma_{\text{lin}}^{\circ} = a_0 [1 + 0.37 \tanh(a_1) \cos(\chi) + 0.62 \tanh(a_2) \cos(2\chi)]^p, \quad (3)$$

where  $p$ ,  $a_0$ ,  $a_1$ , and  $a_2$  are functions of the model inputs and  $\tanh(\ )$  is used to ensure that the expression remains real for all inputs. The values 0.37 and 0.62 are simply scaling parameters chosen to ensure that the expression inside the brackets in (3) remains positive. Their relative values have little importance as the network weights allow rescaling. By taking logs we obtain a model for  $\sigma_{\text{dB}}^{\circ}$

$$\sigma_{\text{dB}}^{\circ} = \frac{10}{\ln(10)} \left\{ a_0 + p \ln [1 + 0.37 \tanh(a_1) \cos(\chi) + 0.62 \tanh(a_2) \cos(2\chi)] \right\}, \quad (4)$$

which we call NN2CMOD. The model is shown graphically in Figure 1. The MLP takes the log of the wind speed and the sine of the beam incidence angle as inputs. The inputs were chosen to reduce the nonlinearity of the mapping and ensure that only positive wind speeds can be retrieved. The outputs are  $[a_0, a_1, a_2, p]$ , which, together with the relative wind direction  $\chi$ , are then used with (4) to yield the backscatter measurement in decibels.

We have chosen to model  $\sigma_{\text{dB}}^{\circ}$  since this transforms the noise on  $\sigma_{\text{lin}}^{\circ}$ , whose variance is a function of  $\sigma_{\text{lin}}^{\circ}$ , into additive noise on  $\sigma_{\text{dB}}^{\circ}$  with a constant variance which makes the cost function more convenient [Stoffelen and Anderson, 1997a]. There remains the question of the form of the noise distribution in  $\sigma_{\text{dB}}^{\circ}$  space. It is assumed that the noise on  $\sigma_{\text{lin}}^{\circ}$  is Gaussian; thus the

Figure 1

noise on  $\sigma_{\text{dB}}^o$  will be non-Gaussian, but for the small signal to noise ratios of  $\sim 7\%$  observed on ERS-2 data the noise distribution of the transformed data is close to Gaussian.

### 3.3. Multibeam Model

During model validation (see section 4.1 and Figure 2) it became apparent that the midbeam antenna  $\sigma^o$  value was not well modeled by either CMOD4 or NN2CMOD. The poor fit to the midbeam  $\sigma^o$  was not due to the inability of the models to represent the relation between  $\sigma^o$  and  $s$ ,  $\chi$  at low incidence angles, as was confirmed by the fact that a more flexible neural network with 12 hidden units performed similarly. A three-beam model was constructed, consisting of two NN2CMOD models with four hidden units in the MLP, one for the midbeam, and one common model for both fore and aft beams, which we denote NN3CMOD.

Figure 2

### 3.4. Bayesian Parameter Estimation in the Presence of Input Noise

NN2CMOD and NN3CMOD depend upon weights in the MLPs,  $\mathbf{w}$ , which are determined from training data. We adopt a pragmatic Bayesian approach for the estimation of the weight vector in the presence of input noise, details of which can be found in the work of *Cornford et al.* [2000]. If the input noise is not properly accounted for, then nonlinear models may learn a biased estimate of the true underlying function [*Wright, 1999*].

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

$$p(\mathbf{w} | \mathbf{D}) = \int \underbrace{p(t_n | \tilde{\mathbf{x}}_n, \mathbf{w})}_{p_1} \underbrace{p(\mathbf{x}_n | \tilde{\mathbf{x}}_n)}_{p_2} \underbrace{p(\tilde{\mathbf{x}}_n)}_{p_3} \underbrace{p(\mathbf{w})}_{p_4} d\tilde{\mathbf{x}}_n, \quad (5)$$

where  $\mathbf{D} = \{t_n, \mathbf{x}_n\}$ ,  $t_n$  are the (noisy) targets in the training data,  $\mathbf{x}_n$  are the corresponding noisy inputs, and  $\tilde{\mathbf{x}}_n$  are the associated noiseless inputs. To train the network, we determine the maximum a posteriori probability (MAP) weight vector and noiseless inputs by minimizing the negative logarithm of (5). This is a suboptimal choice, since ideally we would sample  $\mathbf{w}$  from the distribution  $p(\mathbf{w} | \mathbf{D})$  and use the samples to approximate the predictive integral:

$$p(t^* | \mathbf{x}^*) = \int_{\mathbf{w}} p(t^* | \mathbf{x}^*, \mathbf{w}) p(\mathbf{w} | \mathbf{D}) d\mathbf{w}, \quad (6)$$

where  $\mathbf{x}^*$  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. Thus the cost (or error) function  $E = -\ln[p(\mathbf{w}|\mathbf{D})]$  is the sum of four terms  $E_i = -\ln(p_i)$ .

$E_1$  is 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  $\sigma_{\text{dB}}^o$  space. In the error function for NN3CMOD this will be the sum of three terms, one for each antenna.

$E_2$  is the error due to the modified (noise free) wind vectors differing from the corresponding noisy wind vectors in the data. The distribution  $p_2$  is assumed to be Gaussian in the wind components  $(u, v)$ . 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, 1998].

$E_3$  is derived from 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 uniform distribution in wind speed and direction, which implies no contribution to the cost function. This is reasonable because of the data selection method used.

$E_4$  is proportional to the prior over the weights and controls the complexity of the MLP [Bishop, 1995]. A zero mean Gaussian distribution is assumed. The variance was 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.

From (5) the cost function can be written

$$\begin{aligned} E &= E_1 + E_2 + E_4 \\ &= \sum (f(\tilde{s}, \tilde{\chi}, \theta; \mathbf{w}) - \sigma^o)^2 / (2\sigma_t^2) \\ &\quad + \sum \left( (\tilde{u} - u)^2 + (\tilde{v} - v)^2 \right) / (2\sigma_u^2) \\ &\quad + \sum_{\mathbf{w}} \mathbf{w}^2 / (2\sigma_{\mathbf{w}}^2), \end{aligned} \quad (7)$$

where for clarity we have omitted constant terms (including  $E_3$ ), the first two summations are over all patterns in the training set,  $\sigma_t^2$  and  $\sigma_u^2$  are the variance of the errors in the  $\sigma^o$  (target) and wind vector measurements, respectively,  $f(\tilde{s}, \tilde{\chi}, \theta; \mathbf{w})$  is the output obtained by propagating the modified wind vector  $(\tilde{s}, \tilde{\chi})$  and  $\theta$  through the forward models, and  $\sigma_{\mathbf{w}}^2$  is the weight decay variance, which depends on the weight type.

Equation (7) is similar to the cost function used to determine the parameters of CMOD4 by Stoffelen and Anderson [1997a], with the addition of a prior model for the weight vector. NN3CMOD has the advantage that

during the training process, three  $\sigma^o$  measurements are used to infer the “true” wind vector, as opposed to one  $\sigma^o$  measurement during the training of NN2CMOD.

### 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. We 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 (ECMWF) 10-m wind vectors by the French Research Institute for the Exploitation of the Sea (see <http://www.ifremer.fr/> for details). The ECMWF wind vectors include assimilated CMOD4 retrieved scatterometer winds. If the data assimilation system of the ECMWF model is working, 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  $\sigma^o$  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  $<7\%$  to remove observations with excessive signal to noise ratios.

We make the usual assumption that the observations in the training set are independent, and 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. For each potential sample we computed the variability of the wind field within a circle of radius  $\sim 100$  km. If the variability of the wind components was  $>2.5 \text{ m}^2 \text{ s}^{-2}$ , the central wind vector was not selected, in order 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. A variance of  $2.5 \text{ m}^2 \text{ s}^{-2}$  was chosen on the basis of experimentation.

We further process the training data using an interactive, manual outlier removal procedure. We know that the noise on the  $\sigma^o$  observations is small; thus visualization in  $\sigma^o$  space can quickly identify outliers which can have a large effect on the error term  $E_1$ . By using three linked plots it was possible to eliminate the extreme  $\sigma^o$  outliers present in the data set. Two further linked plots allowed us to examine outliers in wind speed and direction [*Stoffelen and Anderson, 1997b*]. During outlier removal, 1.9% of the training data was rejected.

In order to train models which perform well over all wind speed ranges we select two data sets of equal size, one of which is selected randomly to have an atmo-

spheric distribution in wind speeds, while the other has an almost uniform distribution in wind speed but is random in wind direction. These data sets are combined to provide the training data for the models.

### 3.6. Parameter Estimation

We used estimates by *Stoffelen and Anderson* [1997a] to set the error variances on the scatterometer observations and the wind vectors. We assumed a variance on  $\sigma^\circ$  of  $\sigma_t^2 = 0.04 \text{ dB}^2$  and on  $(u, v)$  of  $\sigma_u^2 = 2.25 \text{ m}^2 \text{ s}^{-2}$  in training. To verify these assumptions, we evaluated trained models on an independent validation set, which had also undergone the process of outlier removal. We used 20,000 iterations of scaled conjugate gradient optimization to determine the MAP weight values to ensure convergence, particularly in the estimation of the noise-free wind vector inputs.

## 4. Validation of Forward Models

There are several measures which one might use when evaluating the performance of the various forward models. The ideal choice, related to the error function used during training, is the root-mean-square error (RMSE) of the  $\sigma^\circ$  observations, given the unknown “true” wind vector. However, the application requirement for accurate wind vector retrieval means that the vector RMSE of the retrieved  $(u, v)$  observations is more important.

Local wind vectors are retrieved by inverting the forward models. The Jacobians of the models were used to determine the local wind vectors which yield the minimum distance to the model manifold using a scaled conjugate gradient minimization algorithm. Since the inverse mapping is typically multivalued, eight starting points for the optimization are chosen, starting with the NWP wind vector and changing the wind direction by  $90^\circ$  at each minimization. This typically retrieves between two and four possible local solutions together with a measure of their distance from the model manifold.

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 retrieval quality of the local models independent of any disambiguation technique, we pick the wind vector from the two to four returned that is closest to the NWP winds [*Offiler, 1994*].

### 4.1. Distance to Model Manifold: Validation in $\sigma^\circ$ Space

In order to obtain quantitative results on the fit of the models in  $\sigma^\circ$  space we have looked at the distance to the



model manifold for a validation set which has the same distribution in wind speed and direction as the training set. Table 1 shows that NN3CMOD fits the model manifold more tightly than CMOD4 and NN2CMOD. The distance from the point on the model manifold corresponding to the retrieved wind vector closest to the NWP wind vector, denoted distance (best), also shows the improvement given by NN3CMOD.

Table 1

Table 1 also contains the standard deviations of the  $\sigma^\circ$  errors, based on distance (best), for each beam; for all models the fore and aft beams have smaller noise levels than the midbeam. For CMOD4 and NN2CMOD, Figure 2 shows that this is related to the poor fitting (bias) of the models to  $\sigma^\circ$  at small incidence angles (the midbeam incidence angles are generally smaller). However, for NN3CMOD the midbeam  $\sigma^\circ$  still has a higher variance, despite an unbiased fit to the  $\sigma^\circ$  observations (Figure 2c). Figure 2c shows no evidence of a systematic dependence on  $\theta$ ; rather, there seems to be a distinctly different variance for the midbeam  $\sigma^\circ$ . This suggests that it is preferable to have a separate model for the midbeam and a joint model for the fore and aft beams, as done in NN3CMOD.

Figure 3 shows the misfit of models in  $\sigma^\circ$  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  $8 \text{ m s}^{-1}$ . The results for NN3CMOD (Figure 3c) suggest that the spread of observations about the manifold decreases with increasing wind speed; however, this may partly reflect the poor fit of all models at small  $\sigma^\circ$  values, which correspond to lower wind speeds.

Figure 3

#### 4.2. 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. As the ECMWF wind vectors used in training the models already have some influence from CMOD4, an independent test set of  $\sigma^\circ$  measurements was used. The test set used UK Meteorological Office (UKMO) “first guess at appropriate time” winds as targets. These are unified model 0- to 6-hour forecast winds [Andrews and Bell, 1998], interpolated to the  $\sigma^\circ$  observation locations. Three days of scatterometer observations, collected from both Northern and Southern Hemispheres, were randomly subsampled to provide the test set of 60,000 measurements with a distribution of wind speed and direction similar to

that observed in the atmosphere.

Table 2 shows the results on the test set. The vector RMSE of the CMOD4+5% retrieval is larger than that of the neural network models by at least  $0.5 \text{ m s}^{-1}$ , which is a large margin, and is certainly statistically significant with over 50,000 observations. This number must be interpreted carefully, since on average, CMOD4+5% returned 2.24 solutions per  $\sigma^\circ$  observation, while NN2CMOD returned 2.36 and NN3CMOD returned 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 parameterizations of the lower boundary layers in the ECMWF and UKMO models are similar, and thus our models could be used consistently with both the UKMO unified model and the ECMWF model.

Table 2

The direction biases are similar and small for all models, but CMOD4+5% has a larger direction standard deviation. Both neural network models have consistently better performance in terms of retrieving the most probable solution (that is, the minimum distance to the model manifold solution) within  $20^\circ$  of the NWP wind vector (which is denoted ONET and expressed as a percentage) compared with CMOD4+5%. This is related to the lower RMSE in  $\sigma^\circ$  space of the hybrid models, particularly NN3CMOD, and illustrates that an improved fitting in  $\sigma^\circ$  space is important for  $(u, v)$  retrieval, particularly for ambiguity removal. The ONET measure will not be affected by the number of solutions returned.

## 5. Discussion

When tuning a nonlinear model, data selection and quality control are very important. Although interactive data manipulation demands a large amount of user time, it can greatly improve the modeling exercise. The more flexible the model, the more important data integrity becomes.

In earlier work a neural network was used to model the backscatter directly, but this was outperformed by the hybrid models on a test set. In particular, the neural network did not capture the manifold shape well at high wind speeds where there is little training data. Using a hybrid model, which contains a priori information, gave better generalisation at high and low wind speeds. This suggests that when a priori information is available it should be used, even if this makes the modeling exercise more complex to implement.

Despite the removal of outliers 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. When standard training, with a sum of squares cost function, was used on the neural network models (i.e., disregarding input noise), the results, both in terms of fit to the  $\sigma^o$  observations and  $(u, v)$  retrieval, were worse.

Results show NN3CMOD fits the  $\sigma^o$  observations better than both NN2CMOD and CMOD4; this is attributed to the use of a separate model for the mid-beam  $\sigma^o$  in NN3CMOD. Local wind vector retrieval is improved using the neural network models. The improvement is related to the hybrid model’s ability to fit the  $\sigma^o$  observations better than CMOD4 and to the Bayesian training procedure used to minimize the impact of input noise on the model parameters. Although NN3CMOD fits much better in  $\sigma^o$  space compared with NN2CMOD, the retrieval in  $(u, v)$  space is only marginally improved.

However, NN3CMOD more often determines which of the ambiguous solutions is the “true” solution. Data assimilation systems which use  $\sigma^o$  rather than  $(u, v)$  will be more accurate if the  $\sigma^o$  fit of the model is more accurate. Even for data assimilation systems which use locally retrieved  $(u, v)$  the accurate fit of NN3CMOD will improve the estimation of  $(u, v)$  directly and will also improve the estimate of the probability of each ambiguous solution which can be fed to the data assimilation system (or ambiguity removal algorithm).

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 their Jacobians, NN3CMOD requires an average of 30% more floating point operations per pattern, compared with CMOD4.

Figure 4 shows the effect of using three different distributions of wind speed in the training set on local wind vector retrieval using NN3CMOD. The results illustrate that models trained with a near-uniform distribution in wind speed (i.e., with more cases in the higher wind speed range) perform better when retrieving winds at higher wind speeds but perform slightly worse when retrieving lower speed winds. It is also clear that mixing the training sets allows the model to fit well at both high and low wind speeds without compromising performance in the midrange. This suggests that online 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 perfor-

Figure 4

mance of NN3CMOD, especially at higher wind speeds.

## 6. Conclusions

This paper has discussed novel neural-network-based scatterometer forward models. Although an interactive outlier removal method was used with careful data selection, a training method that accounted for the input noise in the “cleaned” NWP wind vectors was still needed.

Using the distance to the model manifold measure, we have shown that the hybrid models, particularly NN3CMOD, fit the  $\sigma^\circ$  observations better than does CMOD4. The midbeam antenna has a different response than the fore and aft beams. This strongly suggests that a different model is required for the midbeam  $\sigma^\circ$  measurements, although a common model can be used for the fore and aft beams, as implemented in NN3CMOD.

The hybrid models are shown to be more accurate for wind vector retrieval. The hybrid models are unbiased with respect to wind speed and direction retrieval on the ECMWF and UKMO data sets and have lower direction standard deviations compared with CMOD4+5%. NN3CMOD also improves the number of most probable solutions within  $20^\circ$  of the NWP winds, making ambiguity removal easier.

Future work could consider a better model for the  $\sigma_{dB}^\circ$  error, which should improve the fit of the model, and could consider its inversion for wind vector retrieval. This might be achieved by using learning strategies for the variance as well as the mean of the prediction, but this is difficult when the inputs are also noisy. Online training of the models using multiple data sources, such as buoy and ship observations, as well as NWP winds could produce further improvements to the models.

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

## References

- Andrews, P. L. and R. S. Bell, Optimizing the United Kingdom Meteorological Office data assimilation for ERS-1 scatterometer winds. *Mon. Weather Rev.*, *26*, 736–746, 1998.
- Bishop, C. M., *Neural Networks for Pattern Recognition*, 482 pp., Oxford Univ. Press, New York, 1995.
- Cornford, D., G. Ramage, and I. T. Nabney, A scatterometer neural network sensor model with input noise. *Neurocomput. Lett.*, *30*, 13–21, 2000.
- Daley, R., *Atmospheric Data Analysis*, 457 pp., Cambridge Univ. Press, New York, 1991.
- Dickinson, S. and R. A. Brown, A study of near-surface winds in marine cyclones using multiple satellite sensors, *J. Appl. Meteorol.*, *35*, 769–781, 1996.
- Ebuchi, N., H. Kawamura, and Y. Toba, Physical processes of microwave backscattering from laboratory wind wave measurements, *J. Geophys. Res.*, *98*, 14,669–14,681, 1993.
- Haltiner, G. J. and R. T. Williams, *Numerical Prediction and Dynamic Meteorology*, 477 pp., John Wiley, New York, 1980.
- Janssen, P. A. E. M., H. Wallbrink, C. J. Calkoen, D. van Halsema, W.A. Oost, and P. Snoeij, VIERS-1 scatterometer model, *J. Geophys. Res.*, *103*, 7807–7831, 1998.
- Levy, G., Southern Hemisphere low level wind circulation statistics from the Seasat scatterometer, *Ann. Geophys.*, *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, *Adv. Space Res.*, *13*(5), 19–27, 1993.
- Mejia, C., F. Badran, A. Bentamy, M. Crepon, S. Thiria, and N. Tran, Determination of the geophysical model function of NSCAT and its corresponding variance by the use of neural networks, *J. Geophys. Res.*, *104*, 11,539–11,556, 1999.
- Offiler, D., The calibration of ERS-1 satellite scatterometer winds, *J. Atmos. Oceanic Technol.*, *11*, 1002–1017, 1994.
- Richaume, P., F. Badran, M. Crepon, C. Mejia, H. Roquet, and S. Thiria, Neural network wind retrieval from ERS-1 scatterometer data, *J. Geophys. Res.*, *105*, 8737–8751, 2000.
- Stoffelen, A., Scatterometry, Ph.D. Thesis, Universiteit Utrecht, The Netherlands, 1998.
- Stoffelen, A., and D. Anderson, Scatterometer data interpretation: Estimation and validation of the transfer function CMOD4, *J. Geophys. Res.*, *102*, 5767–5780, 1997a.
- Stoffelen, A., and D. Anderson, Scatterometer data interpretation: Measurement space and inversion, *J. Atmos. Oceanic Technol.*, *14*, 1298–1313, 1997b.
- Thiria, S., C. Mejia, F. Badran, and M. Crepon, A neural network approach for modeling nonlinear transfer functions: Application for wind retrieval from spaceborne scatterometer data, *J. Geophys. Res.*, *98*, 22827–22841, 1993.
- Wright, W. A., Bayesian approach to neural-network modeling with input uncertainty, *IEEE Trans. Neural Networks*, *10*, 1261–1270, 1999.
-

D. Cornford, I. T. Nabney, and G. Ramage, Neural Computing Research Group, School of Engineering and Applied Science, Aston University, Aston Triangle, Birmingham B7 4ET, England, UK. (d.cornford@aston.ac.uk; i.t.nabney@aston.ac.uk)

(Received May 4, 2000; revised May 10, 2001; accepted May 21, 2001.)

Copyright 2001 by the American Geophysical Union.

Paper number 2000JC000417.  
0148-0227/01/2000JC000417\$09.00

**Table 1.** Fit of Models in  $\sigma^\circ$  Space on a Validation Set<sup>a</sup>

	Distance (min) <sup>b</sup>	Distance (best) <sup>c</sup>	$\sigma^\circ$ SD		
			Fore beam	Midbeam	Aft beam
CMOD4	0.29	0.48	0.32	0.43	0.32
NN2CMOD	0.31	0.45	0.24	0.39	0.24
NN3CMOD	0.22	0.29	0.16	0.31	0.16

<sup>a</sup>Validation set comprises 15,000 observations between 4 and 24 m s<sup>-1</sup> which have undergone the manual outlier removal procedure. All units are in decibels.

<sup>b</sup>Average minimum distance to the model manifold is given for all ambiguous wind vectors retrieved.

<sup>c</sup>Average distance to the model manifold is given for the wind vector closest to the numerical weather prediction model wind.



**Table 2.** Performance of Models on UKMO Test Set With an Atmospheric Distribution in Wind Speed and Direction<sup>a</sup>

	vector RMSE	$s$ bias	$s$ SD	$\chi$ bias	$\chi$ SD	ONET <sup>b</sup>
CMOD4+5%	3.26	-0.44	1.75	-0.9	22.4	40.5
NN2CMOD	2.76	-0.09	1.73	0.6	16.7	44.7
NN3CMOD	2.71	-0.19	1.71	0.7	16.3	51.1

<sup>a</sup>Data comprise 50,720 observations in the range 4–24 m s<sup>-1</sup>. Wind speed  $s$  is in m s<sup>-1</sup> and wind direction  $\chi$  is in degrees.

<sup>b</sup>Percentage of the most probable solutions within 20° of the NWP wind vector.

**Figure 1.** Hybrid neural network scatterometer model. Symbols are defined in sections 2 and 3.2.

**Figure 1.** Hybrid neural network scatterometer model. Symbols are defined in sections 2 and 3.2.

**Figure 2.** Residuals  $\delta\sigma^o = \sigma_{\text{predicted}}^o - \sigma_{\text{observed}}^o$  plotted for every tenth point in the validation set as a function of incidence angle. Solid lines give the running mean; dotted lines are  $\pm 1$  standard deviation. Thick lines are midbeam statistics; thin lines are the combined fore and aft beam statistics.

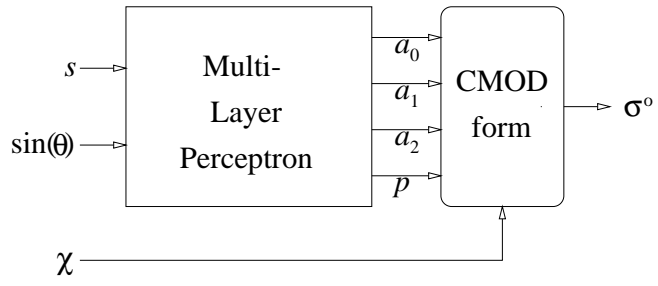
**Figure 2.** Residuals  $\delta\sigma^o = \sigma_{\text{predicted}}^o - \sigma_{\text{observed}}^o$  plotted for every tenth point in the validation set as a function of incidence angle. Solid lines give the running mean; dotted lines are  $\pm 1$  standard deviation. Thick lines are midbeam statistics; thin lines are the combined fore and aft beam statistics.

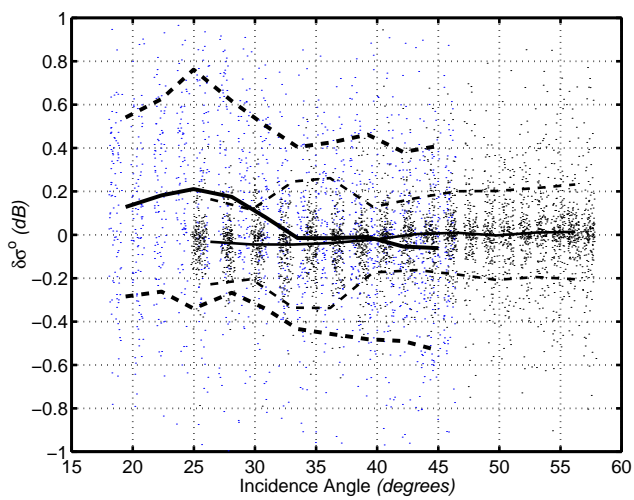
**Figure 3.** Residuals  $\delta\sigma^o = \sigma_{\text{predicted}}^o - \sigma_{\text{observed}}^o$  plotted for every tenth point in the validation set as a function of retrieved wind speed. Solid lines give the running mean; dotted lines are  $\pm 1$  standard deviation. All beams are considered together.

**Figure 3.** Residuals  $\delta\sigma^o = \sigma_{\text{predicted}}^o - \sigma_{\text{observed}}^o$  plotted for every tenth point in the validation set as a function of retrieved wind speed. Solid lines give the running mean; dotted lines are  $\pm 1$  standard deviation. All beams are considered together.

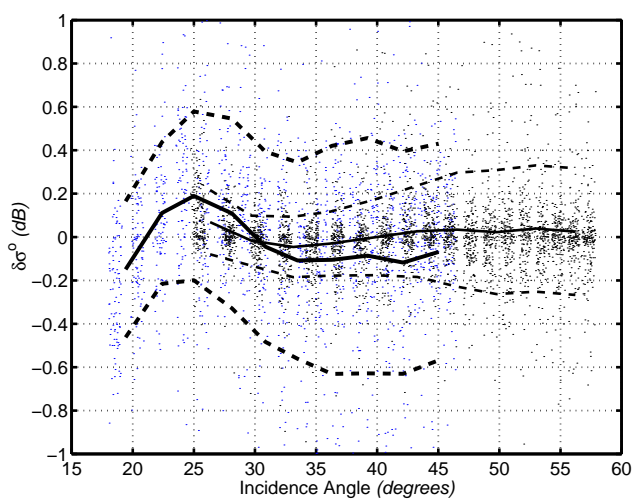
**Figure 4.** Vector RMSE as a function of retrieved wind speed on the UKMO test set for CMOD4+5% and three versions of NN3CMOD trained using a mixed (mix.), atmospheric (atm.), and uniform (uni.) distribution of wind speed in the training set. The test set has an atmospheric distribution in wind speed. Note that the  $y$  axis starts at  $2.0 \text{ m s}^{-1}$ .

**Figure 4.** Vector RMSE as a function of retrieved wind speed on the UKMO test set 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. The test set has an atmospheric distribution in wind speed. Note that the  $y$  axis starts at  $2.0 \text{ m s}^{-1}$ .

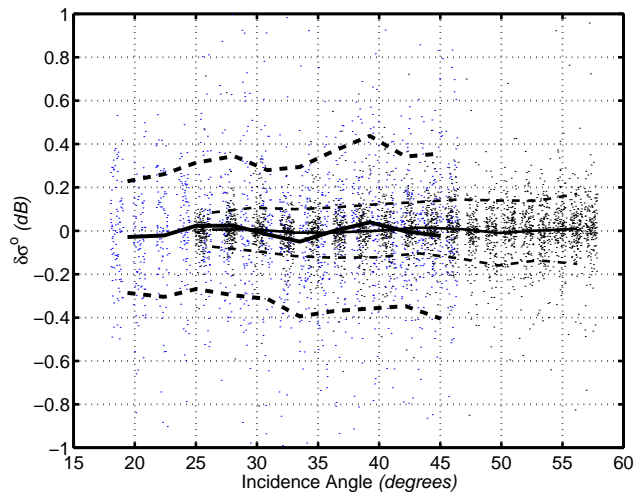




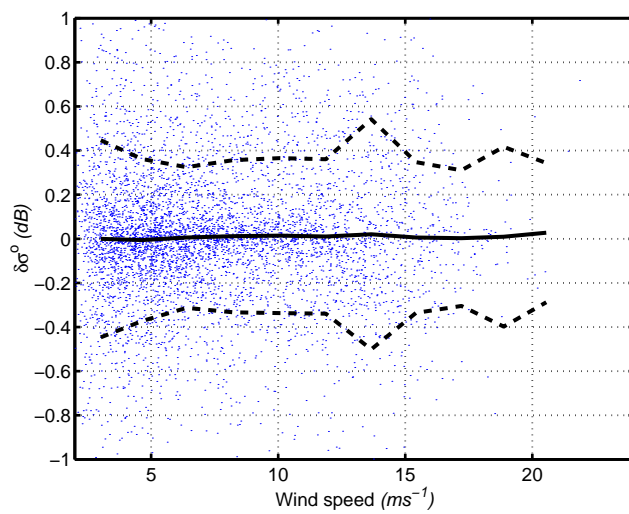
(a) CMOD4



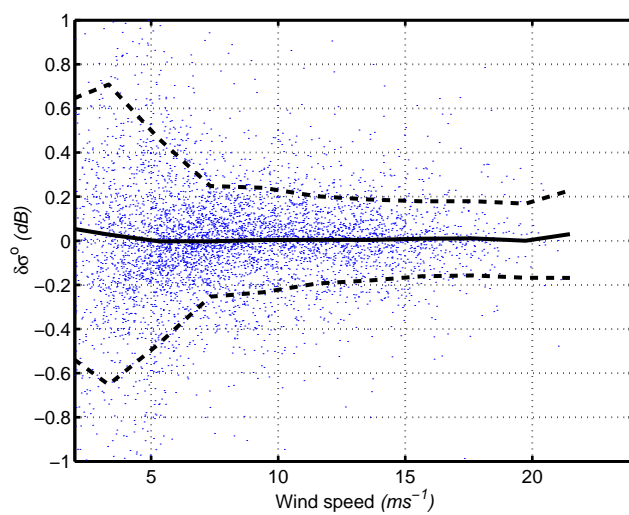
(b) NN2CMOD



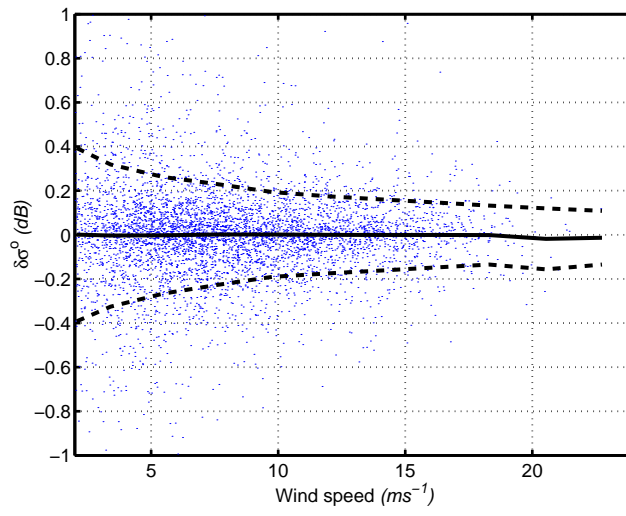
(c) NN3CMOD



(a) CMOD4



(b) NN2CMOD



(c) NN3CMOD

