

# Outlier Detection in Scatterometer Data: Neural Network Approaches

Robert J. Bullen<sup>(1,2)</sup>, Dan Cornford<sup>(1)</sup> and Ian T. Nabney<sup>(1)</sup>

July 2, 2007

**Affiliation:** <sup>(1)</sup>Neural Computing Research Group, School of Engineering and Applied Science, Aston University, Aston Triangle, Birmingham, B4 7ET, UK. <sup>(2)</sup>Advanced Technology Centre, BAE SYSTEMS, Sowerby Building 20R, PO BOX5, FP267, Filton, Bristol, BS34 7QW, UK.

**Acknowledgements:** Robert J. Bullen was supported by an EPSRC MSc studentship. This work was partially supported by the European Union under contract ENV4-CT96-0314.

**Corresponding author:** Dan Cornford

**Address:** Neural Computing Research Group, School of Engineering and Applied Science, Aston University, Aston Triangle, Birmingham, B4 7ET, UK

**Tel., fax, email:** +44 (0)121 359 3611 x4667, +44 (0)121 333 4631, d.cornford@aston.ac.uk.

**Running title:** Outlier detection using neural networks.

# Outlier Detection in Scatterometer Data: Neural Network Approaches

## Abstract

Satellite-borne scatterometers are used to measure backscattered microwave radiation from the ocean surface. This data may be used to infer surface wind vectors where no direct measurements exist. Inherent in this data are outliers owing to aberrations on the water surface and measurement errors within the equipment. We present two techniques for identifying outliers using neural networks; the outliers may then be removed to improve models derived from the data. Firstly the Generative Topographic Mapping (GTM) is used to create a probability density model; data with low probability under the model may be classed as outliers. In the second part of the paper, a sensor model with input-dependent noise is used and outliers are identified based on their probability under this model.

GTM was successfully modified to incorporate prior knowledge of the shape of the observation manifold; however, GTM could not learn the double skinned nature of the observation manifold. To learn this double skinned manifold necessitated the use of a sensor model which imposes strong constraints on the mapping. The results using GTM with a fixed noise level suggested the noise level may vary as a function of wind speed. This was confirmed by experiments using a sensor model with input-dependent noise, where the variation in noise is most sensitive to the wind speed input. Both models successfully identified gross outliers with the largest differences between models occurring at low wind speeds.

**Keywords:** Outliers, Scatterometer, Generative Topographic Mapping, Input-Dependent Noise, Neural Network Sensor Models.

# 1 Introduction

Outliers, that is observations which are widely separated from the main body of points in a sample, are a common feature of many real data sets. The presence of outliers can have a deleterious effect on any further processing or analysis of the data. In this paper we shall use density estimation techniques in order to detect outliers; the methods are tested on a large scatterometer dataset.

## 1.1 Scatterometer Data

Satellite-based scatterometers are active remote sensing instruments which use a radar (C-band, 5.3 *GHz* for the data used in this study) to assess the amount of radiation back-scattered by the target at which they are aimed. In this paper we discuss the detection and removal of outliers from the ERS-2 satellite scatterometer operating in wind mode. The instrument has a large ground footprint of some 50 *km* square and the ultimate goal is to infer the ocean near-surface wind vectors (?). While the paper focuses on this particular application the methods are quite general.

[Figure 1 about here.]

The scatterometer on board the ERS-2 satellite (see Fig. 1) provides measurements of the backscatter,  $\sigma^o$ , from the ocean surface along three different azimuth angles (corresponding to fore-, mid- and aft-facing antennae) at given incidence angles with respect to the local vertical (?). It is then necessary to retrieve information about the wind vector  $\mathbf{v}$  from the vector of measurements,  $\sigma^o$ .

## 1.2 Outlier Identification

One method of estimating a desired state variable (geophysical property) from a remotely sensed observation is to invert a ‘forward’ or sensor model (?). This forward model is a mapping which represents the observation process itself, and is often non-linear (?). The presentation of an outlier in observation space to a

forward model can result in very unlikely geophysical properties being retrieved, particularly for non-linear forward models, such as occur in the scatterometer case. It is thus important to identify automatically these outliers in observation space before attempting a retrieval. Currently, operational methods for this are somewhat *ad hoc* (?).

Another important application for outlier detection occurs when we develop new empirical models of the observation process. In some cases it may be possible to use an understanding of the physics of the observation process to develop a sensor model (?); however often empirical models perform better (?). When training an empirical model care has to be taken over errors in the training set, and in the case of the scatterometer these errors are on both the input and output variables (?). Outliers in the training set may have a significant effect on the resulting sensor model mapping.

When dealing with scatterometer data, current practice is to use a particular two-dimensional projection of the data to be able to visualise the outliers (?). In practice the most straightforward approach is to remove the outliers from the training set, as done manually in (?). The ability to identify outliers automatically would speed up this labour intensive task. This is particularly important because of the extremely large size of most geophysical datasets.

We note that there are alternative methods to working with outliers such as those based on the use of robust estimators (?). However the complexity of the modelling required for scatterometer wind vector retrieval, and our belief that these outliers resulted from errors in the data collection and transmission system, rather than naturally as part of the uncertainty of the observation process, leads to our preference for removal of outliers. In this work, outliers were manually removed by visual inspection from training, validation and test sets, using projections defined in (?).

### 1.3 Processing Scatterometer Data

In the first part of this paper a Generative Topographic Mapping (GTM) is used to model the scatterometer data. GTM is a semi-parametric latent density

model which generates a manifold in data space (?). Individual data points with low probability may be considered outliers.

Secondly, a neural network sensor model is used to represent the forward mapping  $\mathbf{v} \rightarrow \sigma^o$  and the noise on the scatterometer data with respect to the model inputs also learnt. Two networks are used; one to model the conditional mean function and the second to model the noise (conditional variance). The weights of each network are updated iteratively using a Bayesian framework based on that in (?); the treatment of the input noise follows (?).

## 2 Generative Topographic Mapping

### 2.1 Introduction

The principle of a GTM, illustrated in Fig. 2, is that a grid of  $K$  points  $\{\mathbf{x}_1, \dots, \mathbf{x}_K\}$  in a latent space of dimension  $L$  is mapped to the centres,  $\{\mathbf{y}_1, \dots, \mathbf{y}_K\}$ , of a Gaussian Mixture Model (GMM) in a data space of dimension  $D$ . The GMM is constrained in this way so that the centres lie on an  $L$ -dimensional manifold embedded within the  $D$ -dimensional data space. In this instance the latent space is 2-dimensional and the data space 3-dimensional, and we train a separate model for each fixed incidence angle class. Note there are 19 discrete incidence angles at which the scatterometer instrument operates, thus 19 different GTMs are constructed.

[Figure 2 about here.]

The mapping can be performed by any continuous differentiable function but the usual and most convenient form is of a Radial Basis Function (RBF) network (?) with  $M + 1$  fixed basis functions  $\psi_m$ ,  $m = 0, \dots, M$  arranged so as to span the input data and, if desired, take account of variations in its density, and output layer weights,  $\mathbf{w}$ .  $\psi_0$  is a bias term with value 1. The model is thus,

$$y(\mathbf{x}; \mathbf{w}) = \sum_{m=0}^M \psi_m(\mathbf{x}) \mathbf{w}_m, \quad (1)$$

where  $\psi$  is referred to as the *activation function* and is usually a non-linear function such as a Gaussian or Thin Plate Spline (TPS).

The GMM has isotropic Gaussian noise distribution, given by

$$p(\mathbf{t}|\mathbf{x}, \mathbf{w}, \beta) = \left(\frac{\beta}{2\pi}\right)^{D/2} \exp\left[-\frac{\beta}{2} \sum_{i=1}^D (t_i - y_i(\mathbf{x}, \mathbf{w}))^2\right], \quad (2)$$

where the variance of the noise is given by  $\beta^{-1}$ ; this scales the probability of a data point for a given distance from the manifold.

If  $p(\mathbf{x})$  is taken as a set of  $K$  equally weighted delta functions, the data log-likelihood is expressed as

$$l = \sum_{n=1}^N \ln\left(\frac{1}{K} \sum_{k=1}^K p(\mathbf{t}_n|\mathbf{x}_k, \mathbf{w}, \beta)\right), \quad (3)$$

assuming an i.i.d. set of target data points  $\{\mathbf{t}_1, \dots, \mathbf{t}_N\}$ .

When using a mixture of Gaussians and an RBF network it is most convenient to use an Expectation Maximisation (EM) algorithm to fit the manifold to the data (?; ?).

## 2.2 Latent Space Geometry

Earlier work with the scatterometer data has shown that it lies close to a conical manifold, and so the latent space was changed to take advantage of this knowledge. Instead of the usual rectangular grid of latent points, the latent space was laid out on a ‘developed’ (i.e. cut and flattened) cone as shown in Fig. 3(a).

[Figure 3 about here.]

The only change to the GTM algorithm is in calculating distances,  $\delta$ , in latent space for the RBF. The distance across the conical surface in both directions must be considered, so two calculations were made as shown in Fig. 3(b), and labelled  $d$  and  $d'$ , and  $\delta = \min[d, d']$ .

$$\delta = \begin{cases} \sqrt{(r_1 \sin(\phi))^2 + (r_2 - r_1 \cos(\phi))^2} & \text{if } \phi \leq \theta/2, \\ \sqrt{(r_1 \sin(\theta - \phi))^2 + (r_2 - r_1 \cos(\theta - \phi))^2} & \text{otherwise.} \end{cases} \quad (4)$$

## 2.3 Initialisation and Training

A conical set of initial target data for the RBF was created in data space using the Principal Components (PCs) of the data scaled to give reasonable alignment

with the data and the corresponding weights for the RBF network calculated using a matrix pseudo-inversion on (1). The initial noise variance  $\beta^{-1}$  was set as half the distance between GMM centres in data space. The model was then trained using 30 iterations of the EM algorithm. Fig. 4 shows the initialised and trained cone.

[Figure 4 about here.]

### 3 Sensor Model with Input-dependent Noise

#### 3.1 Introduction

In the operational CMOD4 model, it has been assumed that the noise is constant throughout the scatterometer data but previous studies (?), (?) have implied that there may be some dependence of the noise upon the inputs to the scatterometer model, mainly wind speed. In addition, the GTM method does seem to identify more outliers in the area of low wind speeds. With this in mind, this section describes how a Bayesian method for Input-Dependent Noise (IDN) modelling (?) was applied to this problem.

#### 3.2 Input-dependent Noise using Bayesian Inference

We use two networks  $y_n = y(\mathbf{x}_n; \mathbf{w})$  and  $\beta_n = \beta(\mathbf{x}_n; \mathbf{u})$ , where  $\mathbf{w}$  and  $\mathbf{u}$  respectively are the weights to model the conditional mean and conditional variance of the function. These will be described as the  $y$ -network and the  $\beta$ -network. It is possible to use maximum likelihood to determine a single error function for  $\mathbf{w}$  and  $\mathbf{u}$  but this is known to give a biased value for the noise variance on finite data sets (?). Instead, Bayesian inference is used, which involves marginalising over a uniform prior distribution  $p(\boldsymbol{\mu})$ , where  $\boldsymbol{\mu}$  is the mean of the function (modelled by the  $y$ -network) and maximising with respect to  $\sigma^2 = \beta^{-1}$ , the variance, which results in a value for  $\sigma^2$  which is unbiased. Using priors over the network weights which are isotropic and Gaussian, then maximising the posterior distribution of the  $y$ -network weights  $\mathbf{w}$  given the data, an error function

for the  $y$ -network is obtained

$$S(\mathbf{w}) = \sum_{n=1}^N \beta_n E_n + \frac{\alpha_w}{2} \|\mathbf{w}\|^2, \quad (5)$$

where  $E_n = \frac{1}{2}(y_n - t_n)^2$  is the squared error for the  $n$ th point.  $\beta_n$  is the inverse noise variance and  $\alpha_w$  a hyperparameter representing the inverse noise variance of the prior over the  $y$ -network weights  $\mathbf{w}$ .

The error function,  $M(\mathbf{u})$ , for the  $\beta$ -network is developed in a similar way by maximising the marginalised posterior distribution of  $\beta$ -network weights given the data and the hyperparameters

$$M(\mathbf{u}) = \sum_{n=1}^N \beta_n E_n + \frac{\alpha_u}{2} \|\mathbf{u}\|^2 - \frac{1}{2} \sum_{n=1}^N \ln \beta_n + \frac{1}{2} \ln |\mathbf{A}|, \quad (6)$$

where  $\alpha_u$  is a hyperparameter over the  $\beta$ -network weights.  $\mathbf{A}$  is the regularised Hessian of the  $y$ -network, given by

$$\mathbf{A} = \sum_{n=1}^N \beta_n \mathbf{H}_n + \alpha_w \mathbf{I}. \quad (7)$$

$\mathbf{H}_n$  is the noiseless data Hessian matrix for the  $y$ -network for a single input  $\mathbf{x}_n$  and  $\mathbf{I}$  is the identity matrix. The calculation of  $\mathbf{H}$  for the sensor model will be discussed in Section 3.3.

Both these functions can be minimised by non-linear optimisation techniques<sup>1</sup>. In both networks the hyperparameters were fixed on the basis of experimentation, since developmental work had concluded that they could not be reliably estimated in the presence of *input* noise.

### 3.3 Application to Scatterometer Sensor Model

The sensor model used for this project is denoted RBF3CMOD, a hybrid RBF network variation on NN3CMOD (?) with a functional element similar to CMOD4, comprising a Fourier series, although this has been expanded from 3 to 5 terms to render it linear in the RBF outputs. The model is illustrated in Fig. 5.

---

<sup>1</sup>If RBF networks are used, it is normally possible to solve for  $\hat{\mathbf{w}}$  (the MAP solution for  $\mathbf{w}$ ) from (5) using ‘one shot’ matrix inversion techniques but, given the nature of the scatterometer model this was not possible.



[Figure 5 about here.]

The functional form RBF3CMOD is

$$\sigma_{\text{in}}^0 = a_0 + a_1 \cos(\chi) + a_2 \cos(2\chi) + a_3 \cos(3\chi) + a_4 \cos(4\chi) , \quad (8)$$

where  $a_0, a_1 \dots a_4$  are the outputs from the RBF network and  $\chi$  is the wind direction relative to the satellite azimuth angle.

Because of the presence of relatively large noise levels on the wind vector inputs we have extended the (?) models to include the error due to input noise (?) on the wind vectors, so (5) becomes

$$S(\mathbf{w}, \tilde{\mathbf{v}}) = \sum_{n=1}^N \beta_n E_n + \sum \frac{\|\mathbf{v} - \tilde{\mathbf{v}}\|^2}{2\sigma_v^2} + \frac{\alpha_w}{2} \|\mathbf{w}\|^2 , \quad (9)$$

where  $\tilde{\mathbf{v}}$  contains the estimates of the noiseless wind vectors which are learnt during training along with the weights  $\mathbf{w}$ .  $\sigma_v^2$  denotes the variance of the wind vector measurements, which based on the work of (?) is assigned a value of  $2.25 \text{ m}^2 \text{ s}^{-1}$ .

The Hessian is given by,  $\mathbf{H} = \mathbf{g}\mathbf{g}^T$  where  $\mathbf{g}$  is the gradient of the function output with respect to the network weights which, for an RBF,  $\mathbf{y} = \boldsymbol{\psi}\mathbf{w}$ , is simply the activations  $\boldsymbol{\psi}$ . However, the Hessian of the model RBF3CMOD needs to include the Fourier series, every term of which is affected by each activation. So gradients  $\mathbf{g}_m$  have to be calculated for each activation and then concatenated into a single vector

$$\mathbf{g} = \{\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_M\}, \quad \mathbf{g}_m = \boldsymbol{\phi}_m^T \mathbf{F}, \quad m = 0, \dots, M , \quad (10)$$

where  $\mathbf{F} = [\cos(0\chi), \cos(\chi), \dots, \cos(4\chi)]^T$ .

## 4 Results

### 4.1 GTM

A set of validation data with outliers manually removed was used to choose a probability density threshold  $p_t(\mathbf{t}|\mathbf{w}, \beta) = 0.0006$  ( $\mathbf{w}$  and  $\beta$  are the parameters of the GTM) below which data points were deemed to be outliers. Alternatively,

extreme value statistics could be used to estimate the probability of a data point in at least as an extreme location, that is, approximating the integral of the tail of the density as opposed to the probability density at a point, as in (?). Fig. 6 shows a plot of the probabilities of the test data points, the threshold and the outliers with respect to the manifold in data space.

[Figure 6 about here.]

The success of the model in identifying outliers was scored as the number of outliers correctly identified as a proportion of the total identified including erroneous ones (see Table 1). It should be remembered that these outliers are identified on the basis of their probability under the trained model, which is based upon their distance to the manifold. It will be noticed that there are a larger number of identified outliers at low wind speeds around the ‘nose’ of the cone; this is an indication the cone is not so well defined in this area, that is to say that there may be more uncertainty. This is more evidence that the noise may vary over the surface of the cone.

[Table 1 about here.]

## 4.2 Sensor Model and IDN

Initially the RBF3CMOD model was trained, using 50000 iterations of the Scaled Conjugate Gradients (SCG) algorithm and a fixed variance  $1/\beta$ , estimated by (?) as  $0.08 \text{ dB}^2$  for the mid-beam and  $0.04 \text{ dB}^2$  for the fore- and aft-beams. The IDN algorithm was then applied in order to learn an input-dependent variance. Since the scatterometer sensor model is learnt accounting for input noise, in practice it is important to estimate first the ‘true’ inputs before attempting to learn the output noise. Otherwise, the sensor model tended to a poor local optimum, where the  $y$ -network did not fit the data closely, which was compensated for by the higher levels of output noise from the  $\beta$ -network. We used Gaussian activations in the  $\beta$ -network and TPS activations in the  $y$ -network. Both networks contained 12 hidden units. The networks were trained using 10 iterations of 100 cycles of SCG, to learn the most probable weights ( $\hat{\mathbf{u}}$  and  $\hat{\mathbf{w}}$ ).

The estimates of the output noise variance (on  $\sigma^o$ ) following IDN training are illustrated in Fig. 7.

[Figure 7 about here.]

The results of outlier detection can be seen in Table 1 where the threshold for determining whether a point was an outlier was set to be 3 standard deviations (determined from the  $\beta$ - model) from the mean given by RBF3CMOD, so that 99% of observations would theoretically fall within this region.

## 5 Discussion

Having modelled the scatterometer data using a GTM, the following conclusions can be drawn.

- Modifying the latent space using prior knowledge to improve the mapping has been shown to be effective.
- The main problem with this particular application is that the data theoretically lies next to a double skinned cone, and the GTM will only be able to achieve a best fit between the two skins, resulting in identification of spurious outliers and non-identification of valid ones, as illustrated in Fig. 8. Initialisation of the GTM to a target set lying on a double skinned cone was tried but confusion of responsibilities of GMM centres for the same data points resulted in the manifold not retaining the double skinned shape after training.

[Figure 8 about here.]

- We were fortunate to have training, validation and test sets from which outliers had been manually removed. This is because we need to create a model for the true unconditional density of acceptable data for our approach to work. For this reason alternative robust error functions (?), which have been used for outlier detection, might improve the effectiveness of the model in this regard.

- The shape of the GTM manifold seems to be less well defined at each end of the cone (Fig. 6(b)). This is because there are fewer data points at high wind speeds and because the noise is proportionately larger at low wind speeds.
- The iterative IDN algorithm, when applied to the satellite scatterometer data suggests non-constant noise variance. Primarily, from Fig. 7(a), we see that there is a definite increase in the noise at low wind speeds. This is consistent with the evidence from fitting the GTM. There seems to be no correlation between incidence angle and noise.
- The mid-beam has a considerably higher noise variance although there is no obvious physical explanation for this.

## 6 Conclusions

We have presented two methods for determining outliers from remotely sensed data, showing their application to scatterometer observations. Both methods successfully identified the large outliers in the test data set, having been trained on data which had most of the outliers removed manually, using exploratory plots.

GTM built a density model for the scatterometer observations which takes into account the conical nature of the data manifold without any assumptions about these being driven by some geophysical variables. This method has general applicability to almost any type of remote sensing observation, so long as something is known about the dimension and geometry of the underlying data manifold. However, GTM is not a suitable model for high-dimensional (which may be taken as greater than three) manifolds. One improvement that might be considered is to allow variances in the GMM to be set individually, thereby allowing the variance to depend on the position on the manifold.

The IDN method required the construction of a sensor model relating  $\sigma^o$  to the surface wind vectors and the incidence angle of the radar beam. The sensor model could then be used to train a model for the noise on  $\sigma^o$  as a function of

these inputs. The results show that the noise is not constant with respect to the inputs, being particularly sensitive to wind speed. The mid-beam measurement of  $\sigma^o$  appears to have a considerably higher noise level than the fore- and aft-beams. At low wind speeds, the IDN model does not identify points far from the manifold to be outliers, whereas the GTM with fixed noise variance does. This may be related to the natural variability of wind speed and direction at low wind speeds over the footprint of the scatterometer and the use of a logarithmic transformation of the direct power observations which is standard practice in remote sensing (i.e. measurements are in decibels).

It seems likely that the IDN result is more reliable, since the semi-parametric, double skinned cone model for the mean, RBF3CMOD, fits the observations more accurately than the single skinned cone of the GTM. It also allows us to capture the relationship between the wind vector and the corresponding scatterometer measurement; this allows us to detect outliers that lie on the manifold, but in the ‘wrong’ place for the corresponding wind vector.

This work has opened many possibilities for further study. One interesting area would be to investigate the use of robust error functions in training the models (both GTM and RBF3CMOD). This would be computationally expensive, and given the availability of a data set with the worst outliers removed this was not applied here. The GTM modelling could be improved by allowing the GMM variances to be set individually. The IDN modelling has raised an interesting question about the different behaviour of the mid-beam antenna on ERS-2. It is not clear what the physical significance of this is and further work is needed. The interesting behaviour of the variance at very low wind speeds also requires further work to fully understand the mechanism, but may well be related to the relatively large variability of the wind speed and direction within short distances at low wind speeds.

The approach of using probabilistic models to automate the detection and removal of outliers shows promise. Both unconditional and conditional density estimation can be used, and this allows us to incorporate prior knowledge (manifold geometry, in the case of the GTM, and functional form of the sensor

model, in the case of IDN) in a principled way. The use of probability densities, or extreme value theory, is much less *ad hoc* than the current methods involving the ‘distance to cone’ in that they take account of the real shape of the data and the varying noise levels.

## References

- Bishop, C. M. 1995. **Neural Networks for Pattern Recognition**. Oxford: Oxford University Press.
- Bishop, C. M. and C. S. Qazaz 1997. Regression with Input-dependent Noise: A Bayesian Treatment. In M. C. Mozer, M. I. Jordan, and T. Petsche (Eds.), *Advances in Neural Information Processing*, Volume 9, pp. 347–353. MIT press.
- Bishop, C. M., M. Svensen, and C. K. I. Williams 1998. GTM: The Generative Topographic Mapping. *Neural Computation* **10**, 215–234.
- Cornford, D., I. T. Nabney, and G. Ramage 2001. Improved Neural Network Scatterometer Forward Models. *Journal of Geophysical Research - Oceans* **106**, 22331–22338.
- Cornford, D., G. Ramage, and I. T. Nabney 2000. A Scatterometer Neural Network Sensor Model with Input Noise. *Neurocomputing Letters* **30**, 13–21.
- Janssen, P. A. E. M., H. Wallbrink, C. J. Calkoen, D. van Halsema, W. A. Oost, and P. Snoeij 1998. VIERS-1 scatterometer model. *Journal of Geophysical Research* **103**, 7807–7831.
- Nabney, I. T. 2001. **Netlab: Algorithms for Pattern Recognition**. London: Springer.
- Offiler, D. 1987. Wind Measurements from the Earth Remote-sensing Satellite (ERS-1). *Meteorological Magazine* **116**, 279–285.
- Offiler, D. 1994. The Calibration of ERS-1 Satellite Scatterometer Winds. *Journal of Atmospheric and Oceanic Technology* **11**, 1002–1017.

- Qazaz, Q. S. 1996. **Bayesian Error Bars for Regression**. Ph.D. thesis, Aston University, Birmingham, UK.
- Roberts, S. J. 1999. Novelty Detection using Extreme Value Statistics. *IEE Proceedings in Vision, Image and Signal Processing* **146**, 124–129.
- Rousseeuw, P. J. and A. M. Leroy 1987. **Robust Regression and Outlier Detection**. London: John Wiley and Son.
- Stoffelen, A. and D. Anderson 1997a. Scatterometer Data Interpretation: Estimation and Validation of the Transfer Function CMOD4. *Journal of Geophysical Research* **102**, 5767–5780.
- Stoffelen, A. and D. Anderson 1997b. Scatterometer Data Interpretation: Measurement Space and Inversion. *Journal of Atmospheric and Oceanic Technology* **14**, 1298–1313.
- Wright, W. A., G. Ramage, D. Cornford, and I. T. Nabney 2000. Neural Network Modelling with Input Uncertainty: Theory and Applications. *Journal of VLSI Signal Processing Systems* **26**, 169–188.

## 7 Figure Legends

Fig. 1. The observation geometry of the ERS-2 satellite showing the three scatterometer beams (fore, mid and aft).

Fig. 2. The non-linear mapping  $y(\mathbf{x}; \mathbf{w})$  between the latent and data space. A data point (*star*) can be visualised in latent space using the mean of the posterior probability distribution.

Fig. 3. **(a)** Latent points (*crosses*), RBF centres (*circles*) and mean responsibility of the data (*dots*). The two diagonal edges of the developed cone are coincident in data space; therefore there is a space, the equivalent of a column of latent points, inside the right-hand edge to avoid duplication of points along this line. **(b)** The distance between a latent point (*cross*) and an RBF centre (*circle*) is measured as the lesser of distances  $d$  and  $d'$ , calculated using  $\phi$  and  $\theta$ .

Fig. 4. Plot of GTM manifold viewed along the mid-beam. **(a)** Initial conical manifold lying along the main axis (first principal component) with large and small diameters computed using scaled second principal component. The targets (*crosses*) were used to initialise the RBF; these are close to the mixture model centres. **(b)** Manifold aligned to the data (*dots*) after 30 iterations of the EM algorithm.

Fig. 5. Block diagram of the hybrid sensor model RBF3CMOD. The inputs  $s$  (normalised wind speed) and  $\chi$  (wind direction relative to the satellite azimuth angle) are computed from the wind vectors;  $\theta$  is the radar beam incidence angle. The outputs from the RBF network,  $a_0, \dots, a_4$ , are the coefficients of the Fourier series and  $\sigma^o$  is the inferred scatterometer value.

Fig. 6. **(a)** Data points and their GTM probability density with a threshold of 0.0006, from track 10. Those falling below the threshold are classified as outliers. The manually identified outliers, which were used to score the success of the model, are circled. **(b)** Z-scores under the IDN sensor model (difference from the mean divided by the standard deviation). Points falling above the threshold of three standard deviations are considered outliers.

Fig. 7. Input-dependent noise modelling. **(a)** Noise variance as a func-



tion of wind speed ( $ms^{-1}$ ). The lower lines are from the fore- and aft-beam model at fixed incidence angles of 25 degrees (track 1, *circles*) and 59 degrees (track 19 *stars*). The upper lines represent readings using the mid-beam model at incidence angles of 18 degrees (track 1, *crosses*) and 47 degrees (track 19, *diamonds*). **(b)** Noise variance as a function of incidence angle at speeds of  $2ms^{-1}$  and  $24ms^{-1}$  for the fore- and aft-beam (*circles* and *stars*) and mid-beam (*crosses* and *stars*) models respectively.

Fig. 8. True outliers (*stars*) are correctly identified by the GTM if they have low probability under the GTM, shown as a dotted line. This will not always coincide with data points which lie a long way from the theoretical cone. Similarly, in areas where the GTM manifold has to find the best fit to areas where the two surfaces of the cone are far apart, valid scatterometer points may be identified as outliers.

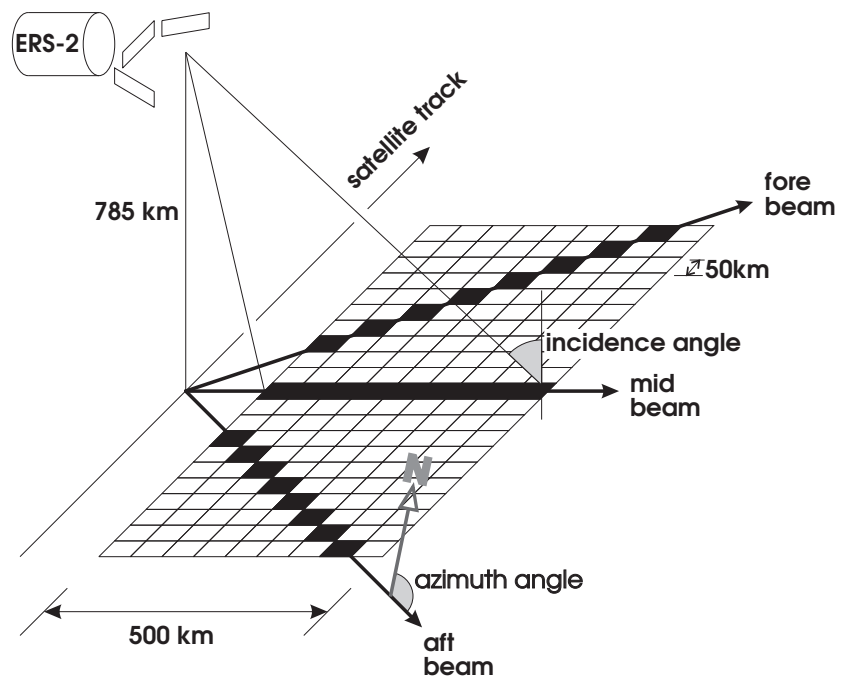


Figure 1:

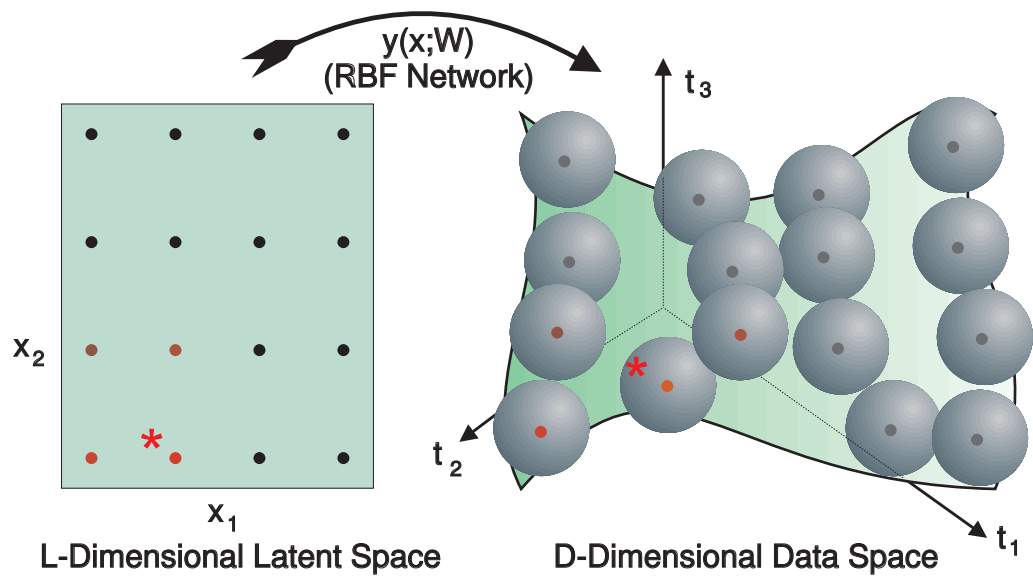


Figure 2:

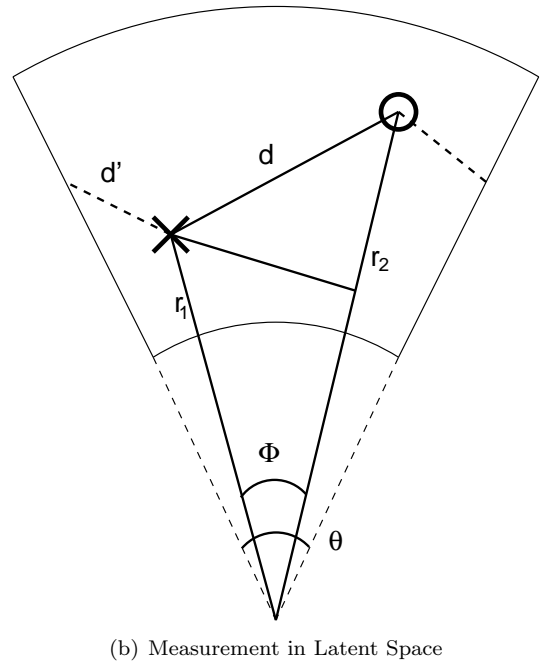
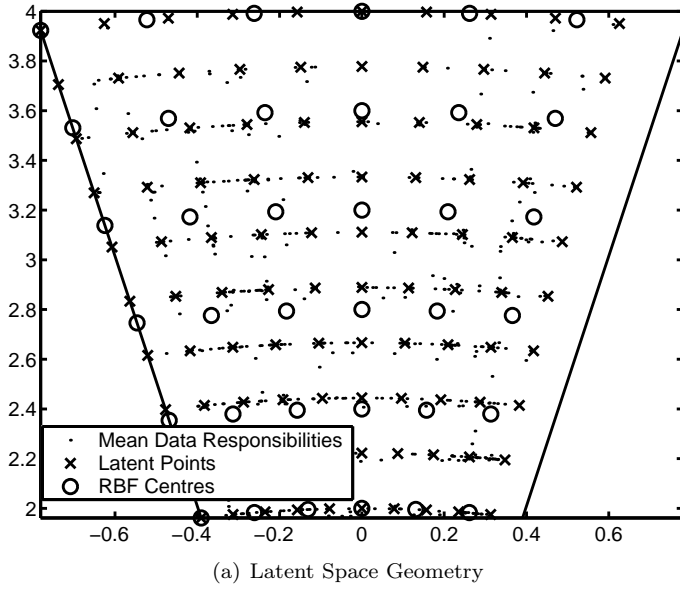
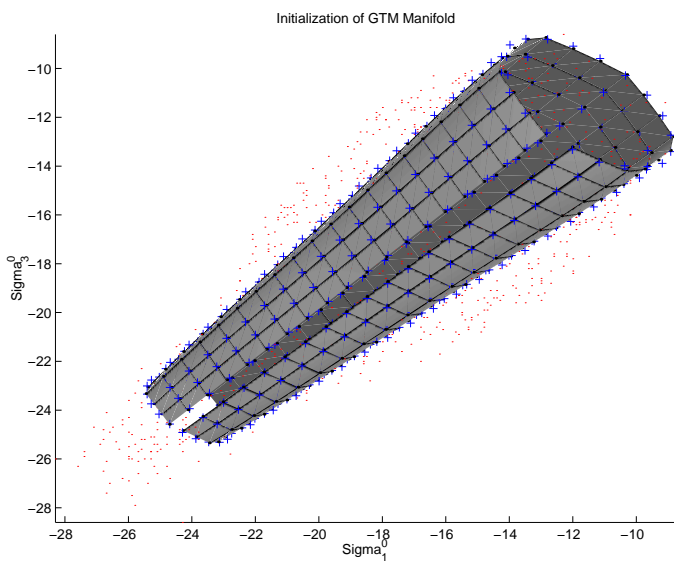
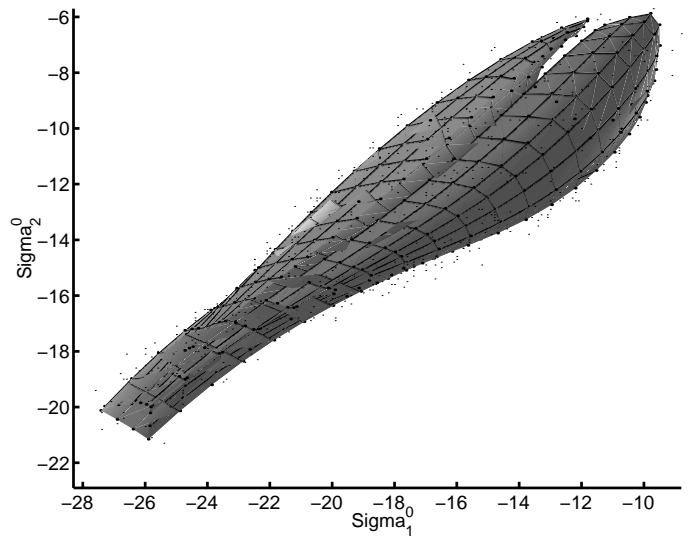


Figure 3:



(a) Initialisation of Conical GTM Manifold



(b) GTM Manifold After Training

Figure 4:

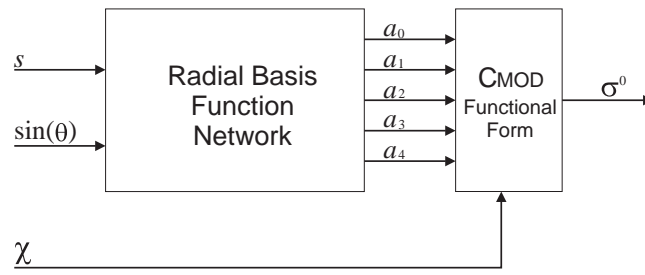
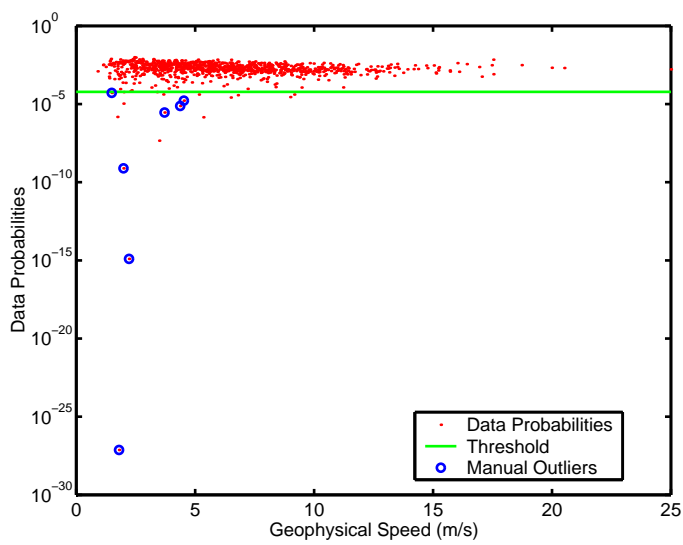
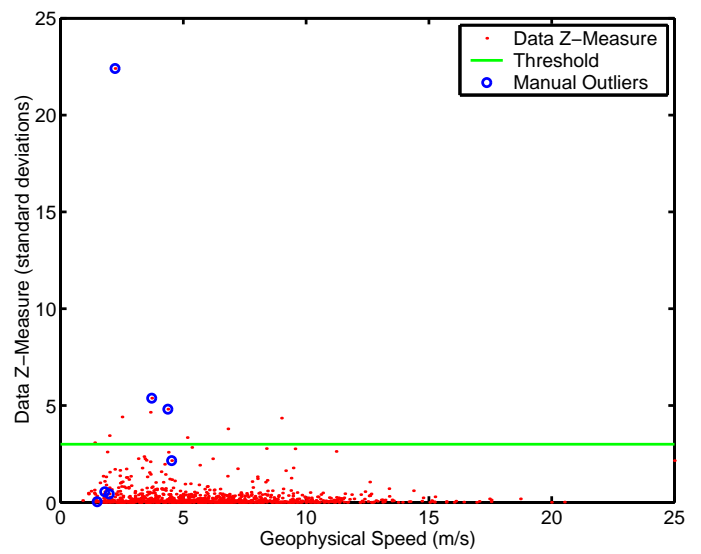


Figure 5:

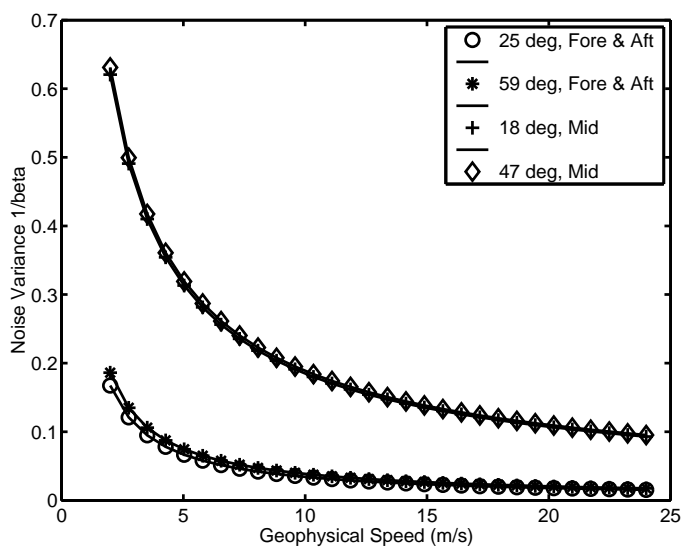


(a) GTM Data Probability, Outliers and Threshold

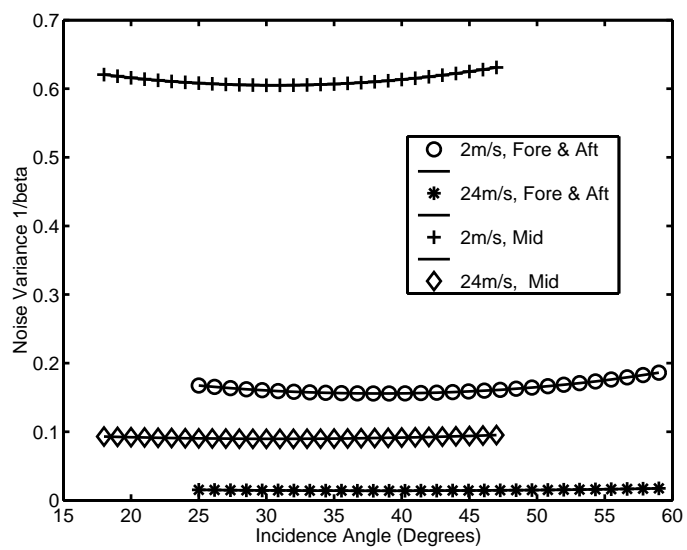


(b) IDN Data Z-score, Outliers and Threshold

Figure 6:



(a) Noise Variance with Geophysical Speed



(b) Noise Variance with Incidence Angle

Figure 7:



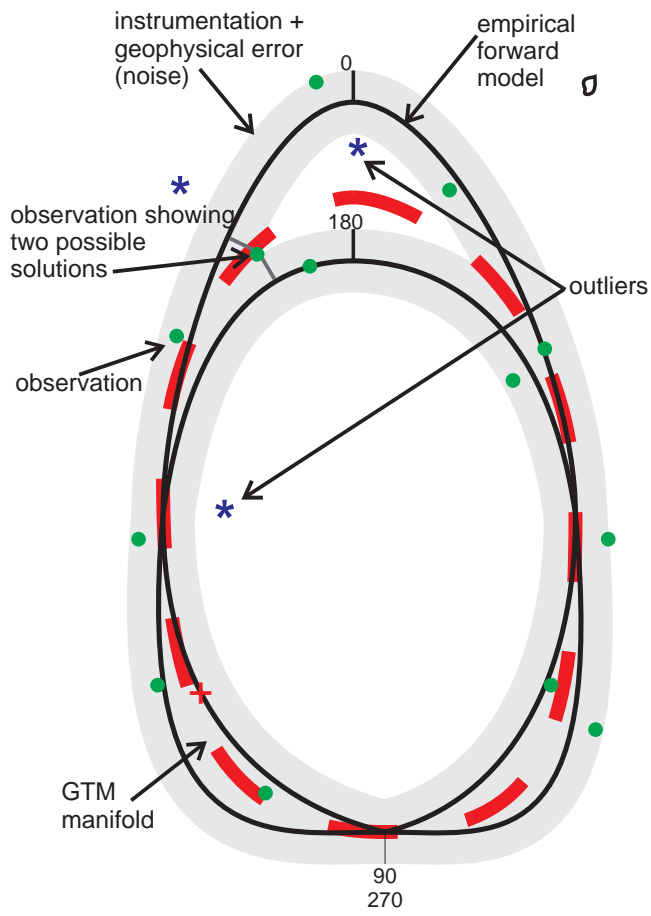


Figure 8:

Table 1: Model evaluation on a test set of 1000 points. The table shows the track number (related to the incidence angle of the beam), the number of outliers identified manually and removed from the test set, the total number of points identified as outliers, and, the number of outliers which agreed with the manually removed ones, for the GTM and sensor model with IDN respectively.

Track	Manual Outliers	GTM Outliers	GTM Score	IDN Outliers	IDN Score
1	11	13	5	23	6
10	7	19	7	10	3
19	10	45	7	47	8