NEURAL COMPUTING
NCRG
RESEARCH GROUP

Neural Computing Research Group
Dept of Computer Science & Applied Mathematics
Aston University
Birmingham B4 7ET
United Kingdom
Tel: +44 (0)121 333 4631
Fax: +44 (0)121 333 4586
http://www.ncrg.aston.ac.uk/

# Bayesian Inference for Wind Field Retrieval

**Dan Cornford and Ian T. Nabney and Christopher K. I. Williams**
d.cornford@aston.ac.uk

Technical Report NCRG/98/023 October 22, 1998

## Abstract

In many problems in spatial statistics it is necessary to infer a global problem solution by combining local models. A principled approach to this problem is to develop a global probabilistic model for the relationships between local variables and to use this as the prior in a Bayesian inference procedure. We show how a Gaussian process with hyper-parameters estimated from Numerical Weather Prediction Models yields meteorologically convincing wind fields. We use neural networks to make local estimates of wind vector probabilities. The resulting inference problem cannot be solved analytically, but Markov Chain Monte Carlo methods allow us to retrieve accurate wind fields.

**Keywords:** Bayesian inference; surface winds; spatial priors; Gaussian Processes

# 1 Introduction

Satellite borne scatterometers are designed to retrieve surface winds over the oceans. These observations enhance the initial conditions supplied to Numerical Weather Prediction (NWP) models (Lorenc *et al.*, 1993) which solve a set of differential equations describing the evolution of the atmosphere in space and time. These initial conditions are especially important over the ocean since other observational data is sparse. This paper addresses the issue of wind field retrieval from scatterometer data using neural networks and probabilistic models alone.

Simply, a scatterometer measures the amount of radiation scattered back toward the satellite by the ocean's surface, which is largely related to the local instantaneous wind stress. Under the assumption of a fixed change in wind speed with height, the surface wind stress can be related to the local 10 $m$ wind vector (Offiler, 1994).

The relation between back-scatter (denoted $\boldsymbol{\sigma^o}$)[1] and the wind vector, $(u, v)$, is complex to model (Stoffelen and Anderson, 1997b). Most previous work has generated forward models (Stoffelen and Anderson, 1997c), that is the mapping $(u, v) \rightarrow \boldsymbol{\sigma^o}$, which is one to one. The inverse mapping, $\boldsymbol{\sigma^o} \rightarrow (u, v)$, is a one to many mapping since the noise on $\boldsymbol{\sigma^o}$ means it becomes very difficult to uniquely determine the wind direction, althought the speed is well defined. There are generally two dominant wind direction solutions which correspond to winds with roughly 180° different directions. The inverse problem can be modelled directly using a neural network approach which produces encouraging results (Richaume *et al.*, 1998). Our work builds on both approaches, unified within a Bayesian framework for combining local model predictions with a global prior model (Tarantola, 1987).

In earlier work using forward models for wind retrieval, including the systems used operationally by the European Space Agency and the UK Meteorological Office, the wind fields are obtained by heuristic methods which rely on a background (or first guess) forecast wind field from a NWP model (Stoffelen and Anderson, 1997a). The method proposed here could include this information, however, due to careful consideration of the prior wind field model, it should be possible to achieve *autonomous* wind field retrieval. This implies that it is only necessary to use the scatterometer observations, which means that wind fields can be retrieved without an NWP model, and also used for validation of NWP models. This is useful since NWP models are not available to all users of scatterometer data due to their massive computational cost.

# 2 Modelling Approach

The polar orbiting ERS-1 satellite (Offiler, 1994) carries a scatterometer which obtains observations of back-scatter in a swathe approximately 500 $km$ wide. The swathe is divided into scenes which are $500 \times 500$ $km$ square regions each containing $19 \times 19$ cells with a cell size of $50 \times 50$ $km$. Thus there is some overlap between cells. We refer to each cell as a *local* measurement, while a *field* is taken to denote a spatially connected set of samples from a swathe. Capital letters are used to denote a wind field $(U, V)$ or a field of scatterometer measurements $\boldsymbol{\Sigma^o}$. Lower case letters are used to represent local measurements in a single cell, and occasionally the subscript $i$ is used to index the cell location.

The aim is to obtain $P(U, V \mid \boldsymbol{\Sigma^o})$, the conditional probability of the wind field, $(U, V)$, given the

[1]The vector $\boldsymbol{\sigma^o}$ consists of three individual back-scatter measurements from different viewing angles. In this article we do not make the dependence on incidence angle or view angle explicit, assuming this forms part of the computational methodology.

satellite observations, $\mathbf{\Sigma^o}$. Using Bayes' theorem:

$$P(U,V \mid \mathbf{\Sigma^o}) = \frac{P(\mathbf{\Sigma^o} \mid U,V)P(U,V)}{P(\mathbf{\Sigma^o})}. \tag{1}$$

Once the $\mathbf{\Sigma^o}$ have been observed, $P(\mathbf{\Sigma^o})$ is a constant and thus we can write:

$$P(U,V \mid \mathbf{\Sigma^o}) \propto P(\mathbf{\Sigma^o} \mid U,V)P(U,V). \tag{2}$$

Since the intention is to use Markov Chain Monte Carlo (Gilks *et al.*, 1996) methods to sample from the posterior distribution (2), $P(U,V \mid \mathbf{\Sigma^o})$, the normalising constant does not need to be evaluated. $P(\mathbf{\Sigma^o} \mid U,V)$ is the likelihood of the observations and $P(U,V)$ is the prior model. The next sections deal with the specification of the models in (2).

## 2.1 Prior wind-field models

The prior wind-field model, $P(U,V)$, is an informative prior, and will constrain the local solutions to produce consistent wind fields. It is required in this approach because the problem is otherwise ill-posed, that is the mapping $\boldsymbol{\sigma^o} \to (u,v)$ is multi-valued so no unique local solution exists. Our solution is to develop a Gaussian Process (GP) based wind-field model that is tuned to NWP assimilated data[2]. This ensures that the wind-field model is representative of NWP wind fields, which are the best available estimates of the true wind field.

GPs provide a flexible class of models (Williams, 1998) particularly when the variables are distributed in space. The wind field data is assumed to come from a multi-variate normal distribution, whose covariance matrix is a function of the spatial location of the observations and some physically interpretable parameters. A particular form of GP was chosen to incorporate geophysical information in the prior model (Cornford, 1998; Cornford, 1997).

We consider a non-zero mean GP model which produces highly realistic wind fields. Since there is also a temporal aspect to the problem, different parameter values are determined for each month of the year. The probability of a wind field, $(U,V)$, is given by:

$$P(U,V) = \frac{1}{(2\pi)^{\frac{n}{2}} \left|K_{uv}\right|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}\left(\begin{bmatrix}U\\V\end{bmatrix} - \begin{bmatrix}m_u\\m_v\end{bmatrix}\right)' K_{uv}^{-1}\left(\begin{bmatrix}U\\V\end{bmatrix} - \begin{bmatrix}m_u\\m_v\end{bmatrix}\right)\right)$$

where $m_u$, $m_v$ are the mean functions for $u$ and $v$ respectively and $K_{uv}$ is the joint covariance matrix for $(u,v)$. The covariance matrices are determined by appropriate covariance functions which have parameters to represent the variance, characteristic length scales of features, the noise variance and the ratio of divergence to vorticity in the wind fields. Tuning these parameters to their maximum *a posteriori* probability values using NWP data produces an accurate prior wind-field model which requires no additional NWP data once the parameters are determined. Since the GP model defines the wind over a continuous space domain it can be used for prediction at unmeasured locations, and can easily cope with missing or removed data.

## 2.2 Likelihood models

There are two approaches to computing the likelihood, based on the forward and inverse models, outlined below.

[2]Assimilated data means we are using the best guess initial conditions of a NWP model.

### 2.2.1 Use of forward models

We have developed a probabilistic forward model, $P(\boldsymbol{\sigma}_i^o|u_i, v_i)$, with methods described in (Ramage *et al.*, 1998, this issue). The wind vector in a given cell is assumed to completely determine the observed $\boldsymbol{\sigma}_i^o$ values, and thus *conditionally* on the wind vectors in each cell the $\boldsymbol{\sigma}_i^o$ values are independent and:

$$P(\boldsymbol{\Sigma}^o \mid U, V) = \prod_i P(\boldsymbol{\sigma}_i^o \mid u_i, v_i) \tag{3}$$

where the product is taken over all the cells in the region being considered. This decomposition is only valid for the *conditional* local model: the observations of $\boldsymbol{\sigma}_i^o$ and $(u_i, v_i)$ are not jointly independent in this way. (2) can now be written:

$$P(U, V \mid \boldsymbol{\Sigma}^o) \propto \left( \prod_i P(\boldsymbol{\sigma}_i^o \mid u_i, v_i) \right) P(U, V) \tag{4}$$

this being referred to as 'forward' disambiguation.

### 2.2.2 Use of inverse models

As shown in (Evans *et al.*, 1998, this issue) the inverse mapping has at least two possible values at each location, thus a probabilistic model of the form $P(u_i, v_i \mid \boldsymbol{\sigma}_i^o)$ is essential to fully model the mapping. Using Bayes' theorem:

$$P(\boldsymbol{\sigma}_i^o \mid u_i, v_i) = \frac{P(u_i, v_i \mid \boldsymbol{\sigma}_i^o) P(\boldsymbol{\sigma}_i^o)}{P(u_i, v_i)}. \tag{5}$$

Having observed the $\boldsymbol{\Sigma}^o$ values $P(\boldsymbol{\sigma}_i^o)$ is constant and (5) can be written:

$$P(\boldsymbol{\sigma}_i^o \mid u_i, v_i) \propto \frac{P(u_i, v_i \mid \boldsymbol{\sigma}_i^o)}{P(u_i, v_i)}. \tag{6}$$

Rewriting (4) using (6) gives:

$$P(U, V \mid \boldsymbol{\Sigma}^o) \propto \left( \prod_i \frac{P(u_i, v_i \mid \boldsymbol{\sigma}_i^o)}{P(u_i, v_i)} \right) P(U, V), \tag{7}$$

this being referred to as 'inverse' disambiguation. Using Bayes' theorem in this form has been called the *Scaled-Likelihood* method (Morgan and Boulard, 1995) although it has not been previously applied in this spatial context.

## 3 Application

When applying the models in an operational environment a modular design is proposed (Fig. 1). Initially the scatterometer data is pre-processed using the prior unconditional density model for $\boldsymbol{\sigma}^o$ to remove outliers. This model is currently under development using a specially modified Generative Topographic Mapping (Bishop *et al.*, 1998) with a conical latent space. The selected $\boldsymbol{\sigma}^o$ data is then forward propagated through the inverse model to produce $P(u_i, v_i \mid \boldsymbol{\sigma}_i^o)$. Some problem specific methods (not discussed here) are used to select an initial wind field. Starting from this point the posterior distribution, $P(U, V \mid \boldsymbol{\Sigma}^o)$, is sampled from using either (4) or (7).

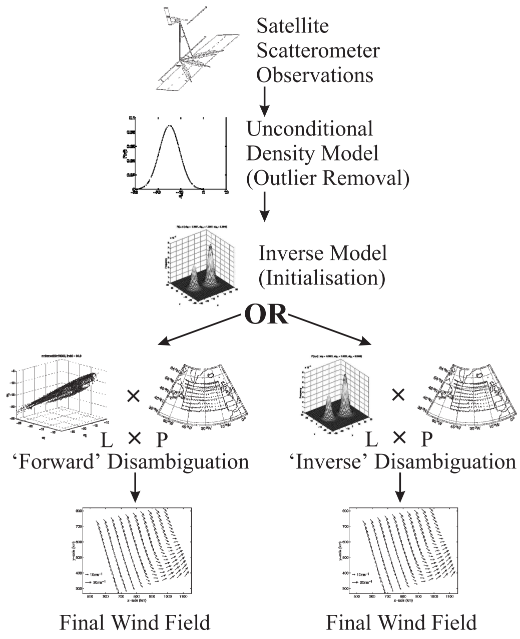


**Figure 1:** A schematic of the proposed methodology for operational wind field retrieval. The likelihood and prior models are denoted by L and P respectively. Either the forward or the inverse models can be used to obtain the final wind field.

A Markov Chain Monte Carlo approach using a modified[3] Metropolis algorithm (Gilks *et al.*, 1996) is applied. We also propose to use optimisation methods to quickly find the modes[4]. The problem is complicated by the high dimensionality of $P(U, V \mid \mathbf{\Sigma^o})$ which is typically of the order of four hundred.

Thus good *local* inverse and forward models are essential to accurate wind field retrieval. This modular scheme makes each part of the modelling process more focussed, while providing an elegant solution to a complex problem. If necessary individual models (such as the forward model) can be updated or changed while the other models can be kept fixed.

# 4 Preliminary Results

At the time of writing, not all components of the operational model proposed above were completed. Fig. 2 shows an example of the results of the application of (7) to a real wind field. The initialisation routine (which is itself complex) produces an excellent initial wind field (Fig. 2a) with a few poor wind vectors. The most probable wind field (Fig. 2b), which occurrs at time = 1323, can be seen to resemble the NWP winds (Fig. 2c), although the trough is more marked and positioned further south and east. It is these situations where the scatterometer data will improve the estimation of NWP initial conditions, since it is very likely that the scatterometer wind field is more reliable than the NWP forecast wind field.

Fig. 2d shows the evolution of the energy (which is equal to the un-normalised negative log probability of the wind field) during Metropolis sampling from the posterior (7). The energy falls

[3]The modifications take advantage of our knowledge of the likely ambiguities.

[4]It is known that the local inverse model has dominantly bimodal solutions, and it is suggested that this is also the case for the global solution. Thus a scaled conjugate gradient algorithm could be used to find both modes.

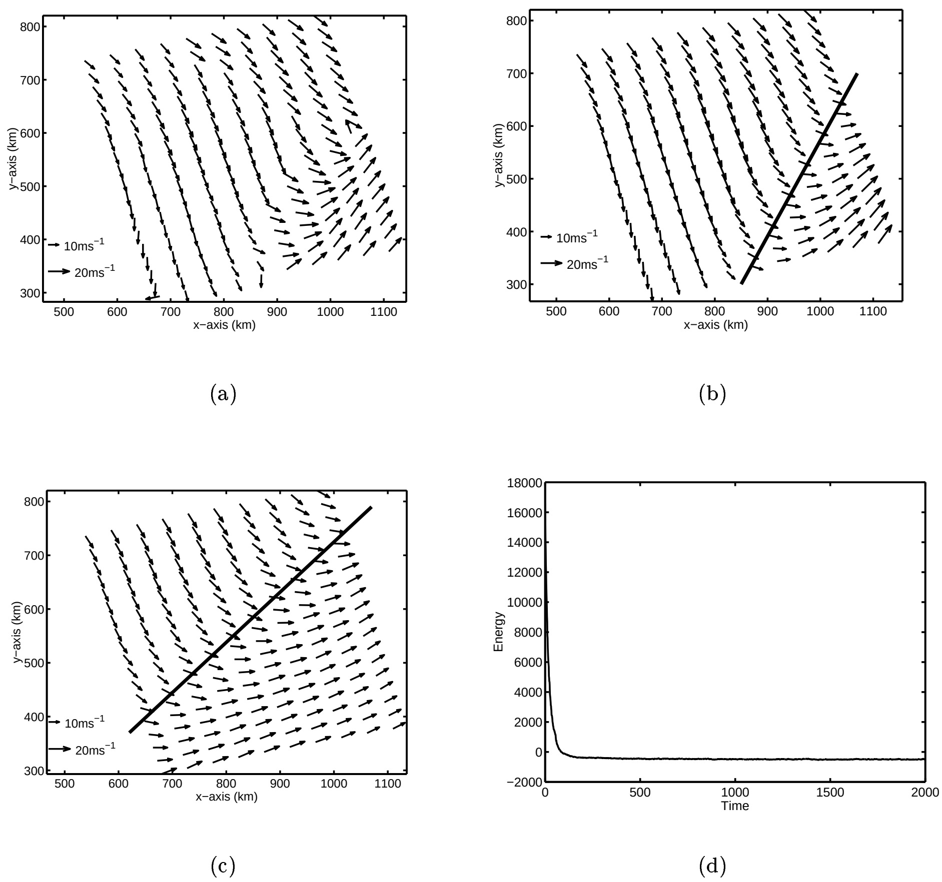


**Figure 2:** Results of running the MCMC sampling on a wind field from the North Atlantic on the 4th of January 1994. (a) the initialised wind field from the inverse model, (b) the most probable wind field from (7), (c) the 'true' NWP wind field and (d) the evolution of the energies in the Markov chain. The position of the trough is marked by a solid line in (b) and (c).

very rapidly in the first 50 iterations, and then remains relatively constant as the sampling explores the mode.

### 4.1 Discussion

The results presented indicate that the methodology is operationally viable. Fig. 2a indicates that there are some cells for which the inverse model does not give a good initialisation. These poor local predictions are believed to result from unreliable $\boldsymbol{\sigma^o}$ measurements and thus the addition of the unconditional prior model for $\boldsymbol{\sigma^o}$ may alleviate these problems.

Comparing Fig. 2b and 2c shows that the NWP winds are not always reliable especially where there are rapid spatial changes in direction. However, NWP winds are used to train all the models, and thus careful data selection is vital to achieving reliable training sets for neural network models for wind field retrieval.

## 5 Conclusions

We have presented an elegant framework for the retrieval of wind fields using neural network techniques from satellite scatterometer data. The modular approach adopted mean that model changes are simple to implement, since each model has a well defined role. In addition to the unconditional density model for $\boldsymbol{\sigma^o}$, we are developing wind field models to include atmospheric fronts (Cornford *et al.*, 1999).

Improving the forward and inverse models using better training data should further improve results. This will be vital when the methods are applied to more complex wind fields than that in Fig. 2. The approach taken will produce reliable, autonomously disambiguated, scatterometer derived wind fields. The use of Bayesian methods allows the assessment of posterior uncertainty.

## Acknowledgements

This work is supported by the European Union funded NEUROSAT programme (grant number ENV4 CT96-0314) and also EPSRC grant GR/L03088 *Combining Spatially Distributed Predictions from Neural Networks.* Thanks to MSc student David Evans for supplying the results of the application of 'inverse' disambiguation. Thanks also to David Offiler for his useful comments during the project.

## References

Bishop, C. M., M. Svensen, and C. K. I. Williams 1998. GTM: The Generative Topographic Mapping. *Neural Computation* **10**, 215–234.

Cornford, D. 1997. Random Field Models and Priors on Wind. Technical Report NCRG/97/023, Neural Computing Research Group, Aston University, Aston Triangle, Birmingham, UK. Available from: `http://www.ncrg.aston.ac.uk/`.

Cornford, D. 1998. Non-Zero Mean Gaussian Process Prior Wind Field Models. Technical Report NCRG/98/020, Neural Computing Research Group, Aston University, Aston Triangle, Birmingham, UK. Available from: `http://www.ncrg.aston.ac.uk/`.

Cornford, D., I. T. Nabney, and C. K. I. Williams 1999. Adding Constrained Discontinuities to Gaussian Process Models of Wind Fields. In *Advances In Neural Information Processing Systems 11*. The MIT Press. Accepted.

Evans, D. J., D. Cornford, and I. T. Nabney 1998. Structured Neural Network Modelling of Multi-valued Functions for Wind Retrieval from Scatterometer Measurements. *Neurocomputing Letters*. submitted.

Gilks, W., S. Richardson, and D. J. Spiegelhalter 1996. *Markov Chain Monte Carlo in Practice*. London: Chapman and Hall.

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

Morgan, N. and H. A. Boulard 1995. Neural Networks for Statistical Recognition of Continuous Speech. *Proceedings of the IEEE* **83**, 742–770.

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

Ramage, G., D. Cornford, and I. T. Nabney 1998. A Neural Network Sensor Model with Input Noise. *Neurocomputing Letters*. submitted.

Richaume, P., F. Badran, M. Crepon, C. Mejia, H. Roquet, and S. Thiria 1998. Neural Network Wind Retrieval from ERS-1 Satterometer Data. *Journal of Geophysical Research*. submitted.

Stoffelen, A. and D. Anderson 1997a. Ambiguity Removal and Assimiliation of Scatterometer Data. *Quarterly Journal of the Royal Meteorological Society* **123**, 491–518.

Stoffelen, A. and D. Anderson 1997c. 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.

Tarantola, A. 1987. *Inverse Problem Theory*. London: Elsevier.

Williams, C. K. I. 1998. Prediction with Gaussian Processes: From Linear Regression to Linear Prediction and Beyond. In M. I. Jordan (Ed.), *Learning and Inference in Graphical Models*. Kluwer.