Modelling Wind Direction from Satellite Scatterometer Data

Ian T Nabney and Christopher M Bishop

Neural Computing Research Group
Aston University, Birmingham B4 7ET, United Kingdom.
Email: bishopcm@aston.ac.uk nabneyit@aston.ac.uk
Tel: +44 (121) 359 3621 x4685
Fax: +44 (121) 333 6215
Neural Computing Research Group Report: NCRG/95/011
Available from http://www.ncrg.aston.ac.uk/

Abstract: *Most of the common techniques for estimating conditional probability densities are inappropriate for applications involving periodic variables. In this paper we apply two novel techniques to the problem of extracting the distribution of wind vector directions from radar scatterometer data gathered by a remote-sensing satellite.*

Introduction

In an earlier paper (Nabney and Bishop, 1995), we have introduced three novel techniques for modelling conditional distributions for periodic variables, and investigated their performance using synthetic data. In this paper we show how these techniques can be applied to the problem of determining the wind direction from radar scatterometer data gathered by the ERS-1 remote sensing satellite.

Extraction of the wind vector from the radar signals represents an inverse problem which is typically multi-valued. A conventional neural network approach to this problem, based on a least squares estimate of the direction, would predict directions which were given by conditional averages. Since the average of several valid wind directions is not itself a valid solution, such an approach would clearly fail. In this paper we show how to extract the complete distribution of wind directions (conditioned on the backscattered power and incidence angle) and hence avoid such problems.

Application to Radar Scatterometer Data

The European Remote Sensing satellite ERS-1 is equipped with three C-band radar antennae which measure the total backscattered power along three directions relative to the satellite track, as shown in Figure 1. When the satellite passes over the ocean, the strengths of the backscattered signals are related to the surface ripples of the water (on a scale of a few centimetres) which in turn are determined by the low level winds. The value σ_0 is the ratio of transmitted to backscattered power signal. Noise is introduced into the signal by the effects of wave breaking, modulation by long waves and rain.

Extraction of the wind vector from the radar signals represents an inverse problem which is typically multi-valued. Although determining the wind speed is relatively straightforward, the data gives rise to 'aliases' for the wind direction. For example, a wind direction of θ will give rise to similar radar signals to a wind direction of $\theta + \pi$, and there may be further aliases at other angles. A conventional neural network approach to this problem, based on a least

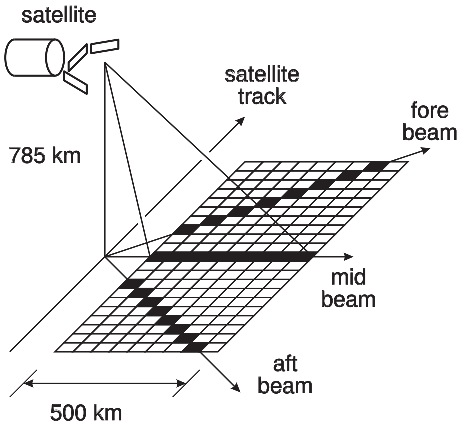

Figure 1: Schematic illustration of the ERS-1 satellite showing the footprints of the three radar scatterometers.

squares estimate of the direction, would predict directions which were given by conditional averages. Since the average of several valid wind directions is not itself a valid solution, such an approach would clearly fail. In this paper we show how to extract the complete distribution of wind directions (conditioned on the three σ_0 values and incidence angle) and hence avoid such problems.

Earlier neural network studies (see, for example, (Thiria et al., 1993)) were based on simulated data. In this paper, the direction was modelled with a target variable discretised into 36 bins. During testing, the estimated direction was calculated by interpolating over the 3 bins surrounding the histogram peak for a given input vector. To improve the disambiguation, inputs were taken from a 3×3 grid of cells around the cell of interest. Our approach is to model the conditional density purely locally (*i.e.* based on the values from a single cell) and then to combine these density functions (for example, using div-curl splines, see (Amodei and Benbourhim, 1991)) using global information about wind field properties. Here we report on the first step in the process.

A large dataset of ERS-1 measurements, spanning a wide range of meteorological conditions, has been assembled by the European Space Agency in collaboration with the UK Meteorological Office. Labelling of the dataset was performed using wind vectors from the Met Office Numerical Weather Prediction model. These values were interpolated from the inherently coarse-grained model to regions coincident with the scatterometer cells. This was the best estimate of the wind vectors that was available in sufficient quantities for developing a neural network solution. It is suitable for predicting a background wind direction, but may be less appropriate in regions where there are smaller scale features, such as frontal zones and other areas of high gradients in wind speed or direction.

The data that was selected for the experiments reported in this paper was collected from low pressure (cyclonic) and high pressure (anti-cyclonic) circulations. These conditions, rather than cases that were homogeneous or with a simple gradient in speed or direction, were chosen to provide a more challenging task to test the modelling techniques. Ten windfields from each of the two categories were used: each windfield contains $19 \times 19 = 361$ cells each of which represents an area of approximately $50 \times 50\text{km}$[1]. This gives a total of 7220 patterns, although the data for some of the cells was missing. When the data was split into three subsets, each contained 1963 patterns.

[1]Given that the total footprint width is 500 km, this implies that the cells overlap to some extent

Density Estimation for Periodic Variables

A commonly used technique for *unconditional* density estimation is based on mixture models of the form

$$p(\mathbf{t}) = \sum_{i=1}^{m} \alpha_i \phi_i(\mathbf{t}) \tag{1}$$

where α_i are called mixing coefficients, and the kernel functions $\phi_i(\mathbf{t})$ are typically chosen to be Gaussians. In order to turn this into a model for conditional density estimation, we simply make the mixing coefficients and any parameters in the kernels into functions of the input vector $\mathbf{x}$. This can be achieved by setting these parameters from the outputs of a neural network which takes $\mathbf{x}$ as input. This technique underlies the 'mixture of experts' model (Jacobs et al., 1991) and has also been considered by a number of other authors ((Bishop, 1994); (?)). In this paper we use a standard multi-layer perceptron with a single hidden layer of sigmoidal units and an output layer of linear units.

In this section we briefly show how we have extended this technique to provide three distinct methods for modelling a conditional density $p(\theta|\mathbf{x})$ of a periodic variable θ. For further details, see (Nabney and Bishop, 1995). For all three models, the error function E which is used is given by the negative logarithm of the likelihood function for the data with respect to the density function given by the network/mixture model combination.

Mixture of Wrapped Normal Densities

The first technique involves a transformation from a Euclidean variable $\chi \in (-\infty, \infty)$ to the periodic variable $\theta \in [0, 2\pi)$ of the form $\theta = \chi \bmod 2\pi$. This induces a transformation of density functions p with domain $\mathbb{R}$ to functions $\tilde{p}$ with domain $[0, 2\pi)$ as follows:

$$\tilde{p}(\theta|\mathbf{x}) = \sum_{N=-\infty}^{\infty} p(\theta + N2\pi|\mathbf{x}) \tag{2}$$

It is clear by construction that the function $\tilde{p}$ is a density function (since its integral is 1) and that it also satisfies the periodicity requirement $\tilde{p}(\theta + 2\pi|\mathbf{x}) = \tilde{p}(\theta|\mathbf{x})$.

Various choices for the component density functions which make up the mixture $\tilde{p}(\chi|\mathbf{x})$ are possible, but here we shall restrict attention to functions which are Gaussian of the form

$$\phi_i(t|\mathbf{x}) = \frac{1}{(2\pi)^{1/2}\sigma_i(\mathbf{x})} \exp\left\{-\frac{\|t - \mu_i(\mathbf{x})\|^2}{2\sigma_i(\mathbf{x})^2}\right\} \tag{3}$$

where $t \in \mathbb{R}$. The transformed density function $\tilde{\phi}_i$ is known as the 'wrapped normal' distribution (Kotz and Johnson, 1992).

The density function $p(\chi|\mathbf{x})$ is modelled using a combination of a neural network and a mixture model as described above.

Mixtures of Circular Normal Densities

The second novel approach is also based on a mixture of kernel functions, but in this case the kernels themselves are periodic, thereby ensuring that the overall conditional density function is periodic. The distribution we use has the form

$$p(\theta) = \frac{1}{2\pi I_0(m)} \exp\{m \cos(\theta - \psi)\} \tag{4}$$

which is known as a *circular normal* or *von Mises* distribution (Mardia, 1972). The normalization coefficient is expressed in terms of the zeroth order modified Bessel function of the first kind, $I_0(m)$, and the parameter m is analogous to the inverse variance parameter in a conventional normal distribution. The parameter ψ corresponds to the peak of the density function.

With this choice of kernel function, we again use a neural network to determine the parameters $\alpha_i(\mathbf{x})$, $\theta_i(\mathbf{x})$ and $m_i(\mathbf{x})$ in a mixture model to generate a periodic conditional density function.

Expansion in Fixed Kernels

The third and final technique introduced in this paper involves a conditional density model consisting of a fixed set of periodic kernels, again given by circular normal functions as in equation (4). In this case the mixing proportions alone are determined by the outputs of a neural network (through a softmax activation function) and the centres ψ_i and width parameters m_i are fixed. We selected a uniform distribution of centres, and $m_i = m$ for each kernel, where the value for m was chosen to give moderate overlap between the component functions.

Clearly a major drawback of fixed kernel methods is that the number of kernels must grow exponentially with the dimensionality of the output space. However, for a single output variable, as in this application, they can be regarded as practical techniques.

Density Modelling

After selection of the wind field datasets, we then trained neural networks to model the wind direction distribution. The inputs used were the three values of σ_0 for the aft-beam, mid-beam and fore-beam, and the sine of the incidence angle of the mid-beam, since this angle strongly influences the reflected signal received by the scatterometer. The σ_0 inputs were scaled to have zero mean and unit variance, while the fourth input value was passed to the network unchanged. The target value was expressed as an angle clockwise from the satellite's forward path and converted to radians. A conjugate gradient algorithm and 'early stopping' were used to train the networks.

Table 1 gives a summary of the preliminary results obtained with each of the three methods. As expected, the complexity of the domain meant that there were many difficulties with local optima. In fact, over 75% of the training runs ended with the network trapped in a local minimum of the error surface. It was found that the validation error for methods 1 and 2 was lowest for models with eight centres, even though fewer centres are actually required to model the conditional density function well. This is also demonstrated by the graph in figure 2 which shows the conditional distribution of wind directions given by this network at a typical data point from the test set, and which is clearly bi-modal. This graph also shows how the fixed kernel model gives rise to a conditional density with a different alignment.

Figure 2 also shows that the density given by the fixed kernel model has four peaks. Note that the extra peaks are not just caused by the variance parameter being too small, since there are kernel centres close to the minima between the two pairs of adjacent local maxima. The

Method	Centres	Hidden Units	Validation Error	Test Error
1	4	20	2474.6	2446.2
2	6	20	2308.0	2337.9
3	36	24	2028.9	1908.9

Table 1: Results on satellite data.
Method 1: Mixture of wrapped normal functions.
Method 2: Mixture of adaptive circular normal functions.
Method 3: Mixture of fixed kernel functions.

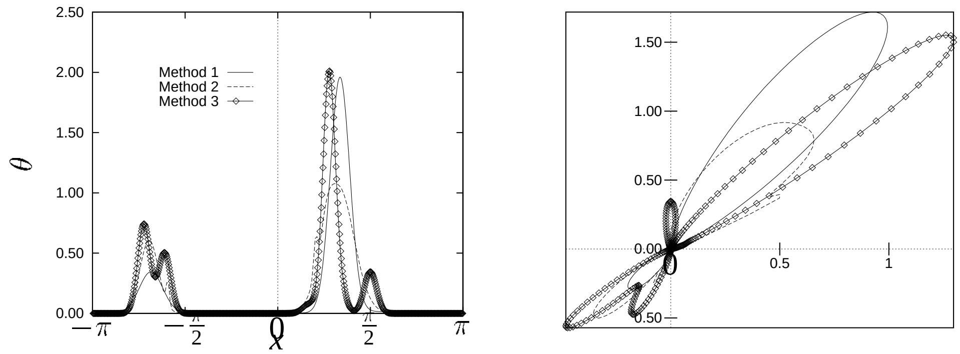

Figure 2: Plots of the conditional distribution $p(\theta|\mathbf{x})$ obtained using all three methods. (a) and (b) show linear and polar plots of the distributions for a given input vector from the test set. The dominant alias at π is evident. In both plots, the solid curve represents method 1, the dashed curve represents method 2, and the curve with circles represents method 3.

differences between the shapes of these distributions, despite the relatively close agreement of the data likelihood values, suggests that more training data is required to obtain better estimates of the density functions.

Discussion

In this paper we have a new approach to modelling the wind direction based on satellite scatterometer data. It involves modelling the conditional density distribution of the direction at a local level, and we have shown how some novel neural network methods for doing this can be successfully applied. A conventional neural network approach, involving the minimization of a sum-of-squares error function, would perform poorly on this problem since the required mapping is multi-valued.

An advantage of the density modelling approach is that it enables a better understanding of the direction ambiguities to be formed. There is the suggestion that there are other aliases than at an angle of π in the distribution. In addition, it provides the most complete information for the next stage of processing, which is to 'de-alias' the wind directions by combining local information to determine the most probable overall wind field.

Ackowledgements

We are grateful to the European Space Agency and the UK Meteorological Office for making available the ERS-1 data. The contributions of Claire Legleye in the early stages of the project are also gratefully acknowledged. We would also like to thank Iain Strachan and Ian Kirk of AEA Technology for a number of useful discussions relating to the interpretation of this data.

References

Amodei, L. and Benbourhim, M. N. (1991). A vector spline approximation. *Journal of Approximation Theory*, 67:51–79.

Bishop, C. M. (1994). Mixture density networks. Technical Report NCRG/4288, Neural Computing Research Group, Aston University, U.K.

Jacobs, R. A., Jordan, M. I., Nowlan, S. J., and Hinton, G. E. (1991). Adaptive mixtures of local experts. *Neural Computation*, 3:79–87.

Kotz, S. and Johnson, N. L., editors (1992). *Encyclopedia of Statistical Sciences*, pages 381–386. Wiley.

Mardia, K. V. (1972). *Statistics of Directional Data.* Academic Press, London.

Nabney, I. T. and Bishop, C. M. (1995). Modelling conditional probability distributions for periodic variables. In *4th International Conference on Artificial Neural Networks*, pages 177–182. IEE.

Thiria, S., Mejia, C., Badran, F., and Crepon, M. (1993). A neural network approach for modeling nonlinear transfer functions: Application for wind retrieval from spaceborne scatterometer data. *Journal of Geophysical Research*, 98(C12):22827–22841.