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/`

Adding Constrained Discontinuities to Gaussian Process Models of Wind Fields

Dan Cornford **Ian T. Nabney** **Christopher K. I. Williams**[*]
`d.cornford@aston.ac.uk` `i.t.nabney@aston.ac.uk` `ckiw@dai.ed.ac.uk`

Technical Report NCRG/98/025 November 27, 1998

Accepted NIPS98

Abstract

Gaussian Processes provide good prior models for spatial data, but can be too smooth. In many physical situations there are discontinuities along bounding surfaces, for example fronts in near-surface wind fields. We describe a modelling method for such a constrained discontinuity and demonstrate how to infer the model parameters in wind fields with MCMC sampling.

---

[*] Now at: Division of Informatics, University of Edinburgh, 5 Forrest Hill, Edinburgh EH1 2QL, Scotland, UK

1 INTRODUCTION

We introduce a model for wind fields based on Gaussian Processes (GPs) with 'constrained discontinuities'. GPs provide a flexible framework for modelling various systems. They have been adopted in the neural network community and are interpreted as placing priors over functions.

Stationary vector-valued GP models (Daley, 1991) can produce realistic wind fields when run as a generative model; however, the resulting wind fields do not contain some features typical of the atmosphere. The most difficult features to include are surface fronts. Fronts are generated by complex atmospheric dynamics and are marked by large changes in the surface wind direction (see for example Figures 2a and 3a) and temperature. In order to account for such features, which appear discontinuous at our observation scale, we have developed a model for vector-valued GPs with constrained discontinuities which could also be applied in domains such as surface reconstruction in computer vision, and geostatistics.

In section 2 we illustrate the generative model for wind fields with fronts. Section 3 explains what we mean by GPs with constrained discontinuities and derives the likelihood of data under the model. Results of Bayesian estimation of the model parameters are given, using a Markov Chain Monte Carlo (MCMC) procedure. In the final section, the strengths and weaknesses of the model are discussed and improvements suggested.

2 A GENERATIVE WIND FIELD MODEL

We are primarily interested in retrieving wind fields from satellite scatterometer observations of the ocean surface[1]. A probabilistic prior model for wind fields will be used in a Bayesian procedure to resolve ambiguities in local predictions of wind direction. The generative model for a wind field including a front is taken to be a combination of two vector-valued GPs with a constrained discontinuity.

A common method for representing wind fields is to put GP priors over the velocity potential Φ and stream function Ψ of the wind field (Daley, 1991). The horizontal wind vector $\boldsymbol{u} = (u, v)$ can then be derived from:

$$u = -\frac{\partial \Psi}{\partial y} + \frac{\partial \Phi}{\partial x}, \qquad v = \frac{\partial \Psi}{\partial x} + \frac{\partial \Phi}{\partial y}. \tag{1}$$

This produces good prior models for wind fields when a suitable choice of covariance function for Φ and Ψ is made. We have investigated using a modified Bessel function based covariance[2] (Handcock and Wallis, 1994) but found, using three years of wind data for the North Atlantic, that the maximum a posteriori probability value for the smoothness parameter[3] in this covariance function was ~ 2.5. Thus we used the correlation function:

$$\rho(r) = \left(1 + \frac{r}{L} + \frac{r^2}{3L^2}\right) \exp\left(-\frac{r}{L}\right) \tag{2}$$

where L is the correlation length scale, which is equivalent and less computationally demanding (Cornford, 1998).

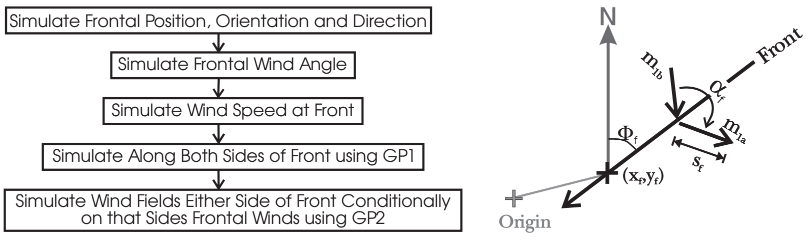
The generative model has the form outlined in Figure 1a. Initially the frontal position and orientation is simulated. This is given by an angle clockwise from north (ϕ_f) that the front makes and a point on the line (x_f, y_f). The present model generates straight fronts. Having defined the

---

[1] See `http://www.ncrg.aston.ac.uk/Projects/NEUROSAT/NEUROSAT.html` for details of the scatterometer work. Technical reports describing in more detail methods for generating prior wind field models can also be accessed from the same page.

[2] The modified Bessel function allows us to control the differentiability of the sample realisations through the 'smoothness parameter', as well as the length scales and variances.

[3] This varies with season, but is the most temporally stable parameter in the covariance function.



(a) (b)

Figure 1: (a) Flowchart describing the generative frontal model. See text for full description. (b) A description of the frontal model.

position of the front, the angle of the wind across the front (α_f) is simulated from a distribution covering the range $[0, \pi)$. This angle is related to the vertical component of vorticity (ζ) across the front through $\zeta = k \cdot \nabla \times \boldsymbol{u} \propto \cos\left(\frac{\alpha_f}{2}\right)$ and the constraint $\alpha_f \in [0, \pi)$ ensures cyclonic vorticity at the front. It is assumed that the front bisects α_f. The wind speed (s_f) is then simulated at the front. Since there is generally little change in wind speed across the front one value is simulated for both sides of the front. These components $\boldsymbol{\theta_f} = (\phi_f, x_f, y_f, \alpha_f, s_f)$ define the line of the front and the mean wind vectors just ahead of and just behind:

$$m_{1a} = (u^m_{1a}, v^m_{1a}) = \left(s_f \sin\left(\phi_f + \frac{\alpha_f}{2}\right), s_f \cos\left(\phi_f + \frac{\alpha_f}{2}\right)\right) \tag{3}$$

$$m_{1b} = (u^m_{1b}, v^m_{1b}) = \left(-s_f \sin\left(\phi_f - \frac{\alpha_f}{2}\right), -s_f \cos\left(\phi_f - \frac{\alpha_f}{2}\right)\right) \tag{4}$$

the front (Figure 1b).

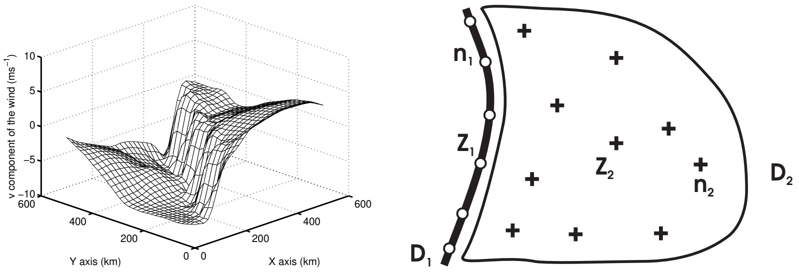
A realistic model requires some variability in wind vectors along the front. Thus we use a GP with a non-zero mean (m_{1a} or m_{1b}) along the line of the front. In the real atmosphere we observe a smaller variability in the wind vectors along the line of the front compared with regions away from fronts. Thus we use a different GP along the front (GP_1), from that used in the wind field away from the front (GP_2), although the same GP_1 parameters are used both sides of the front, just with different means. The winds just ahead of and behind the front are assumed conditionally independent given m_{1a} and m_{1b}, and are simulated at a regular 50 km spacing. The final step in the generative model is to simulate wind vectors using GP_2 in both regions either side of the front, conditionally on the values along that side of the front. This model is flexible enough to represent fronts, yet has the required constraints derived from meteorological principles, for example that fronts should always be associated with cyclonic vorticity with discontinuities at the model scale in wind direction but not in wind speed[4]. To make this generative model useful, we need to be able to compute the data likelihood, which is the subject of the next section.

3 GPs WITH CONSTRAINED DISCONTINUITIES

We consider data from two domains D_1 and D_2 (Figure 2b), where in this case D_1 is a curve in the plane which is intended to be the front and D_2 is a region of the plane. We obtain n_1 variables $\boldsymbol{Z_1}$ at points $\boldsymbol{x_1}$ along the curve, and we assume these are generated under GP_1 (a GP which depends on parameters $\boldsymbol{\theta_1}$ and has mean $\boldsymbol{m_1} = m_1\mathbf{1}$ which will be determined by (3) or (4)). We are interested in determining the likelihood of the variables $\boldsymbol{Z_2}$ observed at n_2 points $\boldsymbol{x_2}$ under GP_2 which depends on parameters $\boldsymbol{\theta_2}$, conditionally on the 'constrained discontinuities' at the front.

---

[4] The model allows small discontinuities in wind speed, which are consistent with frontal dynamics.



(a) (b)

Figure 2: (a) The discontinuity in one of the vector components in a simulation. (b) Framework for GPs with boundary conditions. The curve D_1 has n_1 sample points with values $\boldsymbol{Z_1}$. The domain D_2 has n_2 points with values $\boldsymbol{Z_2}$.

We evaluate this by calculating the likelihood of $\boldsymbol{Z_2}$ conditionally on the n_1 values of $\boldsymbol{Z_1}$ from GP_1 along the front and marginalising out $\boldsymbol{Z_1}$:

$$P(\boldsymbol{Z_2}|\boldsymbol{\theta_2},\boldsymbol{\theta_1}) = \int_{-\infty}^{\infty} P(\boldsymbol{Z_2}|\boldsymbol{Z_1},\boldsymbol{\theta_2},\boldsymbol{\theta_1},\boldsymbol{m_1})P(\boldsymbol{Z_1}|\boldsymbol{\theta_1},\boldsymbol{m_1})d\boldsymbol{Z_1}. \tag{5}$$

From the definition of the likelihood of a GP (Cressie, 1993) we find:

$$P(\boldsymbol{Z_2}|\boldsymbol{Z_1},\boldsymbol{\theta_2},\boldsymbol{\theta_1},\boldsymbol{m_1}) = \frac{1}{(2\pi)^{\frac{n2}{2}}|S_{22}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}\boldsymbol{Z_2^*}' S_{22}^{-1}\boldsymbol{Z_2^*}\right) \tag{6}$$

where:

$$S_{22} = K_{22|2} - K'_{12|2}K_{11|2}^{-1}K_{12|2}, \qquad \boldsymbol{Z_2^*} = \boldsymbol{Z_2} - K'_{12|2}K_{11|2}^{-1}\boldsymbol{Z_1}.$$

To understand the notation consider the joint distribution of $\boldsymbol{Z_1}, \boldsymbol{Z_2}$ and in particular its covariance matrix (assuming that both random fields are generated under $GP2$):

$$K = \begin{bmatrix} K_{11|2} & K_{12|2} \\ K_{21|2} & K_{22|2} \end{bmatrix} \tag{7}$$

where $K_{11|2}$ is taken to mean the $n_1 \times n_1$ covariance matrix between the points in D_1 evaluated under the covariance for GP_2, $K_{12|2} = K'_{21|2}$ the $n_1 \times n_2$ (cross) covariance matrix between the points in D_1 and D_2 evaluated using $\boldsymbol{\theta_2}$ and $K_{22|2}$ is the usual $n_2 \times n_2$ covariance for points in D_2. Thus we can see that S_{22} is the $n_2 \times n_2$ modified covariance for the points in D_2 given the points along D_1, while the $\boldsymbol{Z_2^*}$ is the corrected mean that accounts for the values at the points in D_1, which have non-zero mean.

We remove the dependency on the values $\boldsymbol{Z_1}$ by evaluating the integral in (5). $P(\boldsymbol{Z_1}|\boldsymbol{\theta_1},\boldsymbol{m_1})$ is given by:

$$P(\boldsymbol{Z_1}|\boldsymbol{\theta_1},\boldsymbol{m_1}) = \frac{1}{(2\pi)^{\frac{n_1}{2}}|K_{11|1}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}\left(\boldsymbol{Z_1} - \boldsymbol{m_1}\right)' K_{11|1}^{-1}\left(\boldsymbol{Z_1} - \boldsymbol{m_1}\right)\right) \tag{8}$$

where $K_{11|1}$ is the $n_1 \times n_1$ covariance matrix between the points in D_1 evaluated under the covariance given by $\boldsymbol{\theta_1}$. Completing the square in $\boldsymbol{Z_1}$ in the exponent, the integral (5) can be evaluated to give:

$$\begin{aligned} P(\boldsymbol{Z_2}|\boldsymbol{\theta_2},\boldsymbol{\theta_1},\boldsymbol{m_1}) = {} & \frac{1}{(2\pi)^{\frac{n2}{2}}}\frac{1}{|S_{22}|^{\frac{1}{2}}}\frac{1}{|K_{11|1}|^{\frac{1}{2}}}\frac{1}{|B|^{\frac{1}{2}}} \times \\ & \exp\left(\frac{1}{2}\left(C'B^{-1}C - \boldsymbol{Z_2}'S_{22}^{-1}\boldsymbol{Z_2} - \boldsymbol{m_1}'K_{11|1}^{-1}\boldsymbol{m_1}\right)\right) \end{aligned} \tag{9}$$

where:

$$\begin{aligned} B &= (K'_{12|2}K^{-1}_{11|2})'S^{-1}_{22}K'_{12|2}K^{-1}_{11|2} + K^{-1}_{11|1} \\ C' &= \boldsymbol{Z_2}'S^{-1}_{22}K'_{12|2}K^{-1}_{11|2} + \boldsymbol{m_1}'K^{-1}_{11|1} \end{aligned}$$

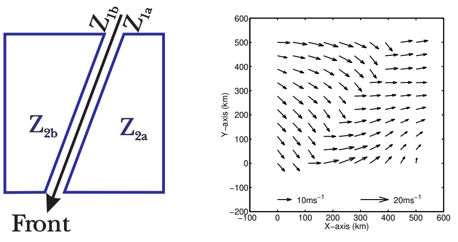
The algorithm has been coded in MATLAB and can deal with reasonably large numbers of points quickly. With $n_1 = 12$ and $n_2 = 200$ for a two dimensional vector-valued GP[5] based on a covariance function given by (2) computation of the log likelihood takes 4.13 seconds on an SGI Indy R5000.

3.1 ESTIMATING POSTERIOR PROBABILITIES OF MODEL PARAMETERS

The mean value just ahead and behind the front define the mean values for the constrained discontinuity (i.e. $\boldsymbol{m_1}$ in (9)). Conditional on the frontal parameters the wind fields either side (Figure 3a) are assumed independent:

$$\begin{aligned} P(\boldsymbol{Z_{2a}}, \boldsymbol{Z_{2b}}|\boldsymbol{\theta_2}, \boldsymbol{\theta_1}, \boldsymbol{\theta_f}) = P(\boldsymbol{Z_{2a}}|\boldsymbol{\theta_2}, \boldsymbol{\theta_1}, \boldsymbol{m_{1a}})P(\boldsymbol{m_{1a}}|\boldsymbol{\theta_f}) \times \\ P(\boldsymbol{Z_{2b}}|\boldsymbol{\theta_2}, \boldsymbol{\theta_1}, \boldsymbol{m_{1b}})P(\boldsymbol{m_{1b}}|\boldsymbol{\theta_f}) \end{aligned}$$

where we have performed the integration (5) to remove the dependency on $\boldsymbol{Z_{1a}}$ and $\boldsymbol{Z_{1b}}$. Thus the likelihood of the data $\boldsymbol{Z_2} = (\boldsymbol{Z_{2a}}, \boldsymbol{Z_{2b}})$ given the model parameters $\boldsymbol{\theta_2}, \boldsymbol{\theta_1}, \boldsymbol{\theta_f}$ is simply the product of the likelihoods of two GPs with a constrained discontinuity which can be computed using (9).



(a) (b)

Figure 3: (a) The division of the wind field using the generative frontal model. $\boldsymbol{Z_{1a}}$, $\boldsymbol{Z_{1b}}$ are the wind fields just ahead and behind the front, along its length, respectively. $\boldsymbol{Z_{2a}}$, $\boldsymbol{Z_{2b}}$ are the wind fields in the regions ahead of and behind the front respectively. (b) An example from the generative frontal model: the wind field used to test the likelihood computations which looks like a typical 'cold front'.

The model outlined above is tested on simulated data generated from the model to assess model sensitivity. We generate a wind field $\boldsymbol{Z^o} = (\boldsymbol{Z^o_{2a}}, \boldsymbol{Z^o_{2b}})$ using known model parameters (e.g. Figure 3b). We then investigate the sampling of model parameters from the posterior distribution:

$$P(\boldsymbol{\theta_2}, \boldsymbol{\theta_1}, \boldsymbol{\theta_f}|\boldsymbol{Z^o}) \propto P(\boldsymbol{Z^o}|\boldsymbol{\theta_2}, \boldsymbol{\theta_1}, \boldsymbol{\theta_f})P(\boldsymbol{\theta_2})P(\boldsymbol{\theta_1})P(\boldsymbol{\theta_f}) \qquad (10)$$

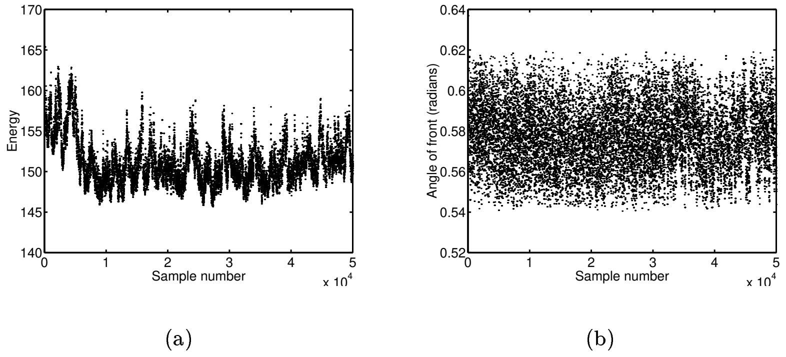
where $P(\boldsymbol{\theta_2}), P(\boldsymbol{\theta_1}), P(\boldsymbol{\theta_f})$ are prior distributions over the parameters in the GPs and front models. This brings out one advantage of the proposed model. All the model parameters have a physical interpretation and thus expert knowledge was used to set priors which produce realistic wind fields. We will also use (10) to help set (hyper)priors using real data in $\boldsymbol{Z^o}$.

MCMC using the Metropolis algorithm (Neal, 1993) is used to sample from (10) using the NETLAB[6] library for MATLAB. Convergence of the Markov chain is currently assessed using visual inspection

---

[5] This is equivalent to $n_1 = 24$ and $n_2 = 400$ for a scalar GP.

[6] Available from `http://www.ncrg.aston.ac.uk/netlab/index.html`.



(a) (b)

Figure 4: Examples from the Markov chain of the posterior distribution (10): (a) the energy = negative log posterior probability. Note that the energy when the chain was initialised was 2789 and the first 27 values are outside the range of the y-axis, (b) the angle of the front relative to north (ϕ_f).

of the univariate sample paths since the generating parameters are known, although other diagnostics could be used (Cowles and Carlin, 1996). We find that the procedure is insensitive to the initial value of the GP parameters, but that the parameters describing the location of the front (ϕ_f, d_f) need to be initialised 'close' to the correct values if the chain is to be run for sensible times. In application some preliminary analysis of the wind field would be necessary to identify possible fronts and thus set the initial parameters to 'sensible' values. We intend to fit a vector-valued GP without any discontinuities and then measure the 'strain' or misfit of the observed winds with the fitted winds. Lines of large 'strain' will be used to initialise the front parameters.