Variational Inference in Stochastic Dynamic Environmental Models

Dan Cornford[1], Manfred Opper[2], John Shawe-Taylor[2], Ian Roulstone[3], Peter Clark[4]

[1] NCRG, School of Engineering and Applied Science, Aston University, Birmingham B4 7ET
[2] ISIS, Computer Science, Southampton University, Southampton SO17 1BJ
[3] Dept. of Maths and Statistics, School of Electronics and Physical Sciences, University of Surrey, Guildford GU2 7XH.
[4] Joint Centre for Mesoscale Meteorology, Met Office, Reading RG6 6BB.

Background

The improvements in computational power that are anticipated over the next decade will enable the development of models that permit the study of emergent behaviour of complex interacting systems, which depend on a *huge* number of degrees of freedom. Environmental forecasting centres have taken strategic decisions to develop high resolution models that will be used to produce detailed regional forecasts. These systems are being designed to predict extreme weather and the output will, in turn, be used to forecast the impact on related phenomena, such as flooding and storm damage, and on the spread of pollutants. The models needed to achieve these capabilities will require fundamentally new approaches to those used in current systems, because the physical assumptions and mathematical approximations employed in current models are likely to be invalid at finer resolutions. Motivated by these considerations, this project will generate an innovative set of machine learning, and other computational developments, to address the above issues.

With finer-resolution models and a greater density of observations, we will never be able to determine the exact state of the atmosphere even at the resolution of the model. This is especially problematic if the simulation is to represent severe, storm-related, events because the progenitors of these phenomena are crucially dependent on discontinuous, nonlinear, dynamical structures with scales at or below the grid-scale of the model. At finer resolutions, not only are the details of the physical interactions more complex, but the basic mathematical description of such processes is much less certain. Hence, the treatment of model error (due to missing or unresolved processes) and the production of probabilistic forecasts are major challenging problems [1].

This project will provide a new principled approach for tackling this problem. We base our research on techniques for approximate inference in large statistical models, which originated in statistical physics but have been further developed and refined in the field of machine learning (for a review, see e.g. [2]). We have shown that such methods provide fast and reliable predictions in a *static* environmental data assimilation problem [3]. Utilising related methods in the dynamic case raises a large range of research questions, discussed in the next section, which this project is uniquely able to address. The project proposes a novel computational approach to inference in complex dynamical systems, which requires interdisciplinary skills from machine learning, computer science, mathematics, engineering and statistical physics, and will deliver a significant and innovative framework for probabilistic environmental prediction. The project brings together machine learning, computational science and meteorological skills at Southampton, Aston and Surrey, strengthened by the involvement of the Met Office as a project partner. The members of this project also have established links with both the Universities Weather Research Network (UWERN) and the NERC Data Assimilation Research Centre (DARC).

Envisaged approach

The VISDEM project will extend recent advances in machine learning to develop a computational inference framework for continuous time stochastic dynamical systems, which will enable:

- the probabilistic treatment of very large dynamical models;
- the inference of the probability distribution function of state variables given observations (data assimilation);
- inference of parameters for representing the stochastic forcings that are not resolved explicitly by the deterministic system.

A recent workshop, funded by the NERC/EPSRC programme on Environmental Mathematics and Statistics, on mathematical techniques for improving the forecasting of hazardous weather, concluded that *probabilistic methods* were going to be essential to developing high resolution models [4]. Further, due to the complexity and to the broad spectrum of scales of the interactions, it would be essential to use high resolution models as part of the data assimilation process. Both of these conclusions imply significant departures from the established methodologies employed in current environmental modelling which will be addressed by major computational developments in machine learning, motivated by this recognised requirement. The major problem in implementing a probabilistic approach to model uncertainties within existing environmental forecast systems such as the ensemble Kalman filter

[5] and 4D VAR [6]), is the vast computational complexity in computing reliable probabilistic forecasts when a sufficiently complex stochastic model is used.

Framework and methodology

We are concerned with the inference of the state of a large, complex dynamical model of an environmental system, such as a weather forecast model. We assume that the state $\boldsymbol{x}$ of the system evolves deterministically in space, $\boldsymbol{s}$, and time, t, according to a partial differential equation (PDE)

$$\frac{\partial \boldsymbol{x}(\boldsymbol{s},t)}{\partial t} = F[\boldsymbol{x}(\boldsymbol{s},t),\boldsymbol{s}] \;, \tag{1}$$

where F represents a functional, typically expressed by a series of (partial) differential operators and boundary forcing functions. We assume that there are no explicit stochastic terms in F.

Integration of (1) is usually only possible by numerical methods, but such methods invariably introduce approximations. These come from i) the finite resolution of the state caused by spatial discretisation of the *functions* $\boldsymbol{x}(\boldsymbol{s},t)$ leading to an approximation of the functional F by a function G of the state *vector* and the replacement of the PDE by a system of ordinary differential equations (ODEs); and from, ii) processes in the system, F, which are not completely known and thus cannot be explicitly included in a deterministic numerical model. Our crucial assumption is that the uncertainties about the evolution of the true state of the system caused by these model approximations can be modelled *probabilistically* by adding a stochastic noise term $\boldsymbol{\eta}$ to G. In environmental models effects which are important to the evolution of the system but are not represented in the finite dimensional dynamics are referred to as *parameterisations*. In most current implementations these *parameterisations* are based around deterministic representations of the unresolved processes in terms of the resolved state variables, but these deterministic parameterisations often do not produce the self-organising structure seen in the real world. Stochastic parameterisations, represented here by $\boldsymbol{\eta}$, have recently received attention [7, 1] because it is believed that deterministic parameterisations eliminate important physical interactions between the modelled flow and the unresolved fluctuations.

Hence, our inference on the state vector of the system will be based on the system of stochastic differential equations (SDEs)

$$\frac{d\boldsymbol{x}(\boldsymbol{s},t)}{dt} = G[\boldsymbol{x}(\boldsymbol{s},t),\boldsymbol{s}] + \boldsymbol{\eta}(\boldsymbol{s},t) \;. \tag{2}$$

Since G is in general nonlinear, the inclusion of the stochastic term does *not* simply lead to a fluctuation of the state around the deterministic path obtained by setting $\boldsymbol{\eta}(\boldsymbol{s},t) = 0$. Thus the calculation of the *expected* (or mean) state vector becomes a highly nontrivial problem. The stochastic term might, in general, be state-dependent, and should be expressed by a multiplicative noise process of the form $\boldsymbol{\eta}(\boldsymbol{s},t) = f(\boldsymbol{x}(\boldsymbol{s},t))\boldsymbol{\xi}(\boldsymbol{s},t)$, where $\boldsymbol{\xi}(\boldsymbol{s},t)$ is defined as a Gaussian process with a specified space-time covariance kernel, $K(\boldsymbol{s},\boldsymbol{s}',t,t')$, and the function f models the state dependence. A Gaussian approximation can be justified by a heuristic central limit theorem argument, using the fact that the stochastic term represents the combination of a very large number of unresolved processes and discretisation errors. The kernel K should be of a simple parametric form, chosen to reflect some of the basic characteristics (e.g. time scales and length scales) of the modes neglected in the construction of G. In the proposed project, we will start assuming the simplest case of additive noise (i.e. assuming $f(\boldsymbol{x}(\boldsymbol{s},t)) = 1$) and then look at some of the more sophisticated representations of noise [8] once the simpler case has been studied carefully. Of special importance is the question of how certain invariants of the original system, such as a hamiltonian structure can be preserved best in a stochastic parametrization.

We will make a *Markov assumption* for the noise process assuming that $\boldsymbol{\eta}$ is a multivariate *Ornstein–Uhlenbeck* process. This makes $\boldsymbol{x}$ non-Markovian, but we can obtain a continuous time Markov process for the augmented state vector $\boldsymbol{\phi} = (\boldsymbol{x},\boldsymbol{\eta})$. This new state vector obeys an *Ito* SDE

$$d\boldsymbol{\phi}(t) = \mathbf{g}(\boldsymbol{\phi})dt + \mathbf{A}(\boldsymbol{\phi})d\mathbf{W}(t) \tag{3}$$

where $\mathbf{g}$ is the nonlinear drift, $\mathbf{A}$ is the diffusion matrix and $\mathbf{W}(t)$ is a vector of independent *Wiener processes*.

Besides these *a priori* statistical assumptions about the state evolution we assume that we have access to a set of data $D = \{y_i\}_{i=1}^M$ which contains noisy information about the state, independently sampled at discrete times τ_i. One may think of noisy samples at certain spatial locations, or of spatially averaged observations. The statistics of the data, for given state variables, will be described by the *likelihood* $p(D|\boldsymbol{\phi}) = \prod_{i=1}^M p(y_i|\boldsymbol{\phi}(\tau_i))$, where $p(y|\boldsymbol{\phi})$ is the conditional density of the observation y given that the true state of the system is $\boldsymbol{\phi}$. Following a *Bayesian approach* to statistical inference [9], we can combine the *prior* statistical assumptions encoded in the Markov process (3) and the likelihood into a *posterior measure* of the random process $\boldsymbol{\phi}(t)$ to obtain a data dependent inference of the state of the system. Note, that the posterior is no longer Markovian but a *hidden Markov* process.

The observations can also be used to estimate the unknown (hyper) parameters which are contained in the essential characteristics of the noise process, like the process kernel K. Their estimation can be based on a maximisation of the *marginal data likelihood* $p(D) = E_{\boldsymbol{\phi}}[p(D|\boldsymbol{\phi})]$ where $E_{\boldsymbol{\phi}}$ denotes an expectation with respect to the *prior* Markov process distribution. The few existing *stochastic parameterisation* schemes are physically justified, but empirically based [10]. Given a prior model for the form of the stochastic forcing, we can estimate the (hyper) parameters in this stochastic scheme using the model and observations.

This probabilistic framework represents a principled *theoretical* approach for state and parameter inference of the model. However, its *practical* realization is highly nontrivial. The computation of any statistics of the prior Markov model (3) (expressed by the joint probabilities of state variables at different times) requires the *transition probability* density $p_{t'-t}(\boldsymbol{\phi}'|\boldsymbol{\phi})$ of the states during time intervals $t' - t$. The exact time evolution of this transition probability follows a multivariate Fokker–Planck equation [11] and the inclusion of frequent data observations makes an approximation by the *stationary distribution* impossible. For very high dimensional systems, numerical solutions of Fokker-Planck equations are extremely time consuming. Alternative Monte Carlo sampling methods for the direct simulation of the SDE (3) may also become impractical. Hence, we will resort to further approximations. Rather than working with the exact process $\boldsymbol{\phi}(t)$, we develop a simpler, tractable approximate process. The most sensible candidate is a Gaussian process.

One method [12] to achieve this is to derive a set of differential equations (ODEs) for the first moments and correlation functions of the state vector $\boldsymbol{\phi}$, together with a Gaussian approximation that ensures a closed system. The disadvantage of this procedure is that we get a set of ODEs for correlation *matrices* and we will have to do multiplications of very large matrices (which for typical operational environmental models will have dimension $O(10^7) \times O(10^7)$ and upwards) for every time step in a numerical integration of the ODEs. For realistic systems the dimension of the state space makes this approach impractical and thus in this project we propose more sensible *variational techniques* which give controlled approximations to the non-Gaussian process. These ideas have their origin in statistical physics (see e.g. [13]), but the technique has also been widely used in machine learning [2], where *Variational approximations* have been frequently applied to nonlinear stochastic processes in discrete time (see, for example the articles of Jaakkola and of Ghahramani and Beal in [2] and references therein). However, their application to continuous time domain problems with a vast number of variables is a major challenge and requires the development of novel machine learning and computational tools.

The advantages of variational methods are that we will be able to control the complexity of the approximation, allowing for *sparse* methods *and* we obtain an approximation on the marginal data likelihood which can be used for a Bayesian maximum likelihood estimation of (hyper) parameters. In these techniques, the dissimilarity between the probability measures dp and dq associated with the true and the approximating processes will be measured by the relative entropy or *Kullback–Leibler* (KL) divergence defined by

$$KL(q||p) = \int dq \ln \frac{dq}{dp} \ . \tag{4}$$

We will mainly work with the *variational approximation* in which (4) is minimised within a family of *tractable distributions*. Applying the variational approximation to the posterior process, a computationally tractable lower bound on the marginal data likelihood can be derived, from which (hyper) parameters in the stochastic forcing can be estimated. We will also consider a variant of the method which combines a variational approximation for the *prior process* with a recently developed technique [14] for sequentially updating a corresponding approximate posterior allowing for a speedy processing of data in the computations of forecasts and data likelihoods. In a static environmental modelling context [3], the sequential update method was found to give efficient and reliable state estimates even for highly non-Gaussian observation models.

To deal with the problem that we do not have simple explicit expressions for the probability measures of the processes involved in the computation of (4), we will apply a functional (or path–) integral (FI) approach [11, 13]. In this case, the variational method will be similar to Feynman's variational approximation used in the path integral approach to quantum physics [13]. The FI approach can be derived by discretising time, setting $t_{j+1} = t_j + \Delta t$, $\boldsymbol{\phi}_i \doteq \boldsymbol{\phi}(t_i)$ for $i = 1, \ldots, k$ and working with finite dimensional joint distributions of the process variables at different times, taking the limit $\Delta t \to 0$, $k \to \infty$ at the end. For the posterior process (the process conditioned on the data) we have

$$p(\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_k|D) = \frac{p(\boldsymbol{\phi}_1)}{p(D)} \prod_{j=1}^{k-1} p_{\Delta t}(\boldsymbol{\phi}_{j+1}|\boldsymbol{\phi}_j) \prod_{i=1}^{M} p(y_i|\boldsymbol{\phi}(\tau_i)) \ ,$$

where for simplicity the observation times, τ_i, are contained in the set $\{t_j\}_{j=1}^k$. Assuming a similar factorisation for the approximating process, using the fact that for diffusion processes, $p_{\Delta t}(\boldsymbol{\phi}'|\boldsymbol{\phi})$ becomes Gaussian as a function of $\boldsymbol{\phi}'$ in the limit $\Delta t \to 0$ [12], one may carefully take the limit $\Delta t \to 0$ to express integrals like (4) by expectations over the *Wiener measure* involving *Ito* stochastic integrals [15]. The FI framework may also be useful to relate

our probabilistic variational approach with the *4D VAR approach* to data assimilation. While the former aims at approximating the *mean* path of a system's state together with its *fluctuations*, the latter may be understood as computing the *most probable path* in the path integral representation of the process.

A full minimisation of (4) in the family of Gaussian processes will again lead to intractable computations. Hence, the development of space–time Gaussian processes with *sparse representations* will be of central importance for our project. For such processes, covariance kernels should be representable in terms of a set of parameters which is small enough to allow for efficient KL optimisations. Sparse kernel methods have been widely studied in machine learning (see, e.g. [14]), but have so far not been extended to the full complexity of spatio–temporal problems (for general kernel methods, see e.g. [14, 16]). The variational Gaussian approximation will not only give a nontrivial prediction for the (mean) state variables (which depends on the noise) but will also allow for an estimate of space time correlations which is of importance for environmental models, where for example, the joint distribution of rainfall over a catchment in space and time will be required for flood forecasting.

To test our approximate inference schemes we use a variety of models of nested complexity. We will principally work on a simplified but relevant dynamical model, described by the Kuramoto-Sivashinsky partial differential equation with periodic boundary conditions [17]. This equation is appropriate because it shares features with the Navier-Stokes equations used widely in real atmospheric forecasting, but is one dimensional and with a controllable complexity. We will also look at simpler models which have been used in meteorology and elsewhere, such as the Hénon-Heiles system and the 5-component Lorenz model [18], and a quasi-geostrophic model [19].

Verifying our approach requires a situation where F, the *system* is known. For this we will utilise the Kuramoto-Sivashinsky model run at very high resolutions. We can then simulate observations from this high resolution run, using realistic observation operators and noise. The approximate *model* G will be a reduced resolution version of the Kuramoto-Sivashinsky model, more representative of the types of models which are operationally used for weather forecasting. In this *perfect model* setting we can validate the principle, which would not be possible in the real world where the state can only observed with error.

Programme of work

Our project will create a novel framework for probabilistic modelling of complex dynamical systems. We will develop methods for the estimation of the probability distribution function of the state of a system over space and time, for non-equilibrium conditions, that are computationally feasible within the time scales required for *operational use*. We will exploit methods from machine learning and statistical physics and in particular the *variational approximations* as outlined above, to produce principled approximations to the evolution of the probability distribution function of the model state, and demonstrate how Bayesian (hyper) parameter estimation can be used to infer information about *stochastic parameterisations* within this framework. We will compare and contrast our methods with the most competitive existing methods – the ensemble Kalman filter and 4D VAR. We have divided the program into several work-packages, which are listed below.

WP1: Framework for approximate inference in continuous time stochastic models. Using functional integral methods and the variational approximation outlined in the *framework and methodology* section, we will develop a theoretical framework for the approximation of multivariate (hidden-) Markov processes by Gaussian processes. The models will be generic in character, and we will focus on a range of models, from a simple Hénon-Heiles system, to a quasi-geostrophic model, and the Kuramoto-Sivashinsky model.

WP2: Sparse approximations for space–time Gaussian processes. In this work package we will develop a theoretical and computational framework for constructing Gaussian processes with flexible sparsity from which tractable approximations to the stochastic dynamics of very large complex systems can be derived based on WP1. We plan to explore a range of methods for sparse representations based around projective ideas we have explored in earlier work [14], the use of space / time limited kernels and representations of the kernels in different bases [16]. We will also utilise the results of WP4 where the symmetries can be exploited to reduce the computational cost of our approach.

WP3: Choice of *a priori* parametric models for additive stochastic forcing. The stochastic model error term is assumed to have Gaussian statistics. The exact form of this process will be defined by the modelling context, through a kernel K. In this work package we will develop sensible parametric forms for the kernels which are appropriate to the specific model.

WP4: Hamiltonian methods in data assimilation. We shall apply techniques that have been developed in geometric integration (such as variational integrators, and symmetry methods for difference equations) to data assimilation treated as a control problem. Hamiltonian methods provide a natural and structured approach to dealing with many problems in data assimilation: the choice of variables for the assimilation (e.g. the use of potential vorticity), the incorporation of balance constraints, and the definition of background error covariances, are important issues that can be studied using the hamiltonian framework. We shall build on the work of [20], and [21], and address related issues such as the choice of boundary conditions for the assimilation.

WP5: Properties of *a priori* parametric models for general stochastic forcing We will investigate the possibility of using state dependent (multiplicative) noise processes based on functional forms derived from the literature [8] and developed from the model equations themselves. We will use ideas for treating model errors developed within the 4D VAR framework. The starting point will be a hamiltonian treatment of the control problem (WP4), because, additional forcing terms and/or constraints must be included in a hamiltonian system in such a way that invariants are preserved. In the control problem, such invariants will be associated with the Euler-Lagrange equations obtained by minimising the cost function, and not those associated with the forward model. Preserving such invariants may have important consequences for issues such as the convergence of the assimilation, in both 4D VAR and our variational approximations. This will also link to the work undertaken by the DARC funded PhD student at Surrey.

WP6: Data assimilation in stochastic dynamical models. We will extend the methods we developed in WP1 to focus on efficient methods for incorporating observations into the estimation of the space time probability distribution function of the state. This will include exploring whether sequential methods, as we have developed for the static time case [3], will be appropriate in the dynamical context. We will also address the issue of how stochastic models can be best handled in a 4D VAR assimilation system and relate this to the variational approximations we are developing. This WP is central to the project and includes all participants.

WP7: Parameter inference in complex dynamical models. The focus of WP1 and WP6 is on the inference of the state of the stochastic dynamical system for *known statistics* of the stochastic driving. These statistics will not be known perfectly for more realistic environmental models, although prior models will be specified in WP3 and WP5. Thus certain parameters (for example the variance and length scales of the kernel defining the forcing stochastic process) of the model can be regarded as *(hyper) parameters* in a Bayesian statistical setting. The data will be used carefully to infer these model parameters in a sequential Bayesian (maximum likelihood) framework.

WP8: Assessing the error of the variational approximation schemes. We will implement a particle filter [22] to provide Monte Carlo estimates of the true probability density of the model state which (for not too large models) are accurate enough to allow for an assessment of the quality of the variational approximation scheme discussed in WP1 and WP6. We will also compare our methods with other leading edge methods for propagating uncertainty in the state; ensemble based systems [23] and parametric approaches based around the extended Kalman filter [24].

WP9: Assessing our methods for parameter inference. Running our *deterministic* Kuramoto-Sivashinsky PDE model at high spatial resolution give us access to the *"true"* state of the system. From this solution, we will generate noisy observations for controlled numerical experiments with our approximate probabilistic inference schemes, where we use a lower spatial resolution Kuramoto-Sivashinsky model. This setting will allow us to verify the ability of our methods to produce sensible estimates of *(hyper) parameters* in the stochastic parameterisations based on observations and the prior model. We will explore the degree to which the learnt model can follow the *"true"* state and the sensitivity of this to the stochastic forcing. By varying the number of observations we will explore how the methods work in data rich and data poor cases.

WP10: Contrasting our methods with existing data assimilation techniques. Using the particle filter, from WP8 as a reference solution, we will compare the merits of our approximate approaches (both the variational approximations outlined in WP1 and WP6, and the 4D VAR with model error from WP4 and WP5) with current state of the art data assimilation methods such as the ensemble Kalman filter and traditional 4D VAR without a model error term. We will perform a range of experiments on a number of different models, with different observation scenarios, to provide an objective comparison.

The work-packages are designed such that the initial theoretical developments are carried out at Southampton and Aston. Work at Surrey will concentrate on Hamiltonian formulations of 4D VAR in the presence of model error and relate this to the variational approximations – which will also inform the development of the variational approximations and the stochastic forcing. The actual implementation and comparison with existing methods is largely carried out at Aston and Surrey. However, there will be significant input from all partners, include the Met Office, to all work-packages ensuring the relevance to the envisaged applications.

Potential Applications

The need for probabilistic models is becoming increasingly recognised in the academic and research community across environmental science, but has yet to make a strong showing in the more operational setting of commercial weather forecasting. Our work will enable a more principled and accurate approach to probabilistic forecasting to be considered. Operational implementation will require further development, but the proof of concept for the numerical methods we will develop in this project is a necessary first step. The generation of improved probabilistic forecasts in meteorology, climatology and oceanography benefits a large range of interests, from individuals, to companies, governments and society as a whole as outlined in a recent review of the benefit of an ensemble approach

to probabilistic prediction in the weather forecasting context [25]. Our methods will offer a (possibly complementary) alternative to ensemble prediction and model output statistics methods which are currently the only approaches for estimating a probability forecast. The strong links with the Met Office will help to disseminate the work to the relevant beneficiaries. Our variational treatment of continuous time processes also has relevance to a wide range of application domains, such as control engineering, chemical and biological sciences and finance.

Summary

In the VISDEM project we will develop a range of approaches to inference in stocahstic dynamic models. In doing so we will link tradiational variational appraoches used with the meteorological community to our fully probabilistic methods through the links to Hamiltonian dynamics. We will also work heavily on the issue of model error and its parameterisation and inference from data. The project will include the testing of the new methods we develop on a number of environmentally relevant, simplified models, and will compare our results to a range of state of the art solutions to assess their potential for further exploitation in real systems.

References

[1] T N Palmer. Representation of model uncertainty in Ensemble Prediction Systems. presentation to ECMWF Ninth Workshop on Meteorological Operational Systems, November 2003. http://www.ecmwf.int/newsevents/meetings/workshops/9th_Meteorological_Operational_Systems/, 2003.

[2] Manfred Opper and David Saad, editors. *Advanced Mean Field Methods: Theory and Practice*. The MIT Press, 2001.

[3] D Cornford, L Csató, D J Evans, and M Opper. Bayesian analysis of the scatterometer wind retrieval inverse problem: some new approaches. *Journal of the Royal Statistical Society B*, 66:1–17, 2004.

[4] EMS. Final report on workshop on mathematical techniques for improving forecasting of hazardous weather. Technical report, EMS programme, 2003. http://www.ma.hw.ac.uk/icms/ems/weather/.

[5] G. Burgers, P. J. van Leeuwen, and G. Evensen. On the analysis scheme in the ensemble kalman filter. *Monthly Weather Review*, 126:1719–1724, 1998.

[6] P Courtier, J N Thepaut, and A Hollingsworth. A strategy for operational implementation of 4D-VAR, using and incremental approach. *Quarterly Journal of the Royal Meteorological Society*, 120:1367–1387, 1994.

[7] J W-B Lin and J D Neelin. Considerations for stochastic convection parameterization. *Journal of the Atmospheric Sciences*, 59:959–975, 2002.

[8] N K Nichols. Treating model error in 3D and 4D data assimilation. To appear in the Proceedings of the NATO Advanced Study Institute on Data Assimilation for the Earth System, Maratea, Italy, June 2002. http://darc.nerc.ac.uk/, 2002.

[9] J M Bernardo and A F M Smith. *Bayesian Theory*. Wiley Series in Probability and Statistics, London, 2000.

[10] R Buizza, M Miller, and T N Palmer. Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. *Quarterly Journal of the Royal Meteorological Society*, 125:2887–2908, 1999.

[11] H Risken. *The Fokker Planck Equation*. Springer Verlag, Berlin, 1989.

[12] J Honerkamp. *Stochastic Dynamical Systems: Concepts, Numerical Methods, Data Analysis*. Wiley, New York, 1993.

[13] H. Kleinert. *Path Integrals in Quantum Mechanics, Statistics and Polymer Physics*. World Scientific, Singapore, 1990.

[14] L Csato and M Opper. Sparse gaussian processes. *Neural Computation*, 14:641–668, 2002.

[15] H. Roepstorff. *Path Integral Approach to Quantum Physics*. Springer Verlag, Berlin, 1994.

[16] N Cristianini and J Shawe-Taylor. *An introduction to support vector machines (and other kernel-based learning methods)*. Cambridge University Press, Cambridge, 2000.

[17] P Holmes, J L Lumley, and G Berkooz. *Turbulence, Coherent Structures, Dynamical Systems and Symmetry*. Cambridge University Press, Cambridge, 2001.

[18] A Hannachi and K Haines. Convergence of data assimilation by periodic updating in simple hamiltonian and dissipative systems. *Tellus*, 50A:58–75, 1998.

[19] T M Hamill, C Snyder, and R E Morss. A comparison of probabilistic forecasts from bred, signular vector and perturbed observation ensembles. *Monthly Weather Review*, 128:1835–1851, 2000.

[20] M J P Cullen. Four-dimensional variational data assimilation: a new formulation of the background error covariance matrix using a potential-vorticity representation. *Quarterly Journal of the Royal Meteorological Society*, 129:2777–2796, 2003.

[21] L R Stanton, A S Lawless, N K Nichols, and I Roulstone. Variational data assimilation for hamiltonian problems. *International Journal for Numerical Methods in Fluids*, 2004. submitted.

[22] A Doucet, N deFreitas, and N. (Eds) Gordon. *Sequential Monte Carlo Methods in Practice*. Springer-Verlag, London, 2001.

[23] F Molteni, R Buizza, T N Palmer, and T Petroliagis. The new ecmwf ensemble prediction system: methodology and validation. *Quarterly Journal of the Royal Meteorological Society*, 122:73–119, 1996.

[24] D T Pham, J Verron, and M C Roubaud. Singular evolutive extended kalman filter with eof initialization for data assimilation in oceanography. *Journal of Marine Systems*, 16:323–340, 1998.

[25] T N Palmer. The economic value of ensemble forecasts as a tool for risk assessment: From days to decades. *Quarterly Journal of the Royal Meteorological Society*, 128:747–774, 2002.