CONTROL ID: 1800127

TITLE: Efficient Mean Field Variational Algorithm for Data Assimilation *(Invited)*

AUTHORS (FIRST NAME, LAST NAME): Michail D. Vrettas[1], Dan Cornford[2], Manfred Opper[3]

INSTITUTIONS (ALL): 1. Earth and Planetary Science, University of California, Berkeley, Berkeley, CA, United States.
2. Computer Science, Aston University, Birmingham, West Midlands, United Kingdom.
3. Artificial Intelligence, TU Berlin, Berlin, Germany.

ABSTRACT BODY: Data assimilation algorithms combine available observations of physical systems with the assumed model dynamics in a systematic manner, to produce better estimates of initial conditions for prediction. Broadly they can be categorized in three main approaches: (a) sequential algorithms, (b) sampling methods and (c) variational algorithms which transform the density estimation problem to an optimization problem. However, given finite computational resources, only a handful of ensemble Kalman filters and 4DVar algorithms have been applied operationally to very high dimensional geophysical applications, such as weather forecasting. In this paper we present a recent extension to our variational Bayesian algorithm which seeks the 'optimal' posterior distribution over the continuous time states, within a family of non-stationary Gaussian processes.

Our initial work on variational Bayesian approaches to data assimilation, unlike the well-known 4DVar method which seeks only the most probable solution, computes the best time varying Gaussian process approximation to the posterior smoothing distribution for dynamical systems that can be represented by stochastic differential equations. This approach was based on minimising the Kullback-Leibler divergence, over paths, between the true posterior and our Gaussian process approximation. Whilst the observations were informative enough to keep the posterior smoothing density close to Gaussian the algorithm proved very effective on low dimensional systems (e.g. O(10)D). However for higher dimensional systems, the high computational demands make the algorithm prohibitively expensive.

To overcome the difficulties presented in the original framework and make our approach more efficient in higher dimensional systems we have been developing a new mean field version of the algorithm which treats the state variables at any given time as being independent in the posterior approximation, while still accounting for their relationships in the mean solution arising from the original system dynamics.

Here we present this new mean field approach, illustrating its performance on a range of benchmark data assimilation problems whose dimensionality varies from O(10) to O(10^3)D. We emphasise that the variational Bayesian approach we adopt, unlike other variational approaches, provides a natural bound on the marginal likelihood of the observations given the model parameters which also allows for inference of (hyper-) parameters such as observational errors, parameters in the dynamical model and model error representation. We also stress that since our approach is intrinsically parallel it can be implemented very efficiently to address very long data assimilation time windows. Moreover, like most traditional variational approaches our Bayesian variational method has the benefit of being posed as an optimisation problem therefore its complexity can be tuned to the available

computational resources. We finish with a sketch of possible future directions.

KEYWORDS: 1873 HYDROLOGY Uncertainty assessment.

(No Image Selected)

(No Table Selected)

Additional Details

Previously Presented Material:

Contact Details

CONTACT (NAME ONLY): Michail Vrettas

CONTACT (E-MAIL ONLY): m.vrettas@berkeley.edu

TITLE OF TEAM: