ISDA-Online

May 08, 2024, 15 – 17 UTC

“Non-linear Data Assimilation”

Organizers and Conveners: *Nachiketa Chakraborty (University of Reading, UK)*
Hristo Chipilski (Florida State University, US), Javier Amezcua (U. Iberoamericana, Mexico; U. of Reading, UK)

Traditional data assimilation methods are optimal under the assumption of linear relationships between the observations and the state variables (observation operator), linear forecast models, as well as Gaussian sources of error. Linearizations are performed when these requirements are not met. However, with increasing resolution and complexity in the models and observations, these linearizations can fail, and one may need to use nonlinear methods. We invite contributions on the areas of particle filters, methods derived from measure transport theory, and methods which leverage tools coming from the AI/ML community.

Program: (UTC)

15:00 – 15:05	**Welcome**
15:05 – 15:25 (17’+3’)	**A localised particle filter for geophysical data assimilation** Eliana Fausti, Dan Crisan
15:25 – 15:45 (17’+3’)	**A localized sequential Markov Chain Monte Carlo algorithm** Hamza Ruzayqat, Omar Knio
15:45 – 16:05 (17’+3’)	**A Non-Linear, Two-Step, Ensemble Data Assimilation Method for Sea Ice** Kate Boden, Ian Grooms
16:05 – 16:25 (17’+3’)	**Adaptive and scalable triangular measure transport filters** Max Ramgraber, Berent Lunde
16:25 – 16:45 (17’+3’)	**A Closed-Form Nonlinear Data Assimilation Algorithm for Multi-Layer Flow Fields** Zhongrui Wang, Di Qi
16:45 – 16:55	**Closing: Information on upcoming sessions**

Please note:

- When you login to the session before 15:00 UTC, and everything could be quiet, this is most likely because we muted the microphones.
- The times in UTC are approximate. In case of technical problems, we might have to change the order of the presentations.
- **Time Zones**: 15 – 17 UTC

Europe:	04 – 06 pm BST (London)	05 – 07 pm CEST (Berlin)	
Asia/Australia:	11 – 01 am CST (Shanghai)	00 – 02 am JST (Tokyo)	02 – 04 am AEDT (Sydney)
Americas:	08 – 10 am PDT (San Fran.)	09 – 11 am MDT (Denver)	11 – 01 pm EDT (New York)

A localised particle filter for geophysical data assimilation

Eliana Fausti[1], Dan Crisan[1]
[1]Imperial College London

Particle filters, compared to other existing data-assimilation (DA) methodologies, are well-suited to deal with non-linear, non-Gaussian models. Their use in geophysical DA for the purpose of weather forecast or ocean prediction, however, is not yet widespread due to the so-called “curse of dimensionality”. In this talk we propose a new strategy for assimilation of high-resolution geophysical fluids data, combining particle filters and ocean models with stochastic transport. The particle trajectories are modelled by stochastic partial differential equations (SPDE), simulated at a coarser resolution than the data. The stochasticity is introduced in the models in a physical way, to capture subgrid-scale processes. To overcome the problem of weight degeneracy in the particle filter, we introduce a localization method using the Gaspari–Cohn localization function, which tapers off the importance of the observations the further away they are from each region of interest. We run a smaller particle filter separately in each region, which, in addition to yielding a parallelizable implementation, also reduces the number of resampling, tempering and jittering steps required to avoid degeneracy of the ensemble. We present some initial results by testing our particle filter in a “twin experiment” for the 2D (stochastic) rotating shallow water equations, where the signal is taken to be an SPDE path, run on the same grid as the particle ensemble. In later experiments we will first generate high-resolution data synthetically using fine grid PDE runs, and, if successful, we will then test our methodology on SWOT satellite data.

A localized sequential Markov Chain Monte Carlo method

Hamza Ruzayqat[1], Omar Knio[1]
[1] King Abdullah University of Science and Technology

I present a local data assimilation technique that uses sequential Markov Chain Monte Carlo (SMCMC) for filtering high-dimensional non-linear, and potentially, non-Gaussian state-space models. Unlike particle filters, which can be considered exact methods and can be used for filtering non-linear, non- Gaussian models, SMCMC does not assign weights to the samples/particles, and therefore, the method does not suffer from the issue of weight-degeneracy when a relatively small number of samples is used. We design a localization approach within the SMCMC framework that focuses on regions where observations are located and restricts the transition densities included in the filtering distribution of the state to these regions. This results in reducing the effective degrees of freedom and thus improving the efficiency. We test the new technique on high-dimensional (10^4 − 10^5) linear Gaussian model and non-linear shallow water models with Gaussian noise with real and synthetic observations. For two of the numerical examples, the observations mimic the data generated by the Surface Water and Ocean Topography (SWOT) mission led by NASA, which is a swath of ocean height observations that changes location at every assimilation time step. We also use a set of ocean drifters' real observations in which the drifters are moving according the ocean kinematics and assumed to have uncertain locations at the time of assimilation. We show that when higher accuracy is required, the proposed algorithm is superior in terms of efficiency and accuracy over competing ensemble methods.

A Non-Linear, Two-Step, Ensemble Data Assimilation Method for Sea Ice

Kate Boden[1], Ian Grooms[1]
[1]University of Colorado, Boulder

The Ensemble Kalman Filter (EnKF) is a powerful tool in the geosciences to integrate real-time observations into dynamical models for an improved estimate of the state. The Gaussian assumptions underlying the EnKF can result in reduced improvements when there are non-Gaussian relationships. Modeled sea ice distributions are highly non-Gaussian due to constraints on the state variables; the CICE/Icepack model divides sea ice within a grid cell into thickness categories and then forecasts the fractional area coverage for each category. The total fractional area, including open water, must sum to 1, requiring the forecasted distribution to live on the simplex, a highly non-Gaussian constraint. There are ways to assimilate sea ice observations using classic EnKF approaches but it requires a post-processing step and results may have increased error due to the Gaussianity assumptions. This work aims to improve the assimilation of sea ice observations into models by building on a non-linear, two-step framework first put forward by Jeff Anderson in 2002.

Given satellite data that provides N total measurements with k of them being open water, the two-step framework works by first applying a scalar Quantile Conserving Ensemble Filter (QCEF) to update the open water fraction. The second step uses a transport-based approach to update the remaining variables in the state vector, i.e the fractional area of ice in the other categories. This second step assumes the forecasted distribution (prior) is a mixed Dirichlet and uses the updated open water fraction from step 1 to transport ensemble members from the prior to the posterior, which is also Dirichlet. We will present details on the method and results from OSSE experiments that compare the EnKF to our non-linear two-step method.

Though this works uses sea ice to develop and test a new non-linear, two-step method, we present it here for potential use in wider applications.

Adaptive and scalable triangular measure transport filters

Max Ramgraber[1], Berent Lunde[2]
[1]TU Delft
[2]Equinor

Ensemble transport filters (EnTFs) leverage triangular measure transport to create true nonlinear generalizations of linear-Gaussian ensemble data assimilation (DA) methods like the EnKF or EnKS. These methods retain the sample-efficient transformative updates of their linear-Gaussian cousins, and can yield substantial improvements in DA performance for nonlinear and non-Gaussian systems. A key consideration in this process is the optimal degree of nonlinearity for the EnTF's constituent map component functions. Insufficient complexity does not capture all statistical features, whereas excessive complexity risks overfitting. In this work, we propose an EnTF parameterization that leverages P-Splines (penalized basis splines). By introducing a penalty term for the splines' smoothness, we can adapt the EnTF's effective degrees of freedom on a continuous scale. Minimizing the corrected Akaike Information Criterion (cAIC) subsequently identifies a parsimonious (as simple as possible, as complex as necessary) parameterization for the EnTF. The result is an adaptive, scalable, nonlinear DA algorithm that operates without manual hyperparameter tuning. We demonstrate the algorithm's performance for nonlinear and chaotic systems, and outline its applications to high-dimensional, discretized models.

A Closed-Form Nonlinear Data Assimilation Algorithm for Multi-Layer Flow Fields

Zhongrui Wang[1], Di Qi[2]

[1]University of Wisoconsin - Madison
[2]Purdue University

State estimation in multi-layer turbulent flow fields with only a single layer of partial observation remains a challenging yet practically important task. Applications include inferring the state of the deep ocean by exploiting surface observations. Directly implementing an ensemble Kalman filter based on the full forecast model is usually expensive. One widely used method in practice projects the information of the observed layer to other layers via linear regression. However, when nonlinearity in the highly turbulent flow field becomes dominant, the regression solution will suffer from large uncertainty errors. In this paper, we develop a multi-step nonlinear data assimilation method. A sequence of nonlinear assimilation steps is applied from layer to layer recurrently. Fundamentally different from the traditional linear regression approaches, a conditional Gaussian nonlinear system is adopted as the approximate forecast model to characterize the nonlinear dependence between adjacent layers. The estimated posterior is a Gaussian mixture, which can be highly non-Gaussian. Therefore, the multi-step nonlinear data assimilation method can capture strongly turbulent features, especially intermittency and extreme events, and better quantify the inherent uncertainty. Another notable advantage of the multi-step data assimilation method is that the posterior distribution can be solved using closed-form formulae under the conditional Gaussian framework. Applications to the two-layer quasi-geostrophic system with Lagrangian data assimilation show that the multi-step method outperforms the one-step method with linear stochastic flow models, especially as the tracer number and ensemble size increase.