ISDA-Online

March 08, 2024, 15 – 17 UTC

"Advancements in Ensemble Data Assimilation"

Organizers / Conveners*: **Hristo Chipilski** (Florida State University, USA), **Laura Slivinski** (NOAA/ESRL, USA), and **Javier Amezcua** (Tec. de Monterrey, Mexico; U. Reading, UK)*

Ensemble forecasting has become an essential tool for prediction in many fields. This practice readily equips predictions with a measure of uncertainty, and with statistical relationships amongst variables which can be used by data assimilation methods. However, statistics coming from raw 'small' ensembles often contain sampling errors, leading to the need for extra steps before using them. The most traditional method in this area is the ensemble Kalman filter, but recent advances go far beyond its original formulation. We encourage submissions on the following areas: challenging problems in localisation and inflation (e.g. time-dependent localisation), extensions beyond the Kalman filter (e.g. filters that use alternatives to linear regression), fully non-linear ensemble methods, multi-scale ensemble data assimilation, data-driven ensemble methods, etc.

Program: (UTC)

15:00 – 15:05	**Welcome**
15:05 – 15:25 (17'+3')	**Filtering Dynamical Systems Using Observations of Statistics** Eviatar Bach, Tim Colonius, Isabel Scherl, Andrew Stuart
15:25 – 15:45 (17'+3')	**Unbiased fully nonlinear data assimilation: the Stochastic Particle Flow Filter** Hao-Lun Yeh, Peter Jan van Leeuwen
15:45 – 16:05 (17'+3')	**DA for multi-time and multi-scale models of the turbulent energy cascade** Vikrant Gupta, Minping Wan, Francesco Fossella, Luca Biferale, Massimo Cencini, Alberto Carrassi, Chunxue Yang
16:05 – 16:25 (17'+3')	**Regularization of the ensemble Kalman filter using a non-stationary, non-parametric spatial model** Michael Tsyrulnikov, Arseniy Sotskiy
16:25 – 16:45 (17'+3')	**Nonlinear Data Assimilation in Chaotic Systems Using Deep Reinforcement Learning** Mohamad Abed El Rahman Hammoud, Naila Raboudi, Edriss S. Titi, Omar Knio, Ibrahim Hoteit
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)

Filtering Dynamical Systems Using Observations of Statistics

Eviatar Bach[1], Tim Colonius[1], Isabel Scherl[1], Andrew Stuart[1]
[1]California Institute of Technology

We consider the problem of filtering dynamical systems, possibly stochastic, using observations of statistics. Thus the computational task is to estimate a time-evolving density $\rho(v,t)$ given noisy observations of the true density $\rho^{\dagger}$; this contrasts with the standard filtering problem based on observations of the state v. The task is naturally formulated as an infinite-dimensional filtering problem in the space of densities ρ. However, for the purposes of tractability, we seek algorithms in state space; specifically we introduce a mean field state space model and, using interacting particle system approximations to this model, we propose an ensemble method. We refer to the resulting methodology as the ensemble Fokker–Planck filter (EnFPF).

Under certain restrictive assumptions we show that the EnFPF approximates the Kalman–Bucy filter for the Fokker–Planck equation, which is the exact solution of the infinite-dimensional filtering problem; our numerical experiments show that the methodology is useful beyond this restrictive setting. Specifically the experiments show that the EnFPF is able to correct ensemble statistics, to accelerate convergence to the invariant density for autonomous systems, and to accelerate convergence to time-dependent invariant densities for non-autonomous systems. We discuss possible applications of the EnFPF to climate ensembles and to turbulence modelling.

Unbiased fully nonlinear data assimilation: the Stochastic Particle Flow Filter

Hao-Lun Yeh[1], Peter Jan van Leeuwen[1]
[1]Department of Atmospheric Science, Colorado State University, Fort Collins, Colorado, USA

The increasing impact of nonlinearities in geoscience numerical models and observation operators necessitates advanced data assimilation methods. The recently developed deterministic version of the Particle Flow Filter (PFF), which is a fully nonlinear and efficient sequential Monte Carlo filter, addresses weight degeneracy yet struggles with biased ensemble spread, particularly in the observed part of the state space in a small number of particles. This issue can be partially mitigated by employing a matrix-valued kernel in the algorithm, but the fundamental challenge persists. To overcome this, we introduce the Stochastic Particle Flow Filter (SPFF), incorporating Gaussian noise in Stein Variational Gradient Descent dynamics. This additional repulsive force between particles ensures an unbiased posterior pdf, even with a limited number of particles.

Through experiments on the 1000-dimensional Lorenz-96 model, the SPFF proves effective in avoiding particle collapse and accurately captures the evolutions of particles. Additionally, the SPFF exhibits faster convergence than the deterministic PFF and thus improves analysis accuracy compared to the PFF with a matrix-valued kernel at the computational cost. The method also shows promise in high-dimensional ocean models, signaling progress in nonlinear data assimilation challenges.

DA for multi-time and multi-scale models of the turbulent energy cascade

Vikrant Gupta[1], Minping Wan[1], Francesco Fossella[2], Luca Biferale[2],
Massimo Cencini[3], Alberto Carrassi[4]

[1]Southern University of Science and Technology, China
[2]University of Rome, Tor Vergata, Italy
[3]Consiglio Nazionale delle Ricerche, Rome, Italy
[4]University of Bologna, Italy

Shell models are low-dimensional dynamical systems that aim to mimic the dynamics of three-dimensional high-Reynolds number turbulence. The system state comprises several shells, each representing a wavenumber, modelling the dynamics from large to small scales. In particular, shell models capture the nonlinear energy transfer across a wide range of scales and non-Gaussianity of small scales. These characteristics make estimation of shell models challenging because data assimilation frequently employ linearisation and are only optimal for systems that are statistically Gaussian. Shell models, therefore, represent an ideal problem to test the applicability of data assimilation in real-world turbulence, such as geophysical flows. We employ ensemble data assimilation (DA) methods to estimate the system state when only a few shells in the intermediate wavenumber range are measured. This means that we aim to estimate the large (mostly Gaussian) as well as the small (intermittent) scale dynamics from the information of intermediate inertial range dynamics. We have three main conclusions. First, we find that DA is only effective when shells spanning a continuous range of scales are observed. This requirement stems from the need to capture the nonlinear interactions for the information to propagate across neighbouring shells. Second, we need regularization to stabilise the ensemble DA methods. This regularization can be interpreted as localisation that essentially excludes the intermittent small scales in the dissipation range from the DA problem. Third, ensemble DA is particularly efficient at estimating the dynamics at large scales from the measurement at smaller scales; given that the measurements are available over a long time period spanning several eddy turnover time of the large scales. These results can guide and thus pave the way for application of DA to turbulent flows.

Regularization of the ensemble Kalman filter using a non-stationary, non-parametric spatial model

Michael Tsyrulnikov[1], Arseniy Sotskiy[1]

[1]Hydrometcenter of Russia

The sample covariance matrix of a random vector is known to be a good estimate of the true covariance matrix if the sample size is much larger than the length of the random vector. In high-dimensional problems, this condition is never met. As a result, the EnKF ensemble does not contain enough information to specify the prior covariance matrix accurately. This necessitates the need for regularization of the analysis algorithm.

Existing regularization approaches like covariance localization are largely ad hoc. We propose a regularization technique based on a new non-stationary, non-parametric spatial model on the sphere. The model is a constrained version of the general Gaussian process convolution model. Constraints on the location-dependent convolution kernel include local isotropy, positive definiteness as a function of distance, and smoothness as a function of location. The model allows for a rigorous definition of the local spectrum, which, additionally, is required to be a smooth function of spatial wavenumber. We regularize the EnKF by postulating that its prior covariances obey this model.

The model is estimated online in a two-stage procedure. First, ensemble perturbations are bandpass filtered in several wavenumber bands to extract aggregated local spatial spectra. Second, a neural network recovers the local spectra from sample variances of the filtered fields. We show that with the growing ensemble size, the estimator is capable of extracting increasingly detailed spatially non-stationary structures.

In simulation experiments, the new technique led to substantially better EnKF performance than covariance localization and mixing of localized sample covariances with climatological ones. Application of the proposed technique to practical high-dimensional problems will be based on a multi-resolution approach, work on which is underway.

Nonlinear Data Assimilation in Chaotic Systems Using Deep Reinforcement Learning

Mohamad Abed El Rahman Hammoud[1], Naila Raboudi[1], Edriss S. Titi[2], Omar Knio[1], Ibrahim Hoteit[1]

[1]KAUST, Saudi Arabia
[2]University of Cambridge, United Kingdom

Data assimilation (DA) plays a pivotal role in diverse applications, ranging from climate predictions and weather forecasts to trajectory planning for autonomous vehicles. A prime example is the widely used ensemble Kalman filter (EnKF), which relies on linear updates to minimize variance among the ensemble of forecast states. Recent advancements have seen the emergence of deep learning approaches in this domain, primarily within a supervised learning framework. However, the adaptability of such models to untrained scenarios remains a challenge. In this study, we introduce a novel DA strategy that utilizes reinforcement learning (RL) to apply state corrections using full or partial observations of the state variables. Our investigation focuses on demonstrating this approach to the chaotic Lorenz '63 system, where the agent's objective is to minimize the root-mean-squared error between the observations and corresponding forecast states. Consequently, the agent develops a correction strategy, enhancing model forecasts based on available system state observations. Our strategy employs a stochastic action policy, enabling a Monte Carlo-based DA framework that relies on randomly sampling the policy to generate an ensemble of assimilated realizations. Results demonstrate that the developed RL algorithm performs favorably when compared to the EnKF. Additionally, we illustrate the agent's capability to assimilate non-Gaussian data, addressing a significant limitation of the EnKF.