ISDA-Online

January 20, 2023, 07 – 09 UTC

“Non-Gaussian Data Assimilation”

Organizers: James Taylor (RIKEN, Japan)
Steven Fletcher (CIRA/CSU, US)

Program:

07:00 – 07:10 UTC	**Welcome**
07:10 – 07:30 (17’+3’)	**A Quantile Conserving Ensemble Filtering Framework: Regressing Probit-Transformed Quantile Increments to Update Unobserved Variables** Jeff ANDERSON
07:30 – 07:50 (17’+3’)	**Revisiting and Repurposing the Gaussian Anamorphosis EnKF** Hristo G. CHIPILSKI, Ian Grooms, Mohamad El Gharamti, Jeffrey Anderson, Ricardo Baptista
07:50 – 08:10 (17’+3’)	**A two-step nonlinear non-Gaussian framework for data assimilation applied to assimilation of wind direction observations** Ian GROOMS
08:10 – 08:30 (17’+3’)	**Improving vortex position accuracy with a new multiscale alignment ensemble filter** Yue (Michael) YING, Jeffrey Anderson, Laurent Bertino
08:30 – 08:50 (17’+3’)	**"Ensemblized" linear least squares (LLS)** Patrick N. RAANES
08:50 – 09:00	**Closing: Information on upcoming sessions**

<u>Please note:</u>

- The times in UTC are approximate. In case of technical problems, we might have to change the order of the presentations.
- **Time Zones**:
  Europe:
  07 – 09 am GMT (London) | 08 – 10 am CET (Berlin)
  Asia/Australia:
  03 – 05 pm CST (Shanghai) | 04 – 06 pm JST (Tokyo) | 06 – 08 pm AEDT (Sydney)
  Americas:
  11pm – 01 am PST (San Fran.) | 00 – 02 am MST (Denver) | 02 – 04 am EST (New York)

A Quantile Conserving Ensemble Filtering Framework: Regressing Probit-Transformed Quantile Increments to Update Unobserved Variables

Jeff ANDERSON
NCAR Data Assimilation Research Section

Ensemble Kalman filters are commonly used for data assimilation in geoscience applications. A novel efficient algorithm that allows the use of arbitrary continuous priors and likelihoods for an observed variable was presented at ISDA in June 2022. The key innovation was to select posterior ensemble members with the same quantiles with respect to the continuous posterior distribution as the prior ensemble had with respect to the prior continuous distribution. While this method allowed the use of arbitrary non-Gaussian distributions and led to significant improvements in analysis estimates for observed variables, those improvements can be lost when using standard linear regression of observation increments to update other state variables. However, doing the regression of observation quantile increments in a probit-transformed bivariate quantile space guarantees that the posterior ensembles for state variables also have all the advantages of the observation space quantile conserving posteriors. For example, if state variables are bounded then posterior ensembles will respect those bounds and eliminate most bias near the boundary. The posterior ensembles also respect other aspects of the continuous prior distributions. Examples are shown for a variety of bivariate prior ensembles including bounded quantities and multimodal distributions. A new release of the Data Assimilation Research Testbed supports many choices for prior distributions including normal, gamma, inverse gamma, beta, uniform, and bounded rank histogram. A comparison of results using these different non-Gaussian distributions in a tracer transport model will be presented; the non-Gaussian distributions can lead to vastly improved estimates of both tracer concentration and the associated flow field.

Revisiting and Repurposing the Gaussian Anamorphosis EnKF

Hristo G. CHIPILSKI
NCAR

Originally proposed by Bertino et al. (2003), the Gaussian anamorphosis ensemble Kalman filter (GA-EnKF) has been successfully applied to a variety of nonlinear estimation problems in oceanography, biogeochemistry, history matching and atmospheric modeling. The computational efficiency of this approach and its ability to respect the physical bounds of model variables have turned it into a scalable alternative to non-parametric, particle-based data assimilation (DA) techniques. However, past studies have also identified pathological cases where GA-EnKF produces larger analysis errors than the benchmark EnKF.

Given the continuing lack of clarity on the performance and use of GA-EnKFs, the first part of our presentation will analyze the Bayesian consistency of these algorithms under several different non-Gaussian regimes. In particular, we will prove that GA-EnKFs recover the true posterior distribution in the special case when joint Gaussianity results from separate transformations of the state and observation variables. Then, anamorphosis theory will be used to establish a formal connection between measure transport and two-step ensemble filtering. Specifically, we will show that this analogy leads to 4 new formulations of the observation-space update within two-step filters, generalizing the rank histogram and more recently introduced quantile-conserving ensemble DA methods. Although the new analysis schemes are theoretically equivalent, we will discuss why their finite-sample performance depends on the precise setup of the DA experiment. Some of the new updates will be also compared against existing formulations via idealized numerical examples.

A two-step nonlinear non-Gaussian framework for data assimilation applied to assimilation of wind direction observations

Ian GROOMS
University of Colorado

The DART software suite is based on a two-step filtering framework where the first step is a Bayesian ensemble update, and the second step is linear regression. The connection between this two-step framework and the Bayesian formulation of the filtering problem has heretofore been murky, except in the Gaussian case, where it is known that the two-step method is equivalent to a one-step EnKF. This presentation shows how to express the full Bayesian problem in a two-step framework without requiring a Gaussian assumption. This new insight enables the development of Bayesian methods that are algorithmically similar to EnKF, but that are appropriate for a very wide class of non-Gaussian distributions.

Wind direction observations are an important data set for historical reanalyses, with accurate instrumental records available over multiple centuries. Wind direction on a 16, 32, or 64-point compass is a strongly nonlinear, piecewise-constant function of a model's state variables though, and standard EnKFs and variational methods struggle to use it effectively. Two methods for assimilating wind direction observations are developed. The first is a sub-optimal method that approximates the observation operator as linear with additive noise, making it amenable to use with EnKF and variational methods. The second is a nonlinear, non-Gaussian two-step ensemble filter. The methods are tested in an idealized problem. The two-step method is able to use wind direction observations successfully. The linear approximation can make effective use of wind direction observations only in situations where other observations have been used to reduce the prior uncertainty in wind direction to a level where the linear approximation is accurate.

Improving vortex position accuracy with a new multiscale alignment ensemble filter

Yue (Michael) YING
NERSC

A multiscale alignment (MSA) ensemble filtering method was introduced by Ying (2019) to reduce nonlinear position errors effectively during data assimilation. The MSA method extends the traditional ensemble Kalman filter (EnKF) to update states from large to small scales sequentially, during which it leverages the displacement vectors derived from the large-scale analysis increments to reduce position errors at smaller scales through warping of the model grid. This study stress-tests the MSA method in various scenarios using an idealized vortex model. We show that the MSA improves filter performance as number of scales (Ns) increases in the presence of nonlinear position errors. We tuned localization parameters for the cross-scale EnKF updates to find the best performance when assimilating an observation network. To further reduce the scale mismatch between observations and states, a new option called MSA-O is introduced to decompose observations into scale components during assimilation. Cycling DA experiments show that the MSA-O consistently outperforms the traditional EnKF at equal computational cost. A more challenging scenario for the MSA is identified when the large-scale background flow and the small-scale vortex are incoherent in terms of their errors, making the displacement vectors not effective in reducing vortex position errors. Observation availability for the small scales also limit the use of large Ns for the MSA. Potential remedies for these issues are discussed.

"Ensemblized" linear least squares (LLS)

Patrick N. RAANES
NORCE

Least-square linear regression (LLS) for model linearization is explicitly or implicitly in use by all ensemble methods. Indeed, by virtue of its very own chain rule this "ensemble gradient" perspective effortlessly unifies many flavors of the EnKF analysis update (Perturbed, EAKF, ETKF). Moreover, by virtue of Stein's identity*, the ensemble gradient can be shown to relate to average derivatives, notably without the awkward use of Taylor expansions.