WG 2 – Data assimilation

Chairs: Rossella Arcucci & Alberto Carrassi

Attendees: Arata Amemiya, Balaji Baduru, Jerome Barre, Kevin Berney, Marc Bocquet, Sid Boukabara, Ashesh Chattopadhyay, Marcin Chrust, Diana Laura Diaz Garcia, Sanjeev Dwivedi, Adam EL-SAID, Stephen English, Ronan Fablet, Alban Farchi, Anthony Fillion, Tobias Finn, Javier Garcia-Pintado, Sam Hatfield, Dorit Jerger, Patrick Laloyaux, Jianyu Liang, Takemasa Miyoshi, Matthieu Plu, Zhaoxia Pu, Manuel Pulido, Mélanie Rochoux, Marco Stefanelli, Jebb Stewart, Aneesh Subramanian, Xuguang Wang, Mingming Zhu

The interplay between DA and ML could (approximately/naively) be divided into three classes:

1. **Integrate ML and DA**

   E.g. using ML to mimic one piece of the DA procedure such as the ADJ, observation operator, ….

2. **Combine ML and DA**

   E.g. DA and ML are talking to each other with one doing one part of the job (DA estimating state) and ML the other (estimate the model, or the model error).

3. **Unify ML and DA**

   E.g. under a Bayesian framework, the two approaches, may be seen as “the same”. A by product of that is to recognise how w-4DVar is already doing ML.

WG 2 – Data assimilation

Chairs: Rossella Arcucci & Alberto Carrassi

Many challenges along the way (theoretical/technical/infrastructural):

- *DATA*: Sparsity, noisy, low-quality
- *REDUCED-SPACE-Representation*: AEs are appealing to address estimation of H, control variable reduction, pre-conditioning etc. Can we make it interpretable? Can we play within the latent space for, e.g., ensemble perturbations. And many more….
- *IMPLEMENTATION*: How to make dialog Python-based ML libraries with (often) Fortran codes. Can new ML library in Julia are paving the way? How to interface highly-parallelised code with ML. Need for developing adequate ”coupler”?

WG 2 – Data assimilation

Chairs: Rossella Arcucci & Alberto Carrassi

More specific challenges:

Improving the solvers

This may imply incorporating physical information.

Mixing different types of physics representation.

Improving/Estimating H

Autoencoders seem promising as they retain nonlinearities but maybe render difficult the interpretability.

This can be facilitated by using physical constraints.

Inpainting techniques from ML may help representation of model filed based on sparse data?

Defining a treatable small control variable

Again AE but a challenge in very high-dimension.

Ensemble construction

This is relevant for ensemble-based DA and prediction systems.

Can we perturb in the latent space (exploit low dimensionality)?

Parametrizing B

Again AU and latent space is worth investigating.

How to render ML training online

Fine tuning