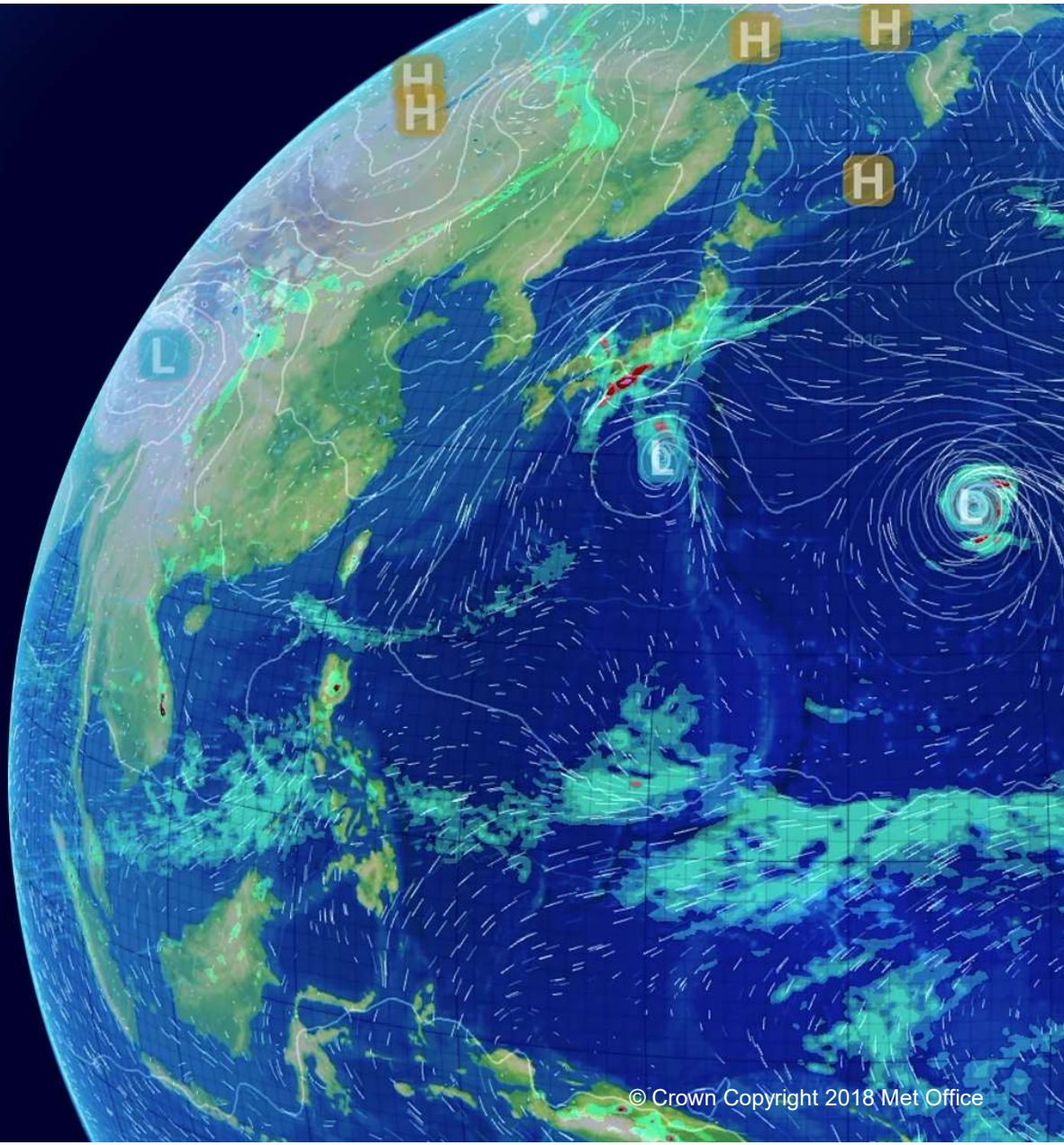
ENSEMBLE-BASED METHODS IN DA
(open discussion)

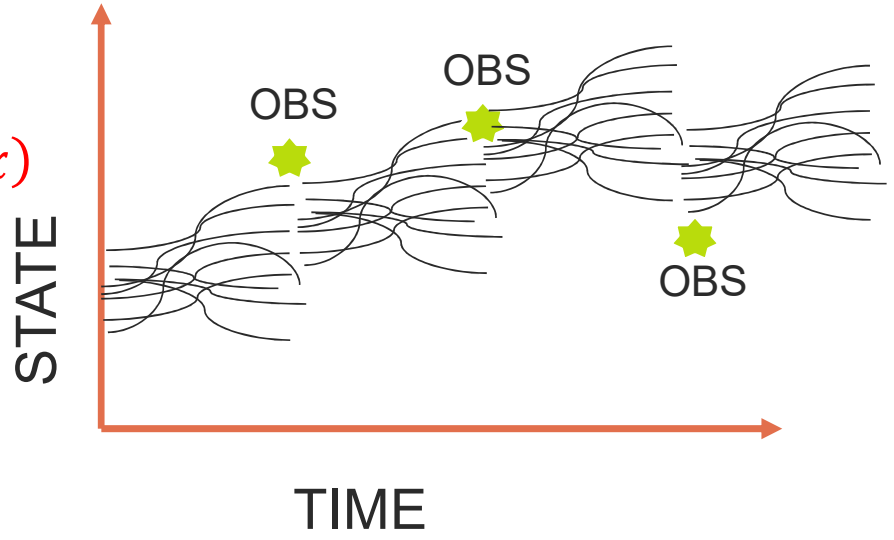
Parallel session on DA
EWGLAM 2023
Reykjavík

Marco Milan, Andrew Lorenc, David Simonin, Mayeul Destouches



GOAL OF THE ENSEMBLE DA

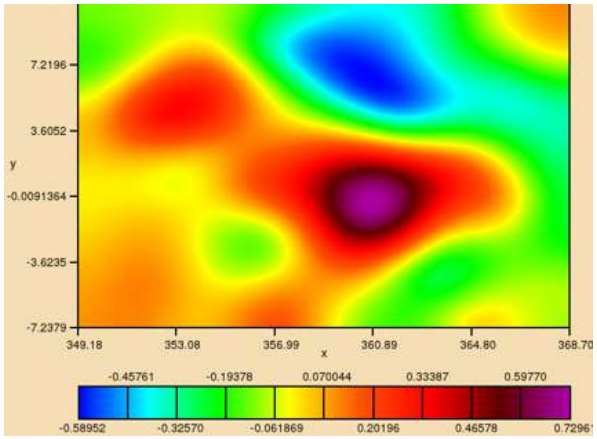
- LAM NWP systems have around 10^9 degrees of freedom.
- Only a limited number of ensemble members is affordable.
- Ensembles provide a gross approximation of the prior distribution $p(x)$
- Enhancing the resolution leads to a lower predictability of the system:
  - Enhance Ensemble members or enhance the resolution?
- How to deal with nonlinearity and non-Gaussianity of the dynamical model?

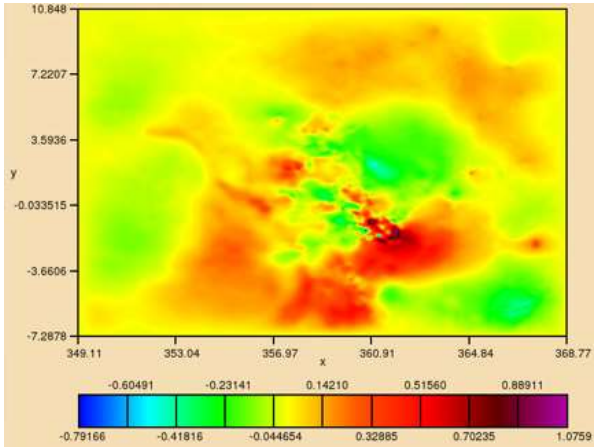


EPS-DA in LAM

- A host model can be used to defined better the large scales. Whilst the regional model can concentrate in determining the small scales.
- Met Office: LSB blending (Milan et al. 2023)
- Other similar systems are used from other Meteorological services/communities. Met No, SHMI, HARMONIE. An alternative solution is mixing the **B** matrix using large-scale forcing (Blendvar). Tudor et al. 2013.
- Different LAM-EPS use perturbations mostly from large-scale systems. (i.e. MOGREPS-UK).

Different ways to add perturbations?





DIFFERENT TYPES LAM-EPS (more popular)

- 3DEnVAR / 4DEnVAR.
- ENKF.
- PARTICLE FILTERS.
- Ideas about combination with data driven systems?
- Other not so popular?
- Bayes approach:
  - Prior distribution (state before new OBS) $p(x)$
  - Marginal PDF of obs (constant) $p(y)$
  - OBS likelihood (state as true) $p(y|x)$
  - Posterior PDF (What we want, update after OBS) $p(x|y)$

VAR approach

- Combines new information from OBS with prior information from forecast based on Bayes Theorem.
- To use Bayes Theorem, we need PDFs describing the likely errors for OBS and for forecast.
- For the forecast models the only practical approach is to assume the PDFs of their errors are [quasi-]Gaussian. Use the mean and the covariances.
- Issue: convective scale model errors need a big approximation.
- VAR can use ensembles in:
  - Hybrid covariances (use of the error of the day). Hybrid 3D/4DVar.
  - 4DEnVAR: Ensemble of 4DVar, doesn't require PF model and the adjoint. The trajectories of the model perturbations are defined using the ensemble forecasts.

How 4DEnVAR deals with LAM?

Met Office

ENKF

- EnKF uses the ensemble for the computation of the covariance matrix.
- Explicit assumption that the prior pdf and the likelihood of the observations as function of the state are [quasi]-Gaussian.
- Background covariances are updated during the cycling.
- ENKF (similar to VAR) can use a regularization, which introduces additional information, e.g.:
  - Localization is used to deal with the limited number of ensemble members (Houtekamer and Zhang, 2016).
  - Hybrid covariances as static covariances are noise free (Penny 2014).
  - Averaging (smoothing) the covariances spatially or temporally (Tsyrulnikov and Rakitko, 2017).
  - Other approaches including the wavelet one (for Var, Fisher 2006) .
- ENKF uses full-model nonlinear H(x), this can allow more flow/situation dependence, and cope better with nonlinearity than Var systems.

How inflation and localization influence the general posterior PDF?

PARTICLE FILTERS (PFs)

- The prior distribution of the state x at time k, is computed using a weighted combination of the ensemble members (particles) using a proposal distribution:

$$p(x_k) \sim \sum_{i=1}^{M} w_{k-1}^{i}\, \delta\left(x_k - x_k^i\right)$$

Leeuwen et al. 2019

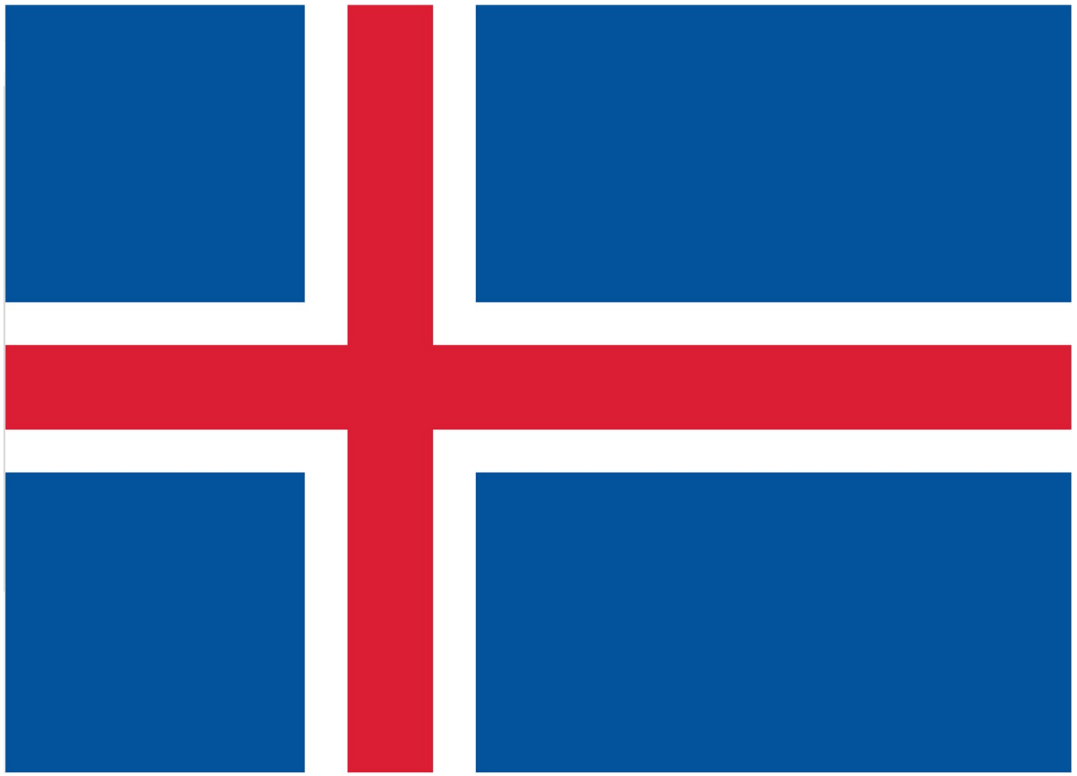
- The weights are updated using the likelihood of the members:

$$w_k^i \;\alpha\; w_{k-1}^i\, p(y_k \mid x_k^i)$$

- There is no assumption about the distribution.
- For high-dimensional systems the weights of many members could collapse to values near zero.
- Needs ad-hoc approaches for resampling and maintain similar weights.
  E.g., localization, use projection of OBS in ensemble space for the weights (Potthast et al. 2019)…

Many ways to develop PFs. Which is the best method for operational purposes?

Met Office



Takk