Recent Development and Evaluation of A Global Atmospheric Ensemble Data Assimilation using NICAM and MLEF with State Space Localization

ISDA 2023, Bologna, Italy

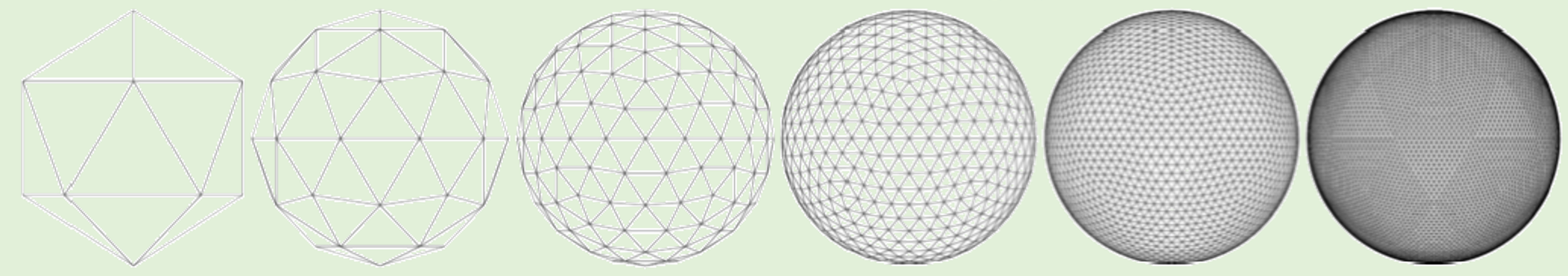
Ting-Chi Wu[1,2], Milija Zupanski[2], and Takemasa Miyoshi[3]

1. Central Weather Administration, Taipei, Taiwan (contact: tcwu@cwa.gov.tw),
2. Cooperative Institute for Research in the Atmosphere, Colorado State University, Fort Collins, CO, USA
3. RIKEN Center for Computational Science, Kobe, Japan

Introduction

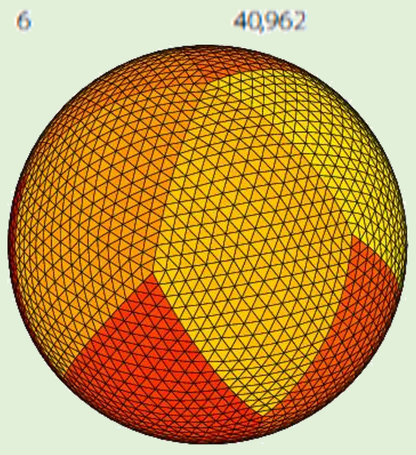
The Maximum Likelihood Ensemble Filter (MLEF) with State Space Localization (MLEFSSL) is an ensemble data assimilation method that incorporates ***state space covariance localization***, ***global numerical optimization***, and implied ***Bayesian inference***. MLEFSSL uses random projection to compute the localized forecast error covariance and reduce the analysis dimensions to a manageable space. MLEFSSL is applied to the Nonhydrostatic Icosahedral Atmospheric Model (NICAM) to explore its capability under a realistic ***high-dimensional*** dynamical application.

Computational Strategy of NICAM



glevel 0	glevel 1	glevel 2	glevel 3	glevel 4	glevel 5
Δx= 7141 km	**Δ**x= 3570 km	**Δ**x= 1785 km	**Δ**x= 893 km	**Δ**x= 446 km	**Δ**x=223 km

g-level	Points N_P	Arcs N_A	Triangles N_T	Average area $\bar{A}$ (km²)	Average interval $\bar{\Delta}$ (km)
6	40,962	122,882	81,920	1,245.361	111.596



- NICAM domain decomposition via rlevel:
  > # of decomposed regions = 10 x 4^{rlevel} (parallel processing)
- In this study, NICAM configured at glevel 6 & rlevel 0 with 38 vertical layers is used.
  > **Δ**x = 111 km with 10 decomposed regions
  > Each ensemble member requires 10 processors (1 region/processor)
  > N ensemble members requires 10 x N processors

The NICAM-MLEFSSL System

$P_f = FF^T$ where $F = (f_1 \dots f_{N_E \times N_{RR}})$; ***E*** = dynamical ; ***RR*** = random reduced-rank

Pre-Calculate localized random sample

$s_n = L^{1/2} r_n \ (n = 1, \dots, N_{RR})$

L = localizing matrix; r_n = uncorr. Gaussian random vector

↓

Data Assimilation cycle:

Calculate localized forecast error covariance

$f_k = diag(p_i) s_n \ (k = 1, \dots, N_E \times N_{RR})$

↓

Minimization iteration (j):

Apply observation operator *h on all ensemble ($N_E \times N_{RR}$)*:

call the NICAM-LETKF ***obsope_driver***

↓

Compute the Z matrix:

$Z_f = (z_1^f \dots z_{N_E \times N_{RR}}^f);\ z_k^f = \mathrm{R}^{-1/2}[h(x_f + f_k) - h(x_f)]$

↓

Compute Hessian matrix using *Cholesky* factorization:

$Q = I + {Z_f}^T Z_f = GG^T$ (only calculate ***G*** at *j* = 0)

↓

Precondition via the triangular system & compute gradient:

$g = w - {Z_f}^T \mathrm{R}^{-1/2}[y - h(x_f + Fw)]$

solve $d_k = \mathrm{d(G, g)}$ **minimization algorithm dependent**

↓

Control variable update (control vars then state vars):

$w_{j+1} = w_j + \alpha_j d_j$ and $x_{j+1} = x_j + Fw_{j+1}$

(→ back to Apply observation operator)

↓

At convergence, update NICAM ensemble

$x^a = x_{jmax}$ and $x_i = x^a + {P_a}^{1/2} \theta_i$

↓

Run NICAM ensemble forecast

call the NICAM-LETKF ***nhm_driver***

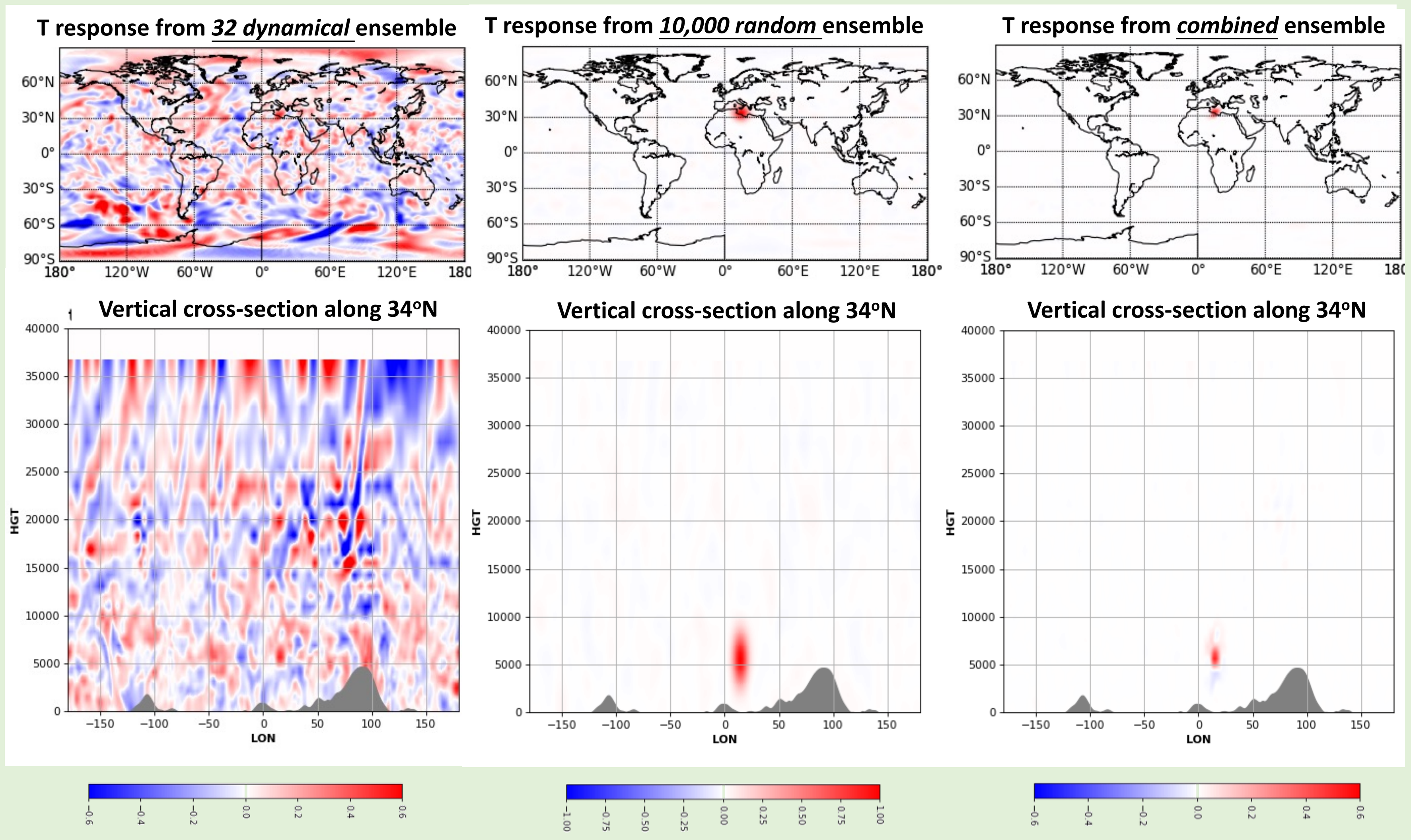
(→ back to Calculate localized forecast error covariance)

Characteristics of MLEFSSL

- A single algorithm with consistent DA methodology and state-space covariance localization
- Maximum a-posterior estimate with global solution (over all points)
- Bayesian inference in terms of the state and the associated uncertainty
- Nonlinearity addressed via numerical optimization

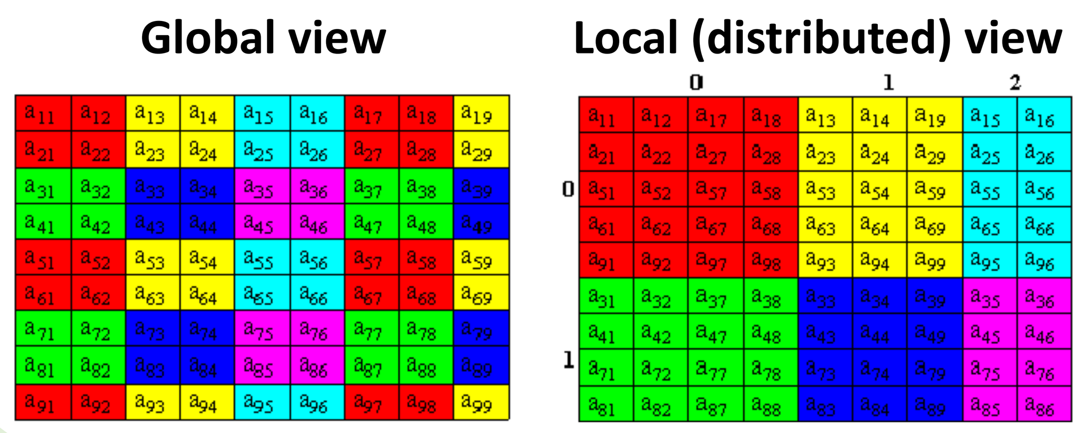
Localized Forecast Error Covariance

- A single observation of temperature at 500 hPa
- 32 dynamical x 10,000 random = 320,000 combined members



Utility of ScaLAPACK

- ScaLAPACK (Scalable Linear Algebra PACKage) is a standard mathematical software that is freely available. It is written in Fortran, C, as well as in Python (PyScalapack).
- Use of ScaLAPACK in MLEFSSL has been proven critical for calculating the Hessian, preconditioning the minimization, control variable update, as well as for the ensemble update, in terms of memory and computational efficiency.
- The relevant ScaLAPACK subroutines used in MLEFSSL are as follows:
  - ***PDPOTRF*** for Cholesky factorization
  - ***PDTRTRS*** for triangular linear system solver
  - ***PDGEMV*** for matrix-vector product
- In MLEFSSL, most vectors and matrices are saved only locally using ScaLAPACK double-cyclic indexing.



- An example illustrating the double-cyclic indexing for a 9 x 9 matrix distributed on 2 x 3 = 6 CPUs.
- Color represents the distribution of matrix elements to the 6 CPUs.

Conclusion & Future Work

- **A first working version of NICAM-MLEFSSL is completed**
  - Calculation of covariance localization is efficient because only random vectors and the localizing matrix are required.
  - Localization is pre-calculated (offline, outside of DA loop)
- ➢ **Continue to increase the size of random ensemble (256, 1024, 2048, and 4096), all using 32 dynamical ensemble, and determine the minimum required size of combined ensemble**
- ➢ **Evaluate assimilation results and further improve code efficiency**

Reference

Zupanski, M., 2021: The Maximum Likelihood Ensemble Filter with State Space Localization, *Mon. Wea. Rev.*, 149, 10, pp 3505-3524, https://doi.org/10.1175/MWR-D-20-0187.1

Acknowledgement

We greatly appreciate the computational resources of the supercomputer ***Fugaku*** provided by the RIKEN Center for Computational Science and the technical support from the NICAM-LETKF team.