Data Assimilation for NWP at ECCC

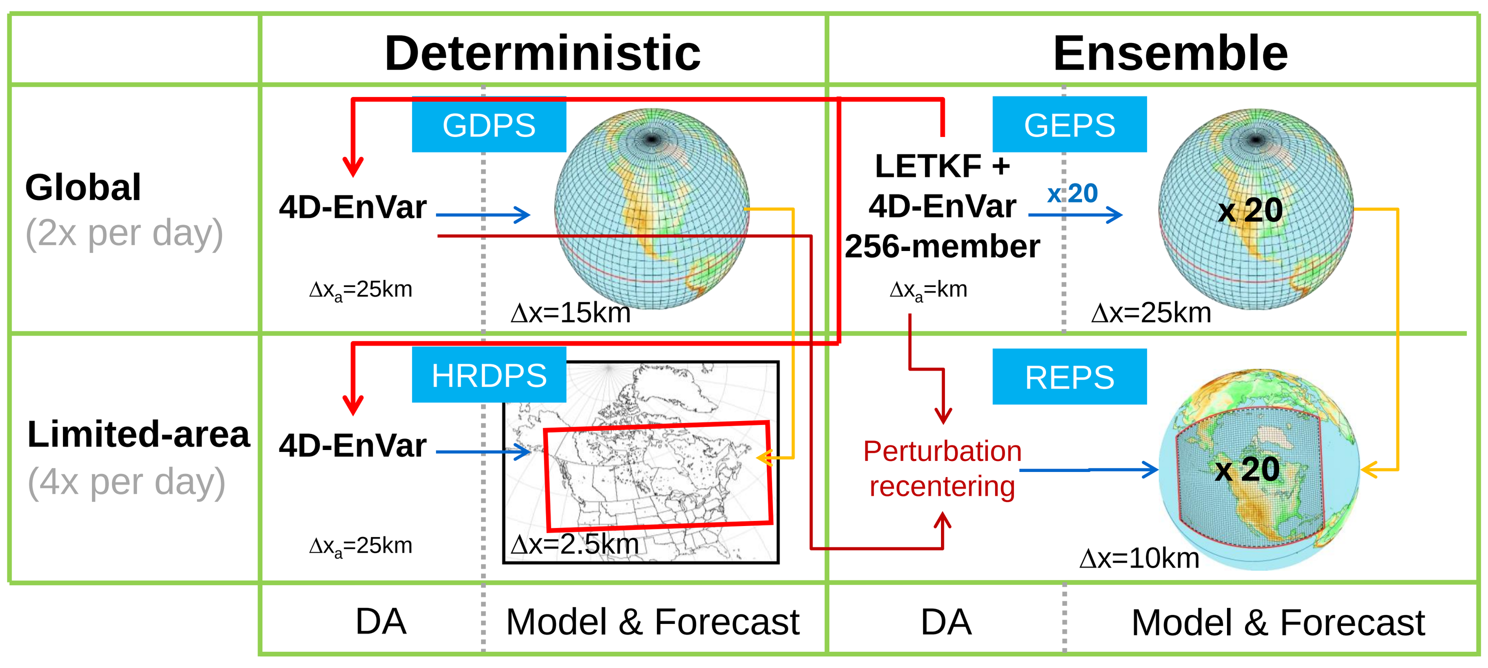
Environment and Climate Change Canada

Jean-François Caron, Mark Buehner, Seung-Jong Baek, Alain Caya, Sylvain Heilliette, Alex Komarov, Sergey Skachko & Seyyedsina Voshtani

Meteorological Research Division

(1) Operational DA systems based on MIDAS

The global ensemble prediction system (GEPS; described below) is the backbone for all atmospheric DA and forecasting operational systems at ECCC.



*Global deterministic and ensemble forecasts are **coupled to ocean-ice models**, with ocean-ice initial conditions derived from uncoupled assimilation systems.*

The Modular and Integrated Data Assimilation System (MIDAS) is a **unified data assimilation framework** for research and operational use at ECCC with the objective of **avoiding redundant work** within the organization and to facilitate the transition towards **coupled data assimilation** of different components of the Earth system.

- All DA algorithms, obs QC, bias correction and thinning, ensemble inflation and recentering all implemented in MIDAS
- Operational sea-ice analysis and experimental SST analysis also use several MIDAS programs
- Forecast Sensitivity to Observation Impact (FSOI) system also implemented in MIDAS

Reference

Complete description of MIDAS available at: https://doi.org/10.5194/gmd-18-1-2025

(2) Global Ensemble Prediction System

Our global ensemble of analysis are generated by a Local Ensemble Transform Kalman Filter (LETKF) that are partially recentered around a 4DEnVar analysis with the hybrid gain approach defined as:

$$\mathbf{x}^a_{i,\mathrm{hyb}} = \mathbf{x}^a_{i,\mathrm{LETKF}} + \gamma\left(\mathbf{x}^a_{\mathrm{EnVar}} - \overline{\mathbf{x}^a_{i,\mathrm{LETKF}}}\right)$$

With γ = 0.5 (p < 50 hPa) and 0.75 (p > 100 hPa)

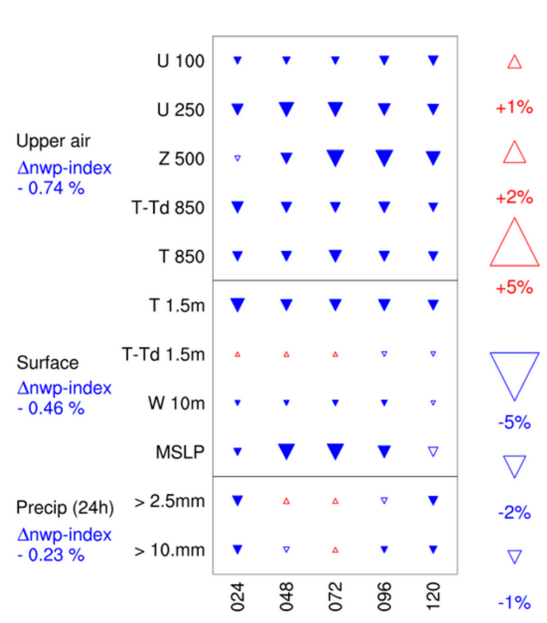
γ = 1 : the analyses are centered on the EnVar analysis
γ = 0 : no recentering takes place

Particular features of this ensemble DA system:

1. The LETKF uses cross-validation (i.e., independent subset of ensemble members for updating each member) to prevent in-breeding and the 8 sub-ensembles are randomly created at every analysis steps to prevent filter divergence.
2. The **background in the 4DEnVar analysis is the ensemble mean background** and not a deterministic forecast. This improves the forecast quality of a) the ensemble mean analysis (control member) and b) the 20-member ensemble as shown below:

Relative changes in CRPS for Boreal Winter 2022

Blue (red) triangles means that using the ensemble mean as the background in 4DEnVar improves (degrades) the ensemble scores.



Ensemble mean background **vs** Deterministic background (i.e., taken from the control member) in 4DEnVar

Reference

Most recent description of the GEPS DA available at: https://doi.org/10.1175/WAF-D-22-0108.1

(3) Scale-Dependent Localization in EnVar

All three (3) NWP systems relying on a 4DEnVar analysis (GEPS, GDPS and HRDPS) adopt a horizontal scale-dependent horizontal localization (hSDhL) defined as :

$$\Delta\mathbf{x} = \sum_k \sum_j \mathbf{e}_{k,j} \circ \left(\mathbf{L}_j^{1/2}\xi_k\right)$$

k: member index
j: horizontal scale index

where $\mathbf{e}_{k,j}$ is horizontal scale j of normalized member k perturbation

Scale and horizontal localization radius (km)	
Small	1000
Medium	2400
Large	3800

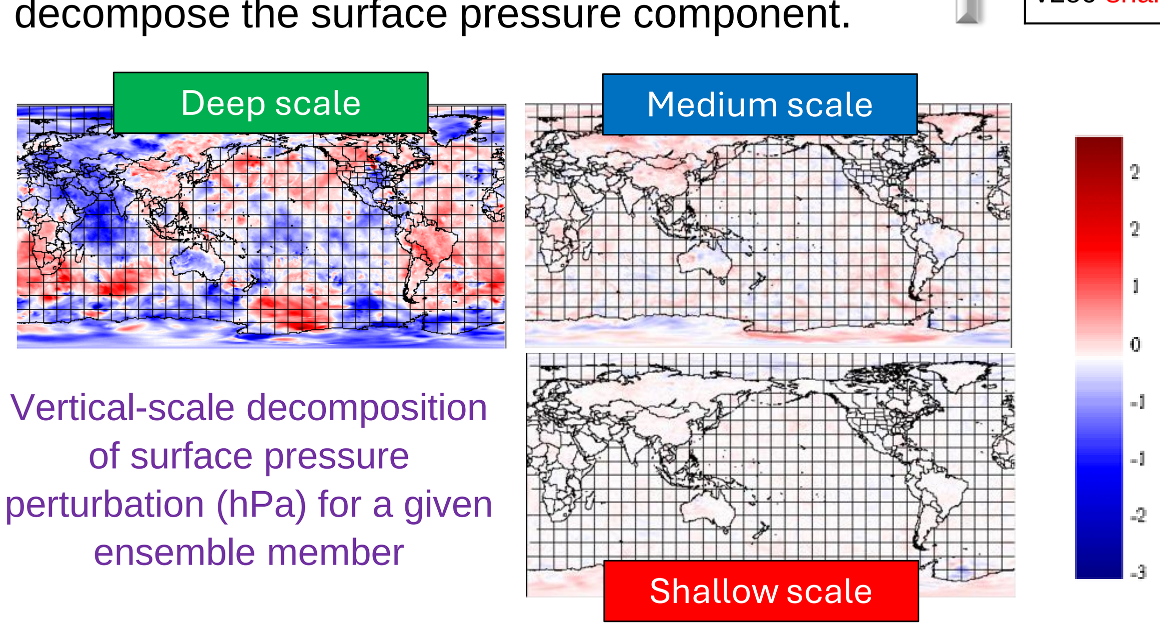
Implementing hSDhL eliminated any benefit from including static covariances. Thus, our DA systems completely rely on flow-dependent covariances since 2024.

Full scale-dependent localization should also take into account vertical-scale dependent vertical localization (vSDvL).

The recent adoption of a height-based coordinate in the Canadian model GEM will (finally) enable testing vSDsL since a meaningful 3D pressure field is needed to vertically decompose the surface pressure component.

Vertical scale ↑			
	hLoc-small vLoc-deep	hLoc-medium vLoc-deep	hLoc-large vLoc-deep
	hLoc-small vLoc-medium	hLoc-medium vLoc-medium	hLoc-large vLoc-medium
	hLoc-small vLoc-shallow	hLoc-medium vLoc-shallow	hLoc-large vLoc-shallow

Horizontal scale →



Vertical-scale decomposition of surface pressure perturbation (hPa) for a given ensemble member

Reference

Implementation and impact of hSDhL in GDPS available at: https://doi.org/10.1175/WAF-D-22-0055.1

(4) Observation Bias Correction Approach

Using the background state to compute observation bias coefficients would reinforce the model biases.

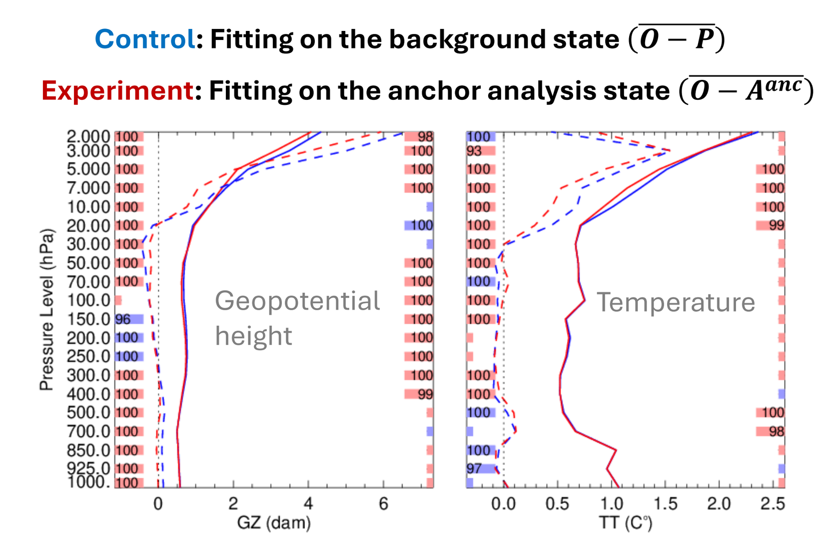
To mitigate this issue, since 2014, the reference state use to compute dynamical bias corrections coefficients in our global systems come from a separate analysis that assimilate only "anchor" observations, i.e. observations that are deemed less prone to biases.

Current approach

- Perform a 3D-EnVar analysis with hSDhL and pure ensemble-derived covariances that assimilates only the following "anchor" observations: **GPS-RO, Radiosonde, AMVs and surface obs.**
- In the GDPS, the background state for this analysis comes from the main 4DEnVar cycle whereas, in the GEPS, the (control) forecast from the LETKF + 4DEnVar ensemble mean analysis is used.
- Use regression to estimate bias model coefficients by fitting the following selected observations families to past 7 days of 3DEnVar analyses: **Satellite radiances, aircraft and ground-based GNSS observations**.

Advantages

- Compared with using the background state, dynamic estimation of bias model coefficients with standalone analysis improves analyses and short-term forecasts.
- Gives added flexibility to optimize for removal of model-induced bias without affecting 4DEnVar and LETKF.



Std dev (solid) and mean (dashed) of GDPS 24h forecast deviation from ECMWF analyses computed over the globe for boreal Summer 2022

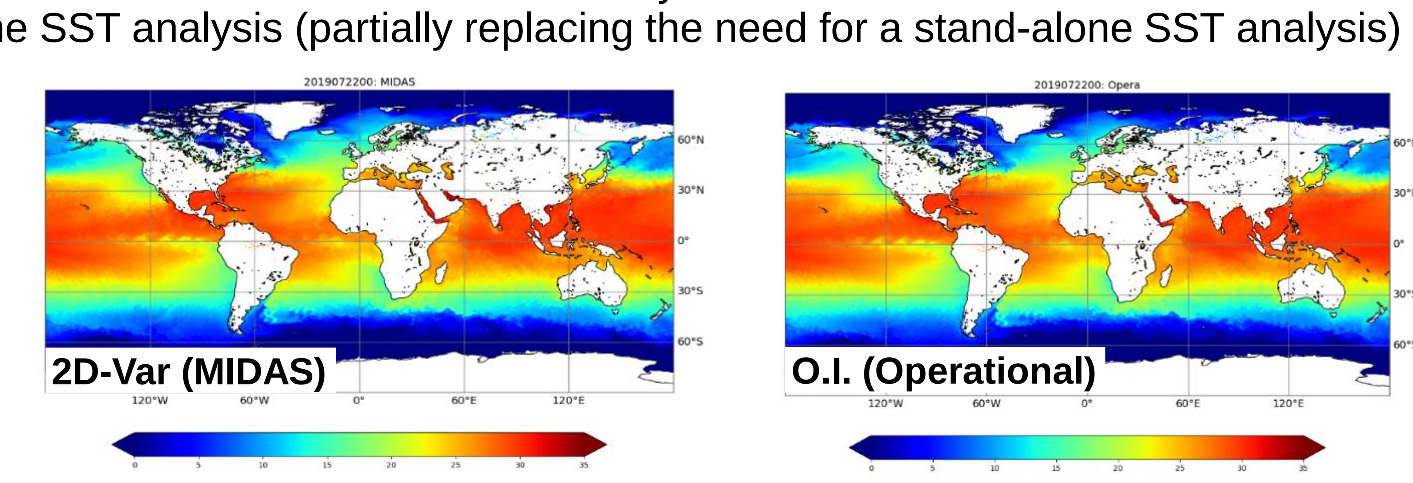
(5) Ocean DA

New MIDAS-based SST analysis

- Operational system based on optimal interpolation (O.I.), difficult to modify/maintain due to old-style code
- New MIDAS-based system uses 2D-Var with diffusion operator **B**-matrix (like sea ice analysis); much code shared with other applications; parallel code is much faster
- Assimilating SST observations in MIDAS facilitates work on 3D ocean DA with MIDAS-LETKF, that will then enable research on fully coupled DA
- A MIDAS-based 3D ocean DA can easily assimilate the same SST observations used in the SST analysis (partially replacing the need for a stand-alone SST analysis)

Reference

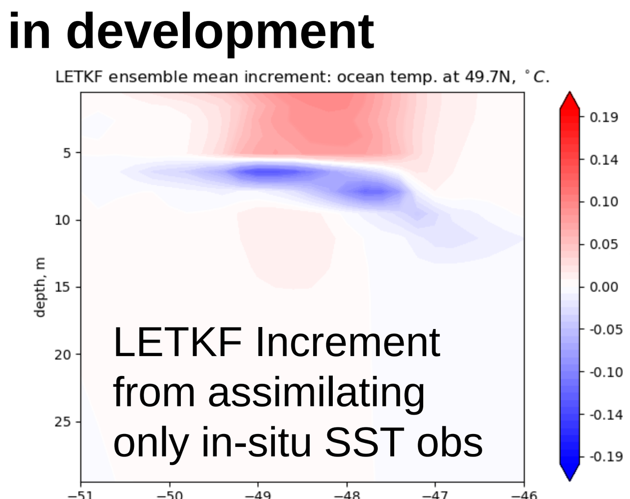
Description of the new SST analysis available at: https://doi.org/10.1002/qj.4796



3D ocean assimilation with MIDAS-LETKF in development

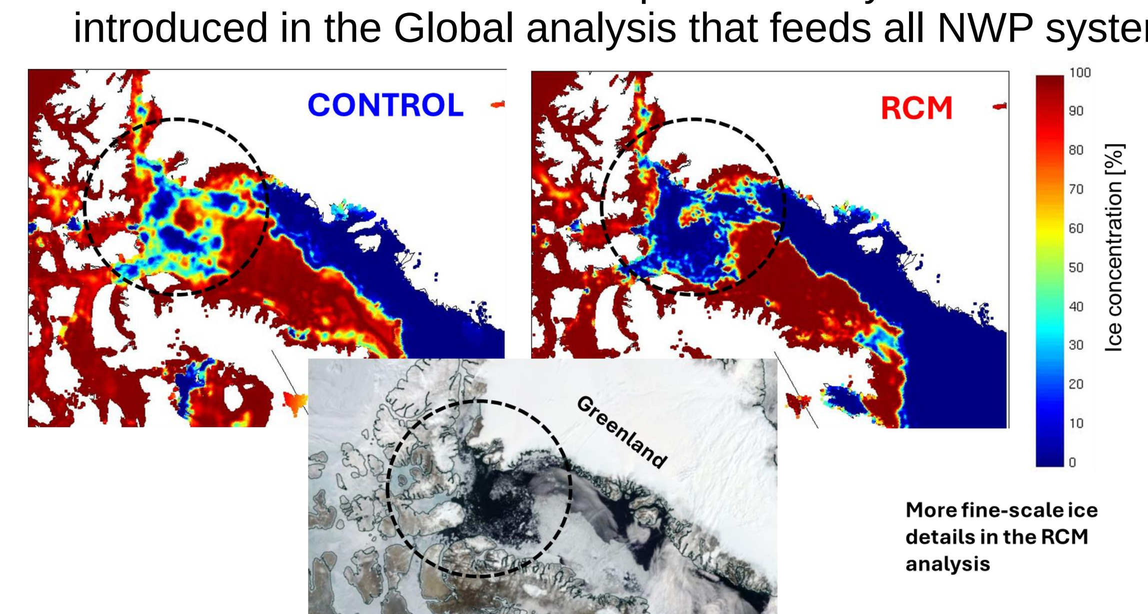
Implementation of LETKF directly on ORCA (tri-polar) grid relatively straightforward, whereas 4D-EnVar more difficult due to localization method

Initial tests used ensembles from existing 20-member GEPS coupled forecasts, recentered on deterministic state → LETKF produces a deterministic analysis



(6) Sea-Ice DA

- Ice concentration analysis produced every 6h on both a 5-km regional and a 10-km global grids, based on 2D-Var approach with horizontal background-error correlations modelled using a diffusion operator
- Assimilate observation types with wide range of spatial resolution for estimating ice concentration: SSM/IS, ASCAT, AMSR2, ice charts, and now SAR (in regional system only) from RadarSat Constellation Mission (RCM)
- For low-resolution observations, observation operator aggregates ice concentration from all grid points within the observation footprint to avoid constraining unresolved small scales
- Produce analysis error estimate using a simple OI approach to provide measure of uncertainty
- Also developing ice type and ice thickness analyses
- Assimilation of RCM data improves ice concentration analyses, particularly in the situations where CIS charts are not available and where high spatial resolution is important: marginal ice zones, inland lakes, areas near the coastline, narrow channels
- Since June 2024, ECCC has become the only center that does operational SAR ice data assimilation in sea ice prediction systems. In the future, RCM data will be introduced in the Global analysis that feeds all NWP systems



References

Description of the sea ice analysis available at: https://doi.org/10.1002/qj.2408 https://doi.org/10.1016/j.rse.2024.114113