Environment and Climate Change Canada   Environnement et Changement climatique Canada

Overview of Observational Aspects of Data Assimilation

Mark Buehner
Meteorological Research Division
Environment and Climate Change Canada

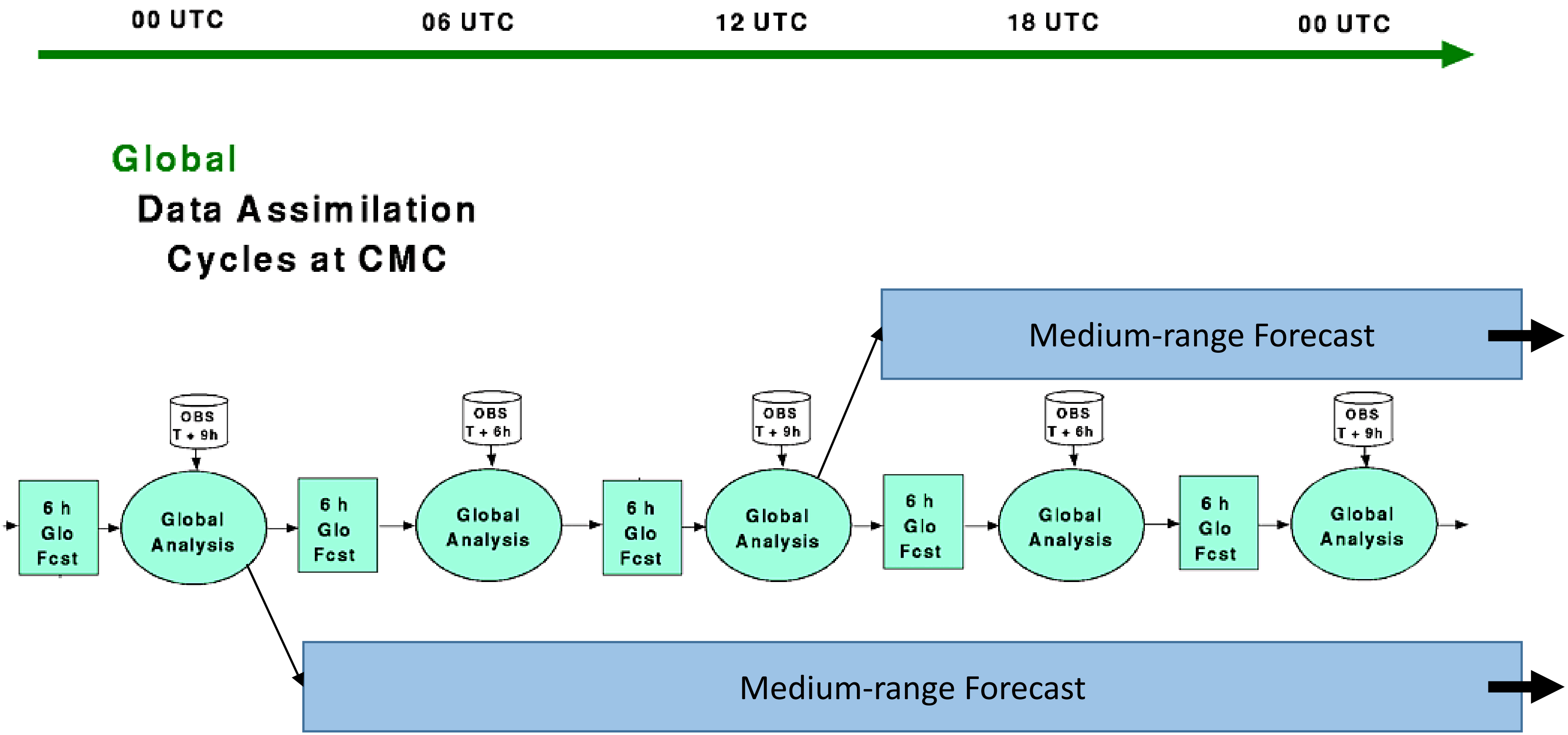
ECMWF Annual Seminar
September 13-17, 2021

Canada

Introduction: Use of observations in DA

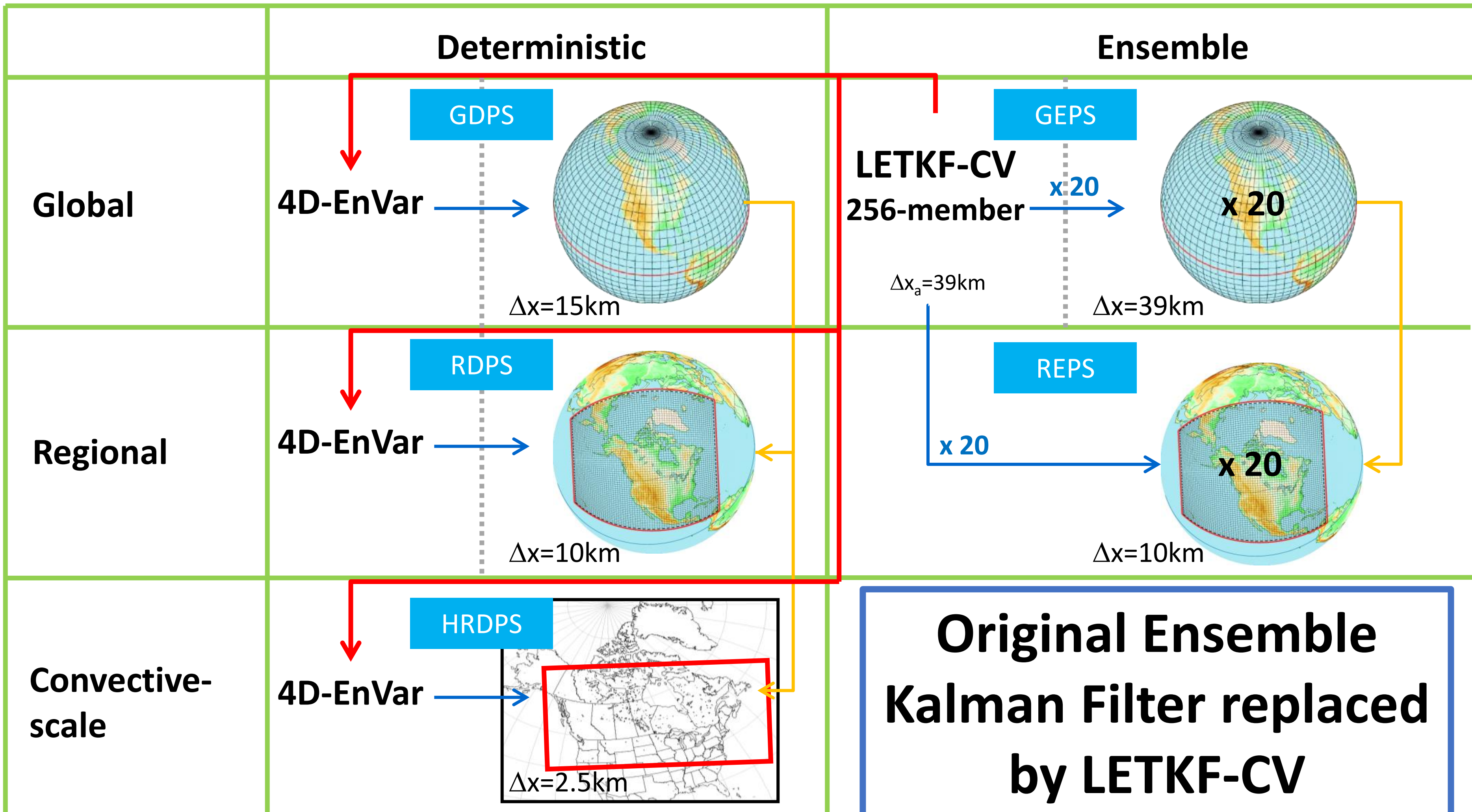
- Various data assimilation (DA) algorithms used for NWP, ocean/sea-ice prediction, and other Earth system components
- Approaches can be derived from Bayes theorem $[p(x|y) \propto p(y|x)p(x)]$ by applying specific simplifying assumptions about errors and models
- Goal is to repeatedly correct short-term forecast (i.e. background state) as best as possible using all available observations
- Deterministic DA produces single "best estimate"; ensemble DA produces ensemble of estimates, approximating probability distribution
- Any observation can be used, even if noisy, biased, indirect, but:
  - Their "optimal" use requires perfect knowledge of both the observation <u>and</u> background error statistics
  - Also requires accurate model of the observing process (i.e. observation operator) and associated uncertainties
  - Incorrect assumptions about observation or background error statistics (bias, covariances, Gaussianity) can result in degraded accuracy from assimilation

Example: Global Deterministic Prediction System



Data Assimilation: Goal is to correct a short-term forecast using information from all available observations

NWP systems at ECCC



Kalman filter – basis of most DA approaches

- The Kalman Filter equations describe the analysis update of the mean and covariance in the case of unbiased Gaussian errors and linear observation operator:

$$\boldsymbol{x}_a = \boldsymbol{x}_b + \mathbf{K}(\boldsymbol{y} - \mathbf{H}x_b)$$
$$\mathbf{P}_a = (\mathbf{I} - \mathbf{KH})\mathbf{P}$$
$$\mathbf{K} = \mathbf{PH}^T(\mathbf{HPH}^T + \mathbf{R})^{-1}$$

$\mathbf{K}$: Kalman gain matrix
$\mathbf{P}$: Background error cov.
$\mathbf{R}$: Observation error cov.
$\mathbf{H}$: Observation operator

- For a single, directly observed, variable:

$$x_a = (1 - k)x_b + (k)y$$
$$\sigma_a^2 = (1 - k)\sigma_b^2$$
$$k = \frac{\sigma_b^2/\sigma_o^2}{\sigma_b^2/\sigma_o^2 + 1} \quad (\text{i.e. between 0 and 1})$$

- Therefore, result of DA is just a weighted average of the background and observation; assimilation always reduces the apparent uncertainty
- Weighting (i.e. k) only depends on ratio of σ_b^2/σ_o^2
- Assumes we know the true background and observation error covariances (which we generally don't)

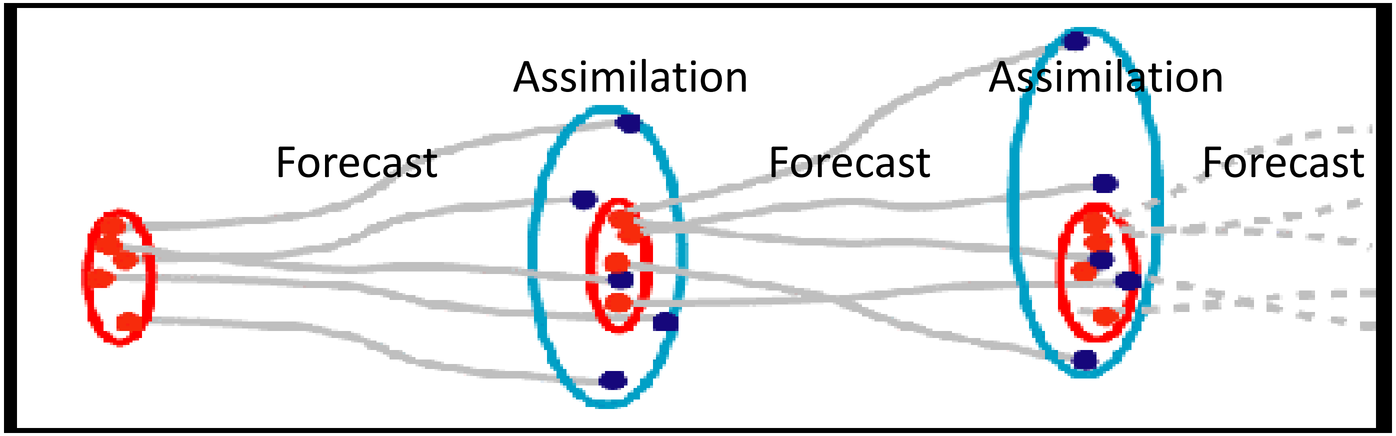
Ensemble Kalman filter: EnKF

- In the EnKF, ensemble of model states are propagated through the forecast-analysis cycle to approximate probability distributions → <u>error of ensemble mean should be consistent with ensemble spread</u>
- Ensemble of short-term forecasts used to approximate the background-error covariances; all members updated during assimilation step to correct mean state and reduce ensemble spread
- Many EnKF flavours; all flavours require "localization" of background-error covariances; original "perturbed obs" EnKF applies the KF equation to each member to assimilate randomly perturbed obs:

$$\mathbf{x}_a^n = \mathbf{x}_b^n + \mathbf{K}_{enkf}(\mathbf{y} + \boldsymbol{\varepsilon}_{obs}^n - H(\mathbf{x}_b^n))$$

$$\mathbf{K}_{enkf} = \mathbf{L} \circ \mathbf{P}_{ens}\mathbf{H}^T(\mathbf{L} \circ \mathbf{H}\mathbf{P}_{ens}\mathbf{H}^T + \mathbf{R})^{-1}$$

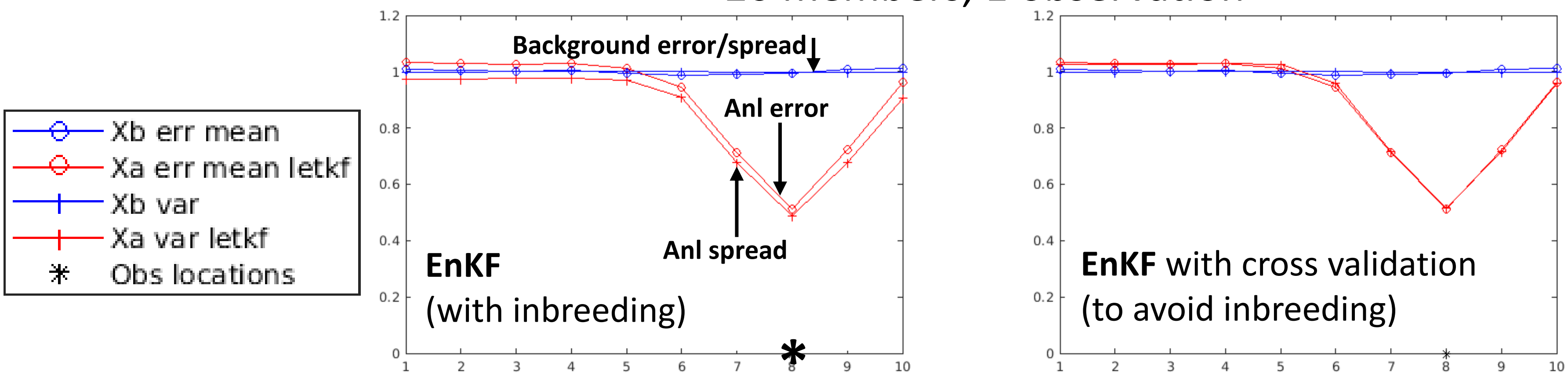
$\boldsymbol{\varepsilon}_{obs}^n$: Random obs perturbation
$\mathbf{L}$: Localization matrix
$\mathbf{P}_{ens}$: Ensemble-based background error cov.



Inbreeding during EnKF analysis

- Most EnKF algorithms do not account for “inbreeding”, caused by dependence between ensemble perturbation being updated and perturbations used to estimate $\mathbf{P}_{ens}$
- Produces too much ensemble spread reduction in assimilation step
- Cross-validation allows inbreeding to be mostly eliminated by using an independent set of perturbations to estimate $\mathbf{P}_{ens}$ for each member update – e.g. use other 19 members to update each of 20 members

20 members, 1 observation



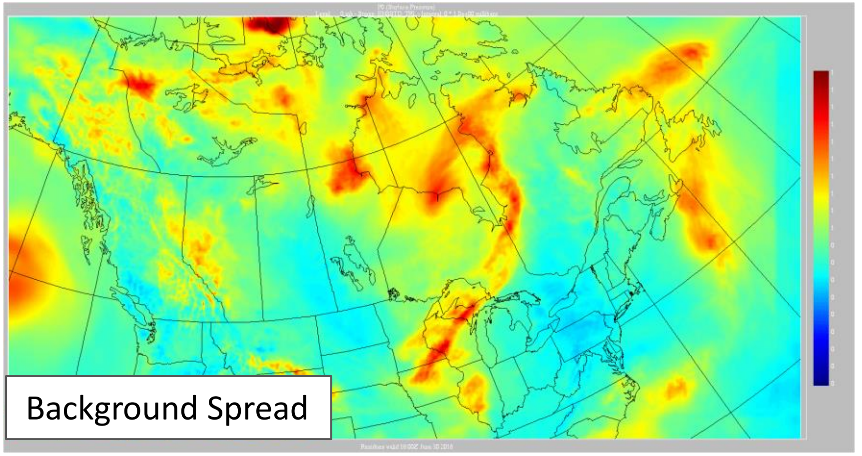
$$\mathbf{x}_a^n = \mathbf{x}_b^n + \mathbf{K}_{enkf}^n(\mathbf{y} + \boldsymbol{\varepsilon}_{obs}^n - \boldsymbol{H}(\mathbf{x}_b^n))$$

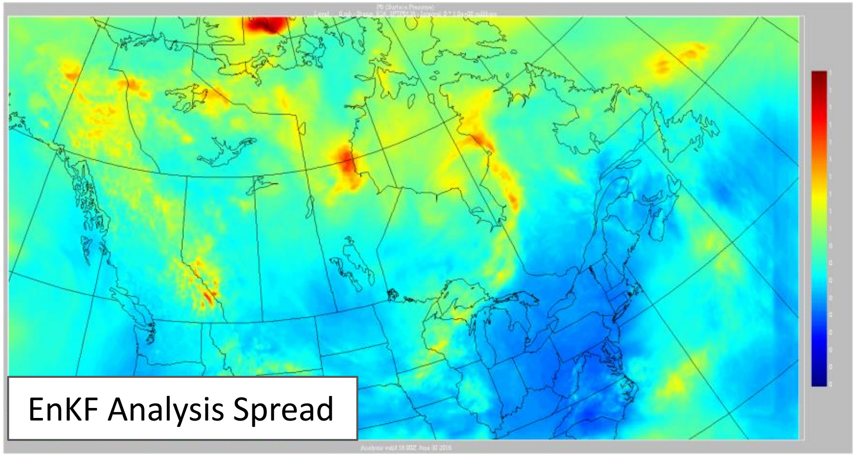
$$\mathbf{K}_{enkf}^n = \mathbf{L} \circ \mathbf{P}_{ens}^n \mathbf{H}^T (\mathbf{L} \circ \mathbf{H}\mathbf{P}_{ens}^n \mathbf{H}^T + \mathbf{R})^{-1}$$

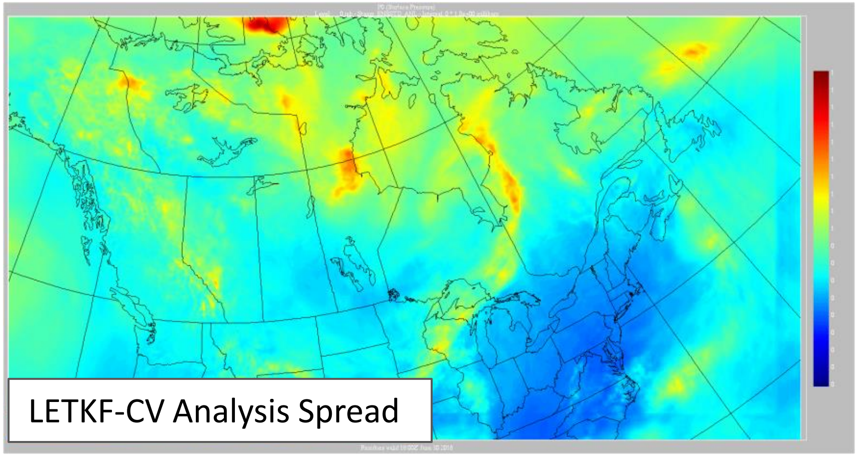
doesn’t use *n*th member

Application of cross-validation to LETKF

- Until now, cross validation only applied to perturbed-observation EnKF, but also possible using "gain form" of LETKF (Bishop et al. 2017)
- Figures show surface pressure ensemble spread (from Buehner 2020)
- Without cross validation, LETKF needed additional multiplicative inflation to maintain reasonable spread (RTPS, α=0.8)
- At ECCC, LETKF-CV is about to replace EnKF in operations







Ensemble-Variational approach: EnVar

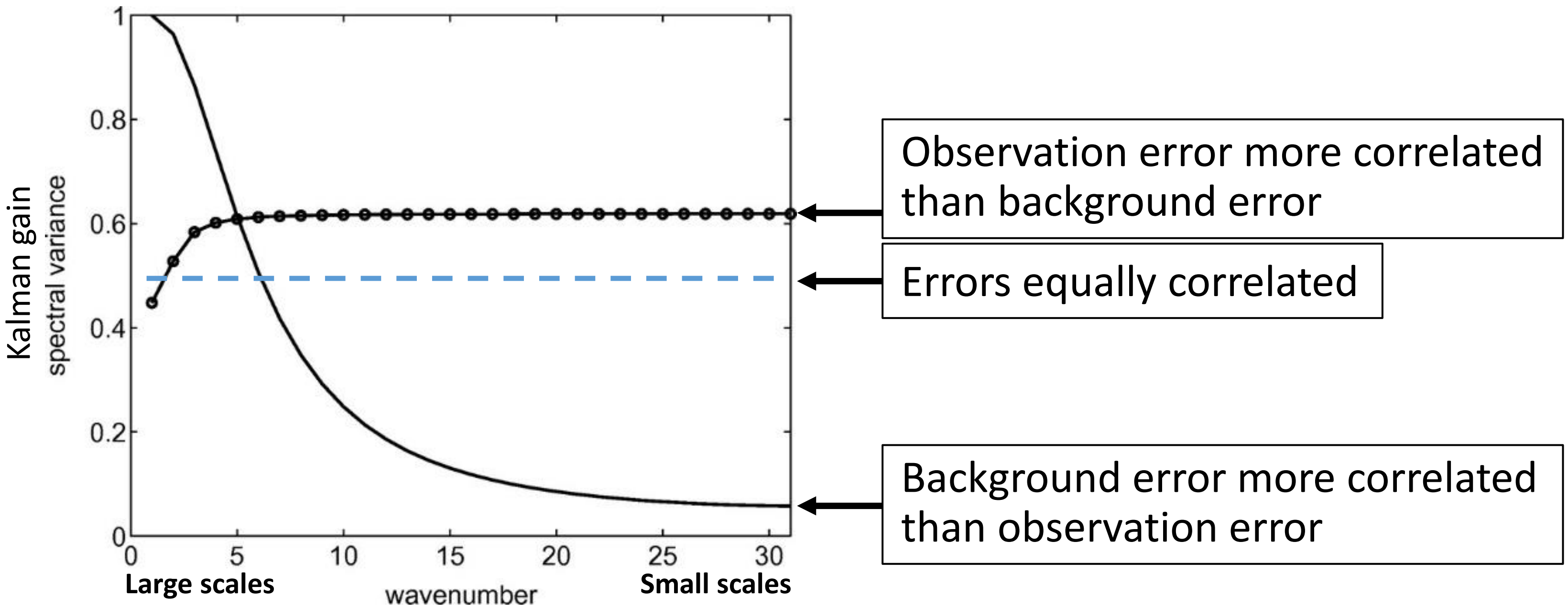
- Ensembles from EnKF also used to specify background-error covariances for deterministic DA with 4D-EnVar
- Similar to EnKF, but unique features of 4D-EnVar approach:
  - Very large number of observations can be assimilated (overall cost not directly related to number of observations)
  - Can use hybrid covariances, blending climatological covariances with ensemble-based covariances, though currently moving away from its use
  - Spatial covariance localization applied in model-space, not observation space → more appropriate for non-local observations (e.g. radiances)
  - Scale-dependent localization (SDL) can be used to better simultaneously fit wide range of scales (Caron and Buehner 2018)
- Analysis obtained by minimizing the pre-conditioned cost function:

$$J(\mathbf{v}) = \frac{1}{2}\mathbf{v}^T\mathbf{v} + \frac{1}{2}[\mathbf{y} - \boldsymbol{H}(\mathbf{x}_b) - \mathbf{H}\Delta\mathbf{x}(\mathbf{v})]^T\,\mathbf{R}^{-1}[\mathbf{y} - \boldsymbol{H}(\mathbf{x}_b) - \mathbf{H}\Delta\mathbf{x}(\mathbf{v})]$$

$$\Delta\mathbf{x}(\mathbf{v}) = \sum_{n=1\rightarrow N_{ens}} \frac{(\mathbf{x}_b^n - \overline{\mathbf{x}_b})}{\sqrt{N_{ens} - 1}} \circ \mathbf{L}^{1/2}\mathbf{v}^n$$

Kalman gain in spectral space

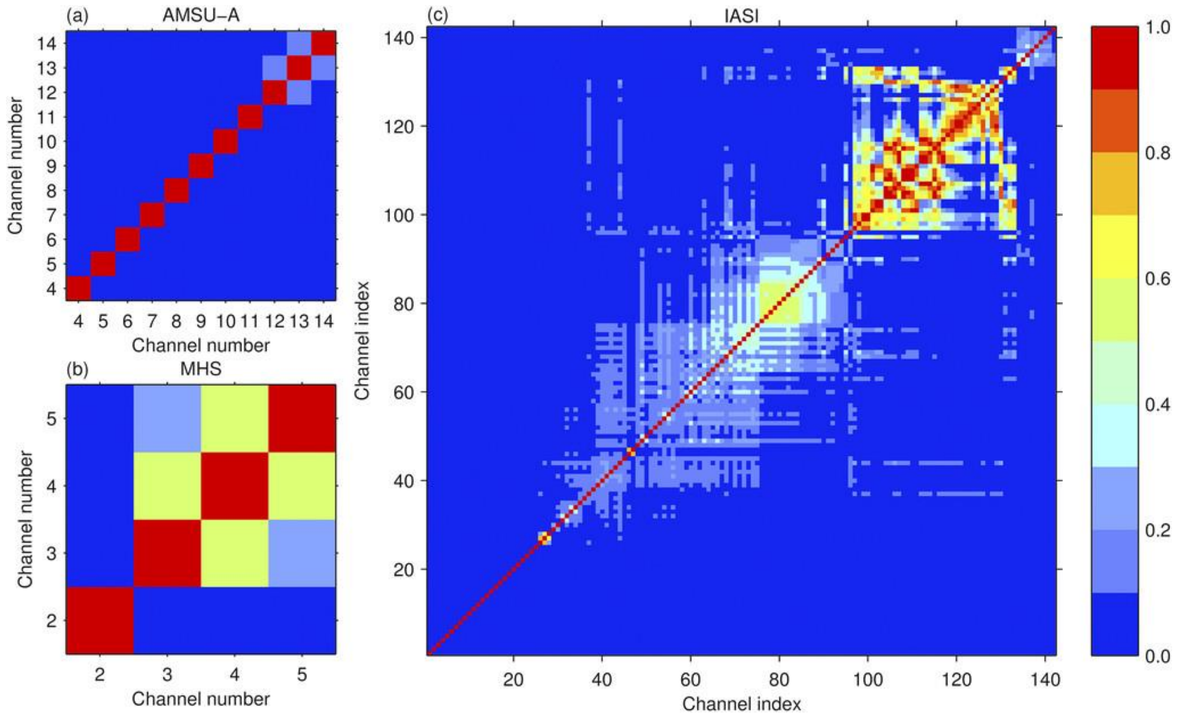
- When considering multiple observations distributed over space/time, error **correlations** can influence weighting as much as variances
- Figure (from Stonebridge et al. 2018) shows Kalman gain ($\mathbf{K}$) for different amounts of observation and background error correlation



- Ignoring real observation error correlations results in giving too much weight to observations at large scales and too little at small scales

Accounting for obs-error correlations

- Can be technically challenging to fully account for observation-error correlations in DA algorithms (requires large matrix inversions)
- Vertical or inter-channel error correlations easier to include than horizontal correlations due to smaller number of observations involved
- Figure shows example of inter-channel radiance error correlations used in ECCC deterministic DA systems



Mis-specified obs error statistics

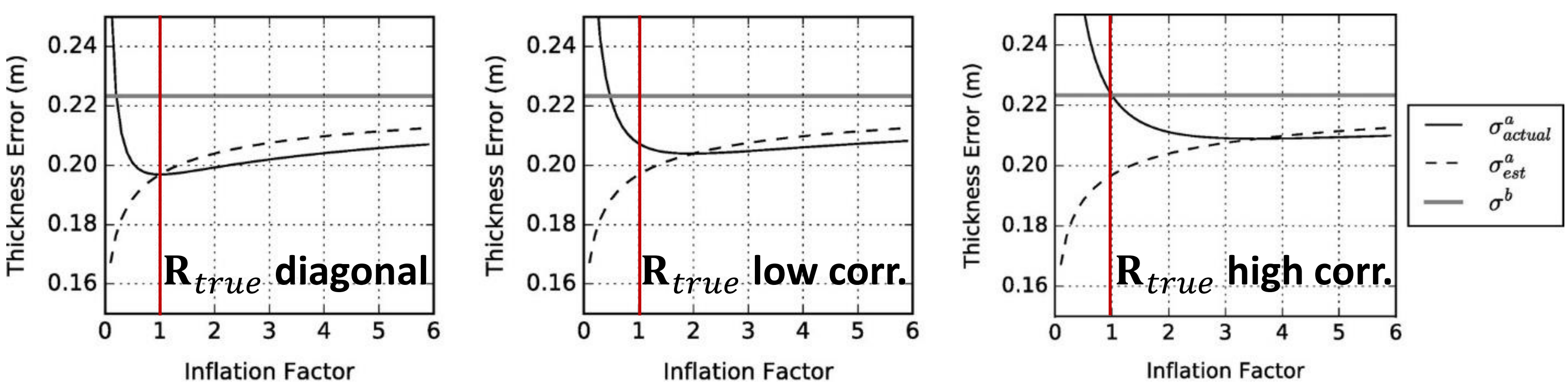
- If $\mathbf{R}$ is mis-specified, the <u>real</u> analysis error covariance is:

$$\mathbf{P}_a = (\mathbf{I} - \mathbf{KH})\mathbf{P}(\mathbf{I} - \mathbf{KH})^T + \mathbf{KR}_{true}\mathbf{K}^T$$

- Instead of the normal <u>estimated</u> value:

$$\mathbf{P}_a = (\mathbf{I} - \mathbf{KH})\mathbf{P}$$

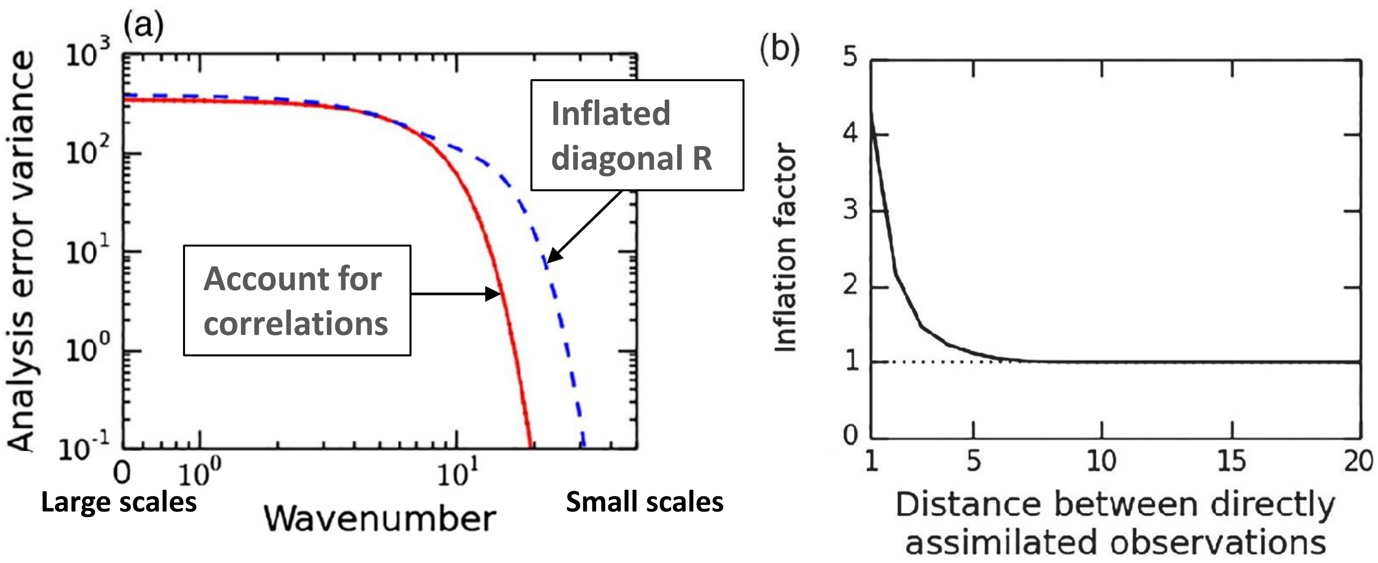
- Common to inflate observation error variance to avoid overfitting large scales from neglecting observation error correlations
- Figure (Stonebridge et al. 2018) shows <u>real</u> and <u>estimated</u> analysis error stddev as a function of inflation factor applied to a diagonal $\mathbf{R}$



- Inflating diagonal $\mathbf{R}$ can reduce both the real analysis error and underestimation of analysis error (e.g. via analysis ensemble spread)

Mis-specified obs error correlations

- Inflation avoids over-fitting large scales, but further under-fits small scales
- Alternatively, spatial thinning of obs can be used to avoid overfitting; also neglects observational information at small scales



(a) Analysis error when assimilating obs with correlated error when either accounting for correlations (red) or using inflated diagonal **R** (blue)

(b) Optimal obs error inflation factor for different amounts of spatial thinning

(Idealized numerical experiments from Bédard and Buehner 2019)

Thinning vs. error inflation in real NWP

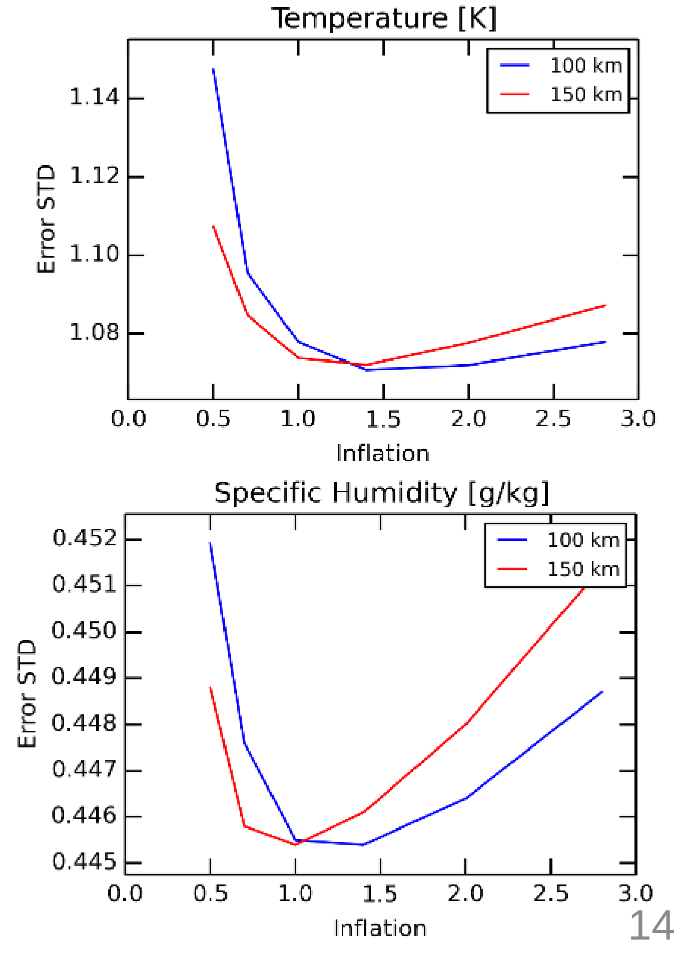
- Series of experiments performed varying both the inflation of the error variance and the spatial thinning for all types of radiance observations

- Decreasing thinning without changing inflation can result in forecast degradation

- Beneficial to increase inflation when thinning reduce from **150km** to **100km**

- However, optimal inflation likely not equal for all sensor types or channels (e.g. humidity vs. temperature sensitive)

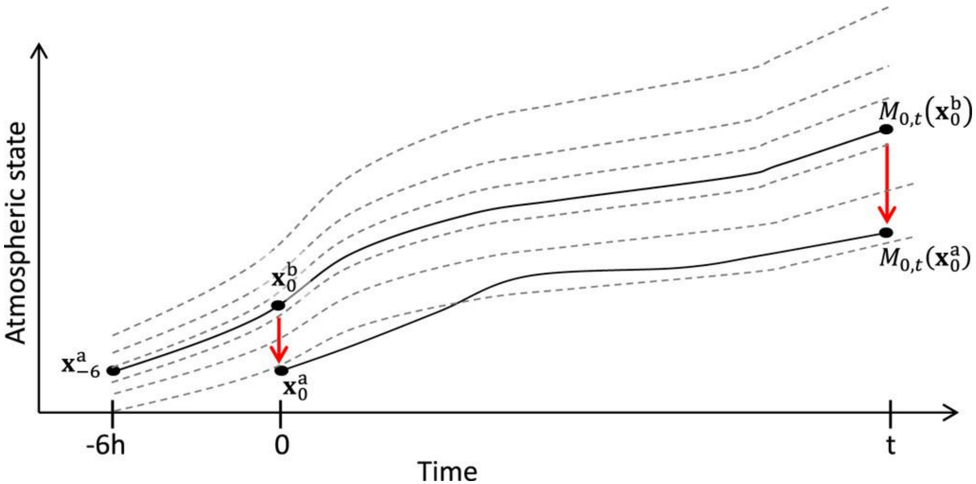
Average global 48h forecast scores (1000 to 1 hPa):



(Courtesy: Joël Bédard and Alain Beaulne)

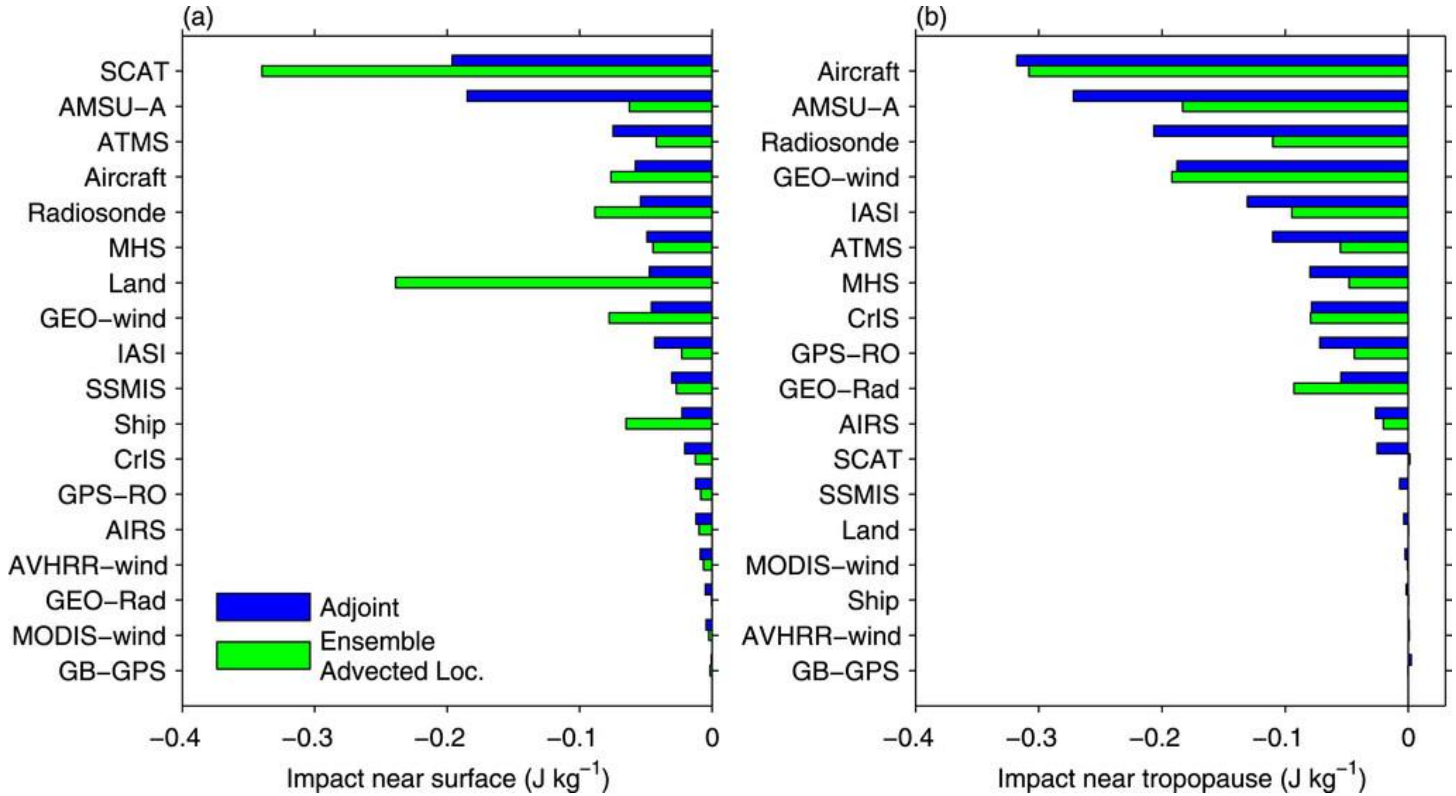
Estimating observation impact

- Modern NWP DA systems are highly complex; measuring the impact of currently assimilated or future hypothetical observations through OSEs or OSSEs can be very costly (and OSSEs difficult to do correctly)
- Several alternative approaches exist for estimating the impact of observation subsets on the analysis or forecast error, including:
  - Impact should be related to change in ensemble spread, which depends on $\mathbf{H}$, $\mathbf{R}$, and $\mathbf{P}_{ens}$, not actual observed values → can be used to estimate impact of new hypothetical observations for observation network design
  - FSOI approximates contribution of currently used observations to error reduction in short-term forecasts by propagating sensitivities from forecast to analysis time using either an adjoint model or ensembles



Estimating observation impact: FSOI

- Estimated relative impact of major observation groups varies with approach used to propagate sensitivity from forecast to analysis time
- Choice of scalar measure of forecast error also affects relative impact, for example: near surface (left) vs. near tropopause (right)



(Figure from Buehner et al. 2018)

The challenge of Earth system DA

- Forecast models are becoming increasingly coupled – e.g., the ECCC medium-range deterministic forecasts include coupled components:
  - Atmosphere
  - Land
  - Ocean
  - Sea ice
  - Ozone
- Forecast model for generating the background state, only couples Atmosphere + Land + Ozone → i.e. *partially* weakly coupled DA
- Data assimilation systems for all components are nearly independent, only some *ad hoc* coupling included where essential
- Use of separate DA systems for each component makes it challenging to obtain consistent analyses for initializing coupled forecasts
- Strongly coupled DA also challenging as it requires all components use same DA system/algorithm, temporal frequency, obs. cutoff time

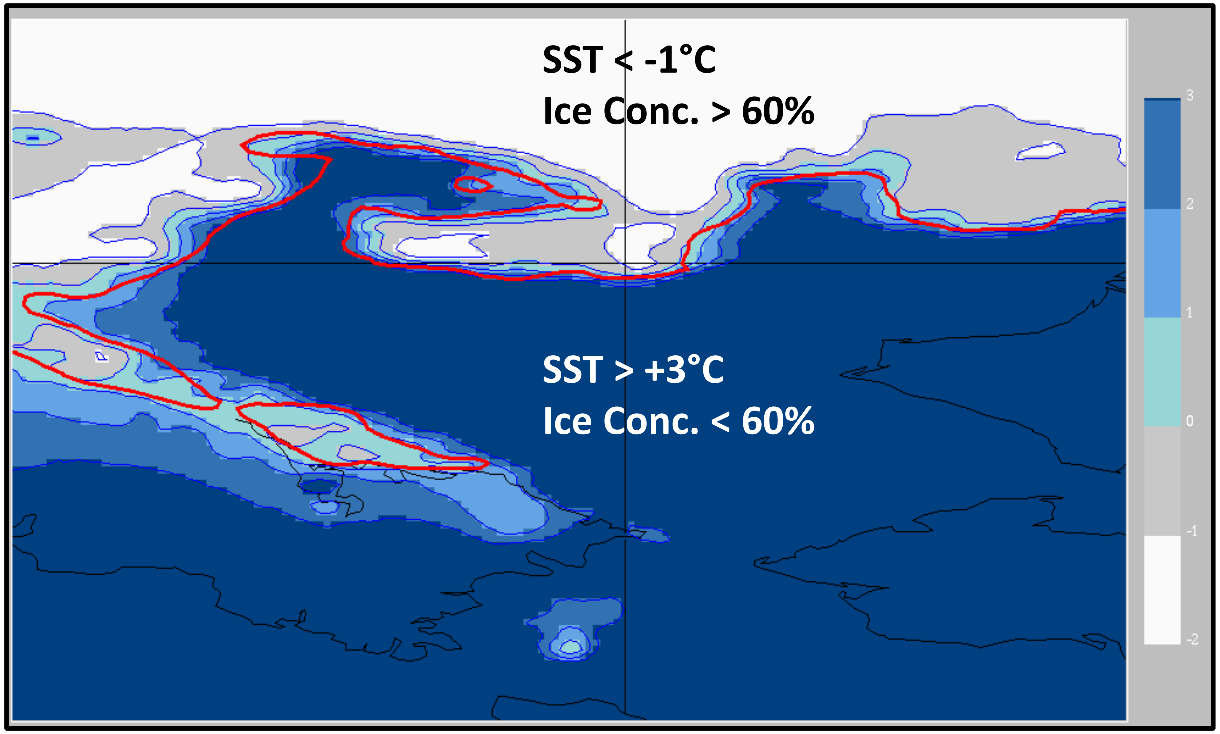
Atmosphere-ice-ocean initial conditions

- Initial condition consistency is required between Atmosphere, Ice and Ocean components to avoid initial shocks and forecast degradations
- Example: ice analysis indicates high concentration, but background state has no ice and SST above freezing → ice added by DA will rapidly melt during subsequent forecast
- Analyses of SST and Ice concentration currently produced with DA systems that cycle through time without using a forecast model
- Since DA components at ECCC are currently separate, *ad hoc* approaches needed to impose consistency:

    1. Consistency between <u>Atmosphere and Ocean</u> from using same gridded SST analysis in both – assimilated in ocean with high weight, specifies lower boundary for atmosphere
    2. Consistency between <u>Ocean and Ice</u> from bogus freezing point SST observations where ice analysis > 60% concentration and remove ice in analysis where SST analysis > 4°C

SST and ice concentration analyses

- Example of analyses of SST (shading) and Ice Concentration (red contour: 60% concentration) in the Bering Sea (July 9, 2021)

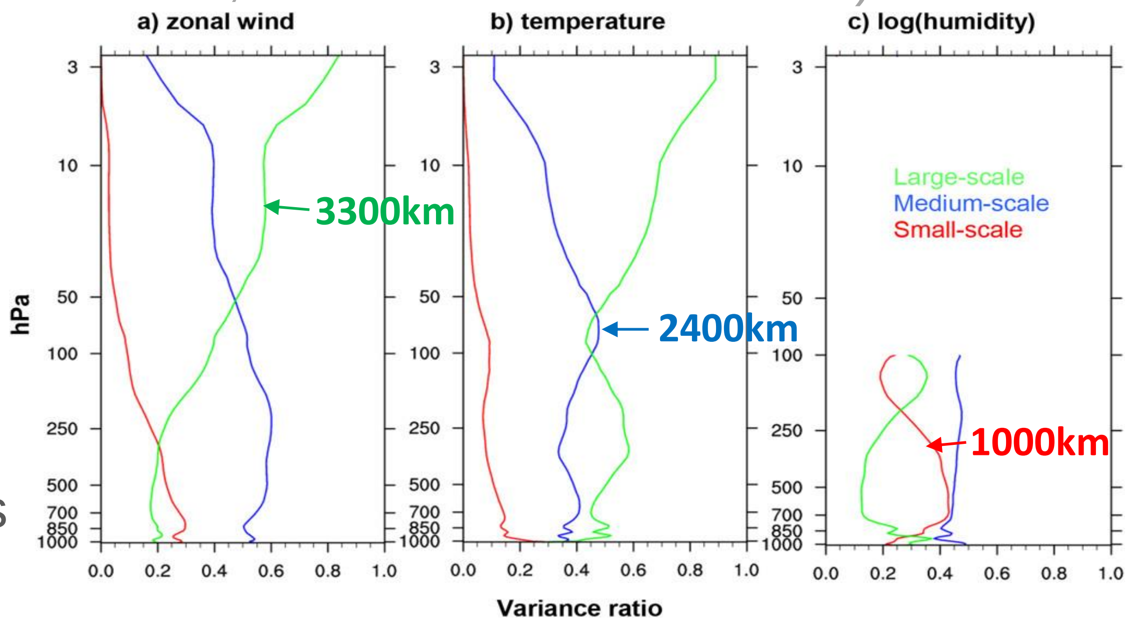


Wide range of spatial scales

- Earth system components may have dominant background errors at very different spatial scales, e.g. in ECCC operational systems:
    - Sea ice 3D-Var correlation scale: **10km** (e-folding distance)
    - SST Optimal Interpolation correlation scale: **85km** to **136km** (e-folding distance)
    - 3D ocean localization scale applied to ensemble of climatological anomalies: **200km** (cut-off radius)
    - Land surface “pseudo-analysis” screen-level Optimal Interpolation correlation scale: **239km**($T\text{-}T_d$) and **425km**(T) (e-folding distance)
    - Atmospheric 4D-EnVar localization scale: **2800km** (zero distance)
- Special treatment required to correctly include between-component cross-covariances with component-dependent length scales
- 4D-EnVar now uses Scale-Dependent Localization with length scales: 1000km, 2400km, 3300km, depending on horizontal scale → significantly improved medium-range forecasts

Scale-dependent localization

- Method: Decompose ensemble perturbations according to specified ranges of horizontal scale and apply different amount of localization to each (Buehner and Shlyaeva 2015; Caron and Buehner 2018)
- Naturally provides varying amount of localization that depends on variable, vertical level, local conditions, etc.
- Figure shows relative amount of ensemble variance in 3 wave bands (Caron & Buehner 2018)



- So far, only use for horizontal-scale-dependent horizontal localization, could also apply to vertical-scale-dependent vertical localization
- If applied to strongly coupled 4D-EnVar with multiple components, also provides component-dependent localization to multi-component ensemble (cross-)covariances while ensuring positive-definite matrix

Overview of MIDAS

- MIDAS: Modular and Integrated Data Assimilation System
- Purpose: provide platform for efficient collaborative development of DA systems at ECCC for various applications and **facilitate future research on strongly coupled DA**
- A set of Fortran programs and modules that implement DA algorithms (3D-Var, EnVar and LETKF) and related procedures
- Evolved out of original variational DA code; never separate from code for operational systems; code improved/refactored *as needed*
- Development is an open collaboration; single code repository managed with *gitlab work flow*: feature/release branches, issues, etc.
- Fortran modules are fairly general, and becoming increasingly so, to facilitate increasing number of applications and scientific innovations
- EnVar and LETKF share a lot of code (e.g. obs operators, file I/O), so when new application is implemented for one, the other is much easier

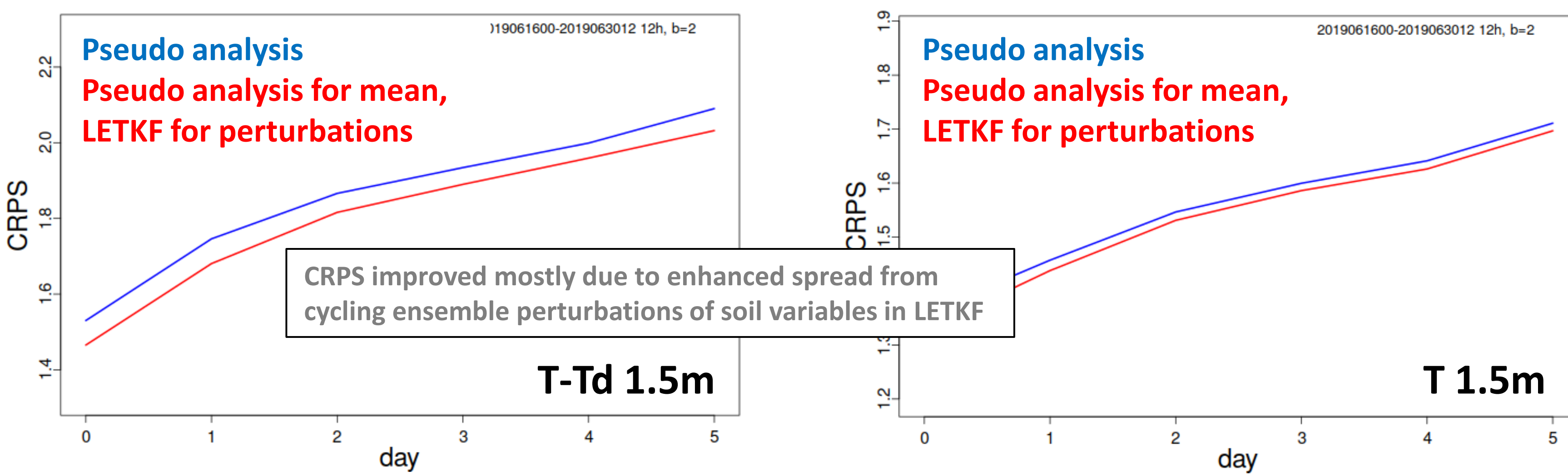
Earth system DA at ECCC

- Currently many separate (uncoupled) DA systems for initializing Earth system components, including atmosphere, land, ocean, sea-ice
- MIDAS used for operational NWP (4D-EnVar, LETKF, Observation QC, thinning and bias correction, soon FSOI)
- Integration in MIDAS of DA for more Earth system components facilitates collaboration on common DA tools/algorithms and essential for research on strongly coupled DA
- Currently testing both sea-ice and SST analysis systems using MIDAS-based 3D-Var (diffusion operator for horizontal correlations)
- Technical work also begun on MIDAS-based DA for the 3D ocean and land surface → starting with LETKF is easiest
- Goal is to first implement in MIDAS stand-alone DA for each Earth system component before exploring strong coupling strategies in LEKTF and 4D-EnVar

Initial experiments for coupled land DA

- Operational “pseudo analysis” produces analysis increments for soil variables based on analysis increments for screen-level temperature and humidity – all ensemble members initialized with same analysis
- Atmospheric LETKF assimilates screen-level temperature and humidity; atmosphere and land-surface models already coupled
- Easy to include soil variables in LETKF analysis → ensemble cov. between soil and atmospheric variables produces soil increments
- Cycling coupled atmosphere-land ensembles, recentered on pseudo analysis, improves screen-level ensemble forecasts in short tests

Screen-level CRPS scores for North America, June 16-30, 2019



ECCC Future directions and challenges

- Remove technical obstacles by developing MIDAS such that it can be used effectively for uncoupled DA of sea-ice, ocean, land, etc.
- Explore pragmatic approaches for strongly coupling DA of selected Earth system components within LETKF and 4D-EnVar
- Related scientific innovations:
  - Address model uncertainty at interface of coupled Earth system models: reduce biases, simulate random errors in ensembles
  - Apply SDL to coupled ensemble covariances of multiple Earth system components with very different length scales
  - Address observation gaps by assimilating obs at interface of Earth system components: e.g. surface-sensitive radiances over land/ice/snow