Motivation and Methods in Earth System Data Assimilation

ECMWF Annual Seminar
Mark Buehner
Data Assimilation and Satellite Meteorology Research
10-13 September 2018
Outline

• Motivation: The need for Earth System Data Assimilation
• Methods:
  – DA Methods used for NWP, including hybrid methods
  – DA Methods for other geophysical components: e.g. Sea ice
  – Methods for coupled DA
• Examples:
  – First attempts at coupled Atmosphere-Ice-Ocean DA at ECCC
• Strategy for Earth System DA at ECCC
  – Use of highly modular common DA software for all components
  – Explore scale-dependent combined with system-dependent ensemble covariance localization
Outline

• Motivation: The need for Earth System Data Assimilation

• Methods:
  – DA Methods used for NWP, including hybrid methods
  – DA Methods for other geophysical components
  – Methods for coupled DA

• Examples:
  – First attempts at coupled Atmosphere-Ice-Ocean DA at ECCC

• Strategy for Earth System DA at ECCC
  – Use of highly modular common DA software for all components
  – Explore scale-dependent combined with system-dependent ensemble covariance localization
Motivation: Earth System Prediction

- Operational forecast models increasingly coupled (2016 Annual Seminar; ECCC global forecasts coupled with Ice-Ocean since November, 2017)

- Benefits from coupled forecasts even at shorter time-scales relevant for medium-range NWP, related to:
  - Tropical convection,
  - Hurricanes, extra-tropical storms
  - Coastal upwelling,
  - Sea ice (polynyas, leads)

- Additional benefits from providing operational ice-ocean forecasts and services – may require new collaborations (e.g. Canadian Ice Service)

- Initialization of these models from independent assimilation systems for each component a challenge
The Need for Coupled Atmosphere-Ice-Ocean Prediction

ECCC requires ice-ocean forecasts and information services for:

• Improved weather prediction
  – Timescales from days to seasons, due to...
  – Sea ice, tropical cyclones, surface interactions

• Sea ice prediction
  – Improved automated analyses and forecasts for the Canadian Ice Service – to complement manual ice chart analyses
  – Identify/predict high pressure areas dangerous for ships

• Emergency response
  – Comprehensive trajectory modelling capacity
  – E.g. dispersion of pollutants
Global Deterministic Prediction System (GDPS)

- **GEM atmospheric model**
  - ECCC's model for global and regional operational forecasts
  - Coupled GDPS 10 day forecasts: atmosphere-ocean-ice (coupled in operations since November 2017)

- **4D-EnVar data assimilation**
  - Variational approach using 4D ensemble covariances from EnKF
  - Hybrid covariances by averaging the ensemble covariances with the static NMC-method covariances

- **Data assimilated by the GDPS:**
  - Radiosondes, Aircraft
  - Surface report (Land, Ship, Buoys)
  - AMSU-A/MHS/ATMS/SSMIS
  - AIRS/IASI/CrIS/Geo-Radiances
  - ASCAT
  - AMVs
  - GPS-RO, ground-based GPS
Ice-Ocean Modelling and Data Assimilation with

- Global Ice-Ocean Prediction System (GIOPS), NEMO-CICE coupled model
  - Used also for Seasonal forecasting
- Produces daily ice-ocean analyses and 10 day forecasts
- Mercator Ocean Assimilation System (SAM2):
  - Sea surface temperature assimilated daily
  - Temperature and salinity profiles weekly
  - Sea level anomaly from satellite altimeters weekly
- 3DVar Ice analysis (6-hourly):
  - SSM/I, SSM/IS, ASCAT, AVHRR
  - CIS charts and image analyses
- SST OI analysis (daily):
  - in situ data, AVHRR, AMSR-E, ATSR
  - foundation SST
  - background: previous day analysis
Coupled Forecasts for Typhoon Neoguri

96h forecasts, valid 00Z, July 10, 2014

SST from drifter obs. Much better agreement of cold wake in Coupled forecasts

Smith et al., 2018, MWR
Coupled Operational Global Forecasts

- **Coupled model:**
  - Atm: GEM 25km
  - Ocean: NEMO-ORCA025 (1/4°)
  - Ice: CICE4
  - Uncoupled DA

- **Evaluation:**
  - 10 day forecasts 15 Jun–31 Aug, 2014
  - Significant forecast improvements over most areas
  - Shown: 850hPa geopotential height versus ERA-Interim

Smith et al., 2018, MWR
The Need for Earth System DA: Example

• Schematic of ECCC coupled global forecast initialization:

- **Atm analysis**
  - 6-hourly 4D-EnVar

- **Land Sfc analysis**
  - 6-hourly OI

- **Sea Ice analysis**
  - 6-hourly 3D-Var

- **SST analysis**
  - Daily OI

- **Daily ocean analysis**

- **Weekly ocean analysis**

- **Coupled Model**

  - Atm
  - Land Sfc
  - Sea Ice
  - Ocean

Adjustments to enforce physical consistency
The Need for Earth System DA: Example

- In principle, coupled DA would be simpler (in practice ?)

**Coupled DA:** 6-hourly 4D-EnVar

- Atm
- Land Sfc
- Sea Ice
- Ocean (incl. SST)

**Coupled Model:**

- Atm
- Land Sfc
- Sea Ice
- Ocean
The Need for Coupled Earth System DA

• Several challenges in initializing coupled models could be handled more directly with coupled assimilation methods.

• Better treatment of physical consistency between component systems:
  – Analysis updates to sea-ice and ocean temperature, consistency essential for even short-term sea ice forecasts.
  – Background errors of near-surface atmosphere can be highly correlated with ocean/land surface errors.

• Accounting for background error correlations between component systems allows observations of one component to correct another.

• Consistent assimilation of "coupled" observations:
  – Many surface-sensitive satellite observations used for extracting sea ice and ocean information also sensitive to atmosphere (e.g. estimating sea ice concentration with an RTM, Scott et al. 2012).
  – Location of sea-ice edge affects selection/usage of surface-sensitive atmospheric and ocean observations.
Example: Atm-Ocean, Atm-Land Coupling

- Background error correlation of near-surface air temperature and surface skin temperature:
- Computed over July 2018 from 48h-24h coupled forecasts (NMC method)
- High positive correlations over land during daytime (many land surface DA systems use atm obs)
- Over ocean: small scale variability, generally positive over north Pacific and Atlantic
Outline

• Motivation: The need for Earth System Data Assimilation

• Methods:
  – DA Methods used for NWP, including hybrid methods
  – DA Methods for other geophysical components: e.g. Sea ice
  – Methods for coupled DA

• Examples:
  – First attempts at coupled Atmosphere-Ice-Ocean DA at ECCC

• Strategy for Earth System DA at ECCC
DA Methods

• For NWP, DA systems based on either:
  – Variational data assimilation (3D-Var, 4D-Var, 4D-Var with hybrid cov., 4D-EnVar, etc.); or
  – Ensemble Kalman Filter (perturbed obs EnKF, EnSRF, LETKF, ensemble of Var's etc.)

• Variational approaches typically used for deterministic prediction, EnKF for ensemble prediction

• For other geophysical DA systems, more variety of methods still used:
  – Optimal Interpolation (OI)
  – Diffusion operator for spatial error correlations of ice and ocean
  – Static SEEK filter (SAM2 ocean DA)
  – Some use persistence of previous analysis as background state (e.g. SST and sea ice analysis systems at ECCC)
Hybrid Methods for NWP DA

- Many flavours of "hybrid" DA approaches:
  - Variational analysis used to recenter the EnKF ensemble (see next few slides)
  - EnKF ensemble used to partially specify background-error covariances in variational systems (especially 4D-EnVar)

- These hybrid systems combine the strengths of both approaches:
  - Variational approaches efficient for producing a single analysis (deterministic, ensemble mean) by assimilating large number of observations and flexible treatment of error covariances
  - EnKF efficient for producing a large ensemble of analyses by assimilating moderately large number of observations, but treatment of error covariances typically more restricted
  - Dual-resolution incremental approach: analysis increment (and ensemble covariances) at lower resolution than deterministic model
Hybrid DA: Operational NWP at ECCC

<table>
<thead>
<tr>
<th></th>
<th>Deterministic</th>
<th>Ensemble</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global</td>
<td>GDPS</td>
<td>GEPS</td>
</tr>
<tr>
<td></td>
<td>4DEnVar $\Delta x_a=39\text{km}$</td>
<td>$\text{EnKF 256-member}$</td>
</tr>
<tr>
<td></td>
<td>$\Delta x=25\text{km}$</td>
<td>$\Delta x_a=39\text{km}$</td>
</tr>
<tr>
<td>Regional</td>
<td>RDPS</td>
<td>REPS</td>
</tr>
<tr>
<td></td>
<td>4DEnVar $\Delta x_a=39\text{km}$</td>
<td>$\Delta x=39\text{km}$</td>
</tr>
<tr>
<td></td>
<td>$\Delta x=10\text{km}$</td>
<td>$\Delta x=39\text{km}$</td>
</tr>
<tr>
<td>Convective-scale (experimental)</td>
<td>HRDPS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\Delta x=2.5\text{km}$</td>
<td>$\Delta x=15\text{km}$</td>
</tr>
</tbody>
</table>

Soon to implement a regional EnKF for initializing regional ensemble forecasts and providing ensemble covariances to regional 4D-EnVar
Hybrid DA: 4D-EnVar for Ensemble Mean
Buehner et al., 2017, MWR

- A simple approach for incorporating more observations in the EnKF with little added cost by using 4D-EnVar to update the ensemble mean and the EnKF to update the perturbations:
  \[ x_k^a = x_k^b + \Delta \bar{x}^a_{\text{envar}} + \Delta x_k^{a'}_{\text{enkf}} \]

- 4D-EnVar has nearly identical configuration as deterministic system
- 4D-EnVar uses direct spatial localization of \( B \) matrix instead of \( BH^T \) or indirectly through \( R \)
- Some centers instead recenter ensemble on deterministic analysis
Hybrid DA: 4D-EnVar for Ensemble Mean

Control member forecasts (deterministic forecast from mean analysis)

72h global forecasts

Using EnVar with all GDPS obs to only update the ensemble mean gives significant improvements for control member vs. Current EnKF
Hybrid Methods for Earth System DA

• Need to consider cost and complexity of expanding such DA systems to directly include other Earth system components:
  – 4D-Var requires coding and maintaining TLM/AD versions of each component model, linearization for some geophysical models challenging due to nonlinearities (e.g. sea ice rheology)
  – EnKF requires large ensemble size (~100 members) to estimate error covariances → lower resolution than deterministic model, not straightforward for ocean/sea-ice (e.g. Arctic archipelago)
  – Additional effort and expertise required to maintain separate DA algorithms and software for each system component
Outline

• Motivation: The need for Earth System Data Assimilation

• Methods:
  – DA Methods used for NWP, including hybrid methods
  – DA Methods for other geophysical components: e.g. Sea ice
  – Methods for coupled DA

• Examples:
  – First attempts at coupled Atmosphere-Ice-Ocean DA at ECCC

• Strategy for Earth System DA at ECCC
Example: ECCC Regional/Global Ice Concentration Analyses (Buehner et al. 2016)

- Regional: ~5 km ; Global: ~10 km resolution
- 4 analyses per day
- background = analysis 6 hours earlier
- total ice concentration (3DVar) and error stddev estimate (simple Kalman filter)
- observations assimilated:
  - CIS ice charts, lake bulletins
  - SSM/I, SSM/IS, AMSR2
  - ASCAT
  - AVHRR (ice/water)
- background error correlations modelled with diffusion operator
- ice is removed where SST > 4°C
- ice field is “corrected” where estimated analysis-error stddev is high

1768 × 1618 grid points
Ice Analysis: Passive Microwave Data
SSMI, SSMIS, AMSR2

• Assimilation:
  • Total ice concentration estimated from NASA Team 2 retrieval algorithm
  • Use "footprint" observation operator that aggregates gridded ice concentration over footprint of instrument

• Quality control - reject data when:
  • Surface Air Temperature > 0°C (melt ponds)
  • Retrieved ice concentration is not zero AND
    • Sea Surface Temperature (SST) is above 4°C OR
    • Historical Frequency of Occurrence of ice is 0 OR
    • Wind speed > 25 knots (Wind filter)
Ice Analysis: Observation Footprints

- Footprint observation operator important for combining information from sensors with such different resolutions
- Observation rejected if footprint touches land, removing most low resolution obs near coast and in narrow channels
Ice Analysis: Impact of Quality Control
Example: July 8, 2007

Without QC

With QC

Necessary due to use of simple ice retrieval algorithm that does not model effect of surface melt ponds
Ice Analysis: Effect of Wind Filter

Necessary due to use of simple ice retrieval algorithm that does not model effect of wind on ocean emissivity.
Correction where $\sigma_a \geq 0.6$
Assimilation of AVHRR Ice/Water Observations

- Cape Farewell, 21 June 2013
- High resolution needed near coastlines, in narrow channels
- Limited coverage due to cloud cover, lack of daylight in winter
SAR Ice/Water Compared with IMS Product
Progress developing a SAR ice/water retrieval algorithm

One of the worst cases: Labrador Sea, May 3, 2013
Retrievals every 5km

Ice is highly dynamic
Many disagreements likely due to low temporal resolution of IMS and difficulty identifying edge

No retrieval produced in areas with high uncertainty

Ice agrees with IMS
Water agrees with IMS
Ice disagrees with IMS
Water disagrees with IMS
Outline

• Motivation: The need for Earth System Data Assimilation
• Methods:
  – DA Methods used for NWP, including hybrid methods
  – DA Methods for other geophysical components: e.g. Sea ice
  – Methods for coupled DA
• Examples:
  – First attempts at coupled Atmosphere-Ice-Ocean DA at ECCC
• Strategy for Earth System DA at ECCC
Cycling Strategies for Earth System DA

- Uncoupled DA:
  
  ![Diagram showing cycling strategies for Earth System DA](image)

- In reality, not possible to be fully uncoupled:
  - Models: land sfc, sea ice, ocean need atm forcing, and vice versa
  - DA: e.g. atmosphere needs SST and sea ice information
  - Some sort of ad hoc coupling needed both for models and DA, but insufficient to ensure physical consistency between components
Cycling Strategies for Earth System DA

- **Weakly coupled DA:**
  - Atm DA
  - Land Sfc DA
  - Sea Ice DA
  - Ocean DA
  - Coupled Model:
    - Atm
    - Land Sfc
    - Sea Ice
    - Ocean

- Physical consistency generated during coupled short-term forecast
- DA may degrade consistency near the component interfaces and not make optimal use of observations affected by multiple components; also difficult to include interactions between analysis systems
Cycling Strategies for Earth System DA

- Strongly coupled DA:

  - DA for all components within the same system allows for coupling between the background errors and within the observation operators.
  - Requires unified procedures: Same DA algorithm, same DA frequency, and likely within same piece of software.
  - Other possibilities: coupling through 4D-Var outer loop (ECMWF)
Outline

- Motivation: The need for Earth System Data Assimilation
- Methods:
  - DA Methods used for NWP, including hybrid methods
  - DA Methods for other geophysical components
  - Methods for coupled DA
- Examples:
  - First attempts at coupled Atmosphere-Ice-Ocean DA at ECCC
- Strategy for Earth System DA at ECCC
Current Operational Uncoupled DA

Models see the same SST and Ice analyses, but otherwise independent, do not evolve in GEM; SAM2 assimilates SST analysis (no diurnal cycle)

assimilation cycle for day j

Atmos → EnVar → GEM → EnVar → GEM → EnVar → GEM → EnVar → GEM

SST OI using data of day (j-2)

Ice 3DVar

Ocean → SAM2 (daily) → NEMO + CICE

18 UTC → 00 UTC → 06 UTC → 12 UTC → 18 UTC
First Weakly Coupled Atm-Ice-Ocean DA

Work of Sergey Skachko

Independent ocean and atmosphere DA; common coupled background state for atmospheric EnVar and oceanic SAM2 DA.

**Diagram:**
- Atmosphere DA: EnVar → GEM → EnVar → GEM → EnVar → GEM
- SST DA: SST OI using data of day (j-1) → Ice 3DVar
- Ocean DA: SAM2 (daily) → NEMO + CICE → NEMO + CICE → NEMO + CICE

**Time Stamps:**
- 18 UTC
- 00 UTC
- 06 UTC
- 12 UTC
- 18 UTC

**Notes:**
- SST and Ice DA cycles still use previous analysis as the background state.
Forecasts from Weakly Coupled DA

Work of Sergey Skachko

Difference of atmospheric temperature StdDev with respect to own analyses as a function of lead time. Region: Northern Extratropics

Coupled forecasts from uncoupled analyses vs. uncoupled forecasts.

Coupled forecasts from weakly coupled analyses vs. uncoupled forecasts.

Coupled forecasts less consistent with own analyses for near-sfc temperature due to use of uncoupled ocean analyses.

Analyses where 4D-EnVar sees the model SST are more consistent with the coupled model forecasts.
Forecasts from Weakly Coupled DA

Work of Sergey Skachko

Atmospheric 1000hPa temperature StdDev and bias with respect to mean analyses as a function of forecast lead time.

Coupled forecasts from Weakly coupled analyses vs. Uncoupled analyses.

Weakly coupled DA produces no significant change in coupled forecast temperature Bias (dashed) or StdDev (solid)
Impact of Coupling on SST Errors

Observations-Background Bias and RMS with respect to SST OI analysis

- Generally better agreement from weak coupling with the foundation SST computed using SST OI data assimilation system
- Validation against raw SST data is under development
Outline

• Motivation: The need for Earth System Data Assimilation

• Methods:
  – DA Methods used for NWP, including hybrid methods
  – DA Methods for other geophysical components
  – Methods for coupled DA

• Examples:
  – First attempts at coupled Atmosphere-Ice-Ocean DA at ECCC

• Strategy for Earth System DA at ECCC
  – Use of highly modular common DA software for all components
  – Explore scale-dependent combined with system-dependent ensemble covariance localization
Strategy: Towards Strongly Coupled DA

• Starting with the deterministic prediction system, migrate all DA systems into a common modular software (MIDAS):
  – Sea Ice 3D-Var
  – SST Optimal Interpolation – implemented as 3D-Var
  – Daily Ocean SEEK filter – implemented as 3D-Var or EnVar

• Step-wise technical and scientific development:
  – Initially, ensure that stand-alone MIDAS versions of these systems provide similar quality as original systems
  – Make stand-alone systems more consistent: common DA frequency (6-hourly upper ocean analysis, including SST)
  – Make it possible to run all systems within the same execution while allowing for different analysis grids to co-exist for each
  – Scientific work to evaluate benefits of including coupling in DA, both in background errors and observation operators
Strongly Coupled Atm-Ice-Ocean DA

Single 4D-EnVar 6-hourly DA for computing analysis of Atmosphere, Sea Ice and Upper Ocean (including SST)

Atmos → GEM → GEM → GEM → GEM → GEM
EnVar: Atm, Ice, Ocean (SST, Mixed-layer)

Ocean Ice → NEMO + CICE → NEMO + CICE → NEMO + CICE → NEMO + CICE → NEMO + CICE

Atm,Oce,Ice
Strategy: Towards Strongly Coupled DA

• Many benefits expected from using highly modular common software (even before coupled DA)

• Modular software components developed for one system can be easily used in another:
  – Diffusion-based $B$ matrix developed for sea-ice analysis can be used for SST/Upper-ocean analysis
  – Horizontal footprint observation operator developed for sea-ice analysis can be used for atmospheric radiance observations

• By using strongly coupled 4D-EnVar for ensemble mean analysis, may allow transfer of most of the benefit to EnKF

• Lots of interesting science to determine practical methods for estimating and modelling coupled background-error covariances: balance operators, scale/system-dependent ensemble covariance localization and coupling, …
System-dependent covariance localization
Also applies to error covariances when they are prescribed

- Positive-semidefiniteness required for physically realizable correlations
- Large differences between the systems in horizontal localization (or correlations themselves) results in reduction of the between-system covariances (e.g. atm-ocean, atm-ice):

**Same** severe horizontal localization for each system, Cross-correlations can be maintained:
A ≠ B and C ≠ D, so A = C and B = D possible

**Very different** horizontal localization for 2 systems, Impossible to maintain cross-correlations:
A ≠ B and C = D, so A = C and B = D not possible
System-dependent covariance localization

1D Idealized System

- System-dependent homogeneous spatial localization functions (Gaussian) are specified with length scales: 10, 3, and 1.5 grid points.
- Localization of between-system covariances constructed to ensure full matrix is positive-semidefinite: $L_{i,j} = (L_{i,i})^{1/2}(L_{j,j})^{1/2} \rightarrow$ btwn systems i & j.
System-dependent covariance localization

1D Idealized System

- Within-system and between-system localization matrices combined into a single “multi-system” localization matrix
- Between-system blocks have diagonal values less than 1
- Could make more sense to apply scale-dependent localization to multi-system coupled ensemble covariances
- Would be interesting to examine between-system error correlation as a function of horizontal scale
Scale-Dependent Localization: Horizontal Scale Decomposition

- Scale-dependent localization could be a convenient approach for treating between-system covariances when dominant scales differ greatly between systems.
- Coupling may only be significant for those scales for which both systems have significant amounts of variance.
- The same basic concept could be applied to prescribed covariances (with multiple length scales; e.g., NEMOVAR) or balance operator approach.

Spectral filters for decomposing atmospheric covariances with respect to 3 horizontal scale ranges.

Caron and Buehner, 2018, MWR
Scale-Dependent Localization: Horizontal Scale Decomposition

Perturbations for ensemble member #001 – Temperature at ~700hPa

Full

Large scale

Small scale

Medium scale

Caron and Buehner, 2018, MWR
Waveband integrated variances

Large scale
Medium scale
Small scale
All the scales

Scale-dependent localization implicitly creates: variable- and level-dependent localization, would also lead to system-dependent localization

6-h perturbation from 256-member EnKF

Caron and Buehner, 2018, MWR
Scale-Dependent Localization: Impact in single observation DA experiments

700 hPa T observation at the center of Hurricane Gonzalo (October 2014)

Normalized temperature increments (correlation-like) at 700 hPa resulting from various B matrices.

Caron and Buehner, 2018, MWR
Scale-Dependent Localization: Impact in single observation DA experiments

700 hPa T observation at the center of a High Pressure

Normalized temperature increments (correlation-like) at 700 hPa resulting from various B matrices.

Caron and Buehner, 2018, MWR
Scale-Dependent Localization:

2D Sea Ice Ensemble

- Ensemble of sea ice concentration background fields (60 members, time-lagged ensemble) from the Canadian Regional Ice Prediction System ensemble of 3DVar analyses experiment

Ensemble mean ice concentration

Ensemble spread

Buehner and Shlyaeva, 2015, Tellus
Scale-Dependent Localization:
Example of one ensemble perturbation

Original perturbation

Scale decomposition with a diffusion operator (that accounts for coastlines) instead of a spectral transform
Assimilation of 2 observations

One obs in area dominated by large-scale error, other in area of small-scale error

30km localization

150km localization

Background field and obs

Scale-dep. localization

Buehner and Shlyaeva, 2015, Tellus
Conclusions: Earth System DA

• Many DA systems for NWP moving towards:
  – Use of large 3D or 4D ensembles with various localization approaches (e.g. scale-dependent, flow following, etc.)
  – Use of increasingly high observation count and spatial resolution
  – Use of hybrid approaches to benefit from advantages of each individual method

• Given this complexity of DA for NWP, move towards strongly coupled DA a scientific and technical challenge:
  – Requires flexible/modular unified DA software for all systems
  – Initial step: use same software for independent systems (forces people to work together and develop flexible/modular code)

• Potential benefits from coupled DA:
  – More consistent initial conditions for coupled forecasts
  – Observations of ice/ocean/land could improve atmosphere
  – Account for coupling in observation operators for "coupled" obs