Multiple timescale coupled atmosphere-ocean data assimilation
(for climate prediction & reanalysis)

Robert Tardif
Gregory J. Hakim
University of Washington

w/ contributions from:
Chris Snyder
NCAR
Context: climate forecasting & reanalysis

- Interannual to decadal: External forcing & initial conditions important (Meehl et al. 2009, Hawkins & Sutton 2009)

- Uninitialized hindcasts: skill limited to externally forced variability over continental & larger scales (Sakaguchi et al. 2012)

- Coupled system: fast atmosphere & slow (deep) ocean

Still unclear how to best initialize the coupled system (Meehl et al. 2014)

- Slow has the memory (source of predictability) but much fewer observations than fast

- Requires coherent analyses of fast & slow components

Strongly (& multiscale) coupled DA?
Challenges / overarching questions

- **Coherence** between initial conditions of slow & fast relies on “cross-media” error covariances
  - **Q1**: What do these look like? How to reliably estimate? Fast component is “noisy” (i.e. high-frequencies)...

- Coupled system with wide variety of scales
  - **Q2**: Any benefits of multi-timescale DA?

- Slow has the memory but fewer observations than in fast
  - **Q3**: What role atmospheric obs. in initializing fast & slow components of a poorly observed ocean? ... a one-way coupling perspective
How to efficiently test ideas, prototype & evaluate strategies?

- **Complex Earth system models problematic for such basic research**
  - Extremely expensive, especially for ensemble DA
    (small ensembles & limited experimentation, realizations, etc.)

- **Motivates using simplified approach:**
  - **Low-order analog** of the coupled N. Atlantic climate system
    -> **few state variables**: obtained from comprehensive AOGCM output
  - **Offline** (i.e. “no cycling”) ensemble DA
    -> prior ensemble members drawn from states of long climate simulations
    -> same prior used at every analysis times
      :: uninformed prior (other than climatology of the model)

Cheap: Allows extensive numerical experimentation
Low-order analogue of N. Atlantic coupled system

- **State variables:**
  - **Atmosphere:**
    - MSLP along 40°N transect
      ("NAO" winds -> gyre)
    - Meridional eddy
      heat flux across 40°N
  - **Ocean:**
    - Subpolar upper temperature & salinity
    - AMOC index (max. overturning streamfunction in N. Atlantic)
      [ taken as unobserved! ]

- Data derived by coarse-graining of state-of-the-art AOGCM gridded output
  -> Simplified system but w/ complex underlying (fast/slow) dynamics

- **Monthly** data for above variables as basis for DA experimentation
  -> truth, observations (truth + random noise) & prior
Low-order analogue of N. Atlantic coupled system

• Analogue data derived from:
  o Community Climate System Model version 4 (CCSM4) gridded output from CMIP5 archives
  o 1000-yr “Last Millennium” simulation (pre-industrial natural variability)
Low-order analogue of N. Atlantic coupled system

**AMOC index**
(Max. value of overturning streamfunction in N. Atlantic)

CCSM4 1000-yr last millennium CMIP5 run: maximum AMOC index

Variability with fast & slow time scales

---

**How much of this *unobserved* component of the coupled system can we recover using coupled multiple timescale DA?***
(by assimilating obs from other components of the low-order analogue)

---

International Workshop on Coupled Data Assimilation 2016, Toulouse
(Strongly) Coupled atmosphere-ocean DA

Ensemble Kalman filter:

\[ x_a = x_b + K(y - Hx_b) \]

\[ \begin{aligned} x &= \begin{pmatrix} MSLP \\ . \\ . \\ T \\ S \end{pmatrix} \\ \text{atmos.} \\ \end{aligned} \]

\[ \begin{aligned} y &= \begin{pmatrix} MSLP \\ . \\ T \\ S \end{pmatrix} \\ \text{atmos.} \]

\[ \begin{aligned} \hat{y}_e &\text{: model estimate of obs.} \\ \end{aligned} \]
Coupled atmosphere-ocean DA

- Ensemble DA & cross-media update
  - Assimilation of atmospheric obs. updating the ocean ...

\[
x_a = x_b + K(y - Hx_b)
\]

**Cross-media covariances:**

- \( y^e \): obs. of fast -> noisy
- \( x \): state vector, including slow variables

Fast noise contaminates \( K \)

Consider assimilation of **time-averaged obs.**

=> **Averaging over the noise** -> increase cov. w/ slow component

=> **Increase “observability”** -> reduce obs. error variance (\( R \)) \(~1/\sqrt{N}\)

[Tardif et al. 2014, 2015; Lu et al. 2015]
Time-average DA

- Assimilation of time-averaged observations
  
  \[
  \begin{align*}
  \mathbf{x} &= \overline{\mathbf{x}} + \mathbf{x}' \\
  \mathbf{y} &= \overline{\mathbf{y}} + \mathbf{y}'
  \end{align*}
  \]

  Time averaging & Kalman-filter-update operators linear and commute

  \[
  \begin{aligned}
  \overline{\mathbf{x}}_a &= \overline{\mathbf{x}}_b + K_A (\overline{\mathbf{y}} - H\overline{\mathbf{x}}_b) \\
  K_A &= \overline{\mathbf{x}}_b \mathbf{y}_e^T [\mathbf{y}_e \mathbf{y}_e^T + \mathbf{R}]^{-1}
  \end{aligned}
  \]

  Time-mean:

  \[
  \begin{aligned}
  \mathbf{x}'_a &= \mathbf{x}'_b + K_P (\overline{\mathbf{y}} - H\overline{\mathbf{x}}_b) \\
  K_P &= \mathbf{x}'_b \mathbf{y}_e^T [\mathbf{y}_e \mathbf{y}_e^T + \mathbf{R}]^{-1}
  \end{aligned}
  \]

  Deviations:

  \[
  \begin{aligned}
  \mathbf{x}'_b \mathbf{y}_e^T \approx 0 \rightarrow \mathbf{x}'_a &= \mathbf{x}'_b \\
  \text{just update time-mean}
  \end{aligned}
  \]

  Full state:

  \[
  \mathbf{x}_a = \overline{\mathbf{x}}_a + \mathbf{x}'_b
  \]

  (Dirren & Hakim 2005; Huntley & Hakim 2010)
Multiple timescale DA

- Assimilate obs. at “appropriate” time scale

\[ x_b = \bar{x}_b^{T1} + x'_b \]

\[ \bar{x}_a^{T1} = \bar{x}_b^{T1} + K_A (\bar{y}_1^{T1} - H \bar{x}_b^{T1}) \]

\[ x_a = \bar{x}_a^{T1} + x'_b \]

\[ x_b = \bar{x}_b^{T2} + x'_b \]

\[ \bar{x}_a^{T2} = \bar{x}_b^{T2} + K_A (\bar{y}_2^{T2} - H \bar{x}_b^{T2}) \]

\[ x_a = \bar{x}_a^{T2} + x'_b \]

If at last step (i.e. \( y_2 \) shares same time scale as \( x_b \)): \( x'_b = 0 \)

End product is final analysis at time scale \( \tau_2 \)
CDA experiments

- **Ensemble square root filter** (Whitaker & Hamill 2002)
- Serial obs. processing
- Low-order system -> no localization
- Offline/no cycling -> no inflation
- “Reanalysis mode” -> all obs. available a-priori

- **Perfect model experiments**, i.e. same model for truth & observations
  -> obs.: random noise added (10% of climatological variance)
- Frequency of obs.: **monthly**

- Generate AMOC analyses over 1000 years
- Run DA experiment w/ **various obs. availability scenarios**
  (atmosphere vs. ocean)
- Consider **2 time scales**: a **slow** ($\tau_1$) and a **fast** ($\tau_2 = $ monthly)
Assimilated obs. vs. time scales

- Covariability w/ **AMOC index** vs **averaging time scale**

![Graph showing covariability with AMOC index](image)

**CCSM4: Covariability with AMOC index**

- MSLP @ 40°N
- Subpolar ocean T, S

Correlation vs **Time scale (year)**:
- Month
- Year
- Decade

Atmospheric meridional eddy heat flux

**Bjerknes compensation**

(?)

(Shaffrey & Sutton 2006)
Single vs. multiple timescale DA

Coupled DA of MSLP (atmosphere) and upper subpolar ocean T, S

Ensemble mean AMOC analyses

2 time scales:
\[ \tau_1 = 20 \text{ yrs} \]
\[ \tau_2 = \text{monthly} \]

International Workshop on Coupled Data Assimilation 2016, Toulouse
Verification metric

Coefficient of efficiency
(Nash and Sutcliffe 1970)

\[ CE = 1 - \frac{\sum_{i=1}^{N} (x_i - x_i^a)^2}{\sum_{i=1}^{N} (x_i - \bar{x})^2} \]

- \( CE \approx 1 \): analysis error variance \(<\) climo. variance
- \( CE = 0 \): no information over climatology
- \( CE < 0 \): Really bad! (…bias)
Verification vs. time scales
(calculated with analyses covering full 1000 yrs)

CE for ensemble mean AMOC analyses

2 time scales:
\( \tau_1 = 20 \) yrs
\( \tau_2 = \text{monthly} \)
Multi-time scale DA: ocean vs. atmosphere-only

2 time scales:

\[ \tau_1 = 50 \text{ yrs} \]
\[ \tau_2 = 1 \text{ month} \]

- monthly only
- 50-yr only
- 50-yr ocean T,S or atmospheric meridional eddy heat flux + monthly MSLP

CE for ensemble mean AMOC analyses

Ocean T,S replaced by atmos. eddy heat flux obs.

Atmosphere-only DA

w/o ocean DA
Multi-time scale DA: ocean vs. atmosphere-only

Atmosphere-only DA: vs. long time scale

- 1 yr + monthly
- 10 yr + monthly
- 20 yr + monthly
- 50 yr + monthly

Averaging interval (yrs)

International Workshop on Coupled Data Assimilation 2016, Toulouse
Toward application to real data...

- **Last Millennium Reanalysis (LMR)**
  - Offline assimilation of paleoclimate data
    - Tree rings
    - Ice core & coral isotope ratios
  - Prior: CCSM4 “Last Millennium”

Takeaways …

• **Q1:** Cross-media covariances, how to reliably estimate?
  
  A: Use **time-averaging** over **appropriate scale**
  
  o Averaging over “noise” in fast atmosphere = > enhances covariability w/ slow ocean

• **Q2:** Benefits from multiple timescale DA approach?
  
  A: Yes!
  
  o More accurate analyses of **fast & slow**
  
  o **Reduced errors** at **intermediate** (~annual) **scales** from DA of monthly & decadal-avg. obs.

• **Q3:** What role atmospheric obs. in initializing ocean’s fast & slow components?
  
  A: Can be **significant**:
  
  o Frequent DA for **fast ocean** component: Fast response to **winds, surface fluxes** etc.
  
  o **Less significant** role for constraining **slow**, if ocean *sufficiently* well-observed
  
  o **Fully coupled DA of time-averaged obs. important** when poorly observed ocean (w/ appropriate choice of assimilated obs.)