Data assimilation in sea ice modeling

François Massonnet
FNRS Research Fellow
Université catholique de Louvain, Belgium
« Back-of-the-envelope » calculation

Sea ice whole state

\[ 50 \times 5000 \times 30 \times 365 = \sim 3.10^9 \text{ data} \]
« Back-of-the-envelope » calculation

Sea ice whole state

\[ 50 \times 5000 \times 30 \times 365 = \sim 3.10^9 \text{ data} \]

Sea ice observations

- Concentration
  \[ 2 \times 5000 \times 30 \times 365 = \sim 1.10^8 \text{ data} \]
- Drift
### Sea ice whole state

<table>
<thead>
<tr>
<th></th>
<th>Variables</th>
<th>Grid Points</th>
<th>Years</th>
<th>Days</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concentration</td>
<td>2</td>
<td>5000</td>
<td>30</td>
<td>365</td>
<td>$\sim 3 \times 10^9$ data</td>
</tr>
<tr>
<td>Drift</td>
<td>1</td>
<td>5000</td>
<td>30</td>
<td>365</td>
<td>$\sim 3 \times 10^9$ data</td>
</tr>
</tbody>
</table>

### Sea ice observations

<table>
<thead>
<tr>
<th></th>
<th>Variables</th>
<th>Grid Points</th>
<th>Years</th>
<th>Days</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concentration</td>
<td>2</td>
<td>5000</td>
<td>30</td>
<td>365</td>
<td>$\sim 1 \times 10^8$ data</td>
</tr>
<tr>
<td>Drift</td>
<td>1</td>
<td>5000</td>
<td>30</td>
<td>365</td>
<td>$\sim 1 \times 10^8$ data</td>
</tr>
<tr>
<td>Thickness</td>
<td>1</td>
<td>5000</td>
<td>10</td>
<td>60</td>
<td>$\sim 3 \times 10^5$ data</td>
</tr>
</tbody>
</table>
Puzzle 1: Antarctic sea ice area is increasing (in a warming world)

Observed changes (1980-2008) in Antarctic sea ice concentration

[data: Comiso et Nishio, JGR, 2008]
Puzzle 2: Seasonal sea ice prediction has random pattern of success

Sea Ice Outlook
Predictions of September Arctic sea ice extent

[adapted from Stroeve et al., GRL, 2014]
Puzzle 2: Seasonal sea ice prediction has random pattern of success

Sea Ice Outlook
Predictions of September Arctic sea ice extent

[adapted from Stroeve et al., GRL, 2014]
Sea ice observations and models are complementary

<table>
<thead>
<tr>
<th>Observations</th>
<th>Models</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incomplete coverage</td>
<td>Complete coverage</td>
</tr>
<tr>
<td>No predictive skill</td>
<td>Predictive skill</td>
</tr>
<tr>
<td>Proxies for reality</td>
<td>Approximations of reality</td>
</tr>
<tr>
<td>Uncertainties</td>
<td>Uncertainties, systematic biases</td>
</tr>
</tbody>
</table>
Sea ice observations and models are complementary

<table>
<thead>
<tr>
<th>Observations</th>
<th>Models</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incomplete coverage</td>
<td>Complete coverage</td>
</tr>
<tr>
<td>No predictive skill</td>
<td>Predictive skill</td>
</tr>
<tr>
<td>Proxies for reality</td>
<td>Approximations of reality</td>
</tr>
<tr>
<td>Uncertainties</td>
<td>Uncertainties, systematic biases</td>
</tr>
</tbody>
</table>

Data assimilation
Data assimilation in sea ice modeling

1. The problem: estimation of the whole sea ice state

2. The limitations: importance of hypotheses

3. The applications: three examples in sea ice modeling
1. The problem: estimation of the whole sea ice state

2. The limitations: importance of hypotheses

3. The applications: three examples in sea ice modeling
NCEP/NCAR 2-m temperatures, 10-m winds
Climatological precipitations, relative humidity and clouds
Just as any model, NEMO-LIM exhibits biases

The model physics can be improved
Rheology, ocean—sea ice interactions, snow are subject of intense research

Atmospheric reanalyses are not perfect, either
Antarctic sea ice concentration trends suspicious in the western seas

All model parameters were tuned by « trial-and-error »
Parameter space may have been underexplored
Observations alone are not sufficient to estimate the whole sea ice state

<table>
<thead>
<tr>
<th></th>
<th>Spatial sampling</th>
<th>Temporal sampling</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concentration</td>
<td>~25 km, everywhere</td>
<td>~daily, 1979-2014</td>
</tr>
<tr>
<td>Thickness</td>
<td>~25 km (tracks, more in Arctic)</td>
<td>~daily (seasonal, intermittent, 2004-2014)</td>
</tr>
<tr>
<td>Drift (large-scale)</td>
<td>~25 km, central Arctic</td>
<td>~daily, 2007-2014</td>
</tr>
<tr>
<td>Deformation</td>
<td>~10 km, central Arctic</td>
<td>~daily (winter, since 1995)</td>
</tr>
<tr>
<td>Snow, melt ponds</td>
<td>very limited</td>
<td>very limited</td>
</tr>
</tbody>
</table>
The ensemble Kalman filter is a data assimilation method

\[ x^a = x^f + K \cdot (d - H x^f) \]
The ensemble Kalman filter is a multivariate data assimilation method.

Analysis: \[ x^a = x^f + K \cdot (d - Hx^f) \]

Kalman gain: \[ K = PH^T (HPH^T + R)^{-1} \]

Model forecast: Observations

[Evensen, 2003]
The ensemble Kalman filter is a sequential ensemble multivariate data assimilation method.

\[
\begin{align*}
    \mathbf{x}^a &= \mathbf{x}^f + K \cdot (d - H \mathbf{x}^f) \\
    K &= P H^T (H P H^T + R)^{-1}
\end{align*}
\]

[Evensen, 2003]
ARE YOU DREAMING?
Data assimilation in sea ice modeling

1. The problem: estimation of the whole sea ice state
   The EnKF is a sequential, ensemble, multivariate data assimilation method

2. The limitations: importance of hypotheses

3. The applications: three examples in sea ice modeling
Data assimilation in sea ice modeling

1. The problem: estimation of the whole sea ice state
   The EnKF is a sequential, ensemble, multivariate data assimilation method

2. The limitations: importance of hypotheses

3. The applications: three examples in sea ice modeling
Not all the hypotheses to reach optimal analysis are fulfilled

1. « The forward model is linear »
   → run it for short time periods (~10 days)
Not all the hypotheses to reach optimal analysis are fulfilled

1. « The forward model is linear »
   → run it for short time periods (~10 days)

2. « Model and obs. errors are uncorrelated »
   → difficult to check, but not impossible
Not all the hypotheses to reach optimal analysis are fulfilled

1. « The forward model is linear »
   → run it for short time periods (~10 days)

2. « Model and obs. errors are uncorrelated »
   → difficult to check, but not impossible

3. « Model and obs. errors are centered around zero »
   → certainly not the case (model has biases)
Not all the hypotheses to reach optimal analysis are fulfilled

1. « The forward model is linear »
   $\rightarrow$ run it for short time periods (~10 days)

2. « Model and obs. errors are uncorrelated »
   $\rightarrow$ difficult to check, but not impossible

3. « Model and obs. errors are centered around zero »
   $\rightarrow$ certainly not the case (model has biases)

4. « The sample model error covariance matrix is a good proxy for model error structure »
   $\rightarrow$ certainly not the case (25 members, only the atmosphere is perturbed)
Can data assimilation provide a consistent and optimal estimation of the whole sea ice state?
Can data assimilation provide a consistent and optimal estimation of the whole sea ice state?

Even if the approximate solution is known to be sub-optimal, can data assimilation help us out to estimate the whole sea ice state?
\[ Q_i = -k \frac{\partial T_i}{\partial z} \]

\[ T(t) = T_i(t, z) \]

\[ Q_OF(t) = \varepsilon \sigma T(t)^4 \]

\[ (1 - \alpha(t))Q_{SW}(t) \]
Model bias is not a major issue for estimating the observed variable.
Model bias is more problematic for initialization of predictions.
Model bias may lead to physical instabilities in the first time steps.
Statistical under-sampling may lead to weak constraints on the non-observed variables.
How are these approximations reflected in the large-scale setup?
The ensemble Kalman filter is a multivariate data assimilation method.

Example of an update in sea surface salinity:
- Analysis saltier than forecast
- Analysis fresher than forecast
In a large-scale setup, the covariances are space- and time-dependent!

**Correlation** between ice concentration and thickness, in an ensemble of 25 members

![Maps showing correlation between ice concentration and thickness on 26th March 2000 and 7th September 2000. The color scale indicates correlation values, with red indicating positive correlation and blue indicating negative correlation. The p-value is marked as greater than 5% for a two-sided test.]
Data assimilation in sea ice modeling

1. The problem: estimation of the whole sea ice state
   The EnKF is a sequential, ensemble, multivariate data assimilation method

2. The limitations: importance of hypotheses
   Not all hypotheses are fulfilled, but the approximate solution can still be satisfactory

3. The applications: three examples in sea ice modeling
Data assimilation in sea ice modeling

1. The problem: estimation of the whole sea ice state
   The EnKF is a sequential, ensemble, multivariate data assimilation method

2. The limitations: importance of hypotheses
   Not all hypotheses are fulfilled, but the approximate solution can still be satisfactory

3. The applications: three examples in sea ice modeling
Observations of Antarctic thickness are too sparse to understand past variability

Distribution of 81 ship cruises between 1981 and 2005

[Worby et al., JGR, 2008]
Antarctic sea ice thickness has lower bias after assimilation of ice concentration.

Mean *bias* in simulated thickness:

- **Weddell Sea**
  - FREE RUN
  - ASSIM
- **Indian Ocean**
- **Pacific Ocean**
- **Ross Sea**
- **Amundsen and Bellingshausen seas**
- **Southern Ocean**

[Antarctic thickness data: Worby et al., JGR, 2008]
State estimation: reconstruction of Antarctic sea ice thickness

Sea ice thickness trends (1980-2008)

[Massonnet et al., Ocean Modell., 2013]
Seasonal «prediction» for 2007 (atmosphere known)

March ice thickness  
September ice concentration
Seasonal « prediction » for 2007 (atmosphere known)

March ice thickness

September ice concentration
Initialization from sea ice concentration improves seasonal Arctic predictions

Error forecast September concentration

RMSE [%]

[Massonnet et al., submitted]
clc; clear all; close all

g=9.81; % accélération de la gravité
h0=0.34; % hauteur initiale du niveau d'eau
dt=0.1; % pas de temps
tf=30; % durée de la simulation

h=zeros(length(0:dt:tf),1) % h(t), à trouver

... 

alpha=1.34 % Coefficient de bidouillage

... 

for t=1:dt:tf
    [a,b,c]=compute_gain(h(t-1))
    ...

clc; clear all; close all

g=9.81; % accélération de la gravité
h0=0.34; % hauteur initiale du niveau d'eau
dt=0.1; % pas de temps
tf=30; % durée de la simulation

h=zeros(length(0:dt:tf),1) % h(t), à trouver

... % Coefficient de bidouillage

alpha=1.34

... % Coefficient de bidouillage

for t=1:dt:tf
    [a,b,c]=compute_gain(h(t-1))
    ...

\[ \vec{F}_{\text{air/ice}} + \vec{F}_{\text{ocean/ice}} + \vec{F}_{\text{ice/ice}} \approx 0 \]

\[ C_a \quad C_w \quad P^* \]

Sea ice drift (model)
\[ \vec{F}_{\text{air/ice}} + \vec{F}_{\text{ocean/ice}} + \vec{F}_{\text{ice/ice}} \approx 0 \]

Sea ice drift (observed)

Sea ice drift (model)
Parameter estimation is a state-augmented data assimilation problem.

**Correlation ($P^*$, ice speed)**

$x = p$-value $> 5\%$ (2-sided test)
\[ \tilde{F}_{\text{air/ice}} + \tilde{F}_{\text{ocean/ice}} + \tilde{F}_{\text{ice/ice}} \approx 0 \]
\[ \vec{F}_{air/ice} + \vec{F}_{ocean/ice} + \vec{F}_{ice/ice} \approx 0 \]

Sea ice drift (observed)

Sea ice drift (model, parameters calibrated)
Estimating *more* parameters is not a guarantee for better solution

Distribution of Arctic sea ice speeds (fall-winter, 2007-2012)
Estimating more parameters is not a guarantee for better solution

\[
\begin{align*}
\tau_a + F_{\text{int}} &= 0 \quad (\text{regime 1, compact ice}) \\
\tau_a + \tau_w &= 0 \quad (\text{regime 2, free drift})
\end{align*}
\]

[budget analysis based on Steele et al., JGR, 1997]
Data assimilation in sea ice modeling

1. The problem: estimation of the whole sea ice state
   The EnKF is a sequential, ensemble, multivariate data assimilation method

2. The limitations: importance of hypotheses
   Not all hypotheses are fulfilled, but the approximate solution can still be satisfactory

3. The applications: three examples in sea ice modeling
   Sea ice data assimilation helps understand past variability, improve short term predictions and calibrate model parameters
Lessons learned

Models or observations alone are not sufficient to address key questions in the polar regions

Even if it is suboptimal, the solution returned by the ensemble Kalman filter has generally an added value

It’s never too late to start playing with very simple models
Thank you!

francois.massonnet@uclouvain.be

www.climate.be/u/fmasson

@FMassonnet
LIM2: no ice thickness distribution
LIM3: ice thickness distribution

Relative abundance [%]

Sea ice thickness [m]

0 1 2 3 4 5

14% 32% 19% 12% 22%
Ensemble spread, restartability and limitations

The distribution of ensemble members should reflect the full model uncertainty

* 25 members with perturbed atmospheric forcing (winds/2m-air temperature)

* Localization [Sakov and Bertino, 2010]
The tricky part: updating the other variable
The tricky part: updating the other variable
Primitive nudging may bring the system into a non-physical state.
Multivariate data assimilation accounts for model state covariance.

20th April 00
The ensemble Kalman filter is a multivariate data assimilation method.

Example of an update in sea surface salinity:
- Analysis saltier than forecast.
- Analysis fresher than forecast.
The ensemble Kalman filter is a forecast-analysis method.
Ensemble spread, restartability and limitations

A «sanity check» for the model is necessary because gaussianity assumption is rarely fulfilled

> Reset negative ice concentrations/thickness to zero
> Bound total ice concentration by 1
> Ice thickness stays within category bounds
Observation (satellite)

Ice concentration [%]

7 September 2000
7 septembre 2000

Observation (satellite)

Model (×25)

Ice concentration [%]

Ice thickness [%]
The ensemble Kalman filter approximates the model error covariance matrix with a finite number of particles.
Simulating the right spread in the marginal ice zone is challenging.
Simulating the right spread in the marginal ice zone is challenging.