Ensemble Kalman Filtering Based on Potential Vorticity for Atmospheric Multi-scale Data Assimilation

Tadashi TSUYUKI
Meteorological College, JMA
Contents

- Introduction
- Method of PV-based EnKF
- Data assimilation experiments using a shallow water model
- Summary
Miura et al. (2007)

Watanabe and Kawatani (2018)
Problems with the nesting

- Since observational data generally have information on both slow and fast variables (or large and small scales), part of the information is discarded.

- A small-area model cannot utilize the information of observational data outside of the model domain in data assimilation.

- The feedback of small-scale observational data to a large-area model is absent unless two-way nesting is adopted.
Multi-scale data assimilation

- Multi-scale data assimilation is free from those problems with the nesting, since observational data of slow and fast variables are simultaneously assimilated into a high-resolution model.

- A problem is that observational data of fast variables are usually available only in limited regions. Then, spurious forecast error correlations between slow and fast variables in EnKF may degrade the analysis of slow variables by assimilating observations of fast variables.
EnKF in coupled Lorenz96 model

**Table 1.** The rms of the error of the free run (FREE) and the five assimilation experiments described in the text

| EXP  | $\text{rms}_X$ | $\text{rms}_Y$ |
|------|---------------|---------------|
| FREE | 6.18          | 0.41          |
| EN1  | 0.60          | 0.05          |
| EN2  | 0.47          | 0.29          |
| EN3  | 0.65          | 0.39          |
| CI1  | 0.48          | 0.02          |
| CI3  | 0.48          | 0.27          |

- Assimilation of all observations (EN1)
- Assimilation of $X$ observations (EN2)
- Assimilation of all $X$ but oversampled $Y$ observations (EN3)

Ballabrera-Poy et al. (2009)
Even if fast variables are strongly controlled by slow variables, forecast error correlations between them should be very weak due to low predictability of fast variables.

\[ x_{\text{slow}}^f : S_1 \quad > \quad S_2 \]
\[ x_{\text{fast}}^f : w_1 \quad > \quad w_2 \]
\[ x_{\text{slow}}^f : S_1 \quad > \quad S_2 \]
\[ x_{\text{fast}}^f : w_3 \quad > \quad w_4 \]
Method of PV-based EnKF

i. Potential vorticity (PV) inversion of large-scale PV anomaly to decompose state variables into slow and fast variables.

ii. Mass adjustment to suppress possible additional sampling errors introduced by the decomposition of state variables.

iii. Approximation to Kalman gain to suppress spurious forecast error correlations between slow and fast variables.
Potential vorticity

- **Conservation of PV**

\[
\frac{\partial P}{\partial t} + \mathbf{v} \cdot \nabla P = S \text{ (diabatic heating and subgrid scales)}
\]
\[
\approx 0 \text{ (in the free atmosphere)}.
\]

- **PV inversion**

  - \(x(r, t)\): state variables.
  
  - \(x_b(r, t)\): balanced part, \(x_u(r, t)\): unbalanced part.

\[
P[x_b] = P[x] \text{ and balance conditions,}
\]
\[
x_b = x_b[P], \quad x_u = x - x_b.
\]
Potential vorticity – cont’d

- Slow and fast variables

\[
\frac{\partial x_b}{\partial t} = \frac{\delta x_b}{\delta P} \frac{\partial P}{\partial t} \approx - \frac{\delta x_b}{\delta P} \mathbf{v} \cdot \nabla P,
\]

\[
\therefore \left| \frac{\partial x_b}{\partial t} \right| \lesssim \left| \frac{\delta x_b}{\delta P} \right| \cdot \| \mathbf{v} \cdot \nabla P \| \sim O \left( \frac{1}{L} \right) .
\]

\((L : \text{horizontal scale of PV})\)

- Slow variables can be derived by PV inversion of large scales. Spatial smoothing of \(x\) is therefore necessary before PV inversion.

- The fastness of \(x_u\) depend on the methods of PV inversion.
Methods of PV inversion

- Bolin-Charney balance (Davis and Emanuel, 1991):

\[
\nabla^2 \phi_b = \nabla \cdot f \nabla \psi_b + 2 \left[ \frac{\partial^2 \psi_b}{\partial x^2} \frac{\partial^2 \psi_b}{\partial y^2} - \left( \frac{\partial^2 \psi_b}{\partial x \partial y} \right)^2 \right],
\]

\[
\nu_b = \left( -\frac{\partial \psi_b}{\partial y}, \frac{\partial \psi_b}{\partial x} \right).
\]

- Hierarchies of PV inversion methods (McIntyre and Norton, 2000). The first-order direct inversion uses the Bolin-Charney balance.

- Optimal PV balance (Viudez and Dritschel, 2004).

- Digital filter initialization (Arbogast et al., 2008).
Primitive equation model

Balance model (B-C balance)

QG model

Whitaker (1993)
Inversion by digital filter initialization

\( w \) (4D-Var analysis)

\( w \) (QG theory)

\( w \) (DFI inversion)

Arbogast et al. (2008)
Mass adjustment

- The Bolin-Charney balance suggests that forecast error correlations between the area-averaged value of balanced mass variable and the balanced horizontal velocity should be very weak:

\[ \langle \Delta \overline{\phi_b} \Delta v_b \rangle \approx 0. \]

However, sampling errors are inevitable in EnKF and this condition generally does not hold.

- Modified mass variables:

\[ \tilde{\phi}_b \equiv \phi_b - \overline{\phi}_b, \quad \tilde{\phi}_u \equiv \phi_u + \overline{\phi}_b, \]

are used instead of the original mass variables. Then, \( \langle \Delta \overline{\tilde{\phi}_b} \Delta v_b \rangle = 0. \)
Approximation to Kalman gain

- Forecast error covariance $P^f = \langle \Delta x^f (\Delta x^f)^T \rangle$ is approximated by

$$P^f \approx P^f_b + P^f_u,$$

where $P^f_b \equiv \langle \Delta x^f_b (\Delta x^f_b)^T \rangle$, $P^f_u \equiv \langle \Delta x^f_u (\Delta x^f_u)^T \rangle$ to suppress spurious error correlations between slow and fast variables.

- Then, Kalman gain is approximated by

$$K \approx K_b + K_u,$$

where

$$K_b \equiv P^f_b H^T (R + H P^f H^T)^{-1},$$

$$K_u \equiv P^f_u H^T (R + H P^f H^T)^{-1}.$$
Approximation to Kalman gain – cont’d

- For **LETKF**, the following Kalman gains are used:
  \[
  K_b = P_b^f H^T (R^{-1} - R^{-1} H P^a H^T R^{-1}), \\
  K_u = P_u^f H^T (R^{-1} - R^{-1} H P^a H^T R^{-1}).
  \]

- For **stochastic EnKF** and **serial EnSRF**, in which observations are serially assimilated in batches, the following approximate decomposition is used:
  \[
  x_b^a \approx x_b^f + K_b \left( y^o - H(x^f) \right), \\
  x_u^a \approx x_u^f + K_u \left( y^o - H(x^f) \right).
  \]
Benefits of PV-based EnKF

- Spurious forecast error correlations between slow and fast variables are suppressed, which increases the accuracy of analysis.

- The balanced and unbalanced parts are separately handled, which makes it possible to appropriately address the balance issue of covariance localization.
Covariance localization and balance

- Single $\phi$ observation
  - No localization
  - Localization in $uv\phi$ space
  - Localization in $\psi\chi\phi$ space

Kepert (2009)
Design of experiments

- Shallow water equations on an $f$-plane with periodic boundary conditions

\[
\frac{\partial \zeta}{\partial t} + R_o^{-1} D = -\nabla \cdot \zeta \mathbf{v} + \nu \nabla^6 \zeta, \\
\frac{\partial D}{\partial t} + R_o^{-1} \nabla^2 h - R_o^{-1} \zeta = -\nabla \cdot (\mathbf{v} \cdot \nabla) \mathbf{v} + \nu \nabla^6 D, \\
\frac{\partial h}{\partial t} + R_o^{-1} D = -\nabla \cdot h \mathbf{v} + \nu \nabla^6 h.
\]

\[
\left( R_o \equiv \frac{U}{f_0 r_d}, \quad F_r \equiv \frac{U}{\sqrt{gH}} = R_o, \quad -\pi < x, y \leq \pi \right)
\]

- PV anomaly

\[
\Delta P = \frac{\zeta - h}{1 + R_o h}, \quad \Delta P_{QG} = \zeta - h.
\]
First-order direct inversion

\[
\zeta_b - h_b = \Delta P (1 + R_oh_b),
\]

\[
\nabla^2 h_b = \zeta_b + 2R_o \left[ \frac{\partial^2 \psi_b}{\partial x^2} \frac{\partial^2 \psi_b}{\partial y^2} - \left( \frac{\partial^2 \psi_b}{\partial x \partial y} \right)^2 \right],
\]

\[
\zeta_b = \nabla^2 \psi_b, \quad D_b = 0.
\]

Quasi-geostrophic inversion

\[
\nabla^2 \psi_b - \psi_b = \Delta P_{QG},
\]

\[
h_b = \psi_b, \quad D_b = 0.
\]
Design of experiments – cont’d

- **Numerical model**
  - Spectral model using double Fourier expansion with $k_{max} = l_{max} = 21$ (degrees of freedom: 5 545).
  - # of grid points: $64 \times 64$.
  - $R_o = F_r = 0.5$.

- **Initial condition of true state**
  - Superposition of a balanced zonal jet and an unbalanced height anomaly.

- **Observational data**
  - **Velocity** ($\sigma^o = 0.05$) or **height** ($\sigma^o = 0.02$).
  - Available on $8 \times 8$ grids every unit time.
Data assimilation

- **Stochastic EnKF** with ensemble sizes of 25, 50, 100, 1000 and ten random number sequences.
- Covariance inflation: multiplicative and constant.
- Covariance localization: *uvh space* with the same localization radius between slow and fast variables.
- Initial forecast ensemble: random superposition of the balanced jet and unbalanced height anomaly.

Decomposition of state variables

- **First-order direct inversion** or **quasi-geostrophic inversion** with the approximate decomposition in serial assimilation.
True state (t=0)

Height and velocity (balanced + unbalanced)

PV inversion: first-order direct inversion with spatial smoothing.
True state \((t=0-200)\)

**Height and velocity**
(balanced + unbalanced)

**Height and velocity**
(balanced)

PV inversion: first-order direct inversion with spatial smoothing.
True state (height and velocity)
True state (divergence)
Decomposition of true state (height)

First-order direct inversion (Bolin-Charney balance)

Quasi-geostrophic inversion (geostrophic balance)
Area-averaged value of balanced height

First-order direct inversion (Bolin-Charney balance)

Quasi-geostrophic inversion (geostrophic balance)
PV-based EnKF w/o mass adjustment

Conventional EnKF

RMSE of analysis and bias in height analysis
- Ensemble size: 50
- Obs.: velocity

PV-based EnKF with mass adjustment
Impact of mass adjustment

Observation: velocity

Observation: height
Impact of spatial smoothing

Obs.: velocity

Obs.: height

Period of statistics
RMSE of analysis (obs.: velocity)
- **Ensemble size:** 25
- **Inflation factor:**
  - PV: 1.05
  - CN: 1.075
- **Localization radius:**
  - PV: 10 grids
  - CN: 10 grids

- **Ensemble size:** 50
- **Inflation factor:**
  - PV: 1.05
  - CN: 1.05
- **Localization radius:**
  - PV: 15 grids
  - CN: 15 grids
- Ensemble size: 100
- Inflation factor:
  PV: 1.025
  CN: 1.05
- Localization radius:
  PV: 25 grids
  CN: 25 grids

- Ensemble size: 1000
- Inflation factor:
  PV: 1.01
  CN: 1.01
- No localization
RMSE of analysis (obs.: height)
- **Ensemble size:** 25
- **Inflation factor:**
  - PV: 1.025
  - CN: 1.075
- **Localization radius:**
  - PV: 10 grids
  - CN: 10 grids

- **Ensemble size:** 50
- **Inflation factor:**
  - PV: 1.05
  - CN: 1.05
- **Localization radius:**
  - PV: 20 grids
  - CN: 15 grids
- Ensemble size: 100
- Inflation factor:
  - PV: 1.025
  - CN: 1.025
- Localization radius:
  - PV: 25 grids
  - CN: 20 grids

- Ensemble size: 1000
- Inflation factor:
  - PV: 1.01
  - CN: 1.01
- No localization
Summary

- PV-based EnKF successfully suppresses spurious forecast error correlations between slow and fast variables in multi-scale data assimilation.

- PV-based EnKF improves the accuracy of analysis much more efficiently than increasing the ensemble size. Only 10% more computational cost is needed in the experiments.

- PV-based EnKF makes it possible to appropriately address the balance issue of covariance localization by handling the balanced and unbalanced parts separately.
Remarks on application to a three-dimensional model

- The decomposition of state variables is conducted to the dynamical state variables in the free atmosphere only, and those in the planetary boundary layer and of water substance are handled as in the conventional EnKF.

- Since the large-scale PV anomaly needs to be inverted, the computational cost of PV inversion can be reduced by using a low-resolution model.

- PV inversion methods based on initialization may be recommended, since those methods do not need to solve ill-conditioned elliptic partial differential equations for state variables.
Future work

- To implement PV-based EnKF to a three-dimensional atmospheric model with a high resolution.
- To investigate the performance of multi-scale covariance localization of PV-based EnKF, including that of the localization of the balanced part in PV space.
- To devise a method to extract the “balanced” part from the state variable of water vapor.
Thank you.