Data Assimilation Studies

Shunji KOTSUKI

Postdoctoral Researcher, Data Assimilation Research Team, AICS

Data Assimilation Team Meeting, AICS, Kobe

Studies about data assimilation using Lorenz-96

Lorenz-96 (Lorenz et al., 1998)

For j=1,…,J, $X_j = X_{j+J}$

$$dX_j / dt = \underbrace{\left(X_{j+1} - X_{j-2}\right) X_{j-1}}_{\text{Advection term}} - \underbrace{X_j}_{\text{Dissipation term}} + \underbrace{F}_{\text{Forcing term}}$$

Lorenz et al. assumed time unit as 5 days (the dissipative decay time).

Properties of system deduced without solving the equations.

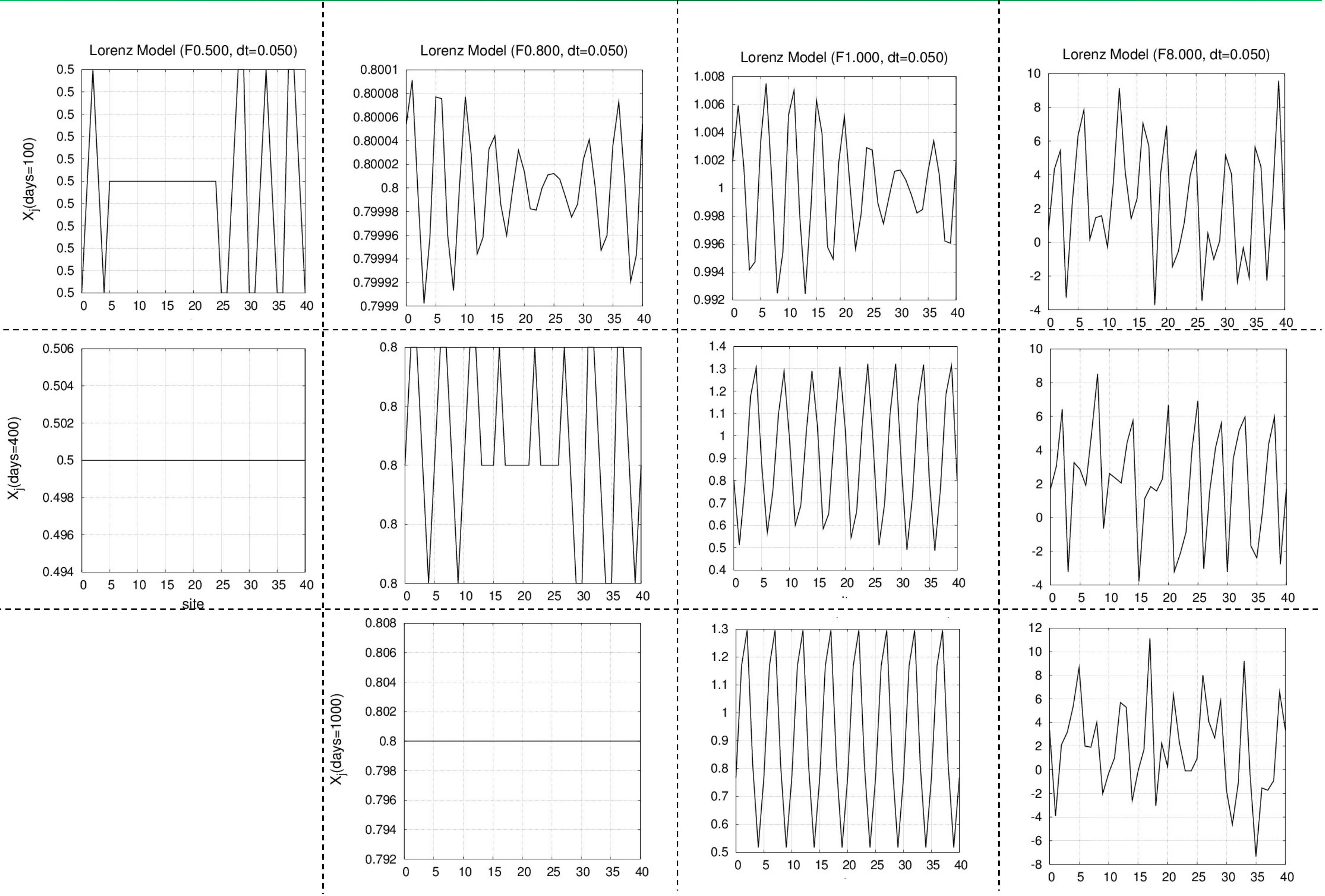
1. When $F > 0.895$, steady solutions is unstable.

Properties of system deduced with solving the equations.

(J:40, dt : 0.05, with 4th order Runge-Kutta scheme)

1. When $F < 4.0$, the perturbations ultimately develop into perfect wave number eight.
2. When $F > 4.0$, the a spatially irregular pattern with chaotic time variations appears.

Ultimate condition (1000 days simulation)



Subject 2: Error propagation analysis

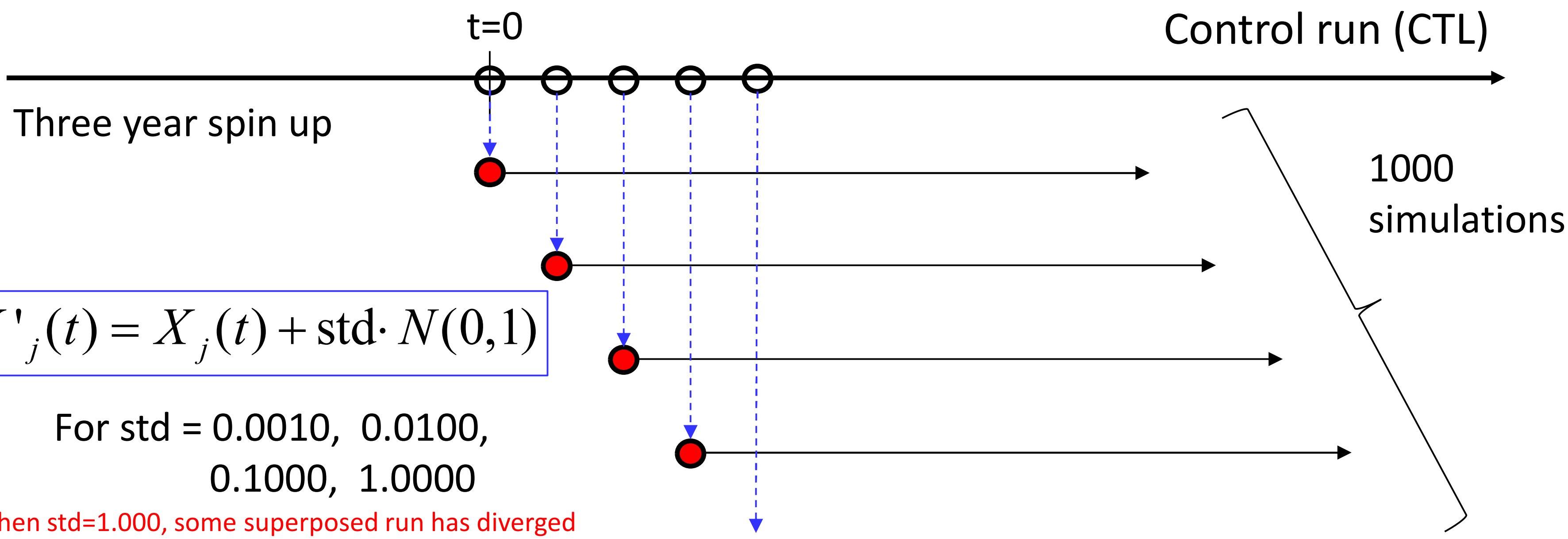
1. Mersenne Twistter method

   : To generate pseudo-random number sampling

2. Box-Muller transformation

   : To generate pairs of normally distributed random numbers

3. Error propagation analysis (average RMSE)

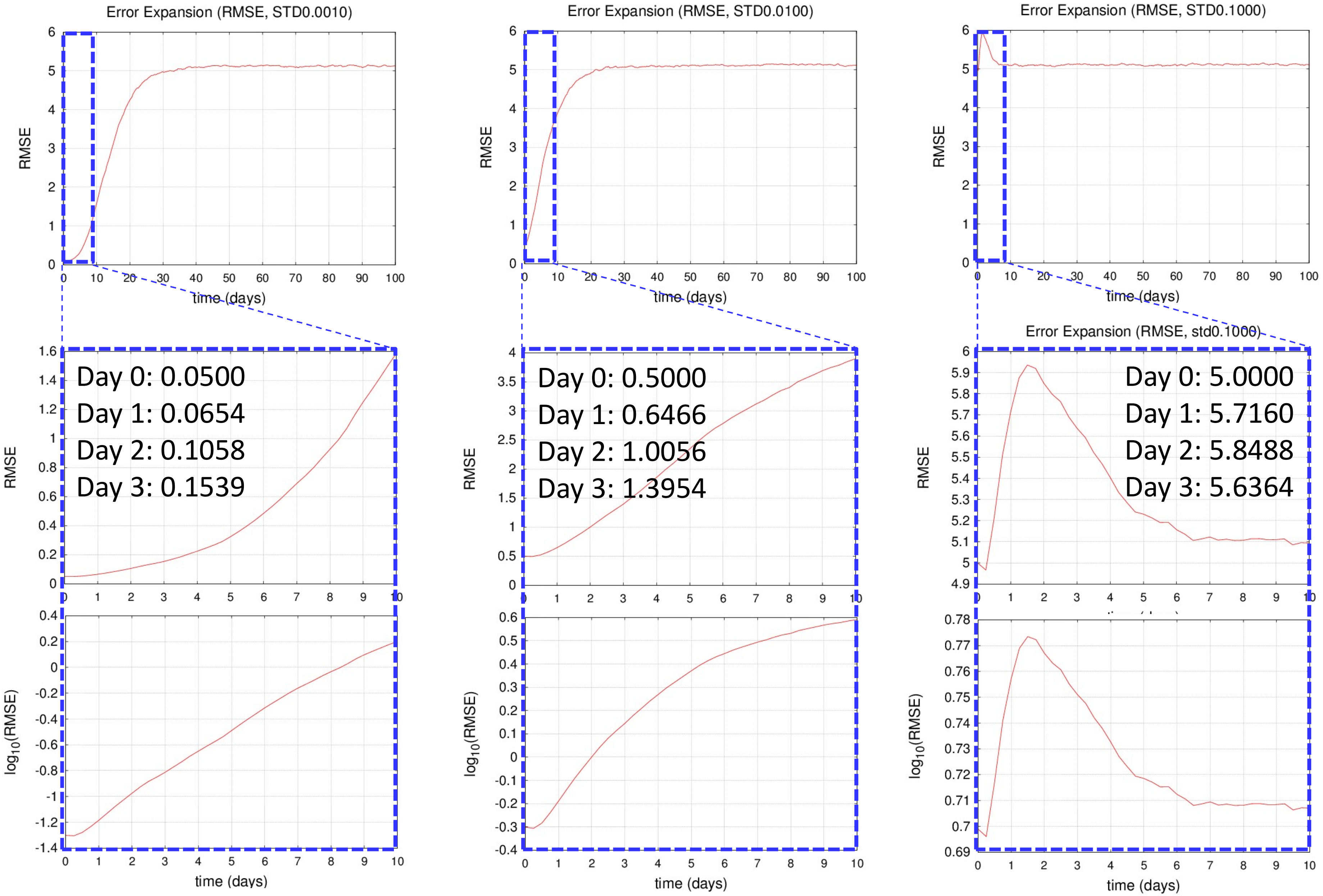


$$X'_j(t) = X_j(t) + \mathrm{std} \cdot N(0,1)$$

For std = 0.0010, 0.0100,
0.1000, 1.0000

When std=1.000, some superposed run has diverged

Subject 2: Error propagation analysis



Subject 3: EKF

Extended Kalman filter (EKF)

$$X_t^f = MX_{t-1}^a$$

$$P_t^f = M' P_{t-1}^a M'^T + Q$$

$$K_t = P_t^f H^T \left[HP_t^f H^T + R \right]^{-1}$$

$$X_t^a = X_t^f + K_t \left[y_t^o - HX_t^f \right]^{-1}$$

$$P_t^a = [I - K_t H] P_t^f$$

Multiplicative covariance inflation

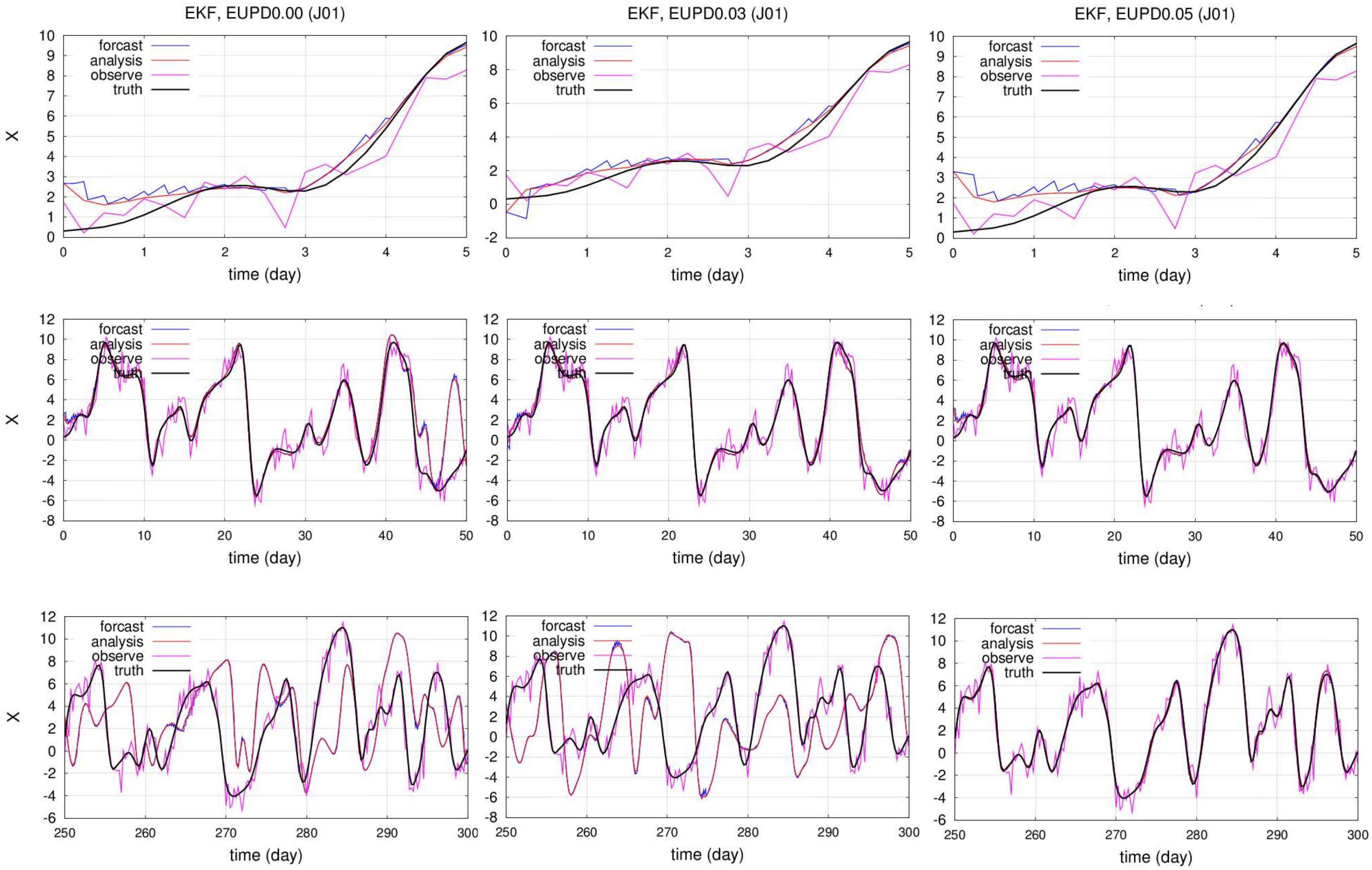
$$P'^a_{t-1} = P^a_{t-1}(1+\delta)$$

Tangent linear method

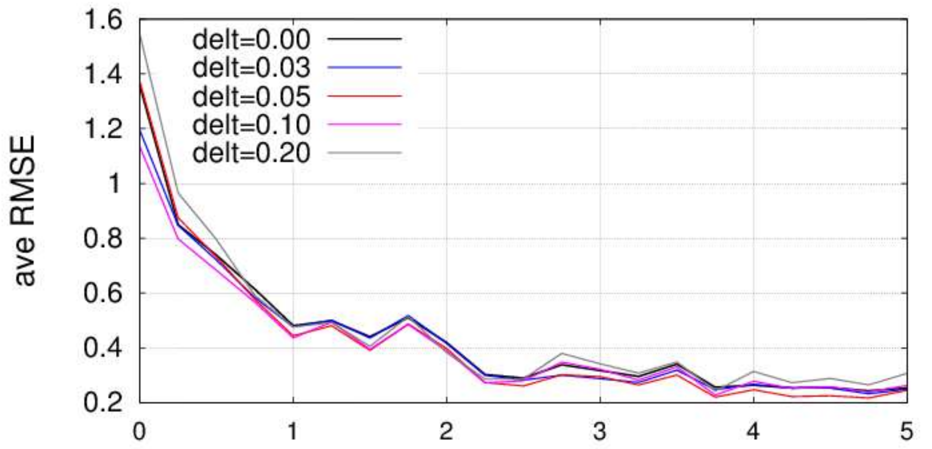
$$M' = \left(\vec{m}_1, \cdots, \vec{m}_j, \cdots, \vec{m}_n \right) \quad \vec{m}_j = \begin{pmatrix} m_{1j} \\ \vdots \\ m_{ij} \\ \vdots \\ m_{nj} \end{pmatrix}$$

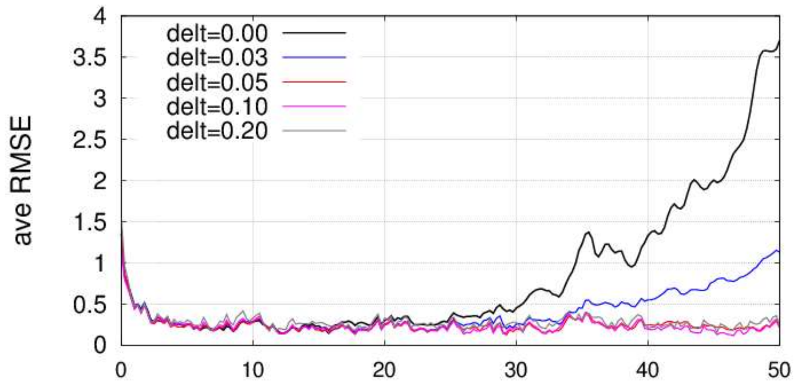
$$\vec{m}_j = \frac{1}{\varepsilon} \left\{ M \begin{pmatrix} x_1^a \\ \vdots \\ x_j^a + \varepsilon \\ \vdots \\ x_n^a \end{pmatrix} - MX^a \right\}$$

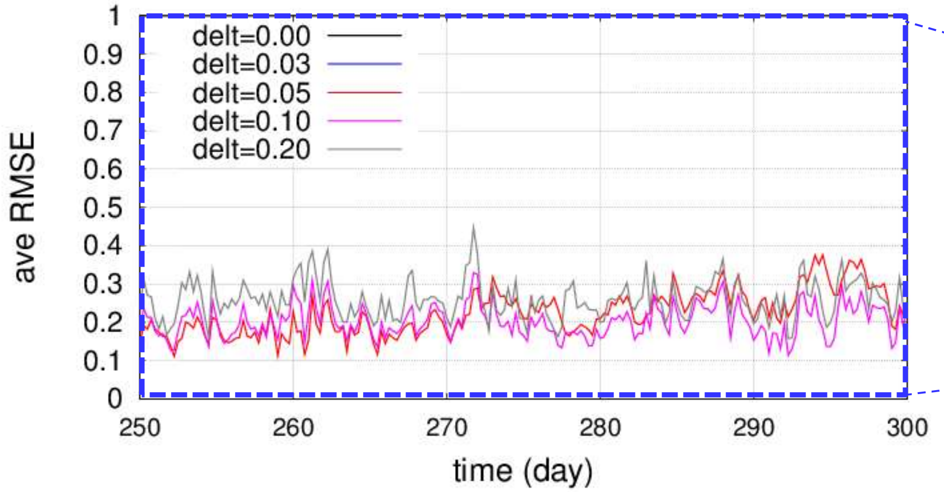
EKF (Full observation)

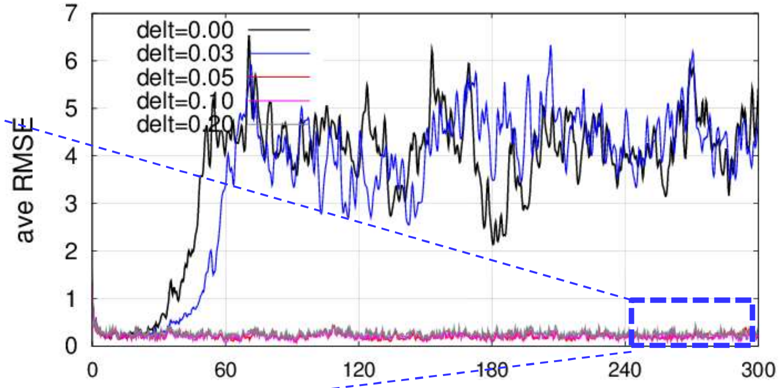


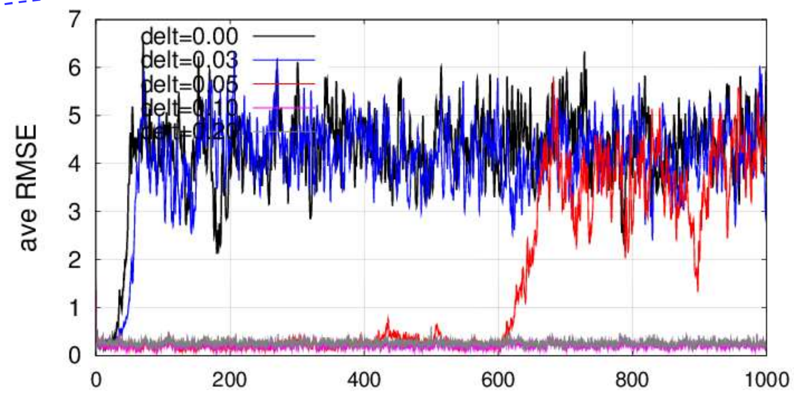
EKF (Full observation)









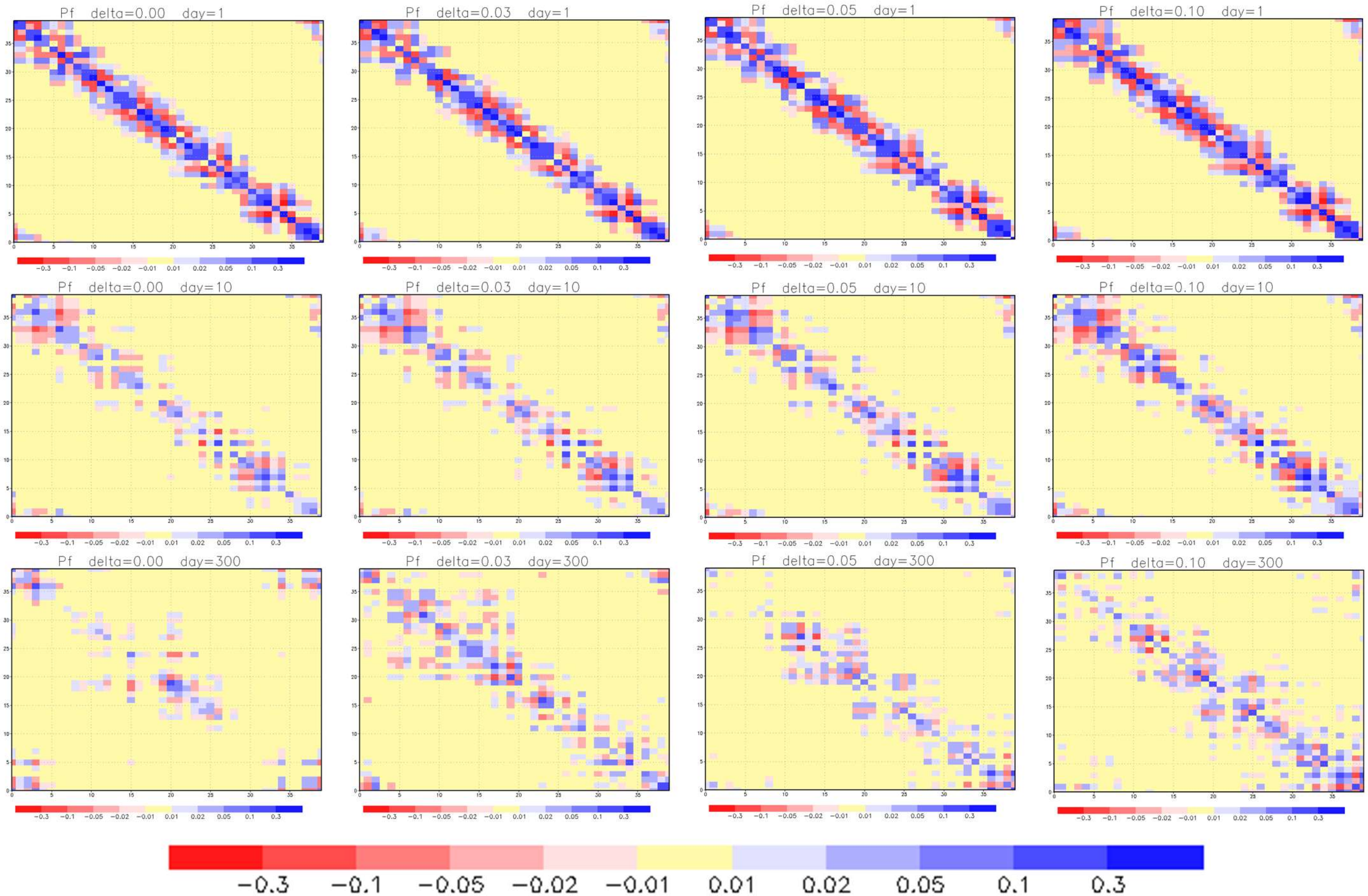


Ave RMSE(from 10th day to 300th day)

δ=0.00 → 3.970 δ=0.03 → 3.970

δ=0.05 → 0.204 δ=0.10 → 0.211

EKF (Full observation)

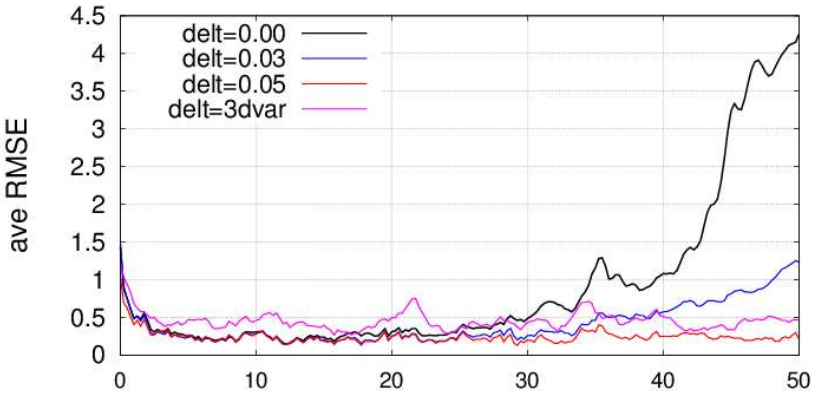


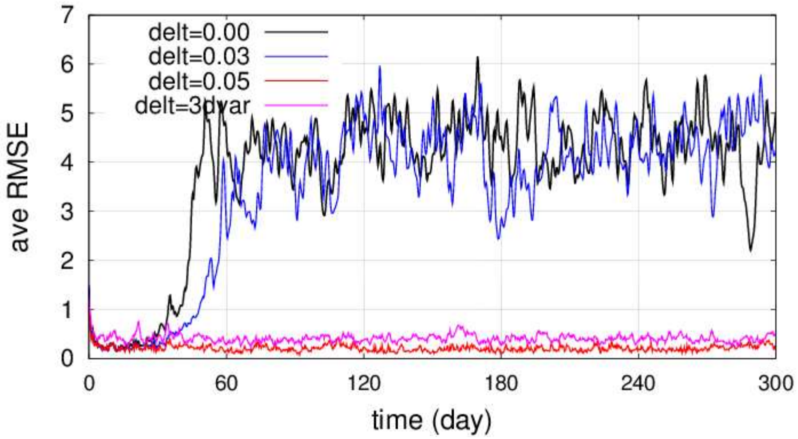
3DVAR (Full observation)

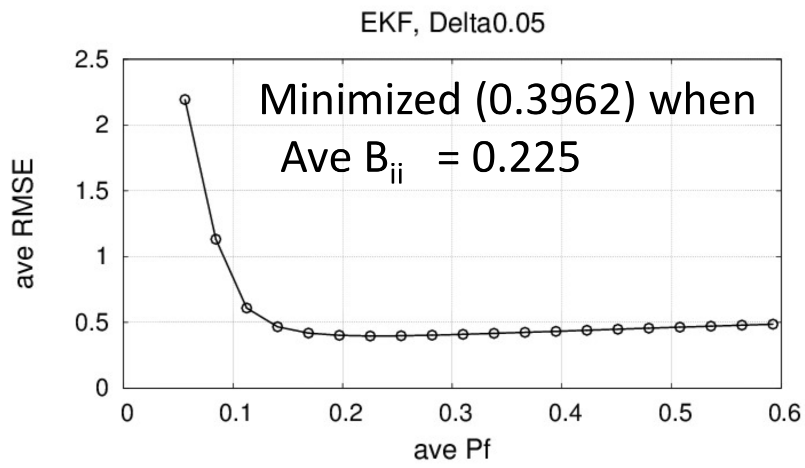
$$X_t^f = MX_{t-1}^a$$
$$K_t = BH^T\left[HBH^T + R\right]^{-1}$$
$$X_t^a = X_t^f + K_t\left[y_t^o - HX_t^f\right]^{-1}$$

B has been defined as diagonal matrix.

$$B = \begin{pmatrix} b_{11} & & & & 0 \\ & \ddots & & & \\ & & b_{ii} & & \\ & & & \ddots & \\ 0 & & & & b_{nn} \end{pmatrix}$$





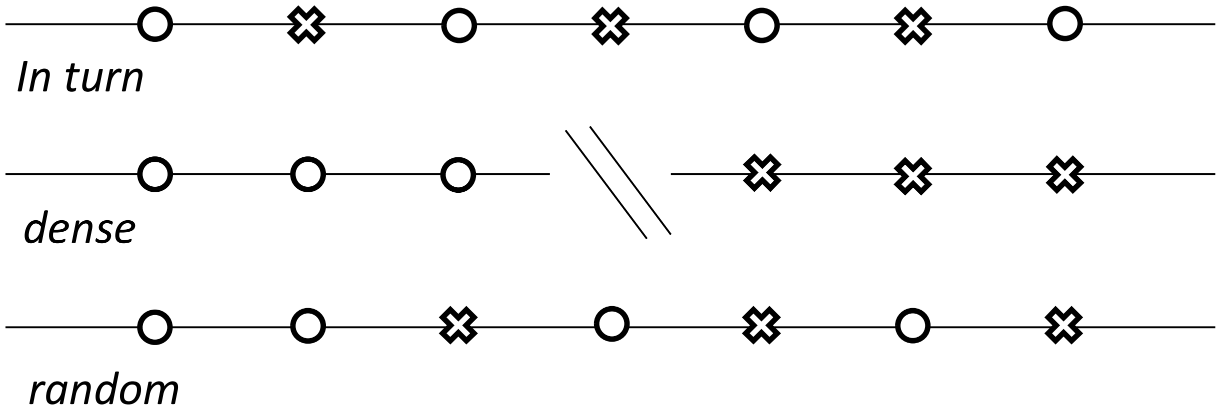


Subject 6: Partial observation

Case 1: Num. Obs. = X

Case 2: Num. Obs. = X

Case 3: Num. Obs. = X



EKF vs 3DVAR

Extended Kalman filter (EKF)	*3DVAR*
$K_t = P_t^f H^T \left[H P_t^f H^T + R \right]^{-1}$	$K_t = B H^T \left[H B H^T + R \right]^{-1}$
$X_t^a = X_t^f + K_t \left[y_t^o - H X_t^f \right]^{-1}$	$X_t^a = X_t^f + K_t \left[y_t^o - H X_t^f \right]^{-1}$

$$P_t^f = \begin{pmatrix} \sigma_{11}^2 & & \sigma_{ji}^2 \\ & \ddots & & \\ & & \sigma_{ii}^2 & & \\ \sigma_{ij}^2 & & & \ddots & \\ & & & & \sigma_{nn}^2 \end{pmatrix} \qquad B = \begin{pmatrix} b_{11} & & 0 \\ & \ddots & & \\ & & b_{ii} & & \\ 0 & & & \ddots & \\ & & & & b_{nn} \end{pmatrix}$$

$$dX_j / dt = \underbrace{\left(X_{j+1} - X_{j-2} \right) X_{j-1}}_{\text{Advection term}} - \underbrace{X_j}_{\text{Dissipation term}} + \underbrace{F}_{\text{Forcing term}}$$

For j=1,…,J, $X_j = X_{j+J}$

- Advection term conserve the total energy defined as $\sum \left(X_j^2 \right) / 2$

Next step

- DA studies
  - Development of EnKF
    - EnSRF, ETKF, LETKF
    - 4DVAR
- To prepare PPT for coming AMS

Appendix

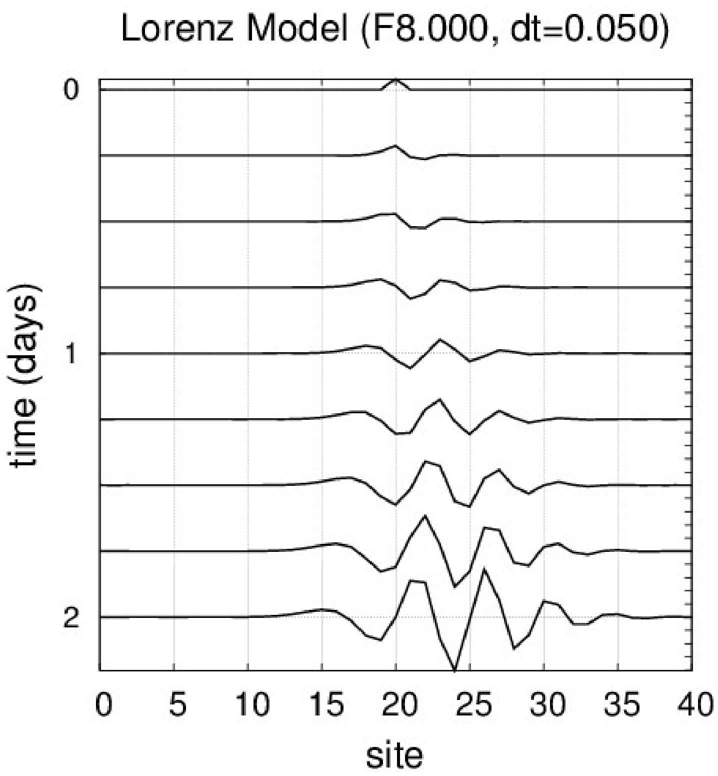
Studies about data assimilation using Lorenz-96

Lorenz-96 (Lorenz et al., 1998)

For j=1,…,J, $X_j = X_{j+J}$

$$dX_j / dt = \underbrace{\left(X_{j+1} - X_{j-2}\right) X_{j-1}}_{\text{Advection term}} \underbrace{- X_j}_{\text{Dissipation term}} + \underbrace{F}_{\text{Forcing term}}$$

Lorenz et al. assumed time unit as 5 days (the dissipative decay time).



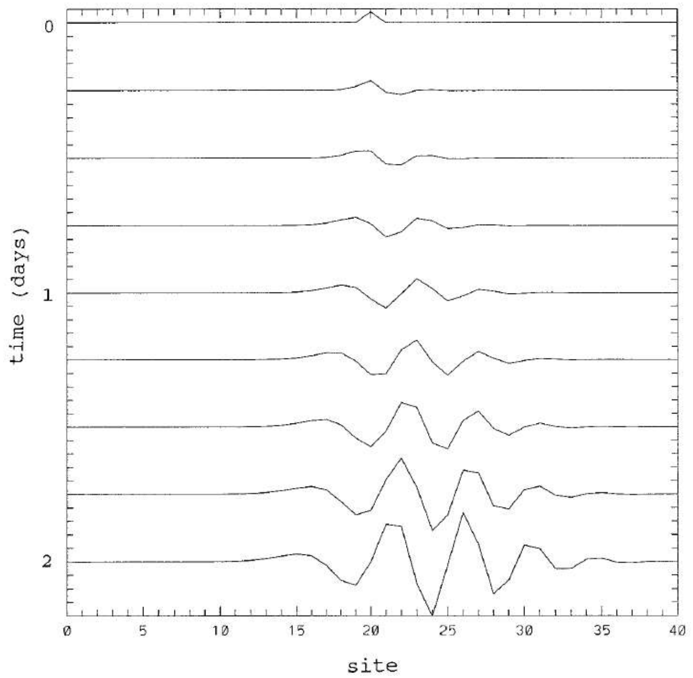
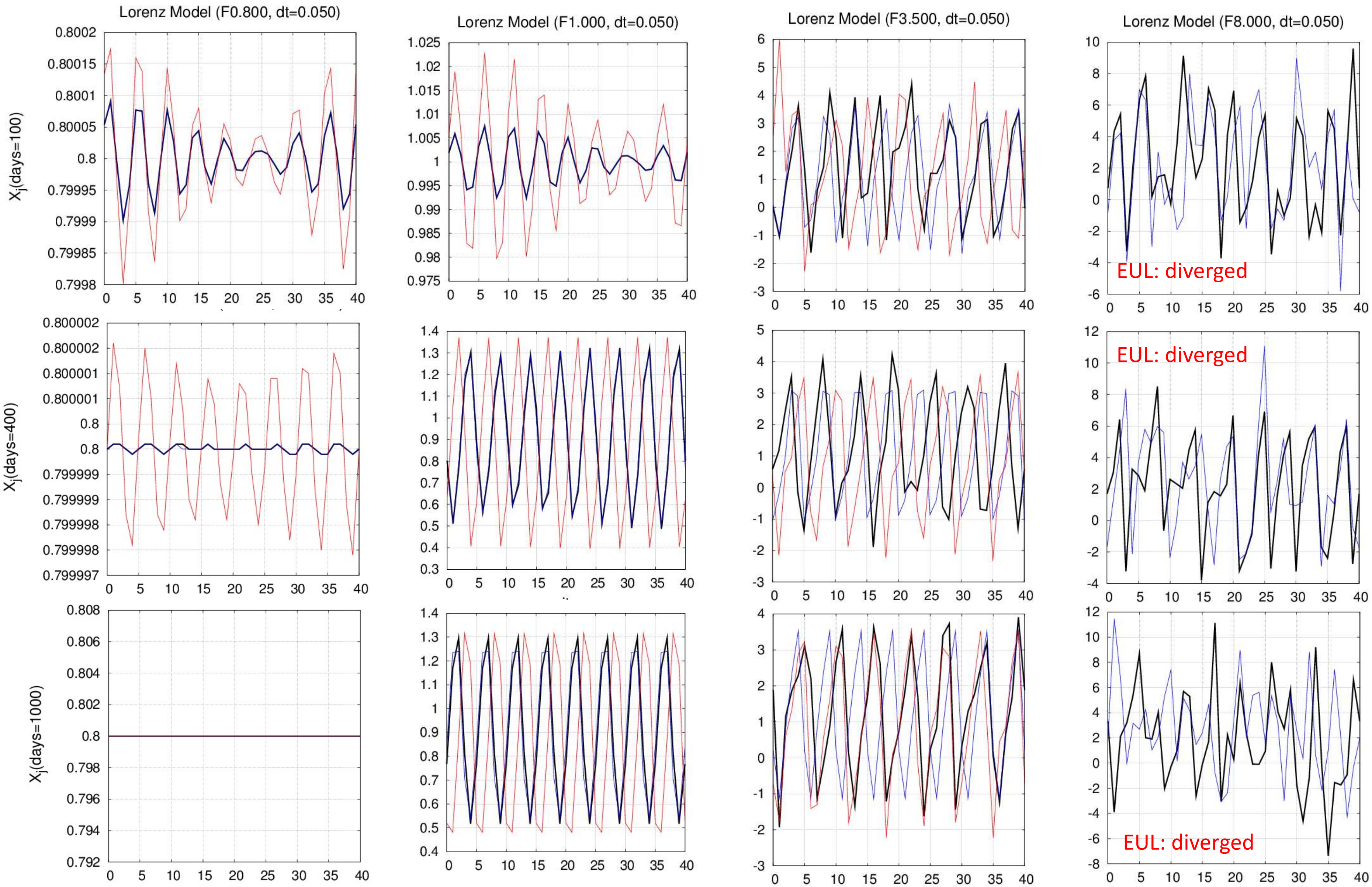


FIG. 1. Longitudinal profiles of X_j at 6-h intervals, as determined by Eq. (1) with $N = 40$ and $F = 8.0$, when initially $X_{20} = F + 0.008$ and $X_j = F$ when $j \neq 20$. On horizontal portion of each curve, $X_j = F$. Interval between successive short marks at left and right is 0.01 units.

Comparison of time integration schemes



—: RK4 (4th order Runge-Kutta), —: RK2 (2nd order Runge-Kutta), —: EUL (Euler method)

Subject 2: Generating random numbers

1. Mersenne Twistter method
   : To generate pseudo-random number sampling

2. Box-Muller transformation
   : To generate pairs of normally distributed random numbers

Box-Muller's method:

When X and Y obey uniform distribution (0,1),
Z_1 and Z_2, defined by following equations,
obey normal distribution N(0,1).

$$Z_1 = \sqrt{-2\log X}\cos 2\pi Y$$

$$Z_2 = \sqrt{-2\log X}\sin 2\pi Y$$