2020.01.30 Data Assimilation Workshop

Application of machine learning methods to model bias correction: Lorenz-96 model experiments

Arata Amemiya*, Takemasa Miyoshi

Data Assimilation Research Team, RIKEN Center for Computational Science

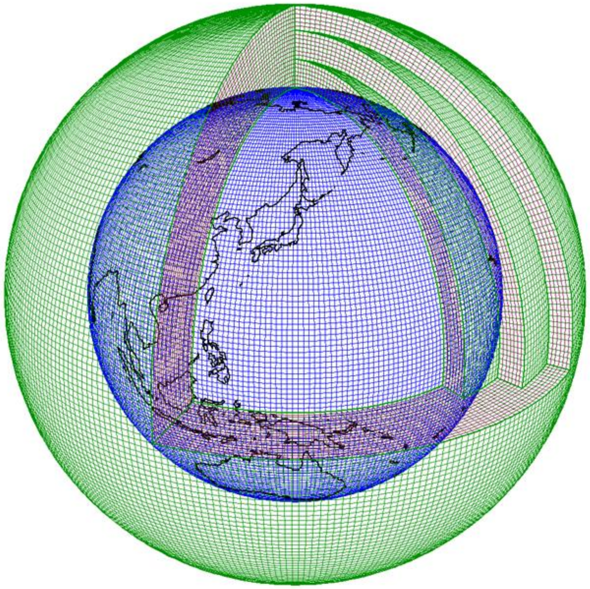
Shlok Mohta

Graduate School of Science, the University of Tokyo

Motivation: bias in weather and climate models

Weather and climate models have model biases from various sources

- Truncation error
- Approximation of unresolved physical processes
  - Convection
  - Small-scale topography
  - Turbulence
  - Cloud microphysics



(From JMA website)

Treatment of forecast error in data assimilation

Kalman filter

Update state and forecast error covariance

$$\boldsymbol{x}_{t+1}^{f} = \boldsymbol{\mathcal{M}}(\boldsymbol{x}_{t}^{a})$$

$$\underline{\boldsymbol{P}_{t+1}^{f} = \boldsymbol{M}\boldsymbol{P}_{t}^{a}\boldsymbol{M}^{T}} \qquad \boldsymbol{M} = \partial\boldsymbol{\mathcal{M}}/\partial\boldsymbol{x}$$

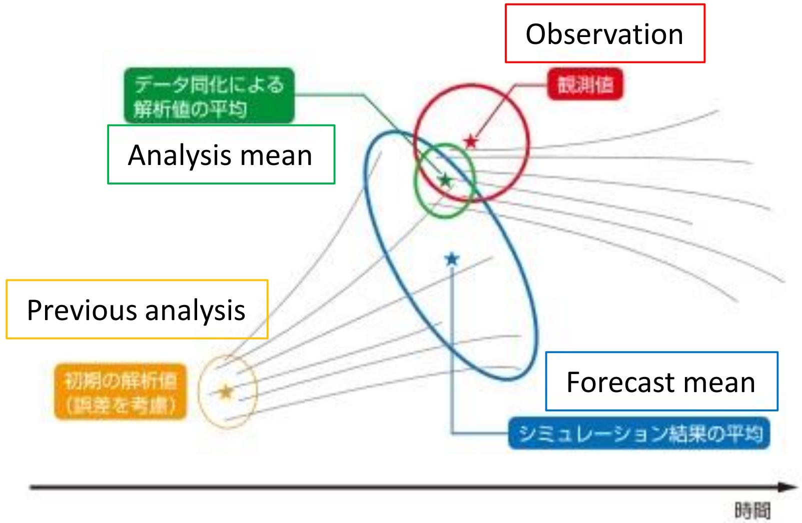
Calculate Kalman gain

$$\boldsymbol{K} = \boldsymbol{P}^{f}\boldsymbol{H}^{T}\left(\boldsymbol{H}\boldsymbol{P}^{f}\boldsymbol{H}^{T} + \boldsymbol{R}\right)^{-1}$$

Calculate analysis state and error covariance

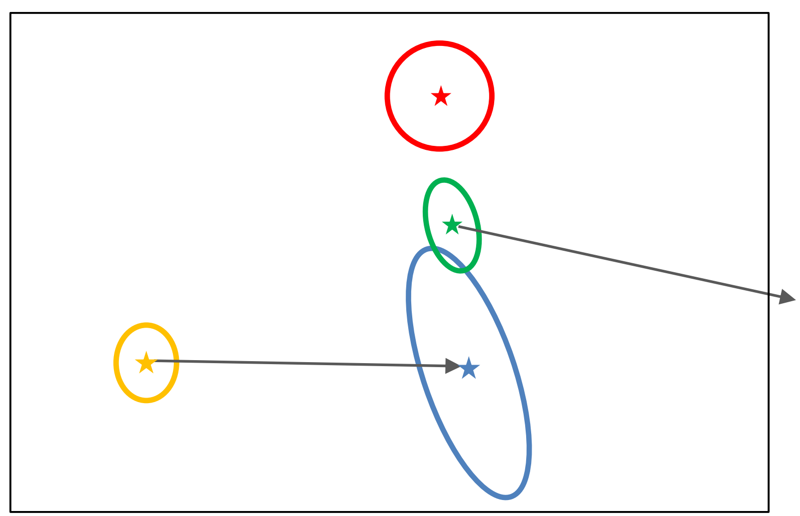
$$\boldsymbol{x}^{a} = \boldsymbol{x}^{f} + \boldsymbol{K}\left(\boldsymbol{y} - H(\boldsymbol{x}^{f})\right)$$

$$\boldsymbol{P}^{a} = (\boldsymbol{I} - \boldsymbol{K}\boldsymbol{H})\boldsymbol{P}^{f}$$



http://www.data-assimilation.riken.jp/jp/research/index.html

Model bias leads to
the underestimation of forecast(background) error



Treatment of imperfect model

Insufficient model error degrades the performance of Kalman filter

1. Covariance inflation

additive inflation

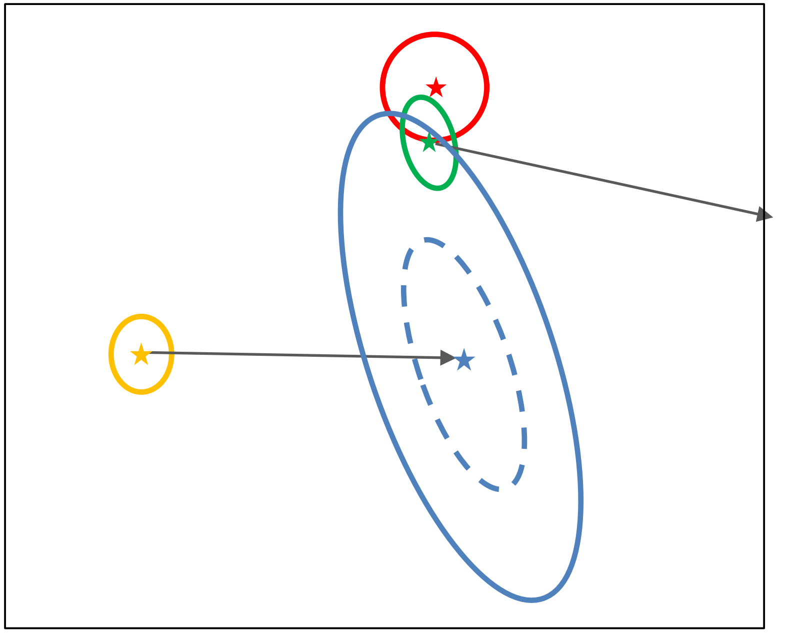
$$\boldsymbol{P}^a \rightarrow \boldsymbol{P}^a + \boldsymbol{Q}$$

multiplicative inflation

$$\boldsymbol{P}^a \rightarrow \alpha\, \boldsymbol{P}^a$$

Relaxation-to-prior

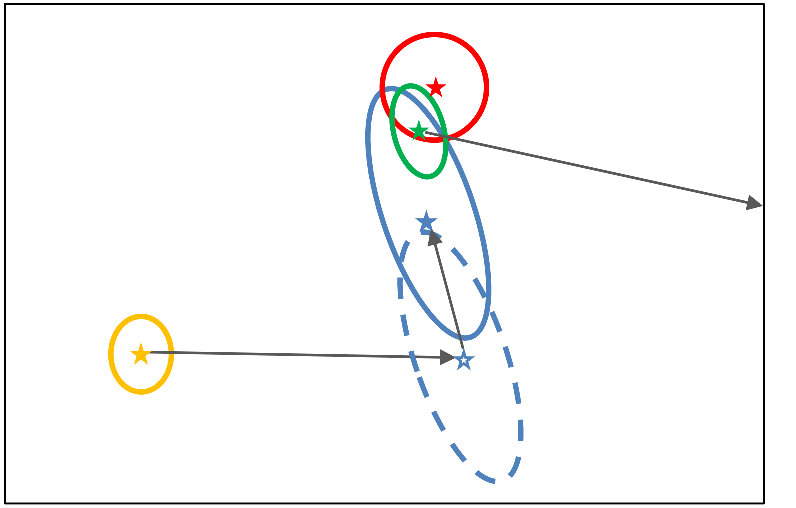
$$\boldsymbol{P}^a \rightarrow (1-\alpha)\, \boldsymbol{P}^a + \alpha \boldsymbol{P}^f$$



2. Correction of systematic bias component

$$\tilde{\boldsymbol{x}}^f_{t+1} = \boldsymbol{x}^f_{t+1} + \boldsymbol{b}$$

$$\boldsymbol{x}^a_{t+1} = \tilde{\boldsymbol{x}}^f_{t+1} + \boldsymbol{K}\left(\boldsymbol{y}_{t+1} - H\left(\tilde{\boldsymbol{x}}^f_{t+1}\right)\right)$$



Bias correction with simple functional form

✓ "Offline" bias correction

$$\frac{d}{dt}\boldsymbol{x} = \boldsymbol{f}_{\mathrm{model}}(\boldsymbol{x}) + \boldsymbol{D}(\boldsymbol{x})$$

Set of training data $\{\delta\boldsymbol{x}, \boldsymbol{x}^f\}$
→ bias correction term $\boldsymbol{D}(\boldsymbol{x})$ estimation

$$\frac{d}{dt}\boldsymbol{x} = \boldsymbol{f}_{\mathrm{true}}(\boldsymbol{x}) \longrightarrow \boldsymbol{x}^t(t), \boldsymbol{x}^t(t+\Delta t) \ldots$$

$$\frac{d}{dt}\boldsymbol{x} = \boldsymbol{f}_{\mathrm{model}}(\boldsymbol{x}) \longrightarrow \boldsymbol{x}^f(t+\Delta t)$$

$$\delta\boldsymbol{x} = \boldsymbol{x}^t(t+\Delta t) - \boldsymbol{x}^f(t+\Delta t)$$

Simplest form: linear dependency

$$\boldsymbol{D}(\boldsymbol{x}) = \boldsymbol{D}_0 + \boldsymbol{L}\boldsymbol{x}'$$

$$\boldsymbol{x}' = \boldsymbol{x}^f - \overline{\boldsymbol{x}}^f$$

$\boldsymbol{C}$: correlation matrix

$$\boldsymbol{D}_0 = \overline{\delta\boldsymbol{x}}/\Delta t$$

Steady component

$$\boldsymbol{L}\boldsymbol{x}' = \boldsymbol{C}_{\delta x, x}\,\boldsymbol{C}_{x,x}^{-1}\boldsymbol{x}' \;/\; \Delta t$$

Linearly-dependent component (Leith , 1978)

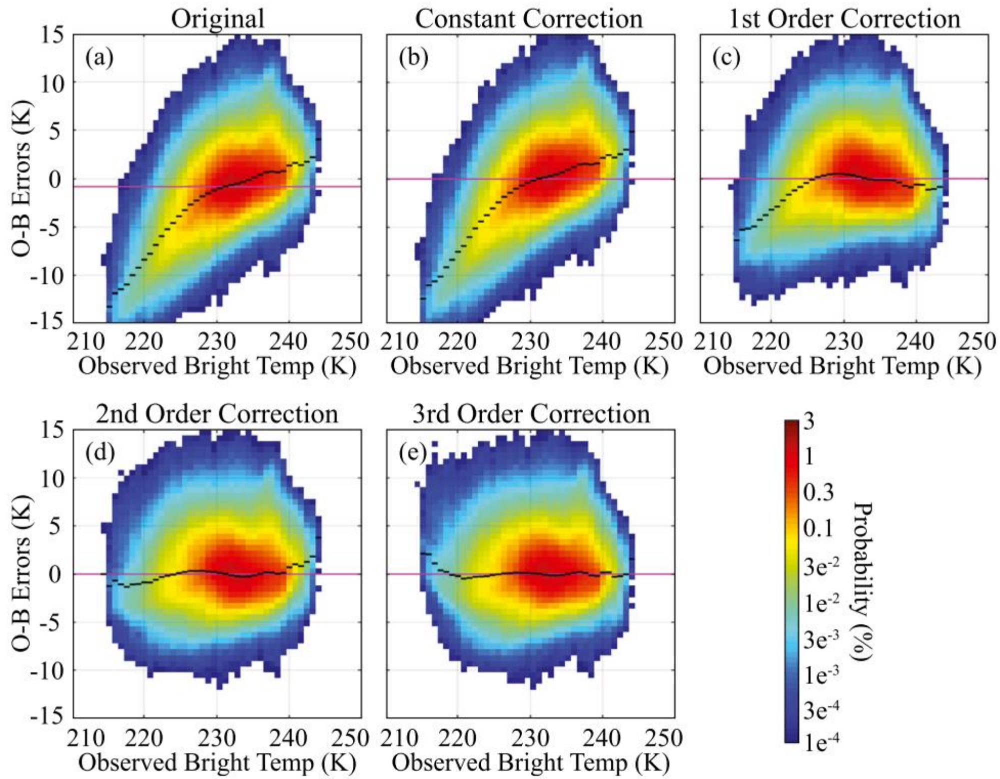
Dimensionality reduction can be applied
using Singular Value Decomposition (SVD) (Danforth et al. 2007)

Bias correction with nonlinear basis functions

Higher order polynomials :

- Coupled Lorenz96 system (Wilks et al. 2005, Arnold et al. 2013)
- Real case: All-sky satellite infrared brightness temperature (Otkin et al. 2018)

Probability of (obs – fcst) vs obs



(Fig.2 of Otkin et al. 2018)

Neural networks :

- Coupled Lorenz96 system (Watson et al. 2019)

Online bias correction

✓ "Online" bias correction

= Simultaneous estimation of state variables and bias correction terms

- Kalman filter

  sequential treatment / augmented state

  - Steady component (Dee and Da Sliva 1998, Baek et al 2006)
  - Polynomials (Pulido et al. 2018)

- Variational data assimilation ("VarBC")

  - Legendre polynomials

    (Cameron and Bell, 2016; for Satellite sounding in UK Met Office operational model)

$$J_o = \frac{1}{2}\sum_k \left(\left(y_k + \sum_{i=1}^{I_k} \beta_i' p_{k,i} - y_k^o \right) R_k^{-1} \left(y_k + \sum_{j=1}^{I_k} \beta_j' p_{k,j} - y_k^o \right) \right)$$

Penalty function for bias-corrected obs error

$$J_\beta = \frac{1}{2}\sum_{i=1}^{I_k} \beta_i'^T V_{(\beta_i)}^{-1} \beta_i'$$

Penalty function for regularization

Localization

In high dimensional spatiotemporal system,
geographically (and temporally) local interaction is usually dominant

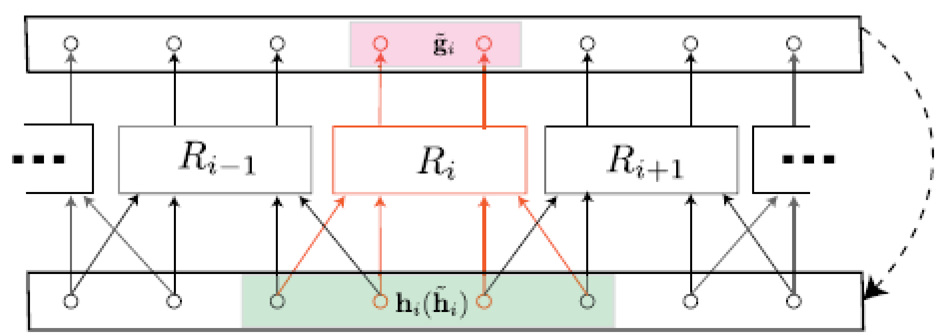
Localization is built-in in LETKF

- ✓ Reduced matrix size -> low cost
- ✓ Highly effective parallelization
  (Miyoshi and Yamane, 2007)

Also used in simultaneous parameter estimation
(Aksoy et al. 2006)

Also in ML-based data driven modelling

(Pathak et al. 2018, Watson et al. 2019)



(Pathak et al. 2018)

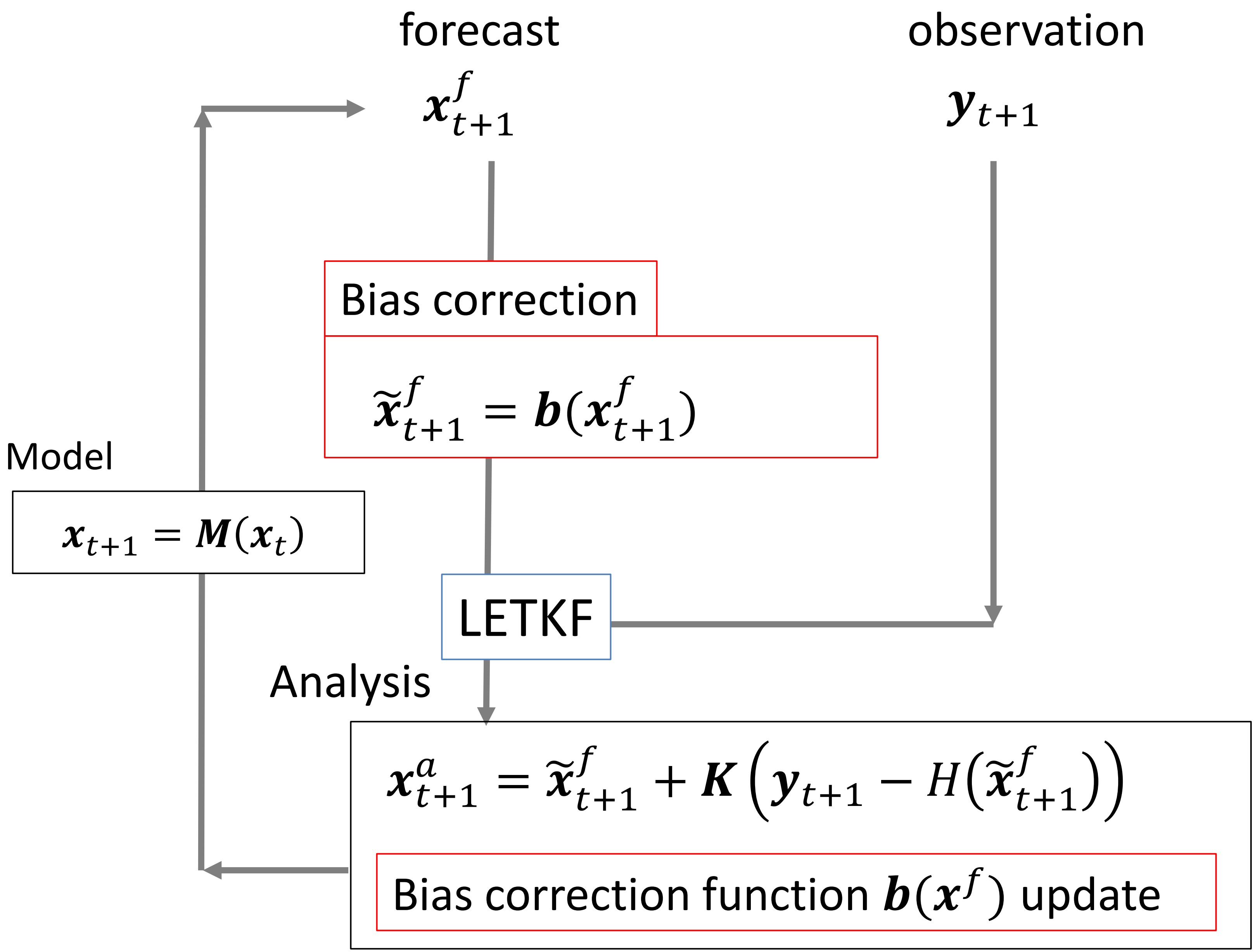
The goal of this study

- ✓ ML-based online bias correction using Long-Short Term Memory (LSTM)
- ✓ Combined with LETKF with similar localization

Test experiments with coupled Lorenz96 model

- Experimental Setup
- Online bias correction with simple linear regression as a reference
- (Online bias correction with LSTM)

bias correction in LETKF system

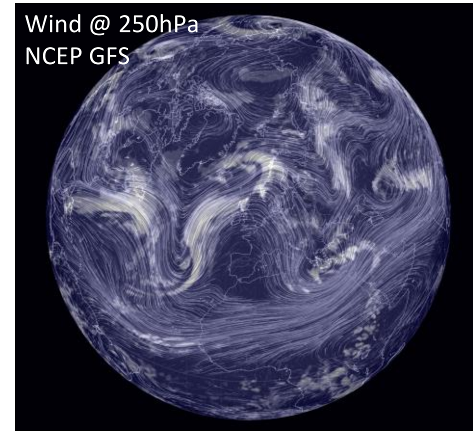


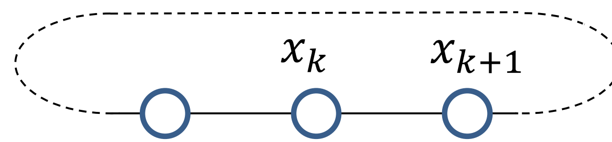
Lorenz96 model

$$\frac{d}{dt}x_k = x_{k-1}(x_{k+1} - x_{k-2}) - x_k + F \qquad k = 1,2,\dots,K$$

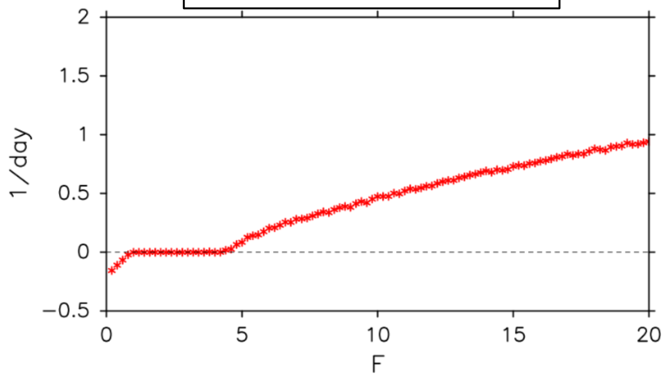
(Lorenz 1996)

- 1-D cyclic domain
- Chaotic behavior for sufficiently large F

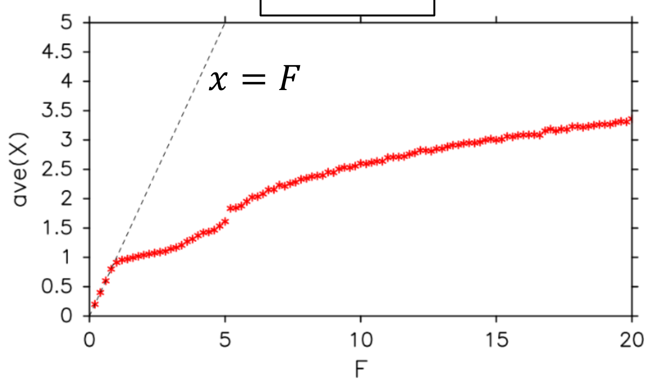




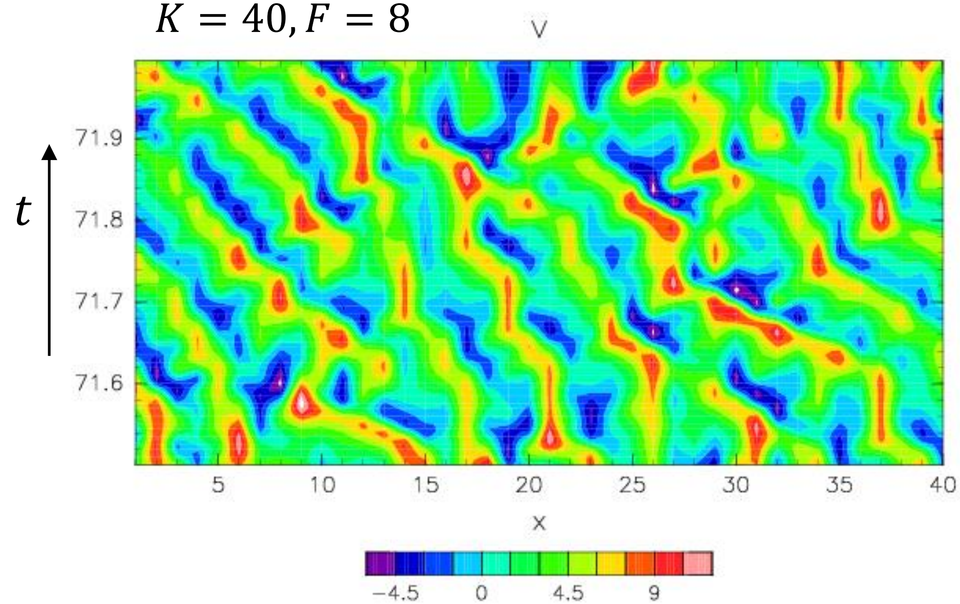
Largest Lyapunov exp.



Mean x



$K = 40, F = 8$



Coupled Lorenz96 model

Multi-scale interaction

"Nature run"

Large scale (Slow) variables

$$\frac{d}{dt}x_k = x_{k-1}(x_{k+1} - x_{k-2}) - x_k + F - \underline{\frac{hc}{b}\sum_{j=J(k-1)+1}^{kJ} y_j} \qquad k = 1,2,\dots,K$$

Small scale (fast) variables

Coupling terms

$$\frac{d}{dt}y_j = -cb\, y_{j+1}\left(y_{j+2} - y_{j-1}\right) - cy_j + \underline{\frac{hc}{b}x_{\mathrm{int}[(j-1)/J]+1}} \qquad j = 1,2,\dots,KJ$$

Forecast model (Danforth and Kalnay, 2008)

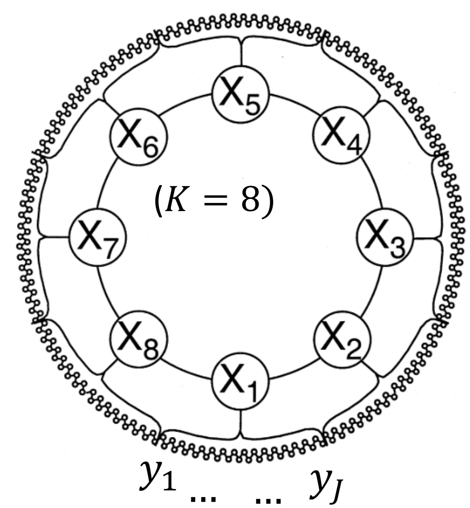
$$\frac{d}{dt}x_k = x_{k-1}(x_{k+1} - x_{k-2}) - x_k + F + A\sin\left(2\pi\frac{k}{K}\right)$$

Parameters used in this study：

$K = 16, J = 16$

$h = 1, b = 20, c = 50$

$F = 16, \mathrm{A} = 1$



(Wilks, 2005)

Data assimilation experiment

“Observation”

“Nature run” + random error

Observation operator：identical （obs = model grid）

Error standard deviation： 0.1

Interval: 0.025 / **0.05** / 0.1 (cf: doubling time $\simeq 0.2$)

LETKF configuration

Member： 20

Localization： Gaussian weighting (length scale = 3 grids)

Covariance inflation: **multiplicative** (factor: α)

“Control run”: Data assimilation without bias correction

Large inflation factor α is needed to stabilize DA cycle

Obs interval	Minimum RMSE	Inflation α
0.025	0.0760	2.9
0.05	0.0851	5.8
0.1	0.0902	15.0

Example: simple linear regression

Online bias correction by linear regression

forecast

$$\boldsymbol{x}^f_{t+1}$$

observation

$$\boldsymbol{y}_{t+1} = H(\boldsymbol{x}^{\text{true}}_{t+1}) + \boldsymbol{\epsilon}^{\text{obs}}$$

Bias correction

$$\tilde{\boldsymbol{x}}^f_{t+1} = \boldsymbol{W}^f_{t+1}\boldsymbol{x}^f_{t+1} + \boldsymbol{b}^f_{t+1}$$

Model

$$\boldsymbol{x}_{t+1} = \boldsymbol{M}(\boldsymbol{x}_t)$$

$$\boldsymbol{b}_{t+1} = (1-\gamma)\boldsymbol{b}_t$$

$$\boldsymbol{W}_{t+1} = (1-\gamma)\boldsymbol{W}_t$$

LETKF

Analysis

$$\boldsymbol{x}^a_{t+1} = \tilde{\boldsymbol{x}}^f_{t+1} + \boldsymbol{K}\left(\boldsymbol{y}_{t+1} - H(\tilde{\boldsymbol{x}}^f_{t+1})\right)$$

$$\boldsymbol{b}^a_{t+1} = \boldsymbol{b}^f_{t+1} + \beta\boldsymbol{K}\left(\boldsymbol{y}_{t+1} - H(\tilde{\boldsymbol{x}}^f_{t+1})\right)$$

$$\boldsymbol{W}^a_{t+1} = \boldsymbol{W}^f_{t+1} + \beta\boldsymbol{K}\left(\boldsymbol{y}_{t+1} - H(\tilde{\boldsymbol{x}}^f_{t+1})\right) \cdot (\tilde{\boldsymbol{x}}^f_{t+1})^{\mathrm{T}} / \left|\tilde{\boldsymbol{x}}^f_{t+1}\right|^2$$

Online bias correction by linear regression

$\beta = 0.02, \gamma = 0.999$ (half-life: ~37)

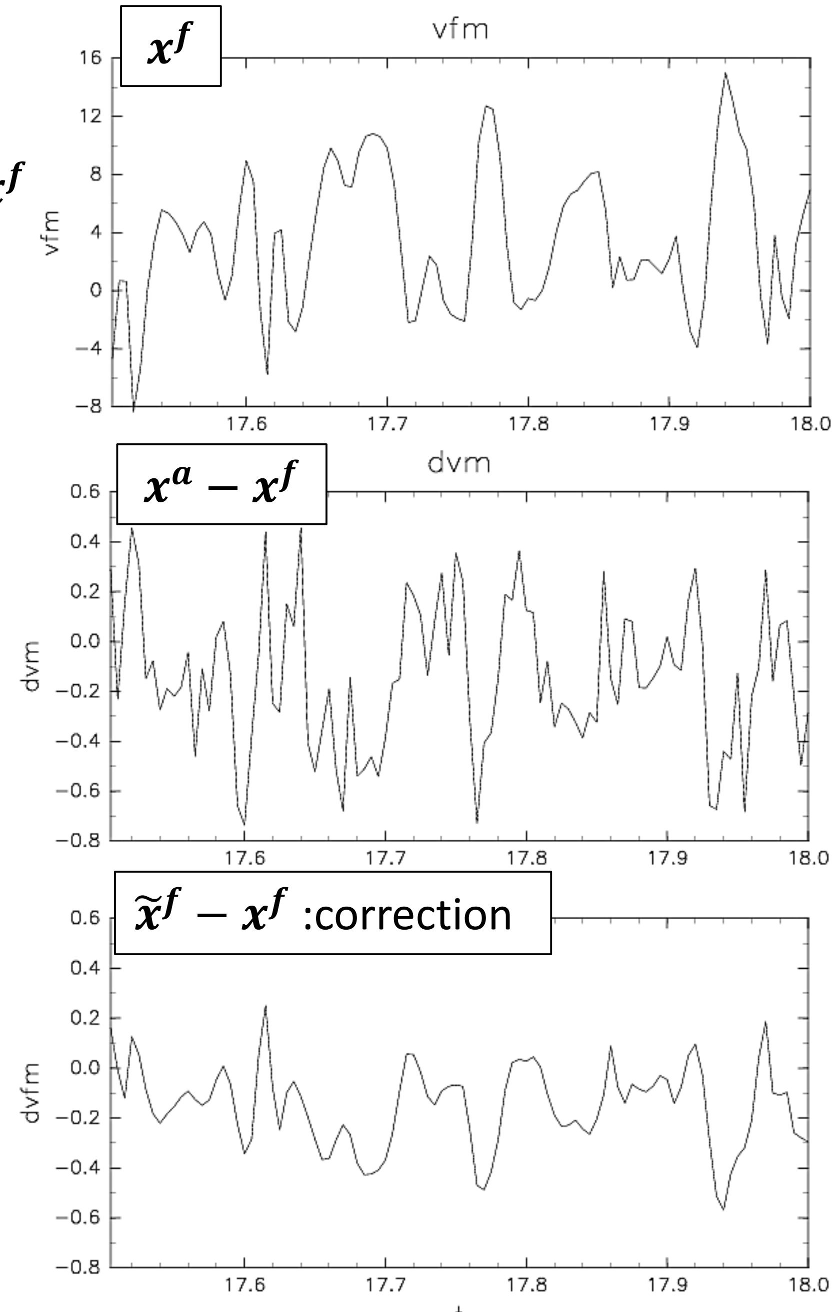
$\boldsymbol{x^a} - \boldsymbol{x^f}$ has large negative correlation with $\boldsymbol{x^f}$ and mostly offset by bias correction term

Stable with smaller inflation factor α

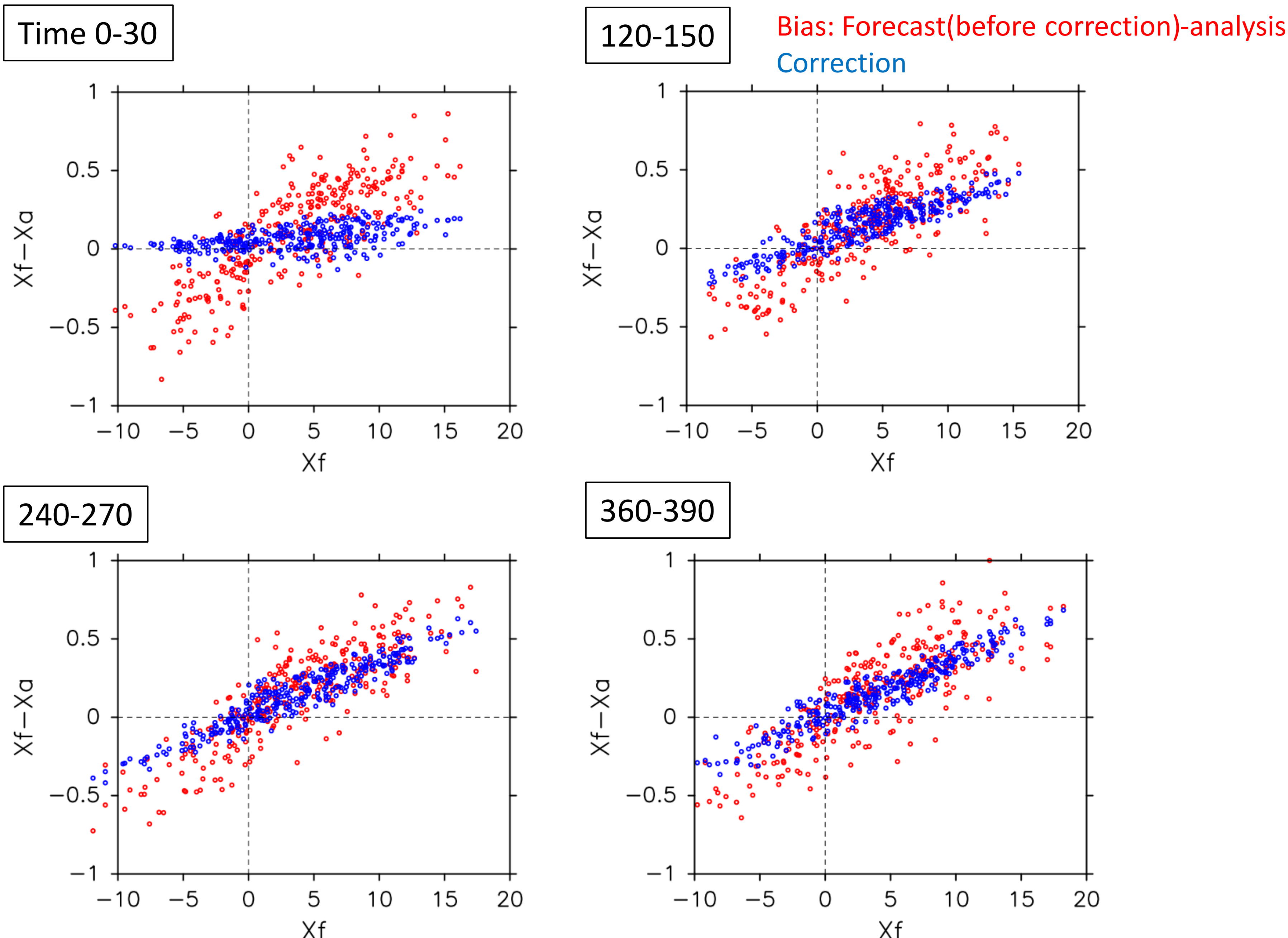
Interval (day)	Min. infl.	RMSE
0.025	2.9	0.0760
0.1	15.0	0.0902

↓

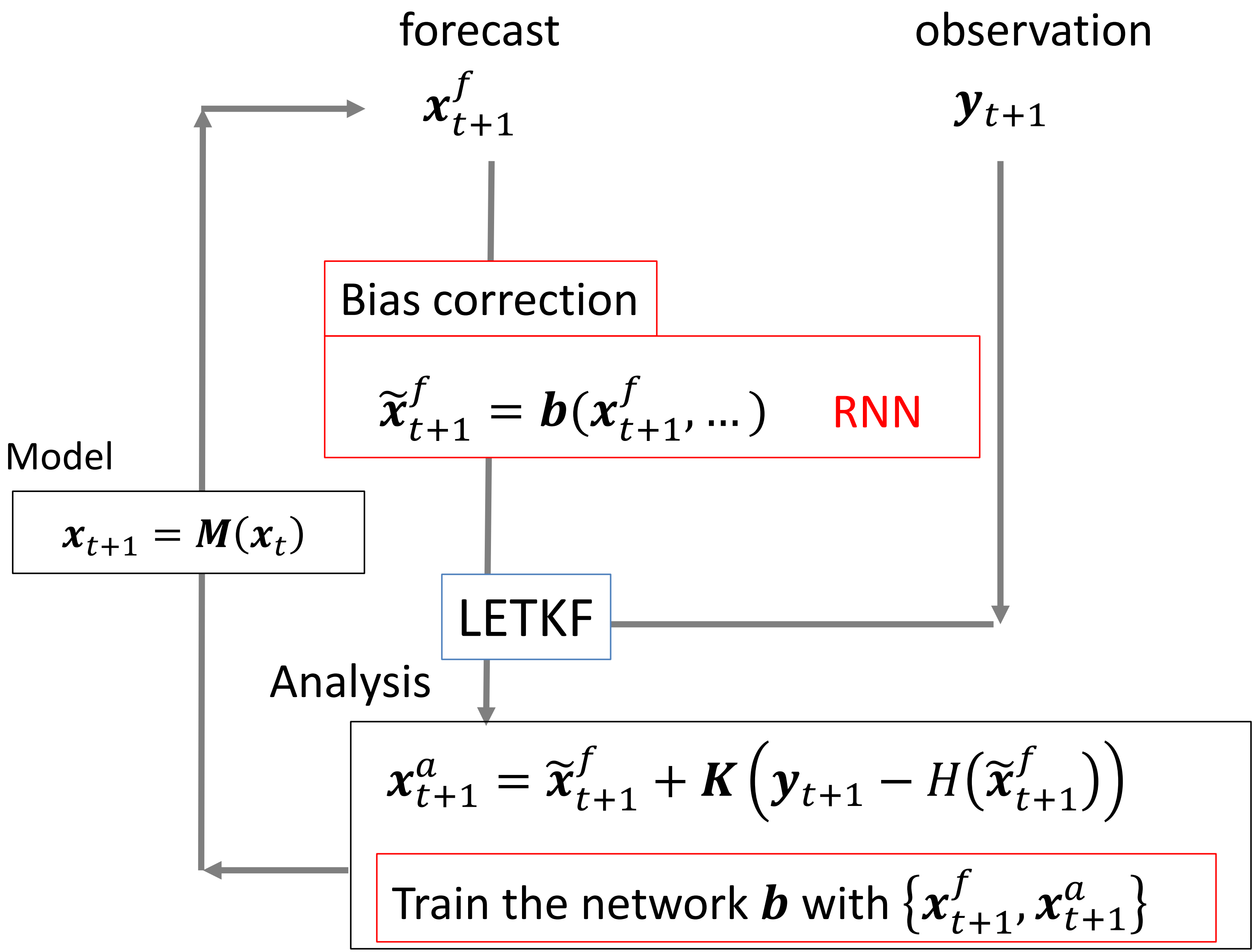
Interval (day)	Min. infl.	RMSE
0.025	2.0	0.0625
0.1	8.2	0.0849



Correction term

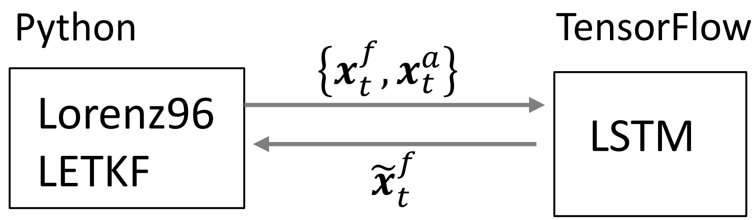


Nonlinear bias correction with ML



Plan: LSTM implementation

Tensorflow LSTM is implemented and integrated with LETKF codes

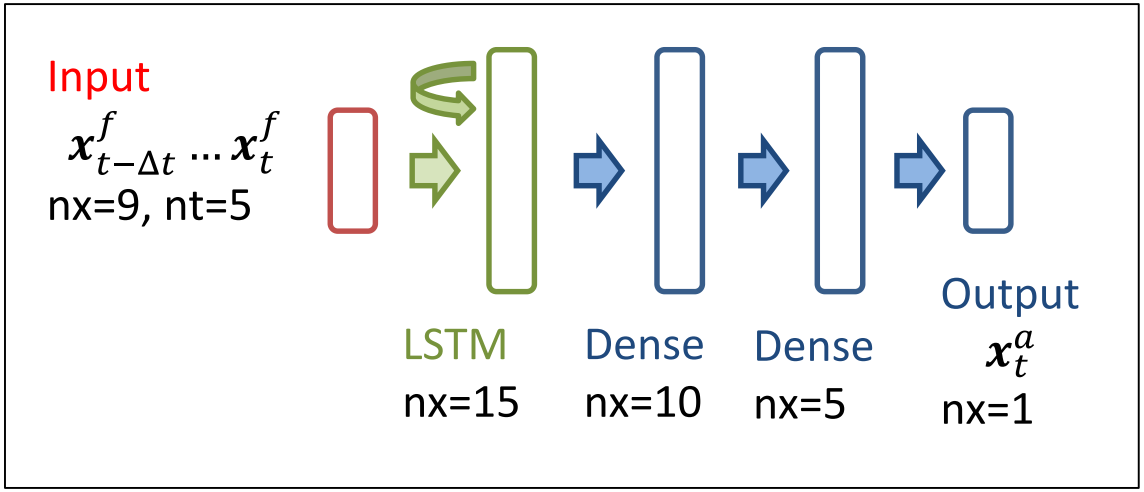


Network architecture

- 1 LSTM + 3 Dense layers
- Activation:
  tanh / sigmoid(recurrent)
- No regularization / dropout
- LETKF-like Localization

  Input : localized area

  Output : one grid point



Summary

- Systematic model bias degrades forecasts and analysis
- "Offline" bias correction can be performed by ML as nonlinear regression
- "Online" bias correction with data assimilation has been studied using a fixed basis function set
- LSTM-based bias correction is implemented and to be tested
    - The efficiency of localization ?
    - Online learning ?