Data Assimilation
- A02. Deterministic Chaos-

Shunji Kotsuki

Center for Environmental Remote Sensing / Institute of Advanced Academic Research

(shunji.kotsuki@chiba-u.jp)

DA Lectures A (Basic Course)

Environmental Prediction Science Laboratory

- (1) Introduction and NWP
- (2) Deterministic Chaos and Lorenz-96 model
- (3) A toy model and Bayesian estimation
- (4) Kalman Filter (KF)
- (5) 3D Variational Method (3DVAR)
- (6) Ensemble Kalman Filter (PO method)
- (7) Serial Ens. Square Root Filter (Serial EnSRF)
- (8) Local Ens. Transform Kalman Filter (LETKF)
- (9) Innovation Statistics
- (10) Adaptive Inflations
- (11) 4D Variational Method (4DVAR)

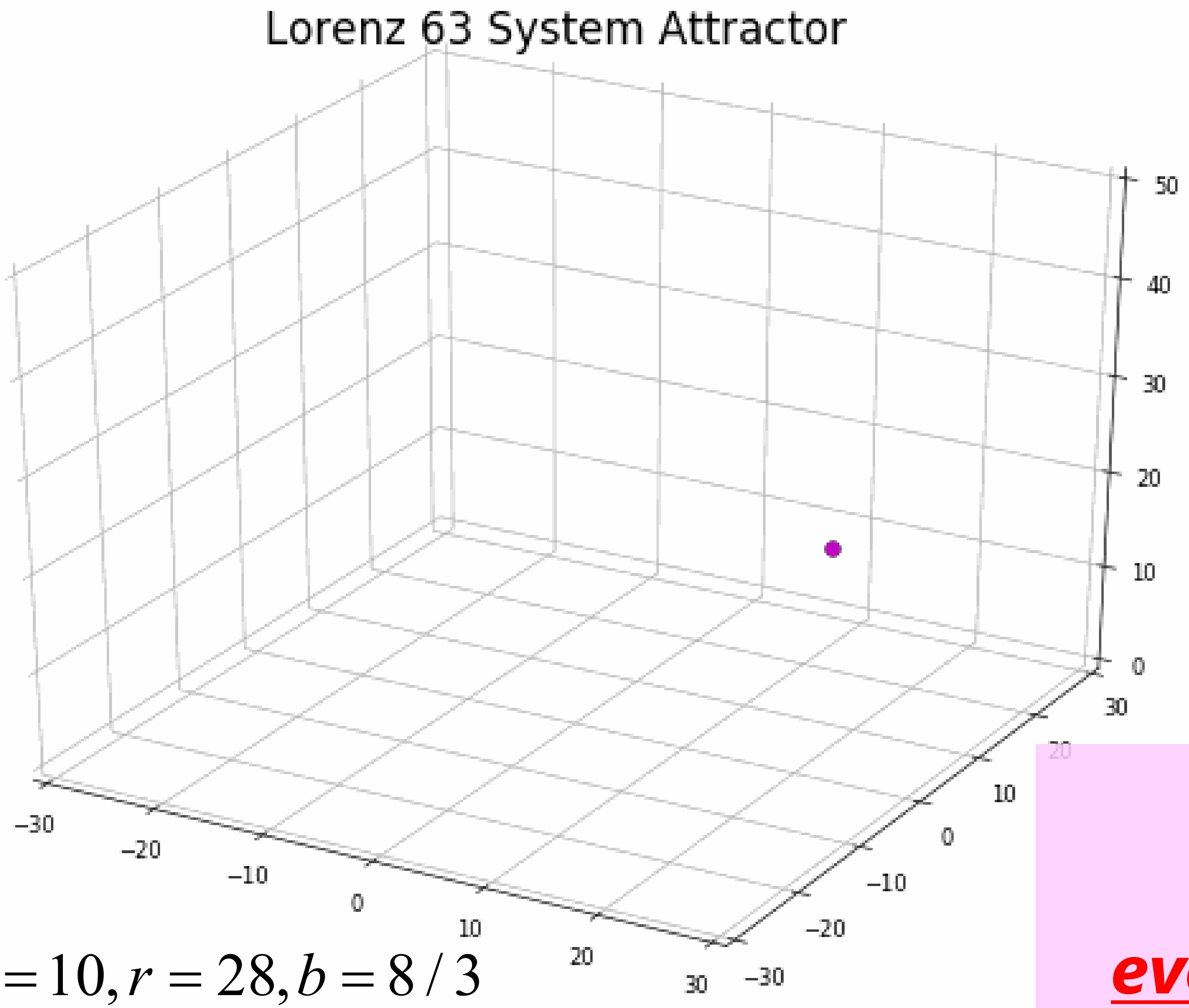
Today's Goal

- **Lecture**
  - To understand the chaotic nature of atmosphere

- **Training Course**
  - To implement Lorenz 96 (any language)
  - To estimate error doubling time

Deterministic Chaos
(Lecture)

Deterministic Chaos and Predictability



$$p = 10, r = 28, b = 8/3$$

Edward Lorenz

Lorenz 63 model

$$\dot{x} = p(y - x)$$

$$\dot{y} = -xz + rx - y$$

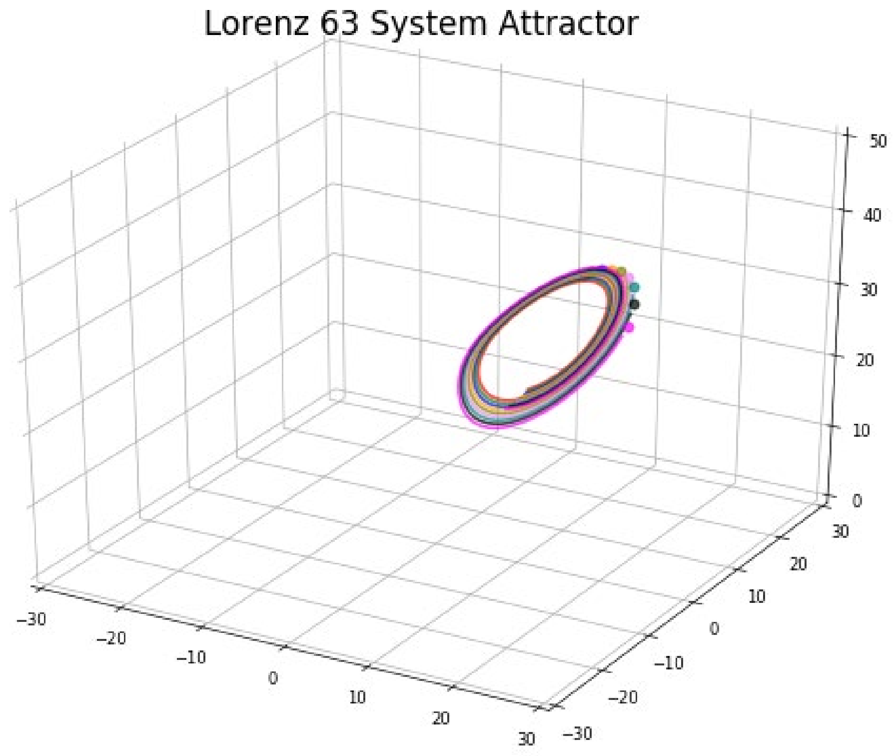
$$\dot{z} = xy - bz$$

Chaotic systems have limits to predictability *<u>even with the perfect model!!</u>*

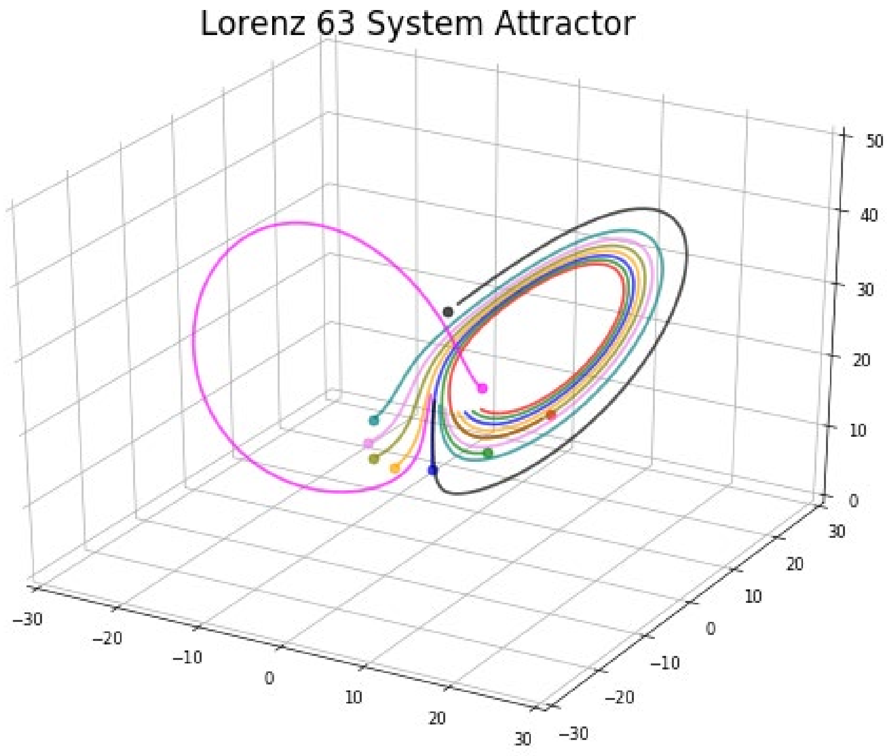
Initial Conditions :: x=y=z=15.000, 15.001, 15.002, ..., 15.009

Predictability differs

more predictable

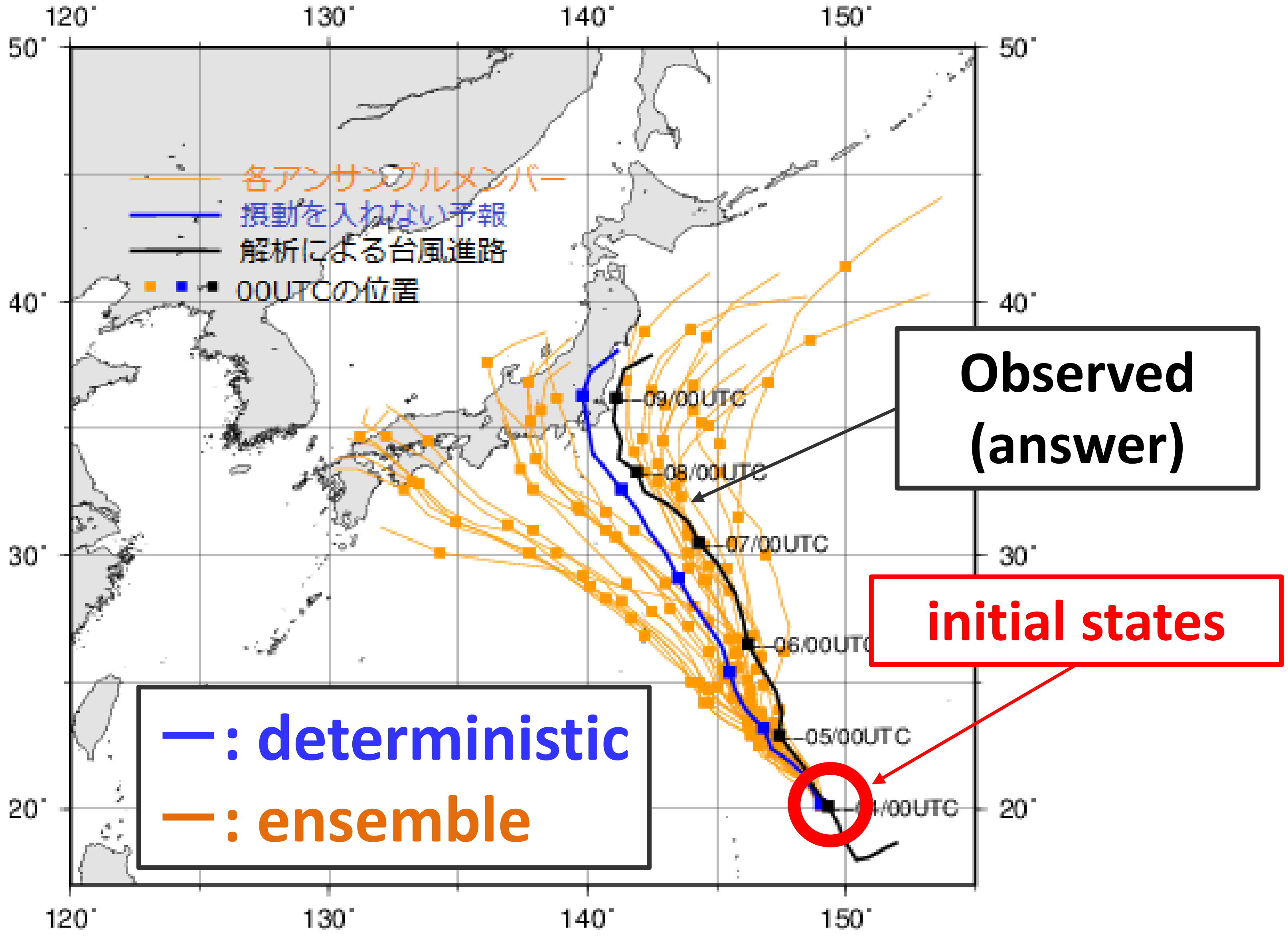


less predictable



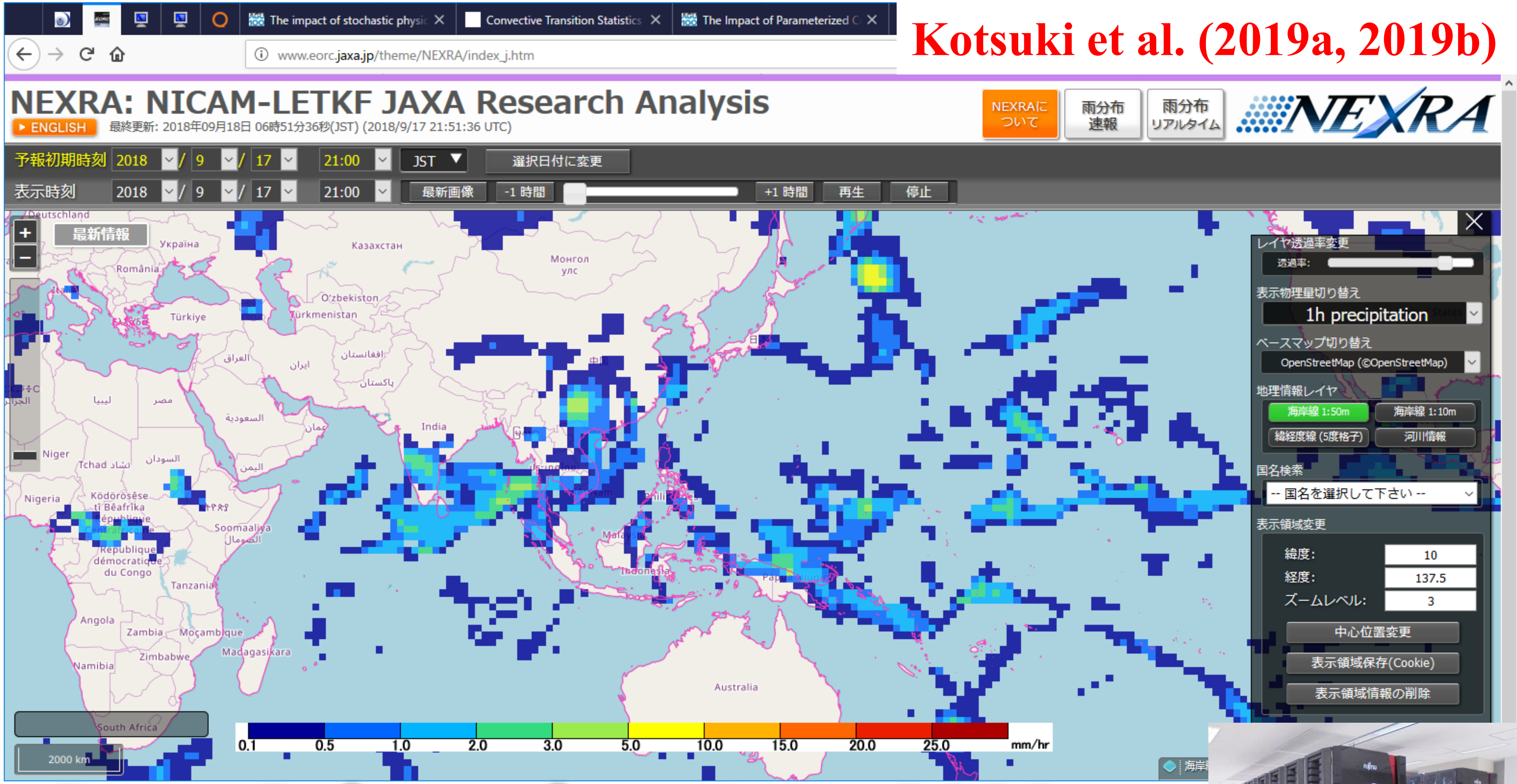
Predictability depends on the initial state!

Ensemble Prediction (e.g. TC)



JAXA's NWP: NEXRA

Environmental Prediction Science Laboratory

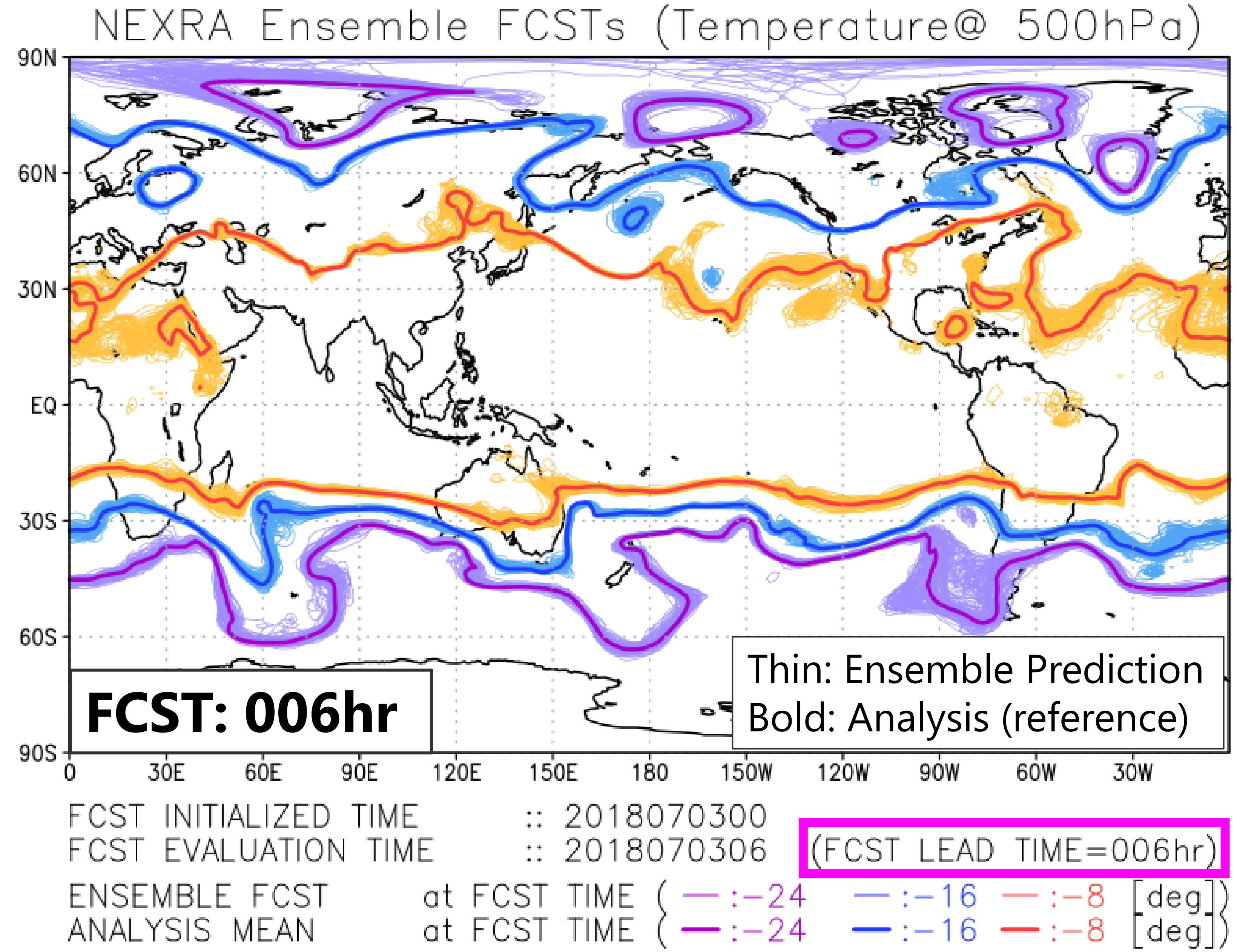


Kotsuki et al. (2019a, 2019b)

Press Released on Aug. 2020

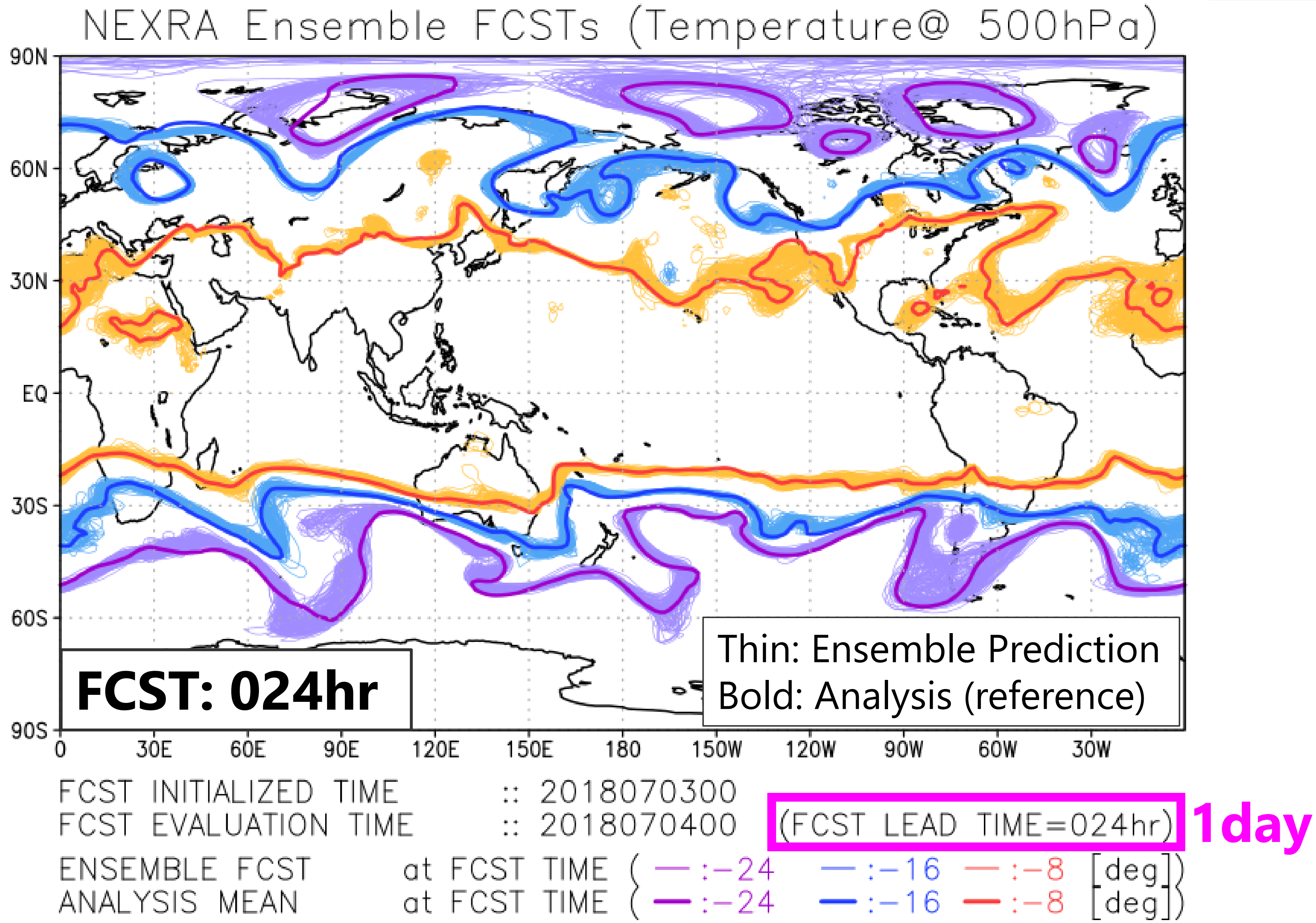
Running Operationally on JAXA's Supercomputing System

Ensemble NWP by NEXRA

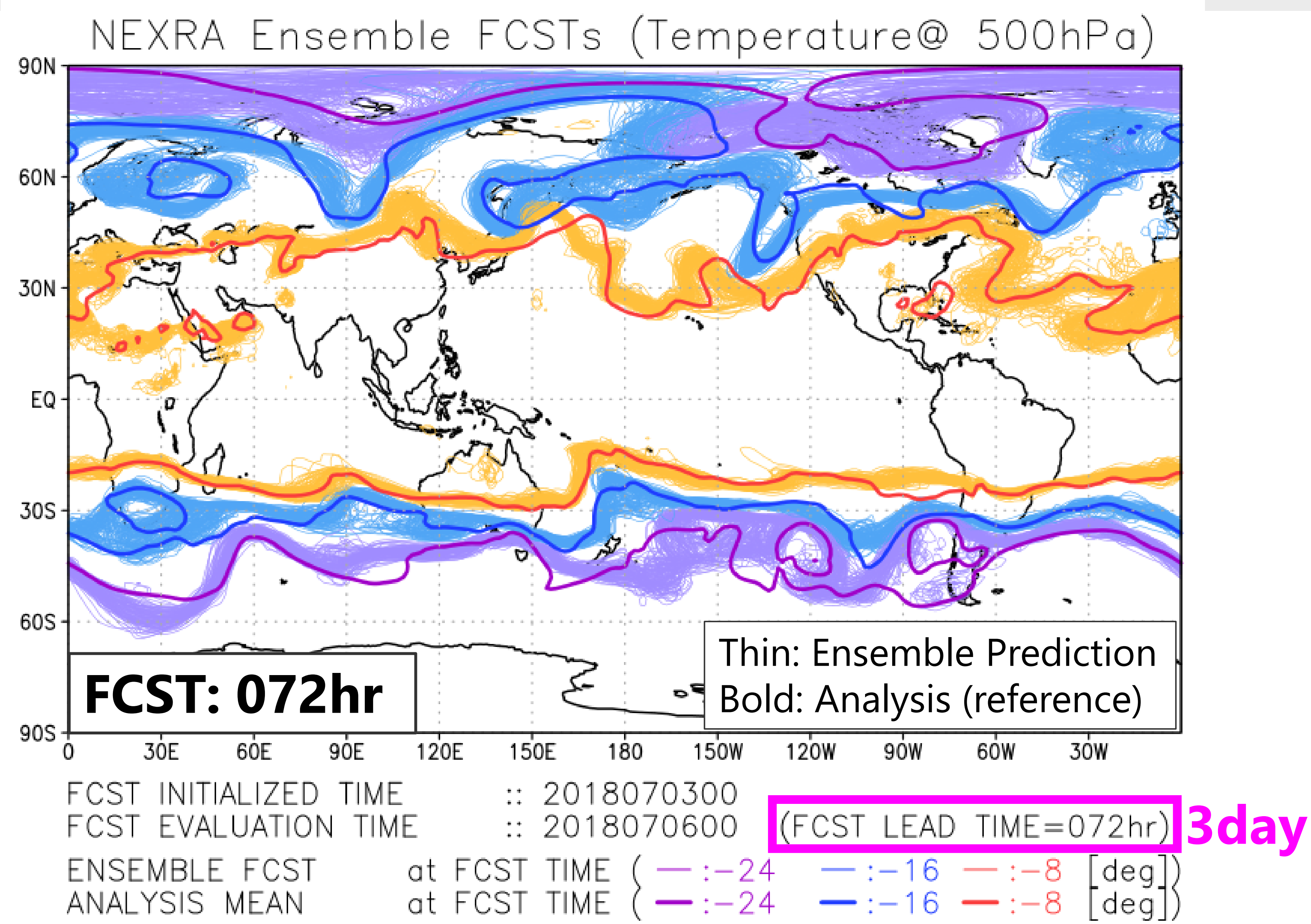


6hr

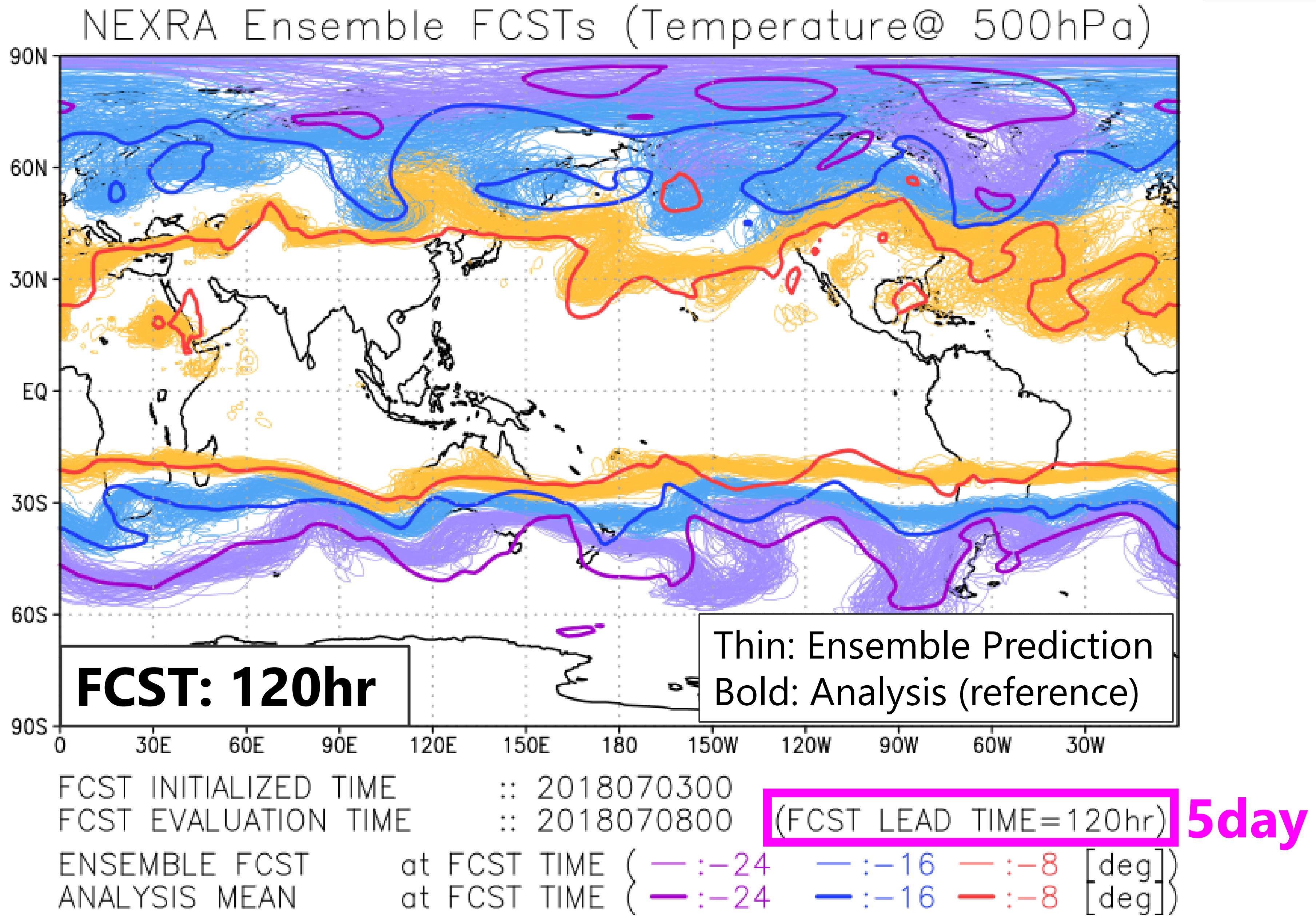
Ensemble NWP by NEXRA



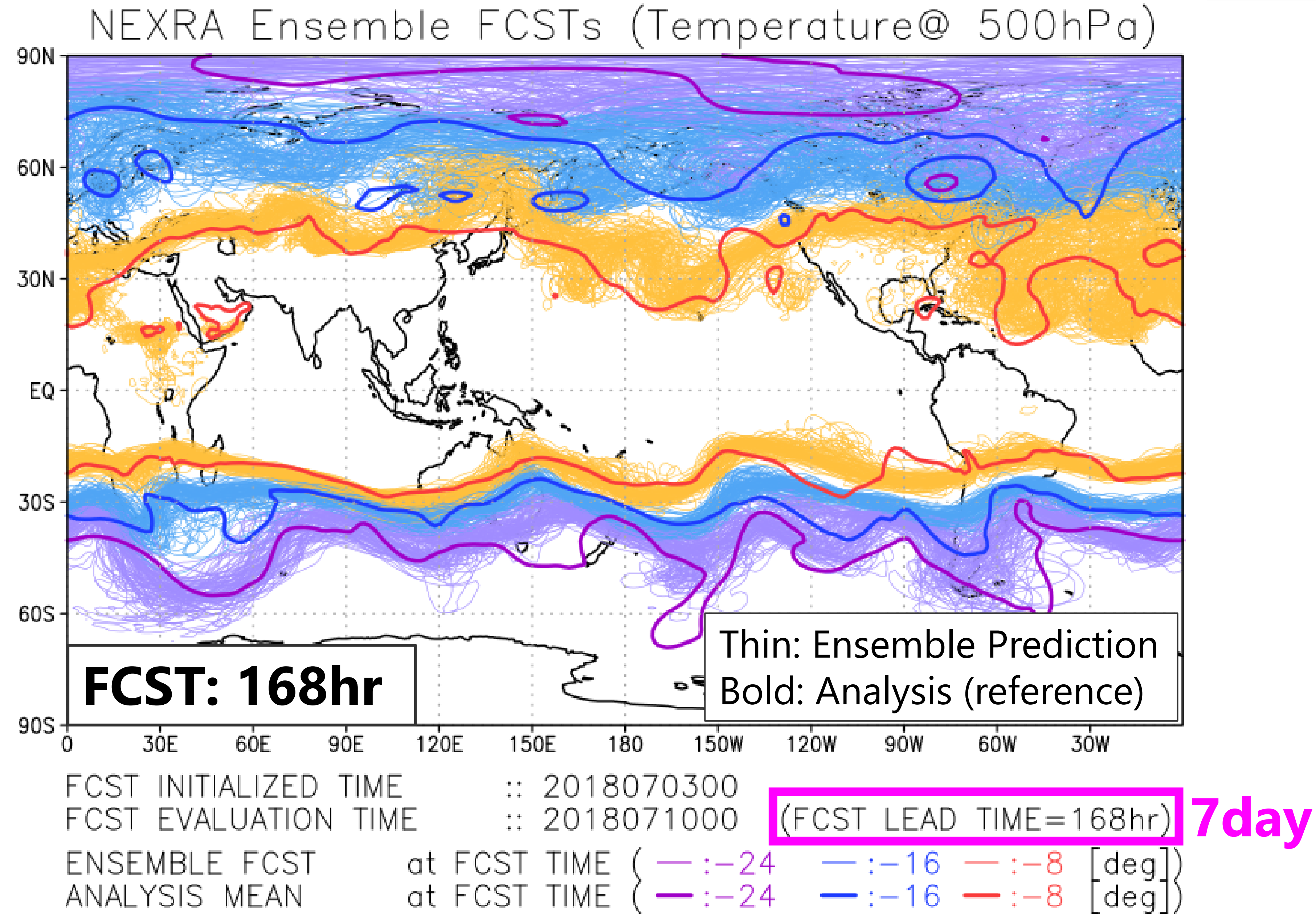
Ensemble NWP by NEXRA



Ensemble NWP by NEXRA

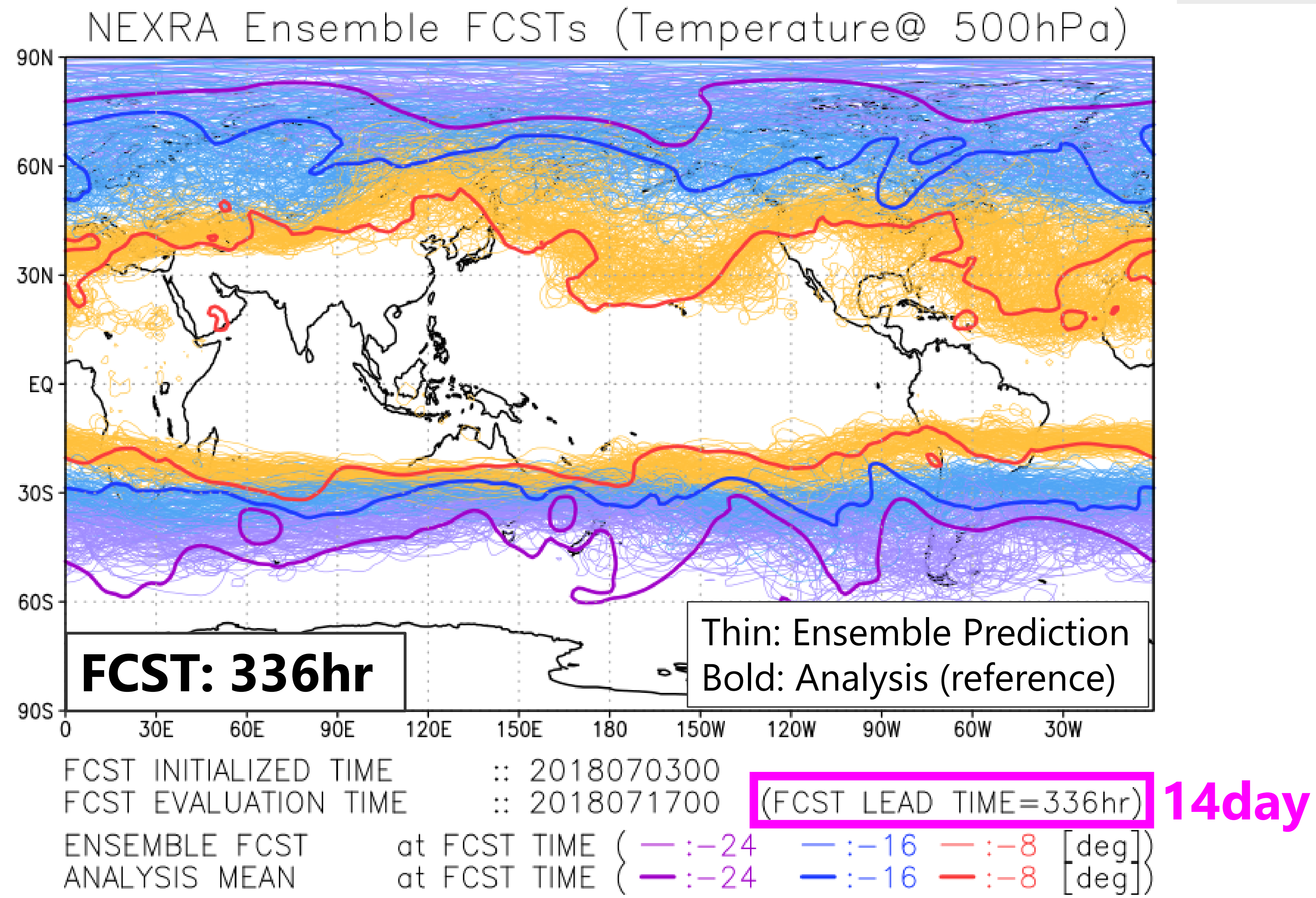


Ensemble NWP by NEXRA

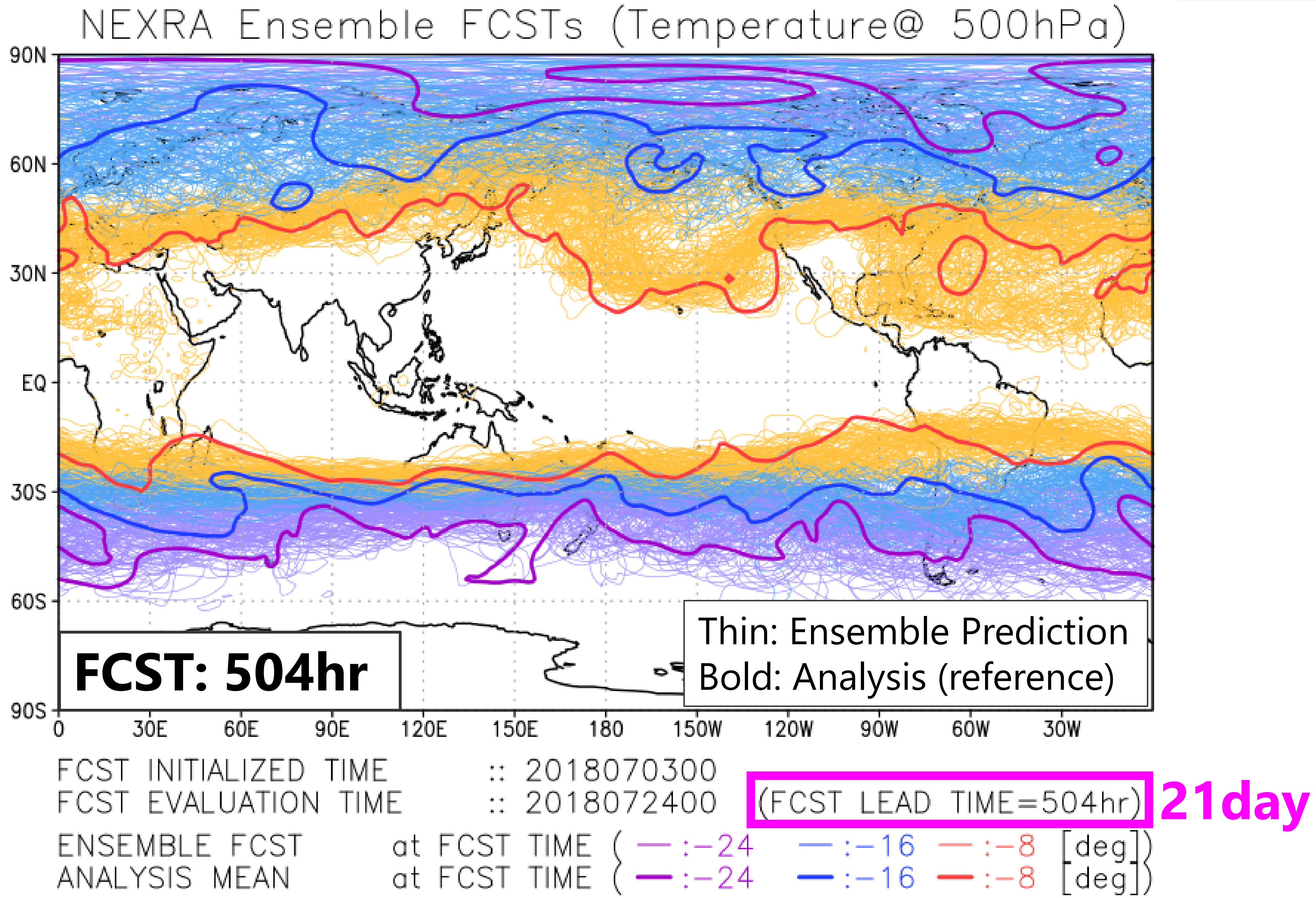


Ensemble NWP by NEXRA

Environmental Prediction Science Laboratory



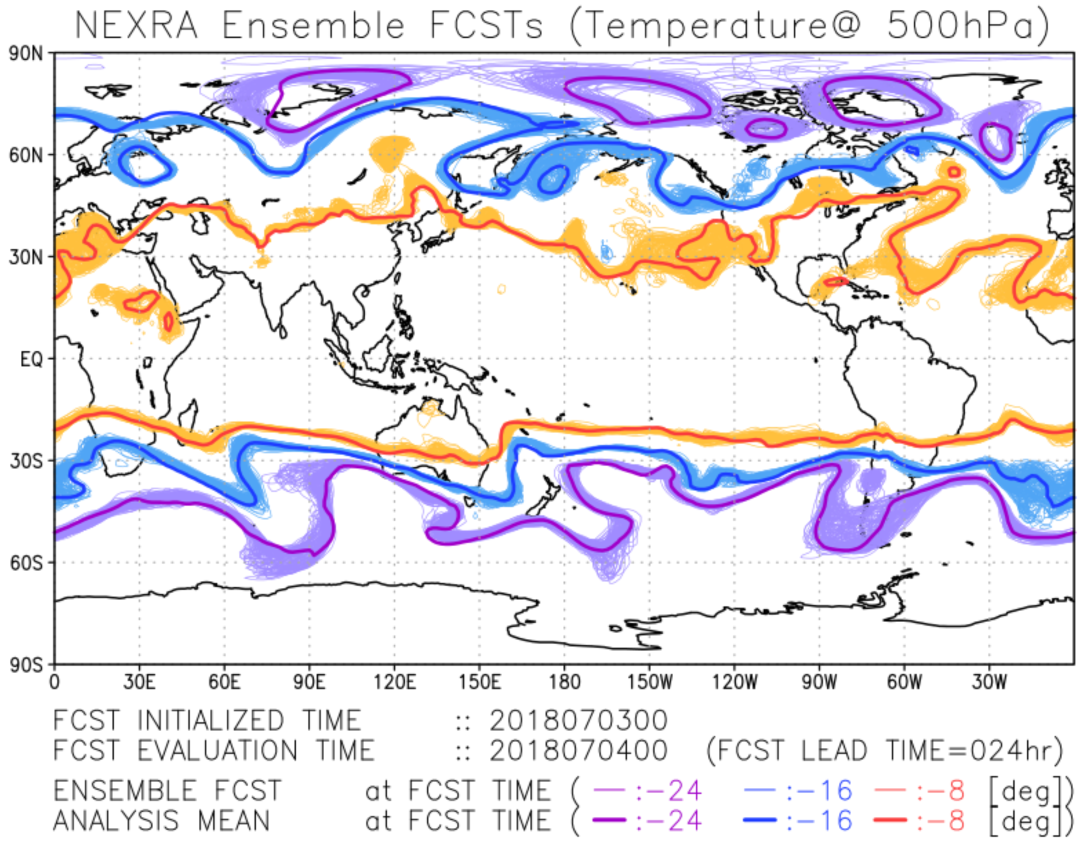
Ensemble NWP by NEXRA



Ensemble NWP by NEXRA

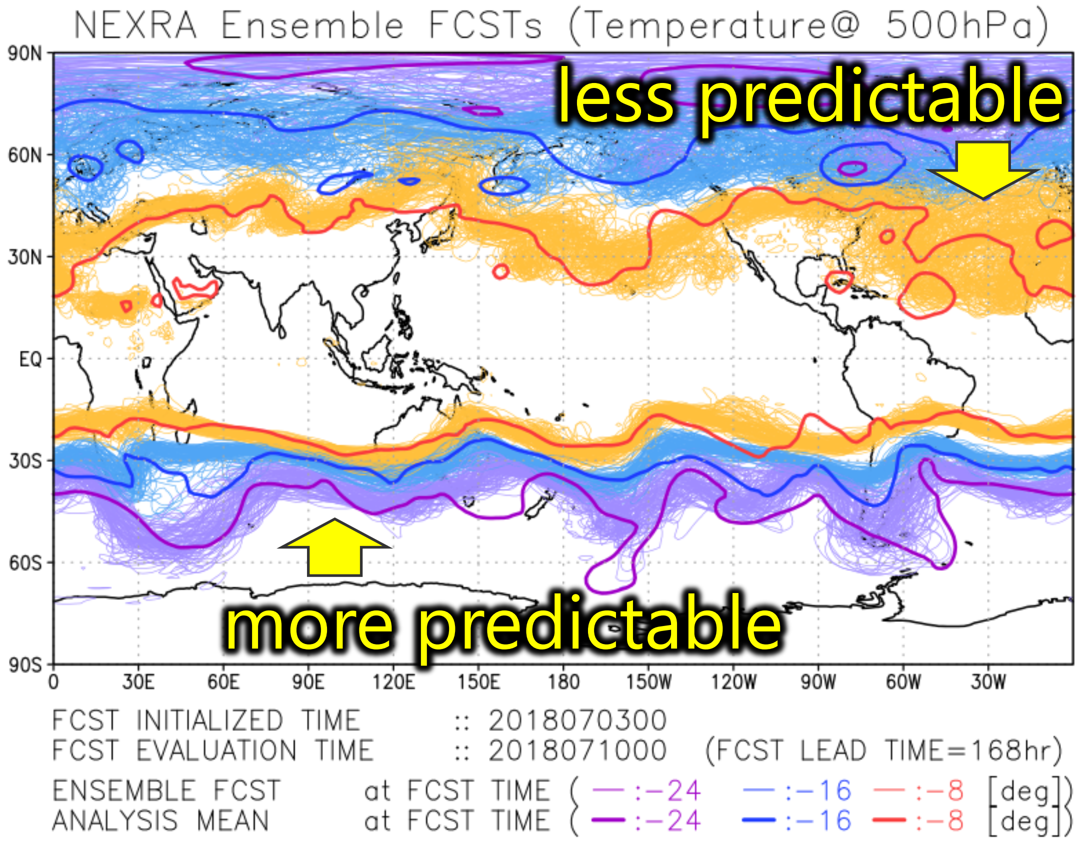
Environmental Prediction Science Laboratory

1-day forecasts



- smaller ensemble spread
- ensemble predictions ≒ reference

7-day forecasts



- larger ensemble spread
- ensemble predictions ≠ reference
- larger spread near extratropical cyclones

Training Course

Environmental
Prediction
Science
Laboratory

DA Study w/ 40-variable Lorenz-96

Lorenz-96 model (Lorenz 1996)

For j=1,…,N, X_j=X_{j+N}

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

力学系モデル・データ同化基礎技術の速習コース

Training Course of Dynamical Model and Data Assimilation

January 31, 2020, Shunji Kotsuki

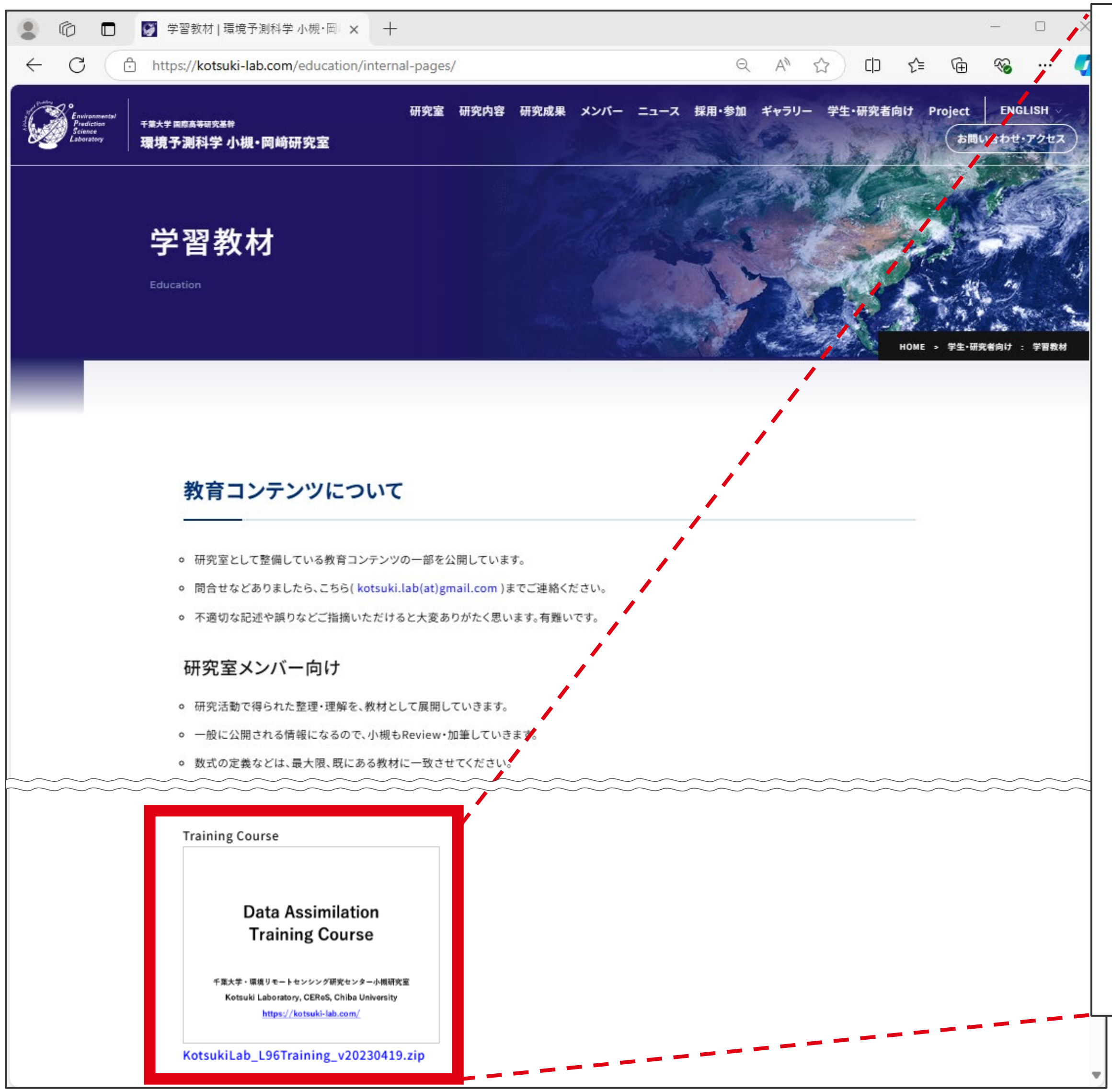
updated 2020/03/19, 2020/06/29, 2021/07/15

目的： 簡易力学モデル Lorenz の 40 変数モデル（以下 L96; Lorenz 1996）を使って複数のデータ同化手法を自ら実装し、様々な実験を行う。データ同化システムを実際に、0 からコーディングすることで、力学モデリングやデータ同化に関する実践的な「使える」基礎技術を体得する。

Purpose: Using the 40-variable dynamical a.k.a. Lorenz-96 (L96; Lorenz 1996), we are going to perform various experiments with multiple data assimilation (DA) methods. By actually coding a data assimilation system from scratch, you will acquire practically "usable" basic techniques related to mechanical modeling and data assimilation.

Text Books

① Training Description



pswd: ceres

力学系モデル・データ同化基礎技術の速習コース

Training Course of Dynamical Model and Data Assimilation

January 31, 2020, Shunji Kotsuki
updated 2020/03/19, 2020/06/29, 2021/07/15

目的： 簡易力学モデル Lorenz の 40 変数モデル（以下 L96; Lorenz 1996）を使って複数のデータ同化手法を自ら実装し、様々な実験を行う。データ同化システムを実際に、0 からコーディングすることで、力学モデリングやデータ同化に関する実践的な「使える」基礎技術を体得する。

Purpose: Using the 40-variable dynamical a.k.a. Lorenz-96 (L96; Lorenz 1996), we are going to perform various experiments with multiple data assimilation (DA) methods. By actually coding a data assimilation system from scratch, you will acquire practically "usable" basic techniques related to mechanical modeling and data assimilation.

方法： 以下の課題を自ら実装し、解決していく。使用言語やプラットフォームは問わない。研究室の MTG において、各自が進捗を報告し、問題点を解消していく。質問は MTG の他も、居室で適宜受け付ける。使用言語については、特に拘りがなければ、行列演算の容易な python が扱いやすい。また、単精度ではなく倍精度でコーディングする事。でないと、既往研究と比較して正しく動作しているか確認できない。

Method: Implement and solve the following problems yourself. Any programing languages or platforms can be used in this exercise. At the Kotsuki Lab. mtg, each personnel will report the progress, and try to solve the problems. Questions are accepted during the MTG as well as at the office when necessary. As for the programing language, python, which is easy to perform matrix operations, is recommended unless specific language is preferred. Also, you should code in double precision instead of single precision. Otherwise, confirming whether performing properly or not compared to the previous studies will not be possible.

‣ https://kotsuki-lab.com/internal-pages/

Basic Task 1

Training Course Problem (1)

Environmental Prediction Science Laboratory

基礎課題：

1． L96 を 4 次の Runge-Kutta 法を用いて実装する。パラメータ値 F を色々と変え、F=8 の時にカオスとなることを確認する。ここでは、Runge-Kutta はライブラリを用いずに自分でコーディングする事。また、オイラー法など、他の積分スキームと比較してみる。
ヒント）まずは、原著論文 Lorenz and Emanuel (1998)の Fig. 1 を再現する。

Basic Tasks：

1. Implement L96 using the 4th-order Runge-Kutta method. Change the parameter value F in various ways and confirm that it becomes chaos when $F = 8$. Here, Runge-Kutta needs to be coded yourself without using any libraries. Also, compare it with other integration schemes such as the Euler method.
Hint) First, reproduce Fig. 1 of the original paper Lorenz and Emanuel (1998).

The Lorenz-96 model (L96)

Journal of the Atmospheric Sciences, **55**

Optimal Sites for Supplementary Weather Observations: Simulation with a Small Model

Edward N. Lorenz and Kerry A. Emanuel

Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts

(Manuscript received 12 December 1996, in final form 22 June 1997)

ABSTRACT

Anticipating the opportunity to make supplementary observations at locations that can depend upon the current weather situation, the question is posed as to what strategy should be adopted to select the locations, if the greatest improvement in analyses and forecasts is to be realized. To seek a preliminary answer, the authors introduce a model consisting of 40 ordinary differential equations, with the dependent variables representing values of some atmospheric quantity at 40 sites spaced equally about a latitude circle. The equations contain quadratic, linear, and constant terms representing advection, dissipation, and external forcing. Numerical integration indicates that small errors (differences between solutions) tend to double in about 2 days. Localized errors tend to spread eastward as they grow, encircling the globe after about 14 days.

In the experiments presented, 20 consecutive sites lie over the ocean and 20 over land. A particular solution is chosen as the true weather. Every 6 h observations are made, consisting of the true weather plus small random errors, at every land site, and at one ocean site to be selected by the strategy being considered. An analysis is then made, consisting of observations where observations are made and previously made 6-h forecasts elsewhere. Forecasts are made for each site at ranges from 6 h to 10 days. In all forecasts, a slightly weakened external forcing is used to simulate the model error. This process continues for 5 years, and mean-square forecast errors at each site at each range are accumulated.

Strategies that attempt to locate the site where the current analysis, as made without a supplementary observation, is most greatly in error are found to perform better than those that seek the oceanic site to which a chosen land site is most sensitive at a chosen range. Among the former are strategies based on the "breeding" method, a variant of singular vectors, and ensembles of "replicated" observations; the last of these outperforms the others. The authors speculate as to the applicability of these findings to models with more realistic dynamics or without extensive regions devoid of routine observations, and to the real world.

The Lorenz-96 model (L96)

Environmental Prediction Science Laboratory

and at short or extended range. A search for an answer has constituted a part of the recent Fronts and Atlantic Storm-Tracks Experiment (FASTEX) (e.g., Snyder 1996; Joly et al. 1997).

Various strategies for locating the new observations suggest themselves. Some seek the regions where the analyses, as performed without the new observations, will be most greatly in error. Others try to target the locations where the present weather conditions will most strongly influence the subsequent weather. Each strategy possesses many possible variants. To test a large number of them adequately in the field within a reasonable period, once the platforms are available, seems to be out of the question. In these days when mathematical models of the weather are rife, it is virtually an axiom that one or more of them should be used for our first tests. This can be done even before any platforms are ready; nevertheless, we should anticipate that, whatever strategy we may decide upon, the need for modifications, at least of the details, will become apparent as soon as the new observations become a reality.

Because of the large number of specific procedures that might be tested, and the considerable number of simulated weather situations to which each must be applied before a definitive choice among them can be made, a full-scale experiment using a reasonably sophisticated model, such as the operational model of a

We have chosen a model with J variables, denoted by $X_1, \ldots, X_J$; in most of our experiments we have let $J = 40$. The governing equations are

$$dX_j/dt = (X_{j+1} - X_{j-2})X_{j-1} - X_j + F, \qquad (1)$$

for $j = 1, \ldots, J$. To make Eq. (1) meaningful for all values of j we define $X_{-1} = X_{J-1}$, $X_0 = X_J$, and $X_{J+1} = X_1$, so that the variables form a cyclic chain, and may be looked at as values of some unspecified scalar meteorological quantity, perhaps vorticity or temperature, at J equally spaced sites extending around a latitude circle. Nothing will simulate the atmosphere's latitudinal or vertical extent.

We know of no way that the model can be produced by truncating a more comprehensive set of meteorological equations. We have merely formulated it as one of the simplest possible systems that treats all variables alike and shares certain properties with many atmospheric models, namely,

1) the nonlinear terms, intended to simulate advection, are quadratic and together conserve the total energy, defined as $(X_1^2 + \cdots + X_J^2)/2$;
2) the linear terms, representing mechanical or thermal dissipation, decrease the total energy;
3) the constant terms, representing external forcing, prevent the total energy from decaying to zero.

Time Integration Schemes

$$y' = f(t, y)$$

Euler method (EUL)

$$y(t+1) = y(t) + f(t, y)h$$

Runge-Kutta 2nd order (RK2)

$$y(t+1) = y(t) + (k_1 + k_2)h/2$$

$$k_1 = f(t, y)$$

$$k_2 = f(t + h, y + hk_1)$$

Runge-Kutta 4th order (RK4)

$$y(t+1) = y(t) + (k_1 + 2k_2 + 2k_3 + k_4)h/6$$

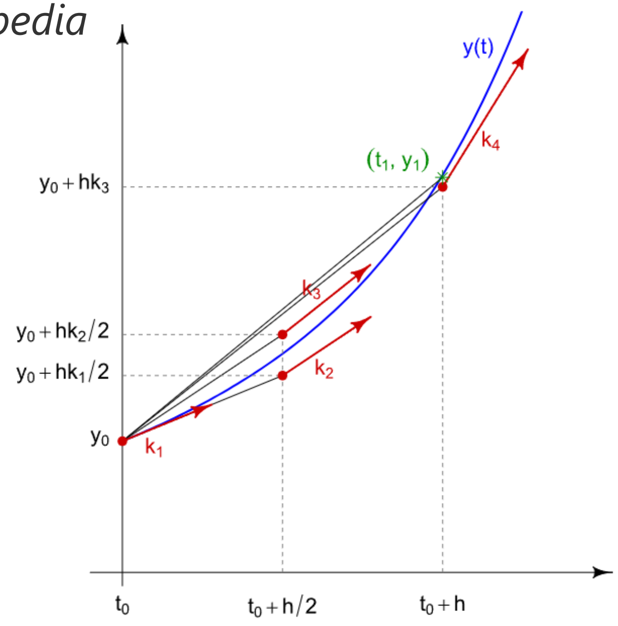
$$k_1 = f(t, y)$$

$$k_2 = f(t + h/2, y + hk_1/2)$$

$$k_3 = f(t + h/2, y + hk_2/2)$$

$$k_4 = f(t + h, y + hk_3)$$

from wikipedia



DA Study w/ 40-variable Lorenz-96

Lorenz-96 model (Lorenz 1996)

For $j=1,\ldots,N$, $X_j = X_{j+N}$

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

Initial Condition

$$X_i = \begin{cases} 1.001F & \left(i = \frac{N}{2}\right) \\ F & \text{(otherwise)} \end{cases}$$

DA Study w/ 40-variable Lorenz-96

Lorenz-96 model (Lorenz 1996)

For $j=1,\ldots,N$, $X_j=X_{j+N}$

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

Lorenz et al. assumed 1 time unit = 5 days based on error doubling 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.

(N: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.

DA Study w/ 40-variable Lorenz-96

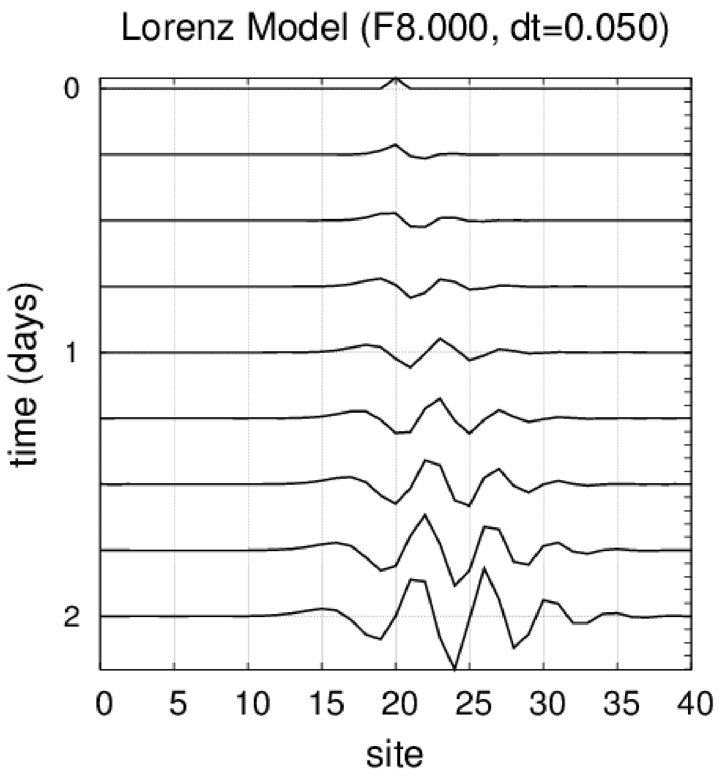
Lorenz-96 model (Lorenz 1996)

For j=1,…,N, $X_j = X_{j+N}$

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

Advection term Dissipation term Forcing term

Lorenz et al. assumed 1 time unit = 5 days based on error doubling time.



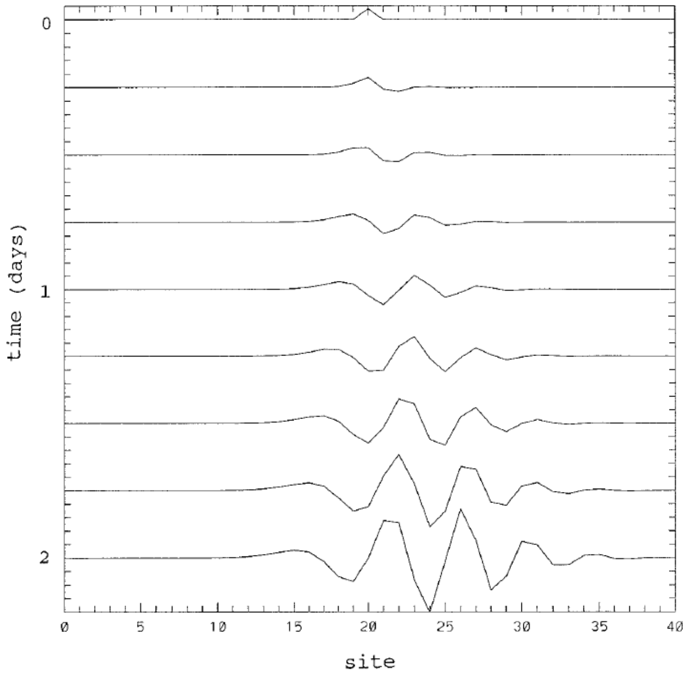
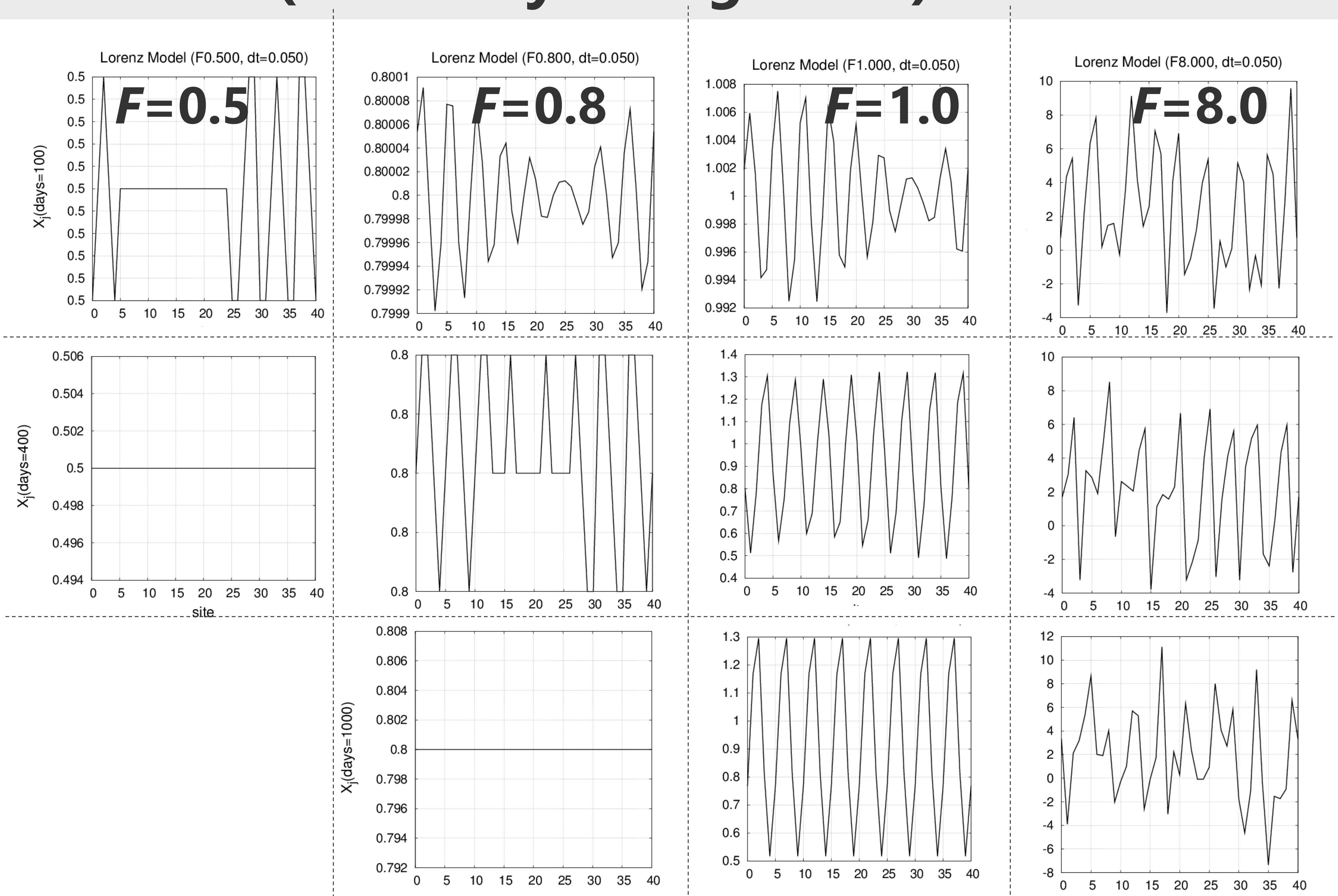


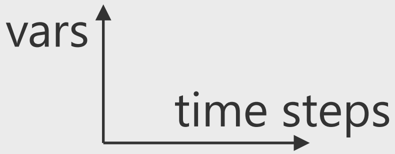
FIG. 1. Longitudinal profiles of X_j at 6-h intervals, as determined by Eq. (1) with $N = 40$ and $F = 8.0$, when initially $X_{20} = F + 0.008$ and $X_j = F$ when $j \neq 20$. On horizontal portion of each curve, $X_j = F$. Interval between successive short marks at left and right is 0.01 units.

Ultimate States (1000-days integration)

Environmental Prediction Science Laboratory

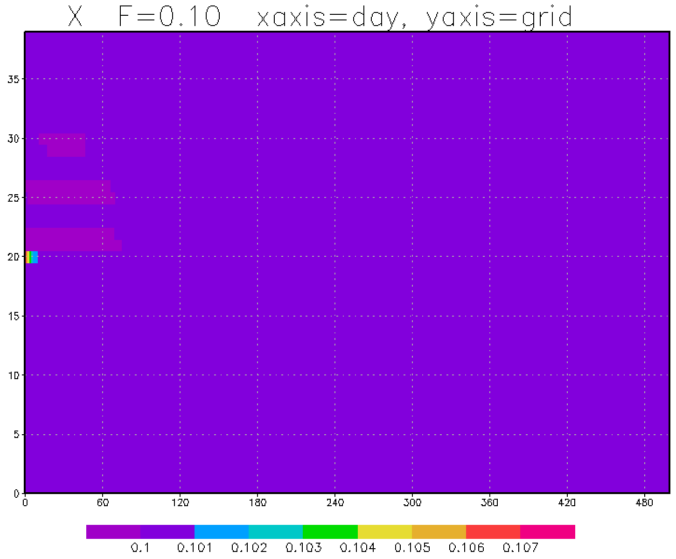


Hovmöller Diagram

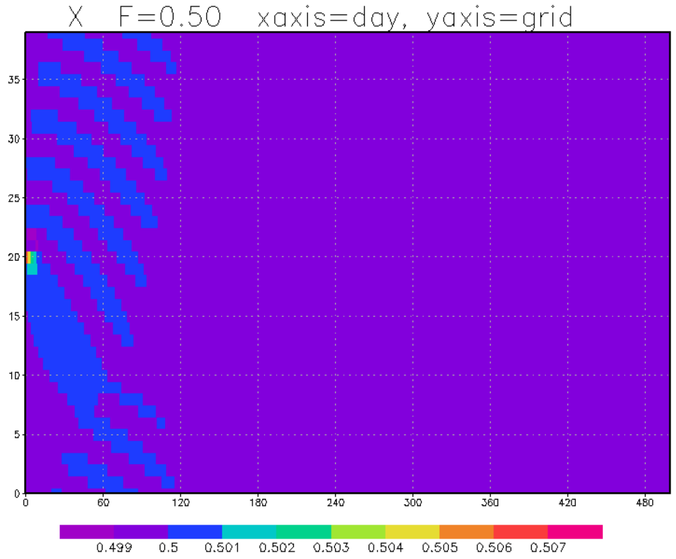


Hofmüller

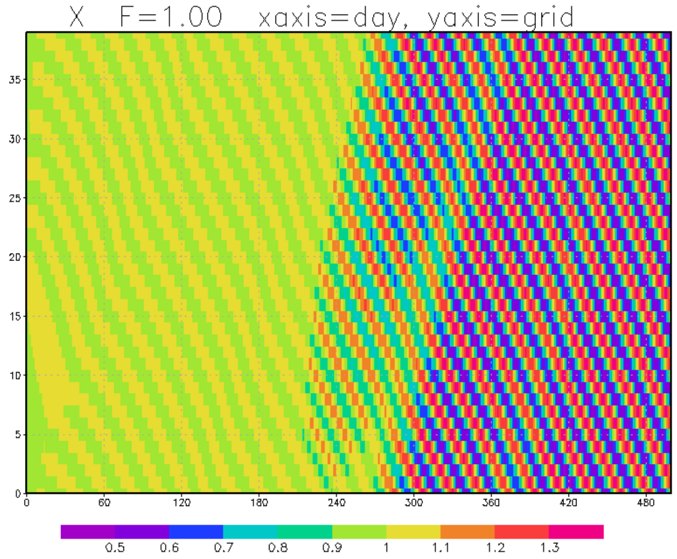
F=0.10



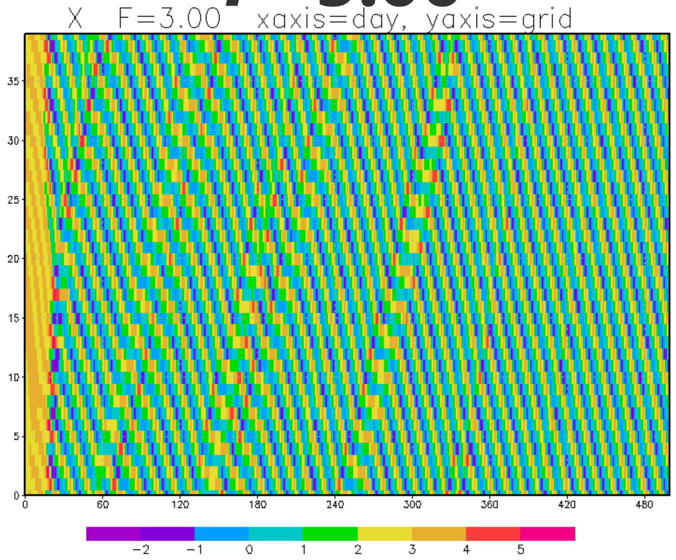
F=0.50



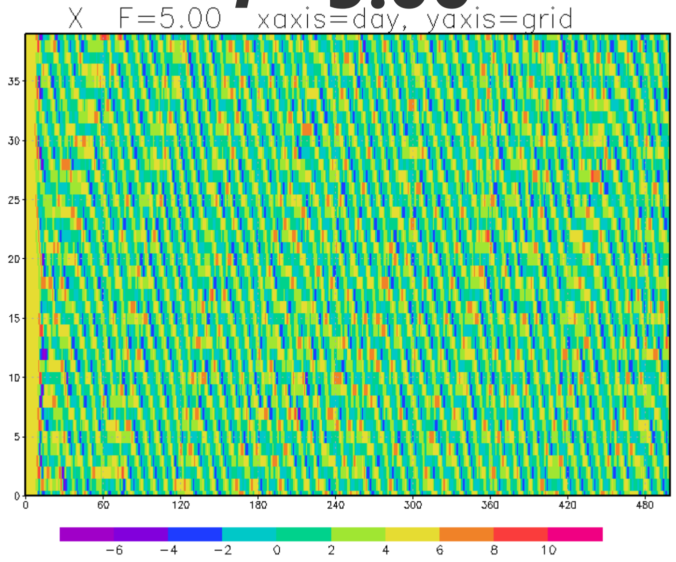
F=1.00



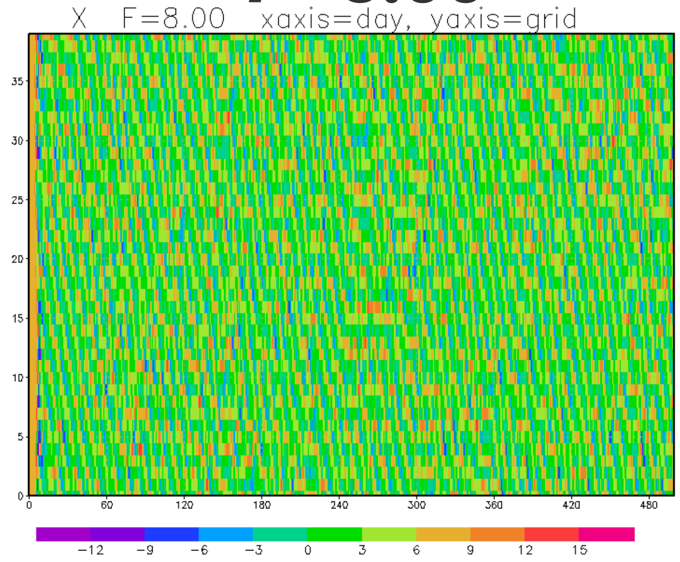
F=3.00



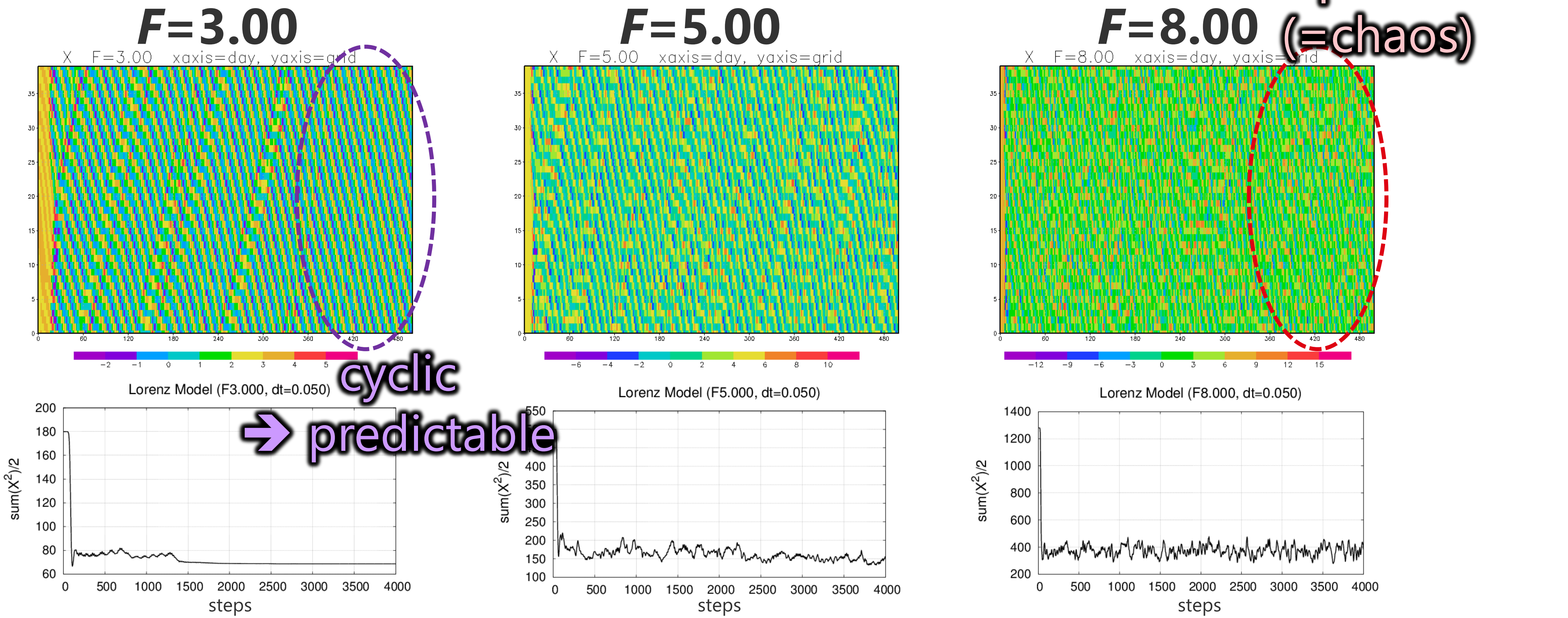
F=5.00



F=8.00



Stable or Chaotic?



$$dX_j/dt = (X_{j+1} - X_{j-2})X_{j-1} - X_j + F$$

Advection term　Dissipation term　Forcing term

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

Time Integration Schemes

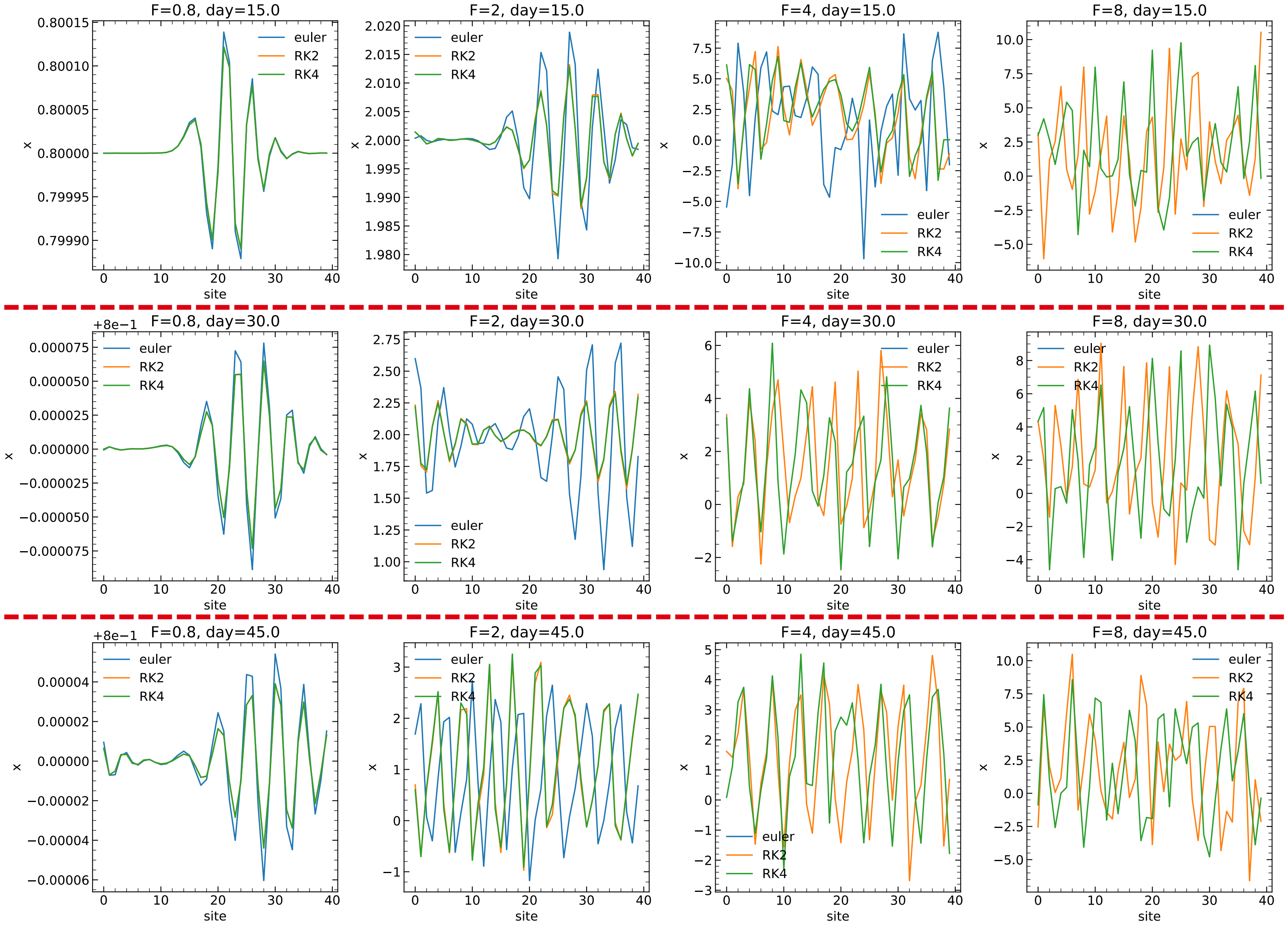
Environmental Prediction Science Laboratory

$F = 0.8$ | $F = 2.0$ | $F = 4.0$ | $F = 8.0$

day 15

day 30

day 45



Basic Task 2

Training Course Problem (2)

2．パラメータ値 F=8 とする。誤差の平均発達率について調べ、0.2 時間ステップを 1 日と定義することの妥当性を確認する。

ヒント）Lorenz (1996)の"error doubling time"の議論をフォローすると良い。データ同化コミュニティでは誤差は通常、root mean square error (RMSE)で評価するので、以後 RMSE で評価すること。

2. The parameter F is set to be 8. Investigate the average of error generating ratio and validate the definition of a 0.2-hour step as one day.
Hint) Follow Lorenz's (1996) discussion of "error doubling time". In the data assimilation community, the error is usually evaluated by root mean square error (RMSE), so it should be evaluated by RMSE hereafter.

To Calc. Error Doubling Time

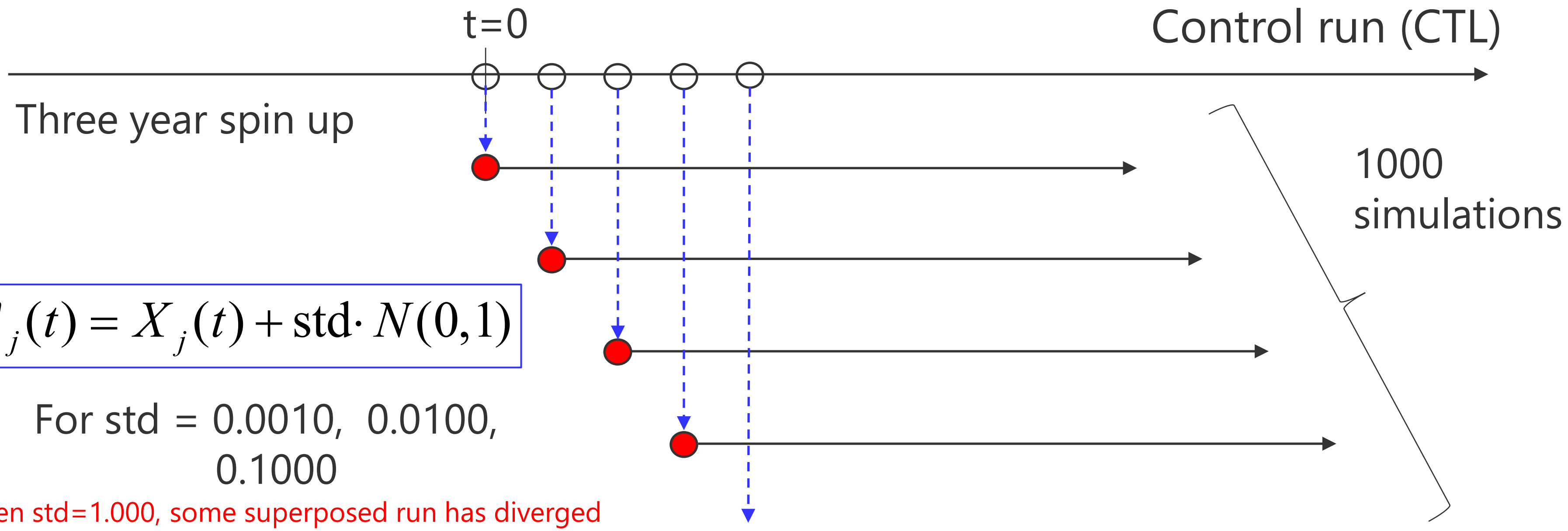
1. Mersenne Twistter method

   : 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

When std=1.000, some superposed run has diverged

To Generate *N*(0,1) Numbers

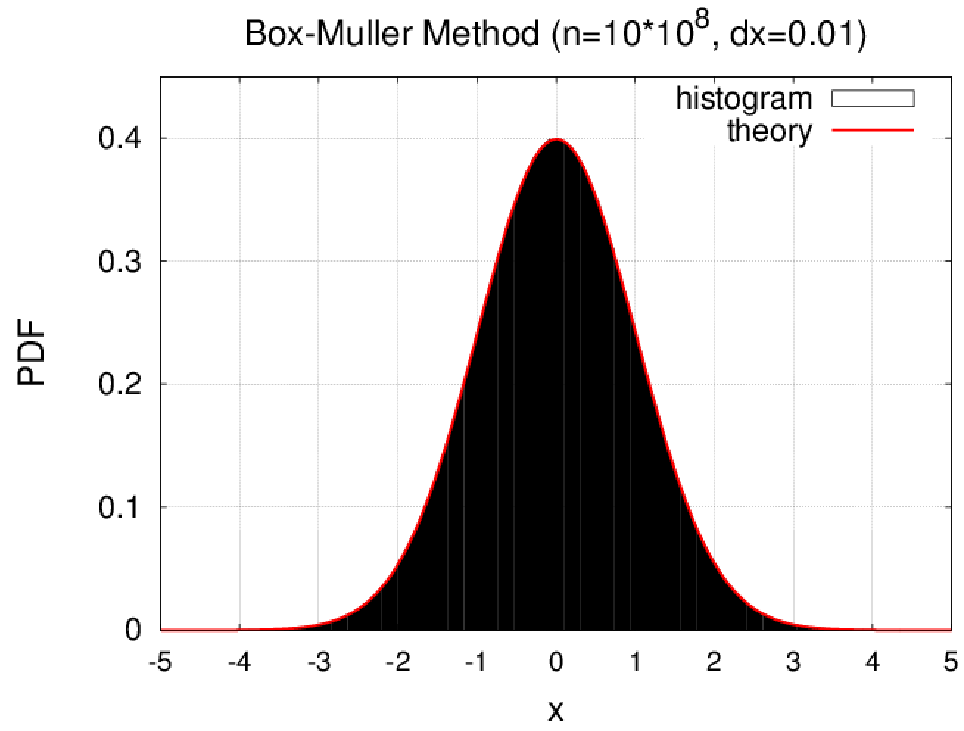
1. Mersenne Twistter method
   : To generate pseudo-random number sampling

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

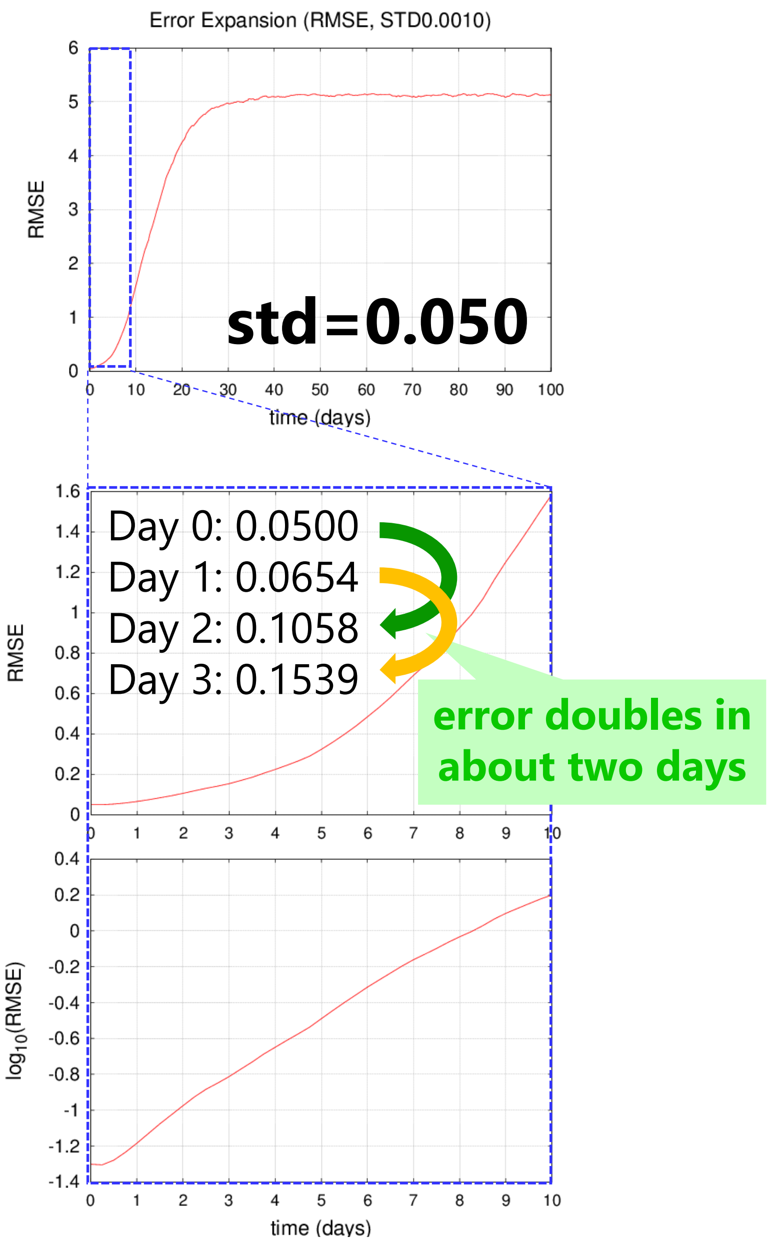
Box-Muller's method:

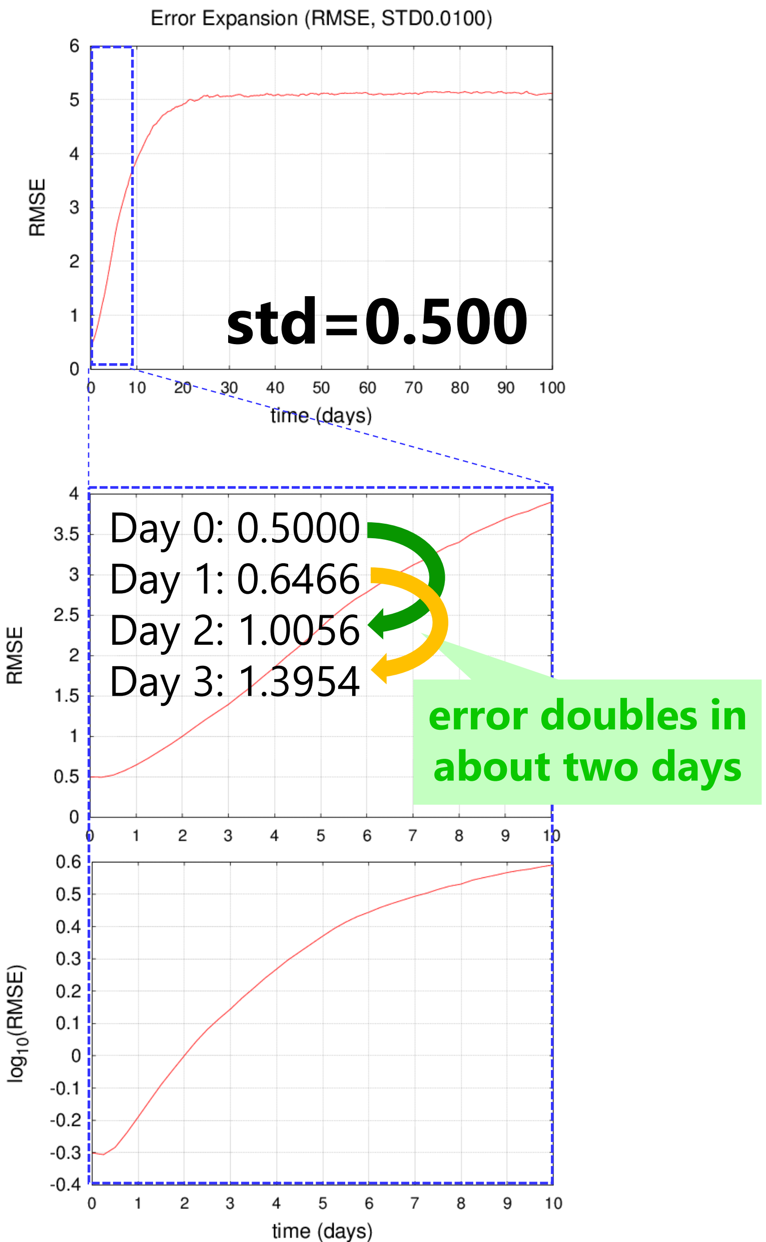
When X and Y obey uniform distribution (0,1), Z_1 and Z_2, defined by following equations, obey normal distribution N(0,1).

$$Z_1 = \sqrt{-2\log X}\cos 2\pi Y$$
$$Z_2 = \sqrt{-2\log X}\sin 2\pi Y$$

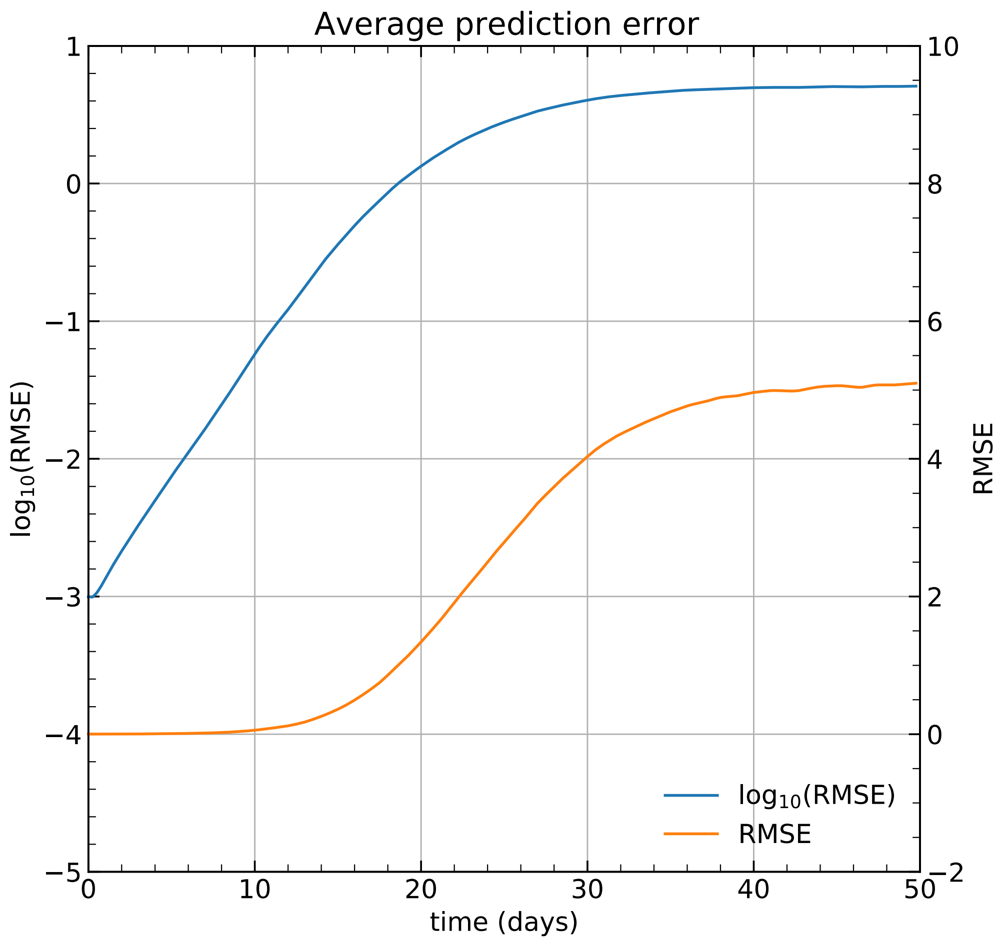


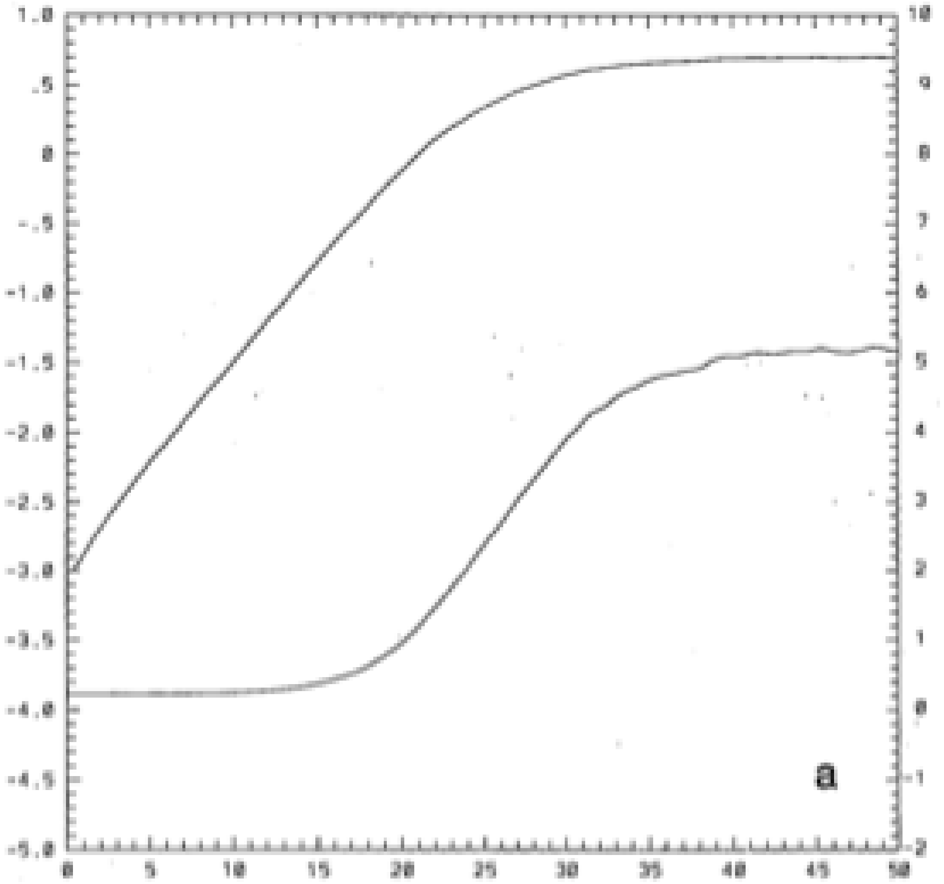
Error Propagation Analysis





Error Propagation Analysis





Lorenz (1996) Fig.2a std=0.001

Thank you for your attention!

Presented by Shunji Kotsuki
(shunji.kotsuki@chiba-u.jp)

Further information is available at
https://kotsuki-lab.com/

Environmental Prediction Science Laboratory