



# 「観測の価値」を最大化する データ同化手法の開発 ～最先端の天気予報研究より～

**Shunji Kotsuki**  
[\(shunji.kotsuki@chiba-u.jp\)](mailto:shunji.kotsuki@chiba-u.jp)

Associate Professor  
Center for Environmental Remote Sensing (CEReS), Chiba Univ.

\* PRESTO, Japan Science and Technology

**2021/02/19 設計と運用に活かすデータ同化研究会**

# Who am I ? :: Shunji Kotsuki (小槻 峻司)

高知

(1986)

京都大学・工学部

(2005)



理研・計算科学

(2014)

Postdoc



文科省・卓越研究員  
(2017-2019)

千葉大・CEReS  
(2019)

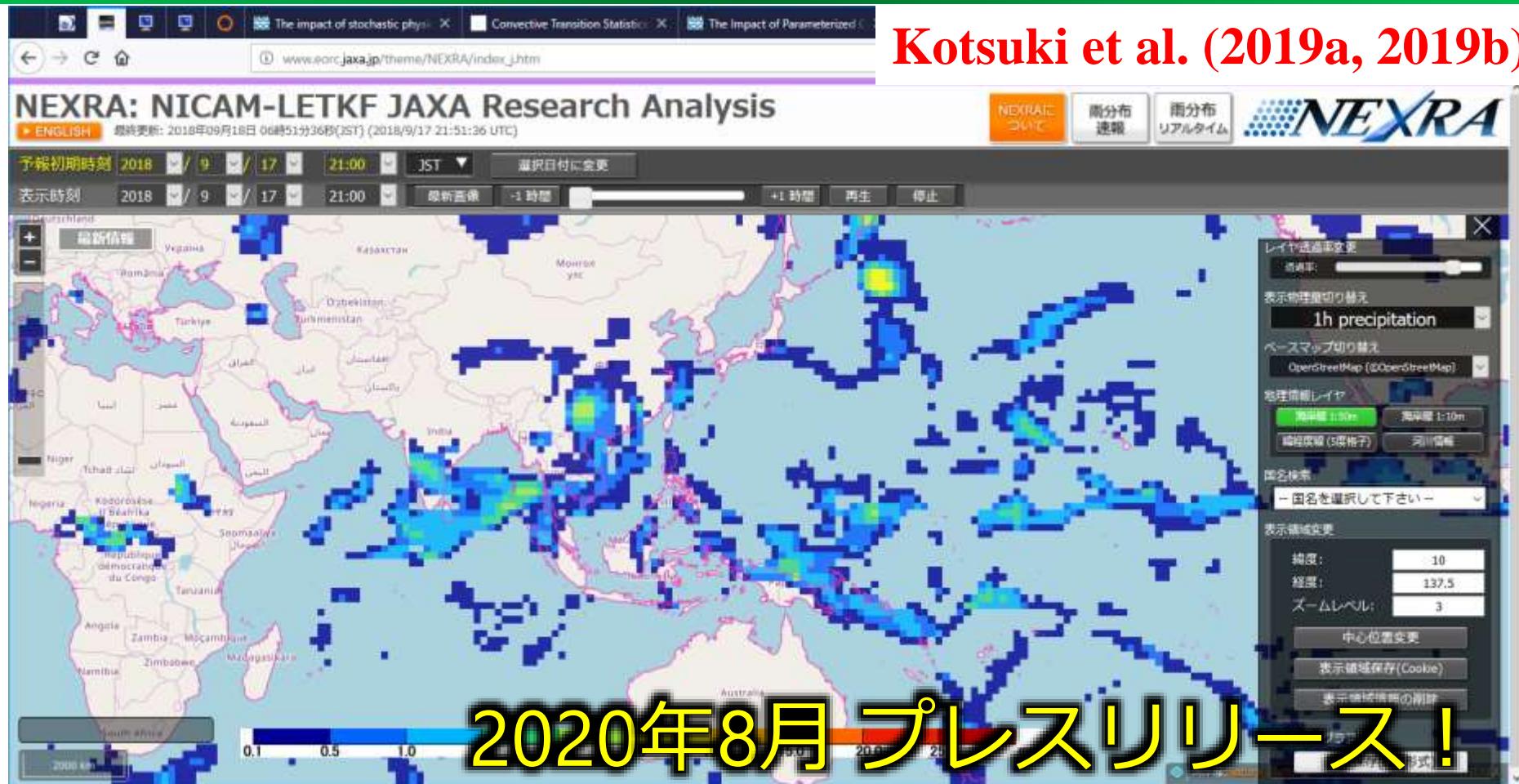
Associate Prof.



JST・さきがけ研究員 (2019-)  
理研計算・客員研究員 (2019-)

# コアアプリ: JAXA全球天気予報

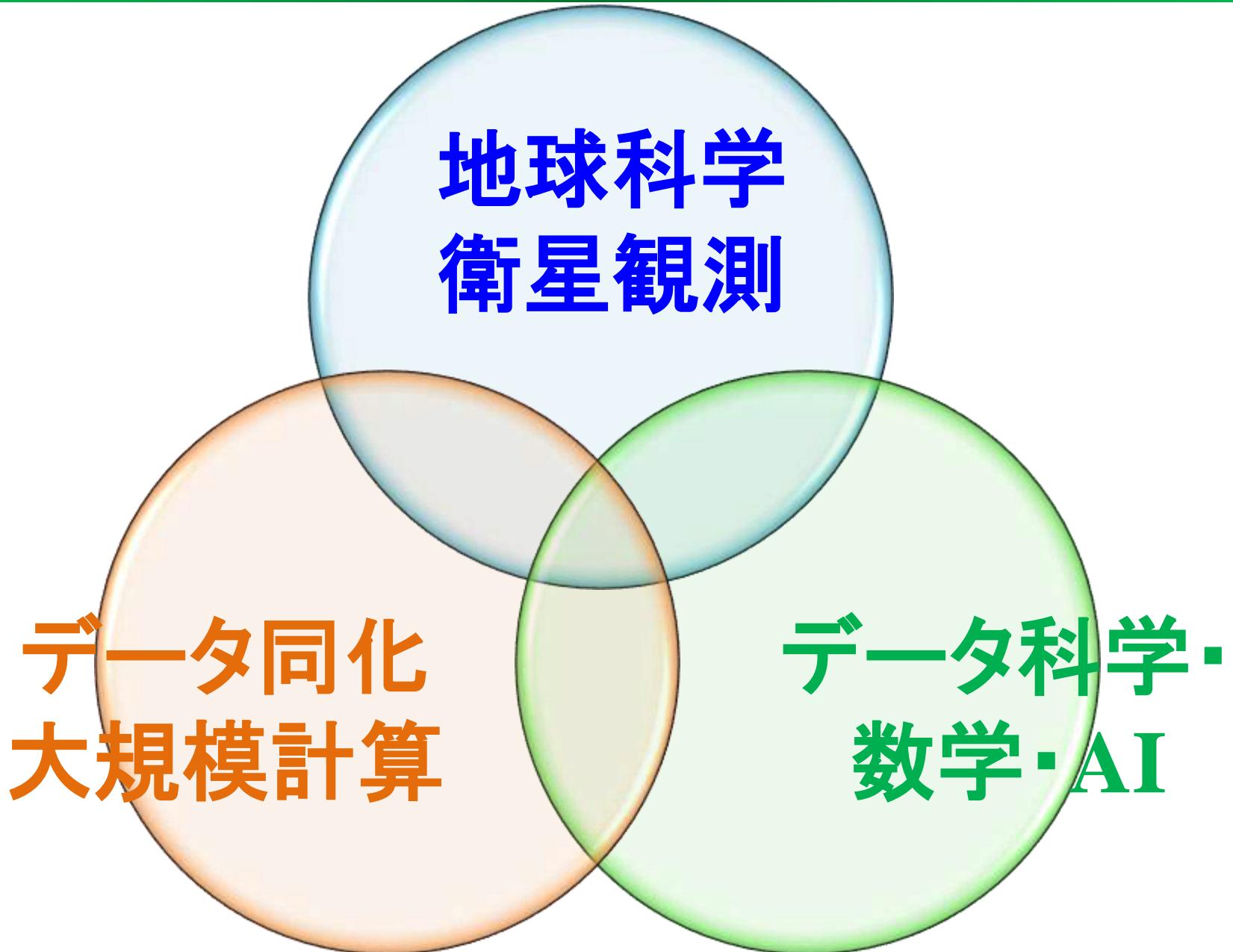
Kotsuki et al. (2019a, 2019b)



- JAXA,理研, 東大と共同開発。
- スパコン「富岳」も使い研究推進。
- 社会生活に直結するDEEPな研究！

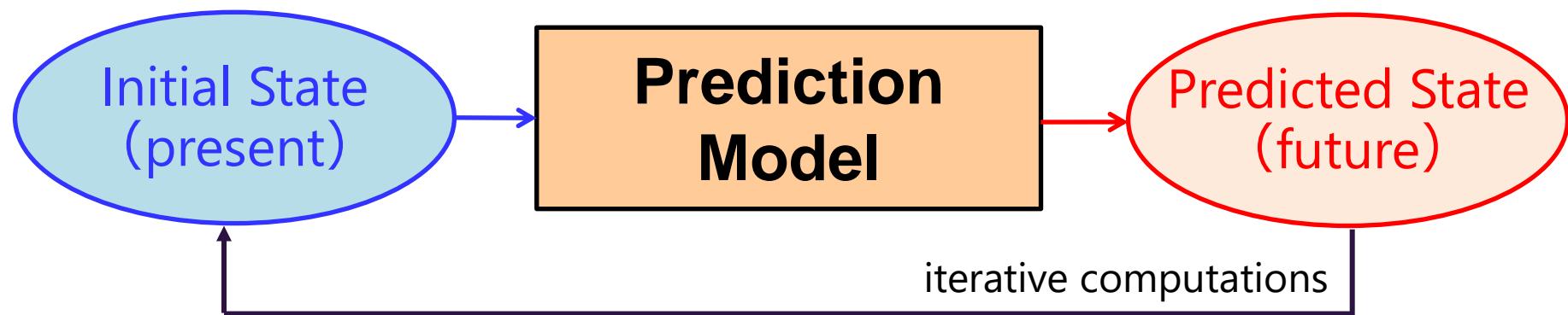


# 研究室の3本柱

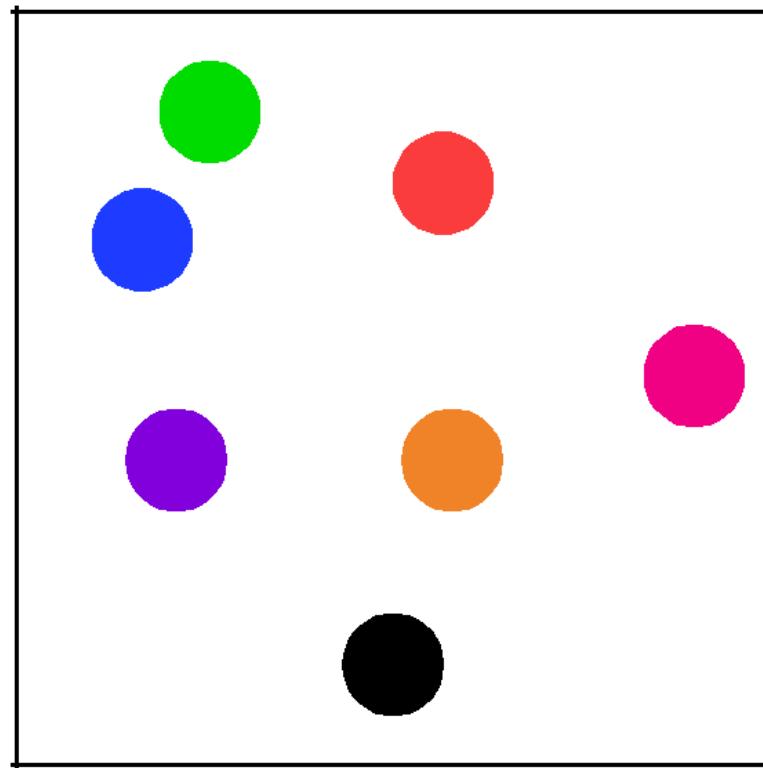


# **データ同化について 導入**

# Numerical Simulation w/ Computers

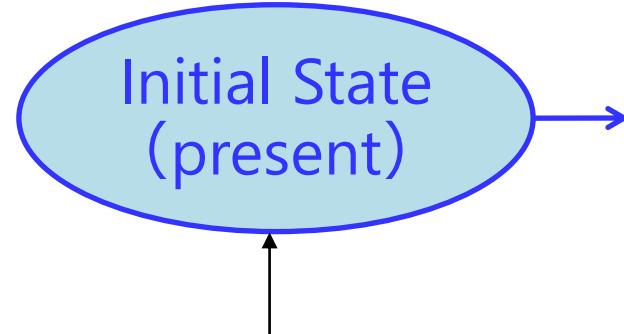


*an example  
of billiards*



# Numerical Weather Prediction

wind, temperature, humidity, pressure



wind, temperature, humidity, pressure

Predicted State (future)

iterative computations

・水平方向の運動方程式

$$\frac{\partial u}{\partial t} - 2\Omega \sin \theta v + \frac{1}{a \cos \theta} \frac{\partial \phi}{\partial \lambda} = -\mathbf{V} \cdot \nabla u - \omega \frac{\partial u}{\partial p} + \frac{\tan \theta}{a} uv + F_u \quad (1)$$

$$\frac{\partial v}{\partial t} - 2\Omega \sin \theta u + \frac{1}{a} \frac{\partial \phi}{\partial \theta} = -\mathbf{V} \cdot \nabla v - \omega \frac{\partial v}{\partial p} - \frac{\tan \theta}{a} uu + F_v \quad (2)$$

・熱力学第一法則の式

$$\frac{\partial c_p T}{\partial t} + \mathbf{V} \cdot \nabla c_p T + \omega \frac{\partial c_p T}{\partial p} = \omega \alpha + Q \quad (3)$$

・質量保存則

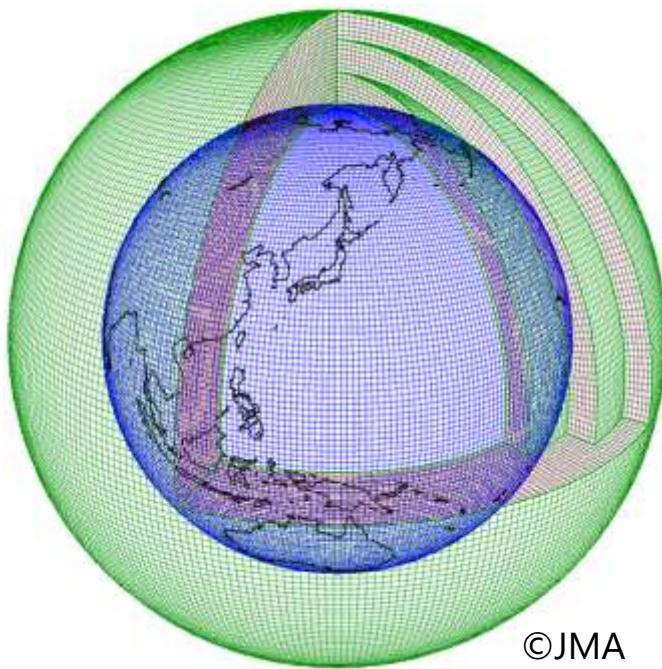
$$\frac{1}{a \cos \theta} \frac{\partial u}{\partial \lambda} + \frac{1}{a \cos \theta} \frac{\partial v \cos \theta}{\partial \theta} + \frac{\partial \omega}{\partial p} = 0 \quad (4)$$

・状態方程式

$$p\alpha = RT \quad (5)$$

・静力学平衡の式

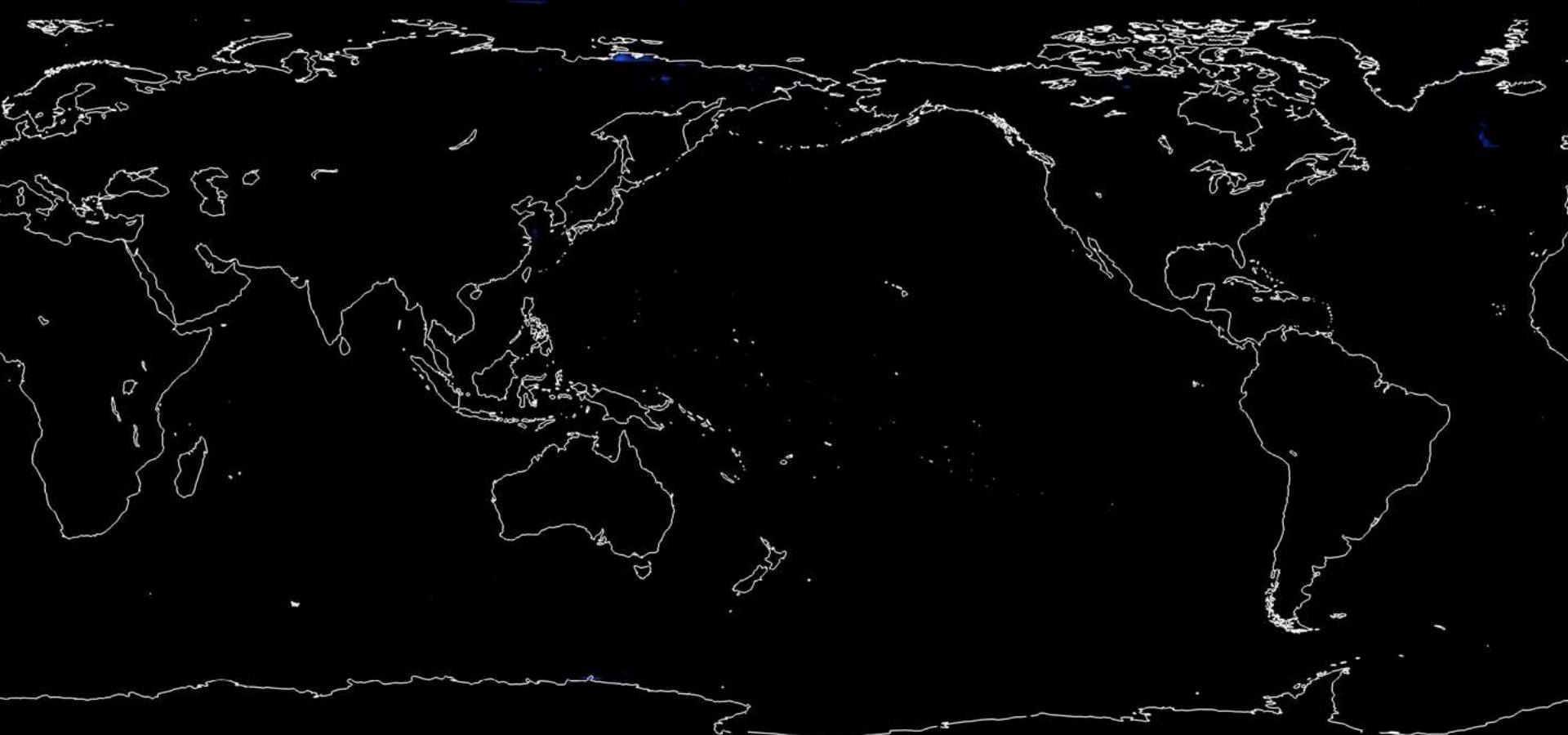
$$\frac{\partial \phi}{\partial p} = -\alpha \quad (6)$$



©JMA

3-dimensional grids of atmosphere

# Simulated Global Precipitation



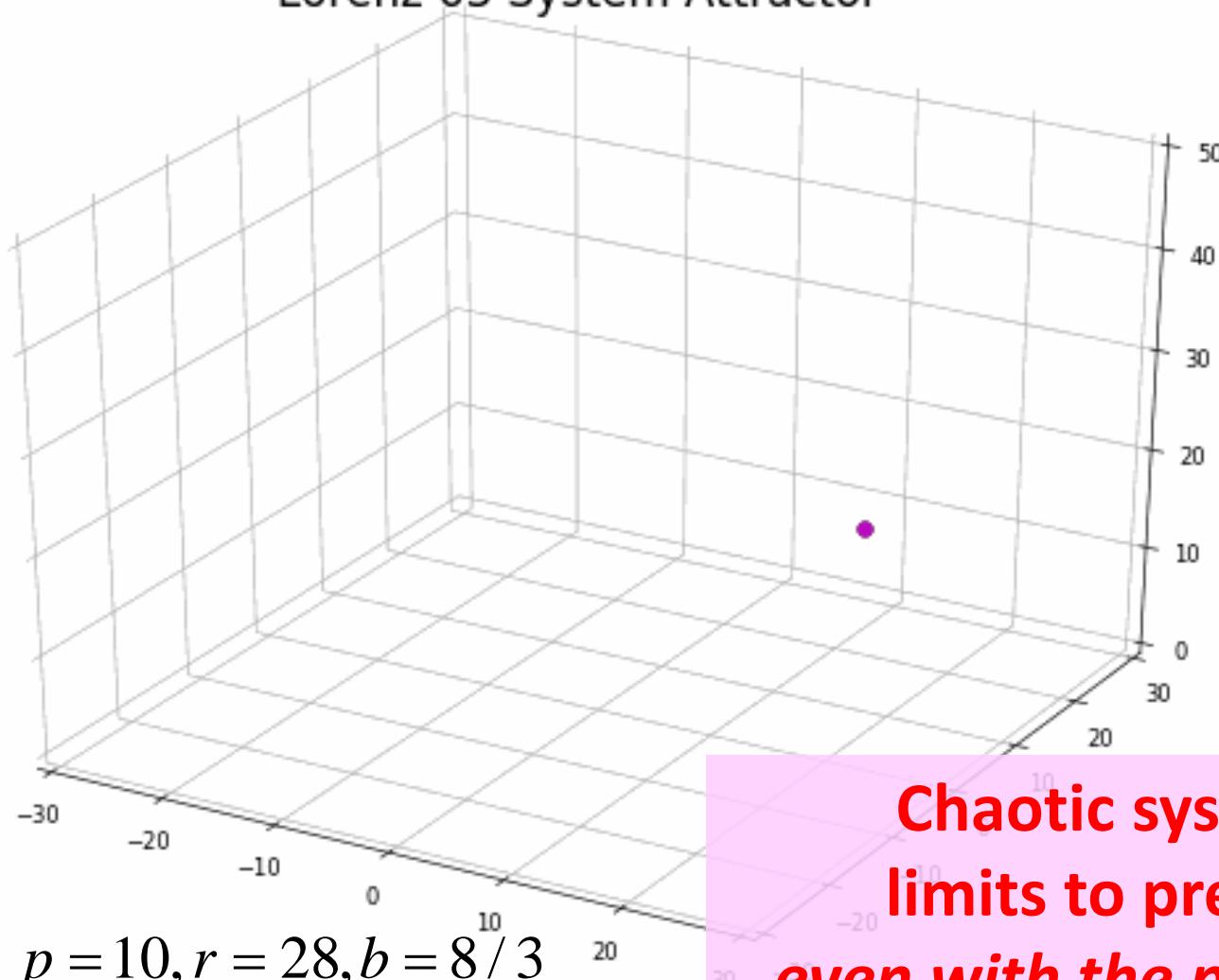
2014/05/25 00:00

# Deterministic Chaos & Predictability

Edward Lorenz



Lorenz 63 System Attractor



Lorenz 63 model

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

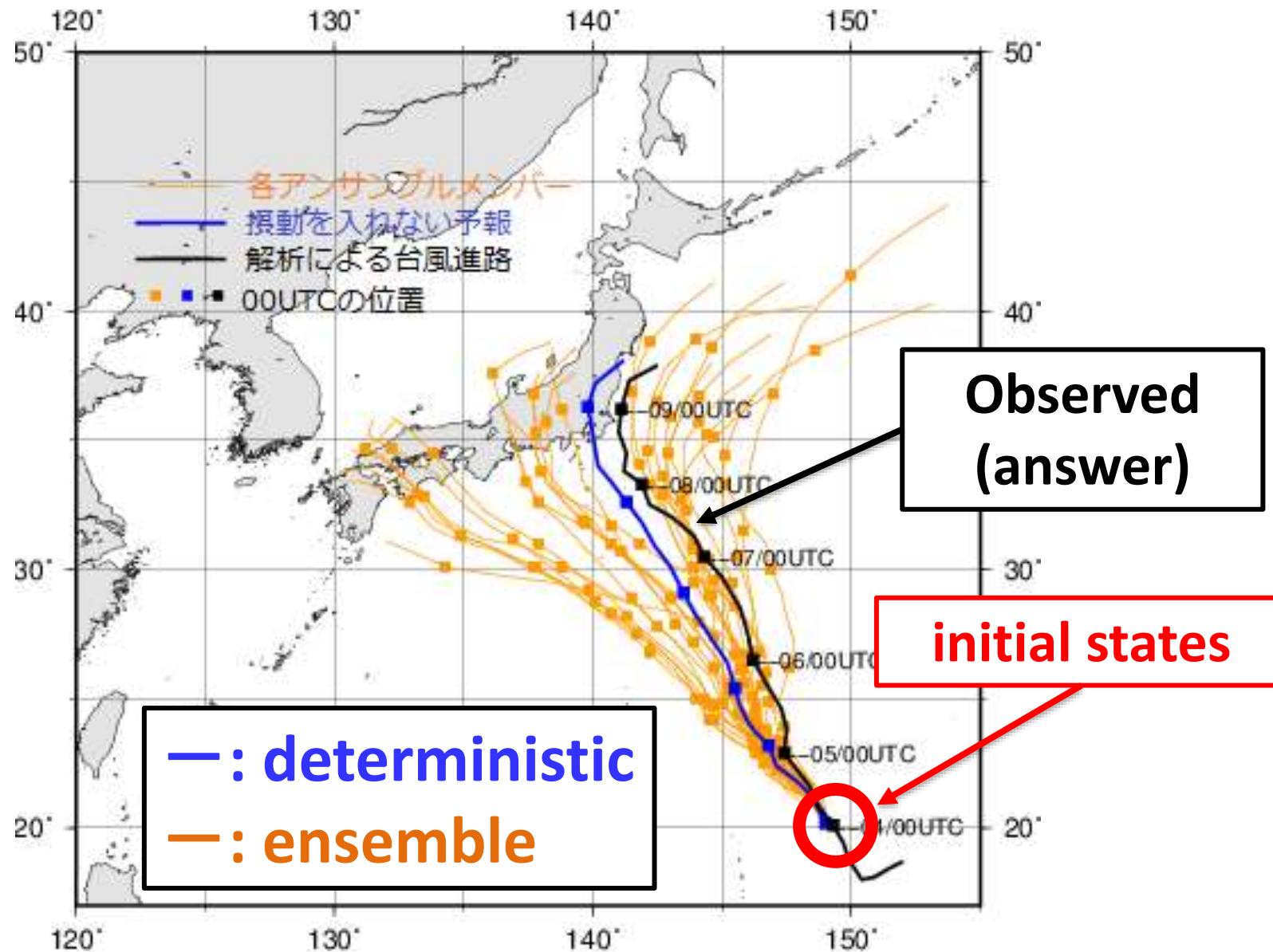
$$\dot{y} = -xz + rx - y$$

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

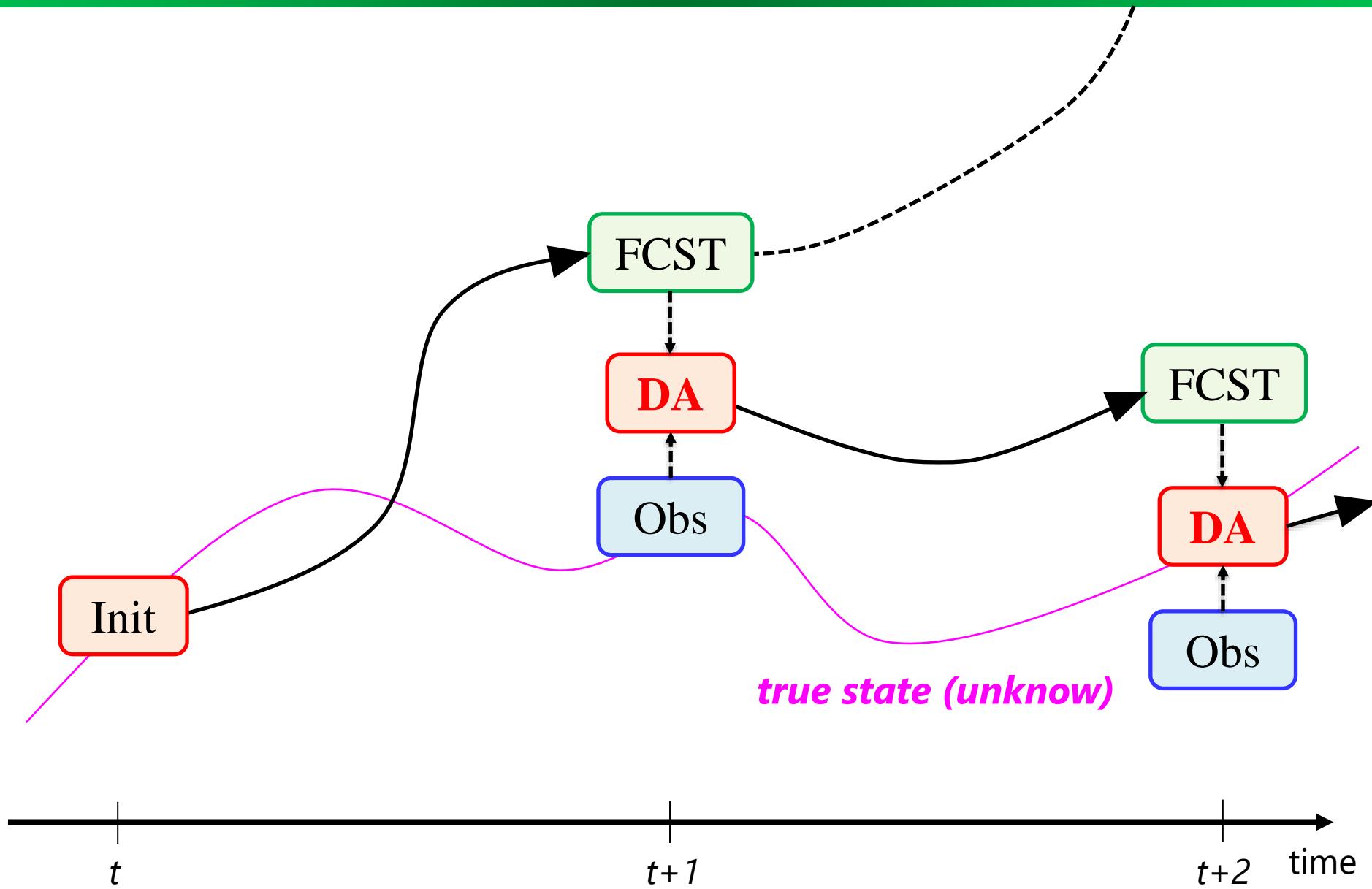
**Chaotic systems have  
limits to predictability  
even with the perfect model!!**

Initial Conditions ::  $x=y=z=15.000, 15.001, 15.002, \dots, 15.009$

# Ensemble Prediction: an example of typhoon

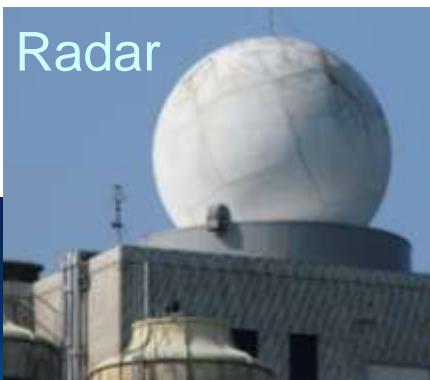


# Numerical Weather Prediction



# Global Observing System

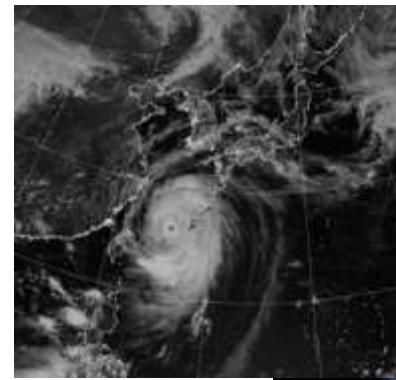
Radar



Aircraft



Satellite



Ship



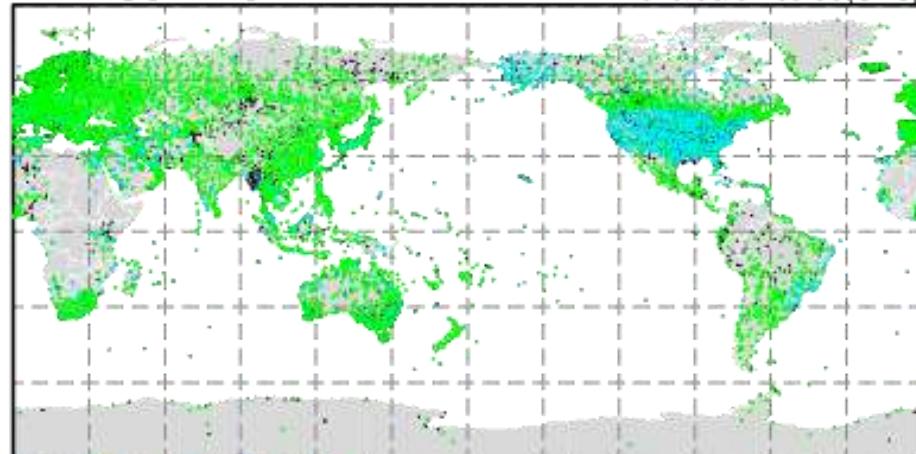
Buoy



# Observation Data in NWP: world's efforts!

LAND SURFACE

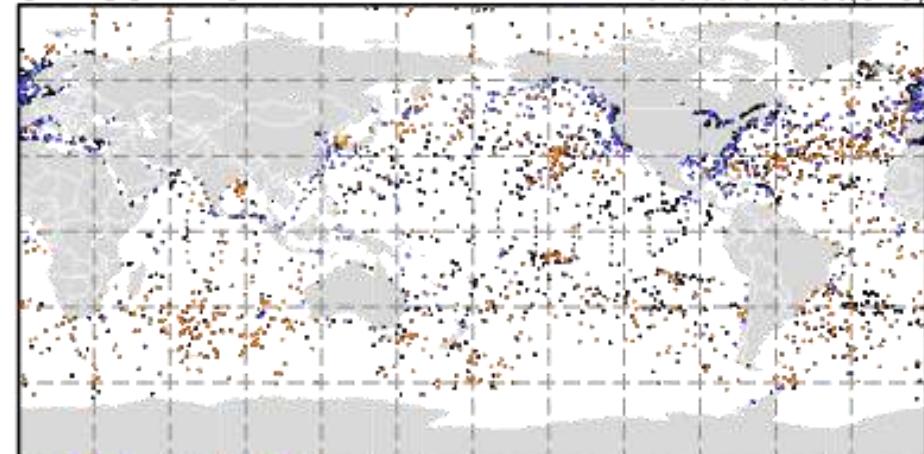
2019/05/01 00:00(UTC)



SYNOP(●): 4163      METAR(●): 1178  
NOUSE(●): 18901      NOUSE(●): 42966  
ALL: 23064      ALL: 44144

SEA SURFACE

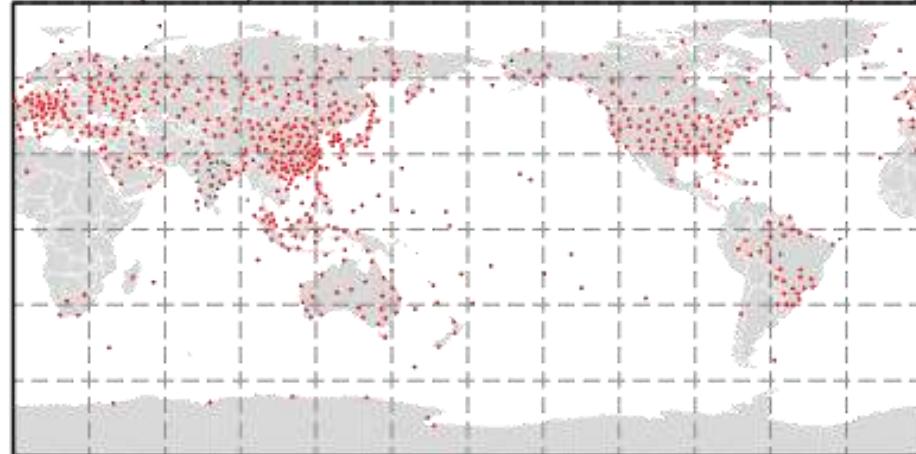
2019/05/01 00:00(UTC)



SHIP(●): 378      DRIFTER(●): 709  
NOUSE(●): 5668      NOUSE(●): 10206  
ALL: 6046      ALL: 10915

UPPER(TEMP)

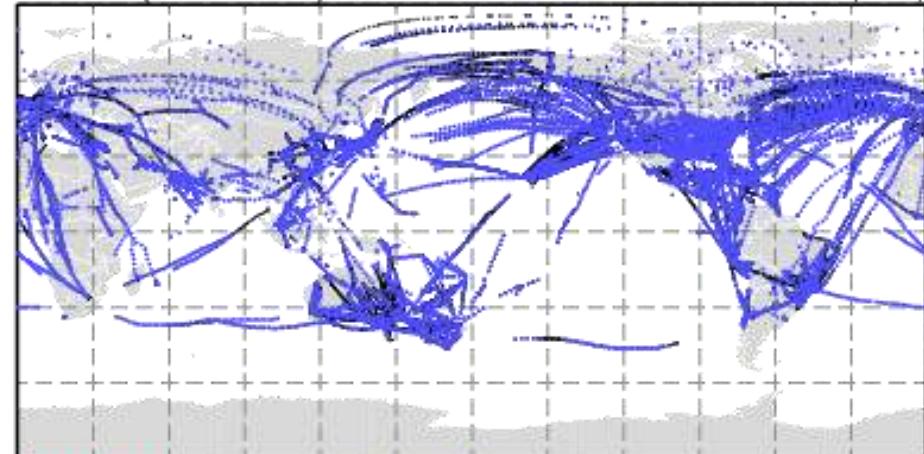
2019/05/01 00:00(UTC)



TEMP(●): 632  
NOUSE(●): 24  
ALL: 656

UPPER(AVIATION)/BOGUS

2019/05/01 00:00(UTC)

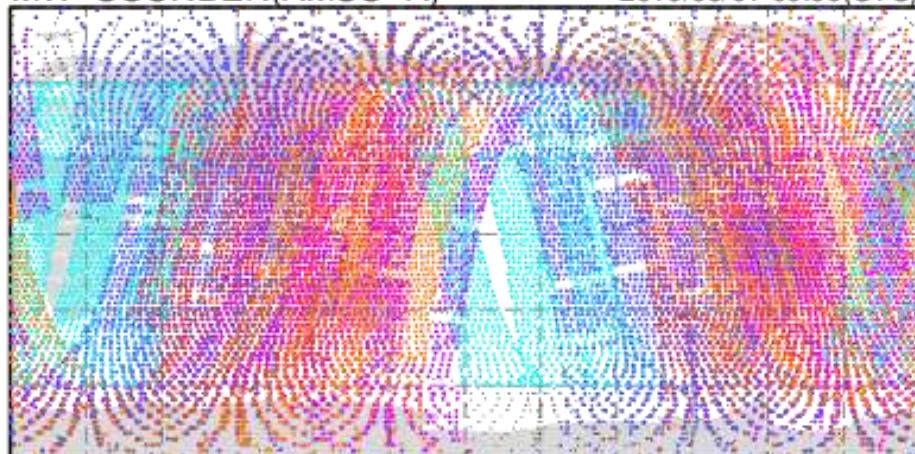


TYBOGUS(●): 0      YHTC  
NOUSE(●): 0      NOUSE(▼): 0      NOUSE(●): 88767  
ALL: 0      ALL: 0      ALL: 98686

courtesy of JMA (2019/05/01 00:00 UTC)

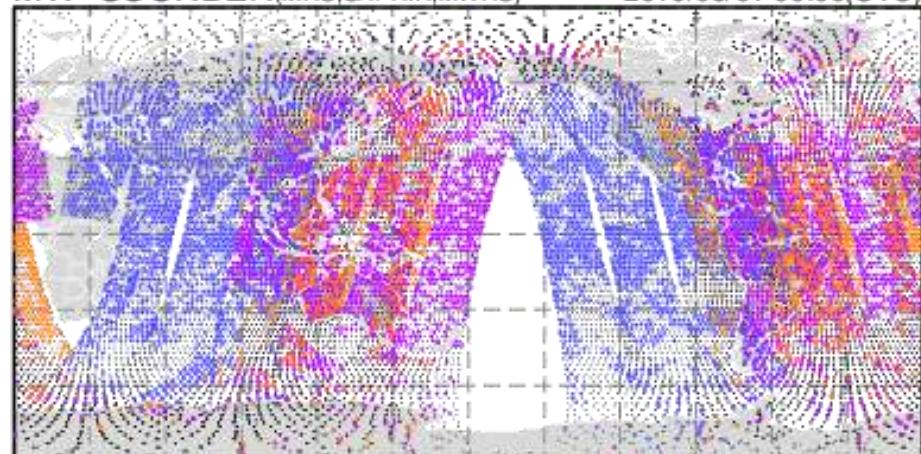
# Observation Data in NWP: world's efforts!

MW-SOUNDER(AMSU-A)



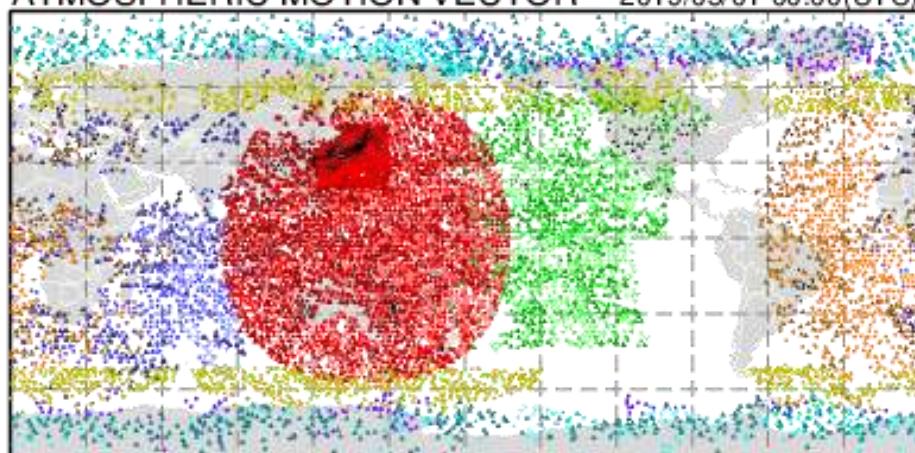
2019/05/01 00:00(UTC)

MW-SOUNDER(MHS,SAPIR,MWHS)



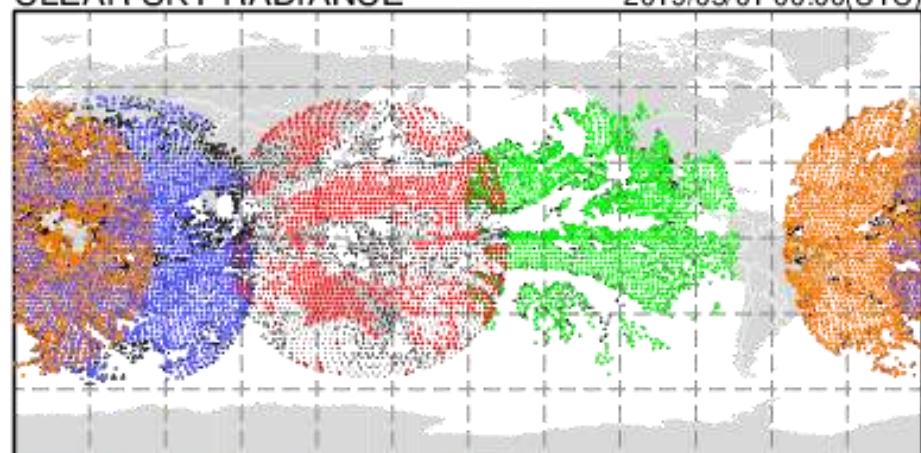
2019/05/01 00:00(UTC)

ATMOSPHERIC MOTION VECTOR



2019/05/01 00:00(UTC)

CLEAR SKY RADIANCE



2019/05/01 00:00(UTC)

2019/05/01 00:00(UTC)

Satellite	Count
NOUSE(●)	9881
NOUSE(▲)	133
NOUSE(■)	83
NOUSE(●)	132
NOUSE(●)	49
NOUSE(●)	146
NOUSE(●)	36
ALL:	13263
ALL:	1539
ALL:	961
ALL:	1266
ALL:	1501
ALL:	1865
ALL:	349

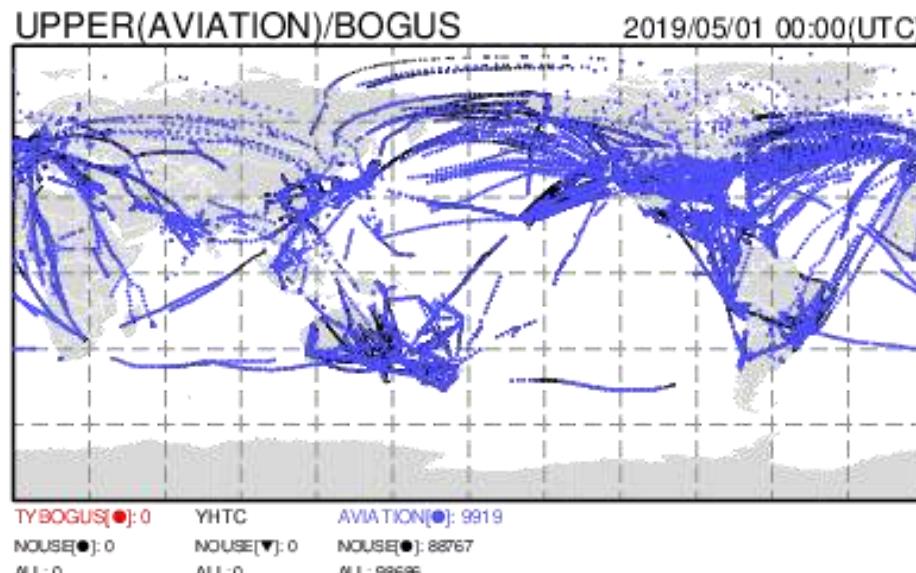
2019/05/01 00:00(UTC)

Satellite	Count
NOUSE(●)	8031
NOUSE(▲)	2109
NOUSE(●)	5723
NOUSE(●)	4329
ALL:	10843
ALL:	4366
ALL:	9817
ALL:	10184

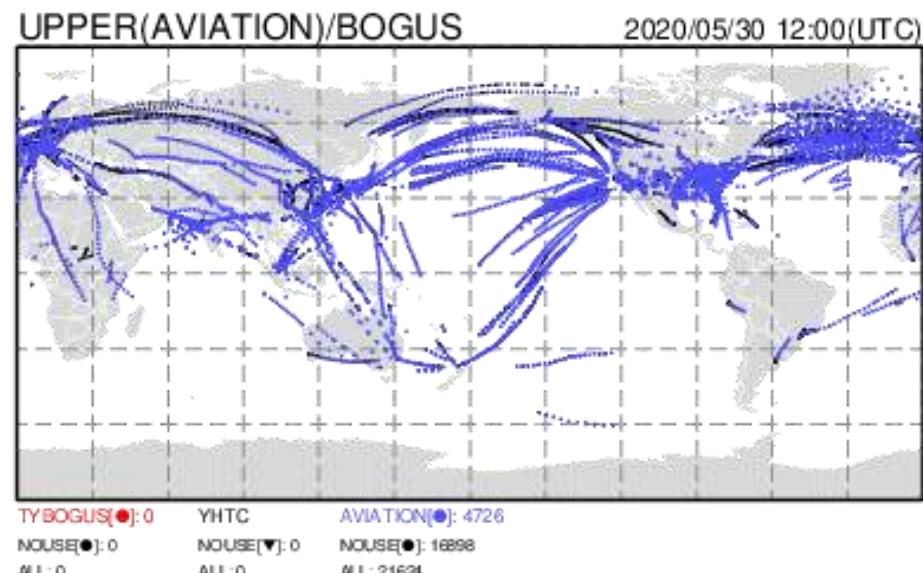
courtesy of JMA (2019/05/01 00:00 UTC)

# Observation Data in NWP: world's efforts!

***Before COVID-19***

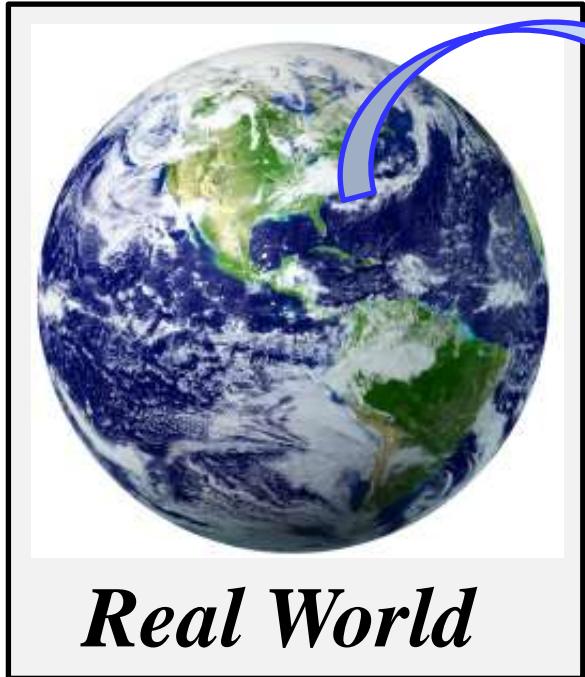


***in COVID-19 restriction***

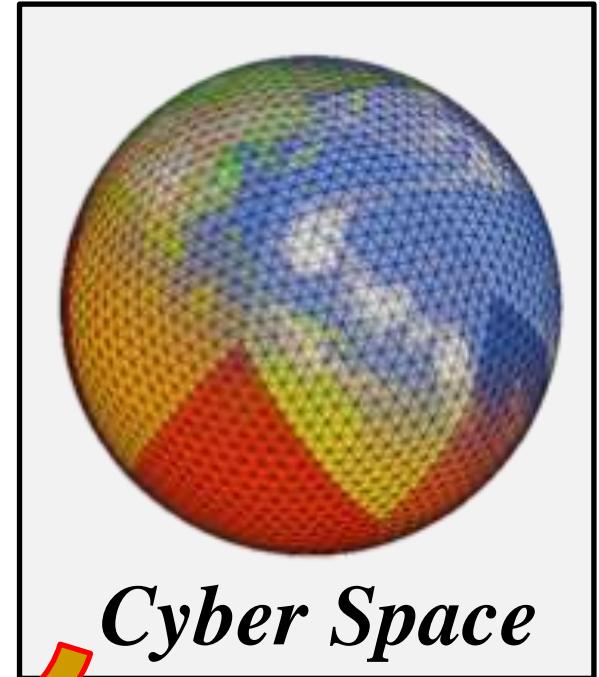


# Data Assimilation (DA)

Observation



**DA**  
(best estimates)



Prediction

Data-driven  
(inductive)

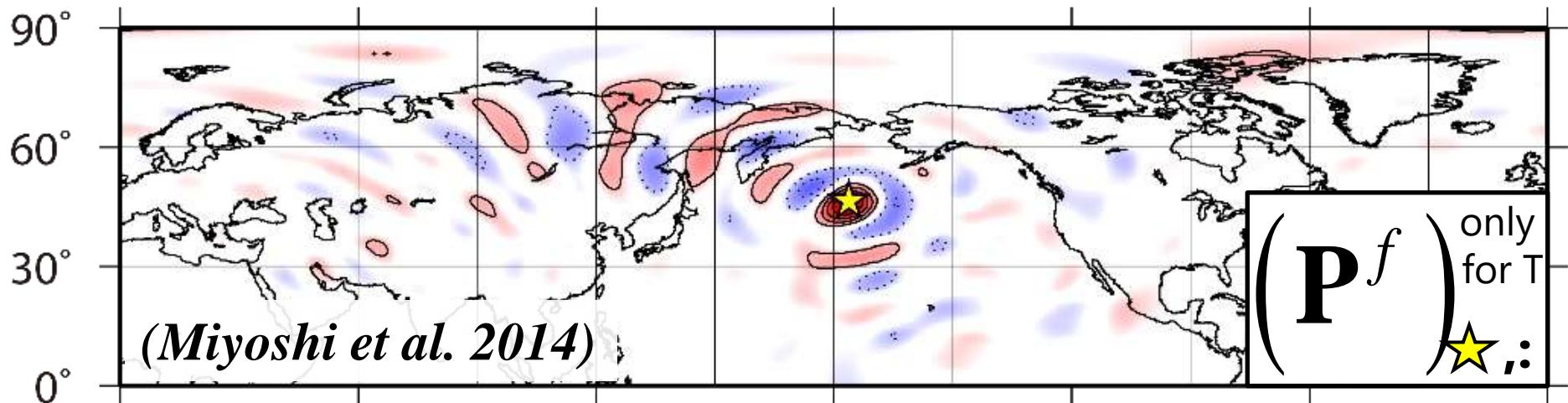
Process-driven  
(deductive)

Sparse & Infrequent  
(low-dimensional)

Dense & Frequent  
(high-dimensional)

# Error “co-variance”; an example of mid. troposphere

## An error covariance structure with 10,000-member SPEEDY



$$\mathbf{P}^f = \frac{1}{m-1} \delta \mathbf{X}^f \left( \delta \mathbf{X}^f \right)^T$$

model dimension

$$\delta \mathbf{X}^f$$

ensemble size

Simply to say, covariance sees the error “**correlation**”.

covariance

$$\sigma_{XY} = E[(X - E[X])(Y - E[Y])]$$

correlation

$$\rho = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

variance  
=std<sup>2</sup>

$$\sigma_X^2 = E[(X - E[X])^2]$$

$$\sigma_Y^2 = E[(Y - E[Y])^2]$$

# How can we assimilate dense observations?

$$\mathbf{P}^f = \frac{1}{m-1} \delta \mathbf{X}^f (\delta \mathbf{X}^f)^T \quad n \times n$$

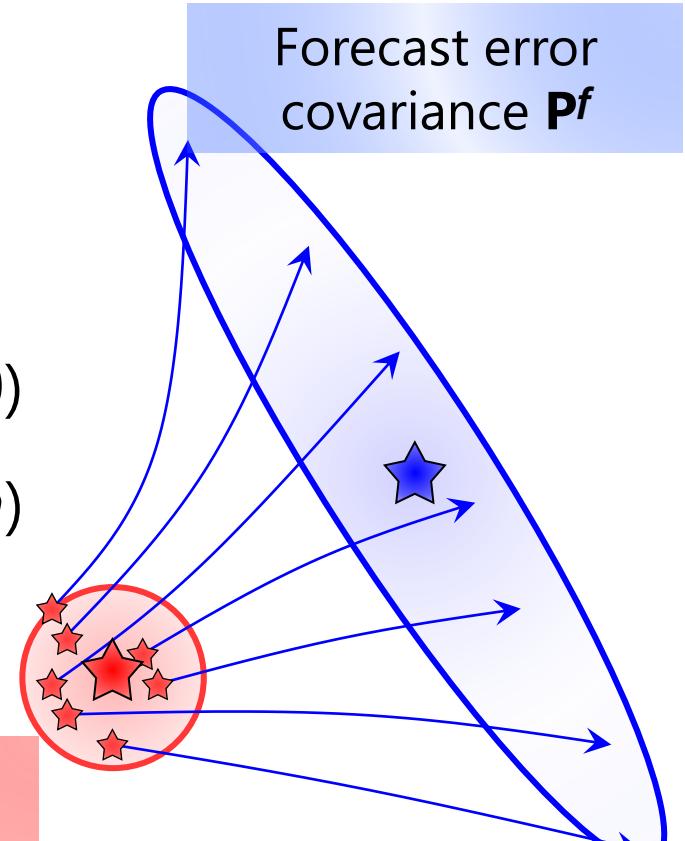
$n$  : # of model variables ( $> 10^8$ )

$m$  : ensemble size (usually  $< 1,000$ )

$\delta \mathbf{X}^f$  : ensemble perturbation ( $n \times m$ )

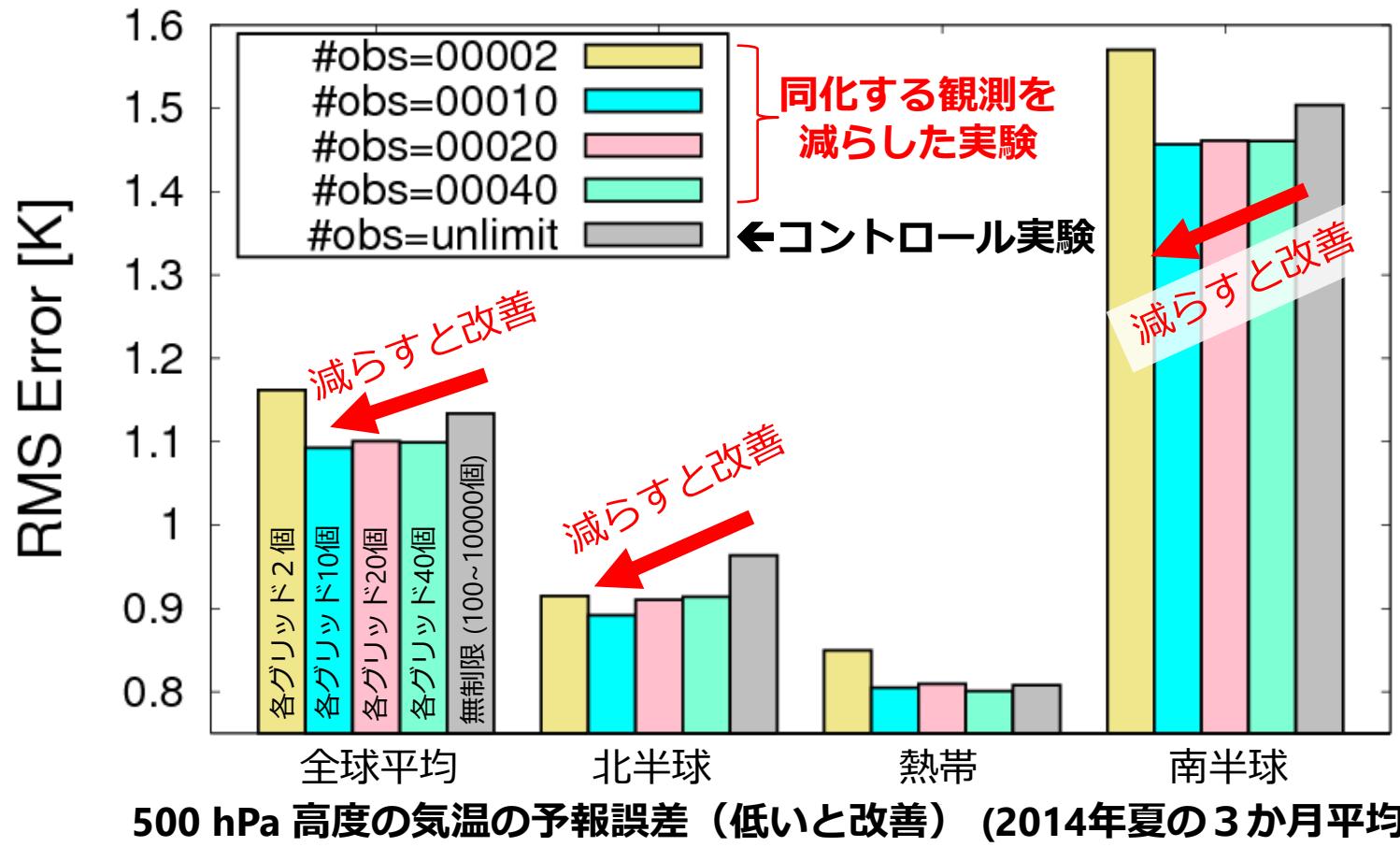
$$\text{rank}(\mathbf{P}^f) \leq \min(n, m) \approx m$$

Analysis error covariance  $\mathbf{P}^a$



# 問題意識: 観測ビッグデータの利用限界

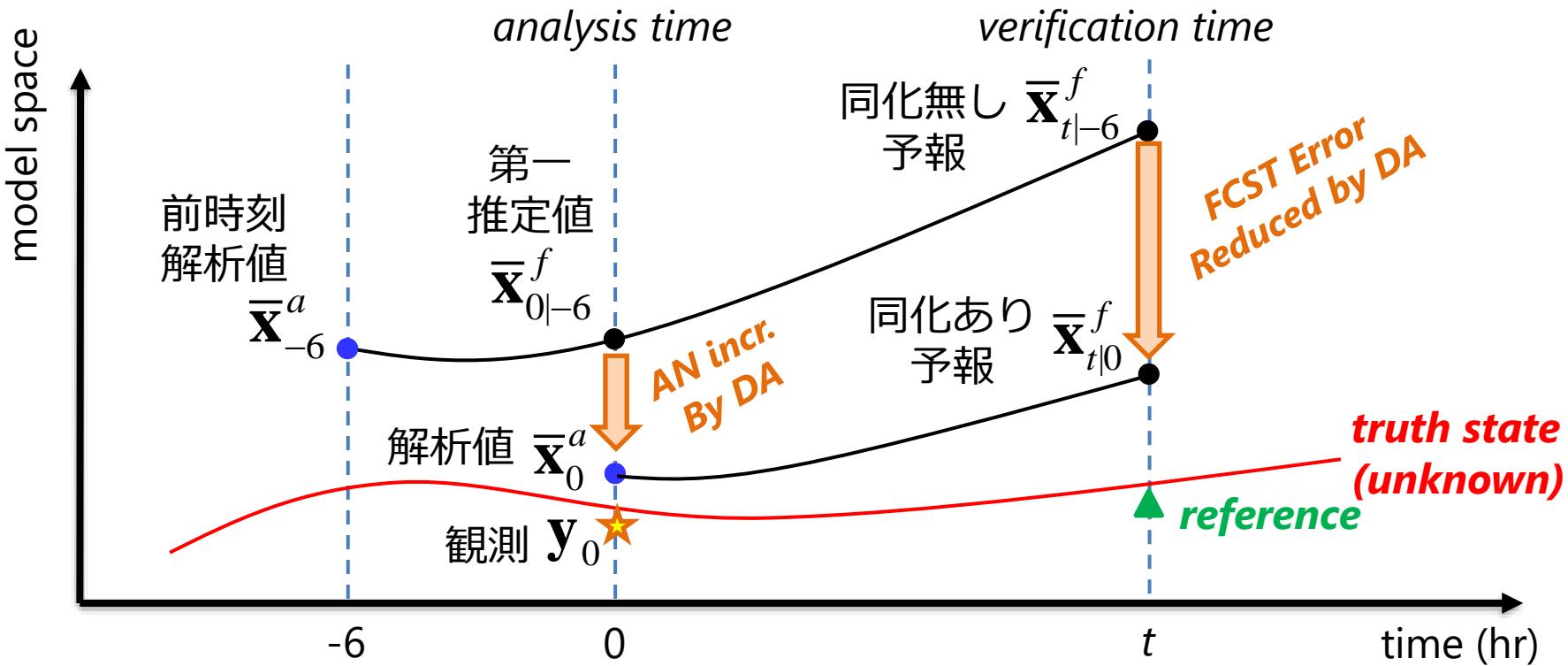
同化する観測を減らすと予報精度が改善！



- (1)得られる観測の数%しか利用できていない  
(2)観測が多くれば多いほど良い、とは限らない

**「観測の価値」を推定**

# 観測インパクト推定手法：EFSO



$$\Delta e_{MTE}^2 = \left( \mathbf{e}_{t|0}^T C \mathbf{e}_{t|0} - \mathbf{e}_{t|-6}^T C \mathbf{e}_{t|-6} \right) / 2$$

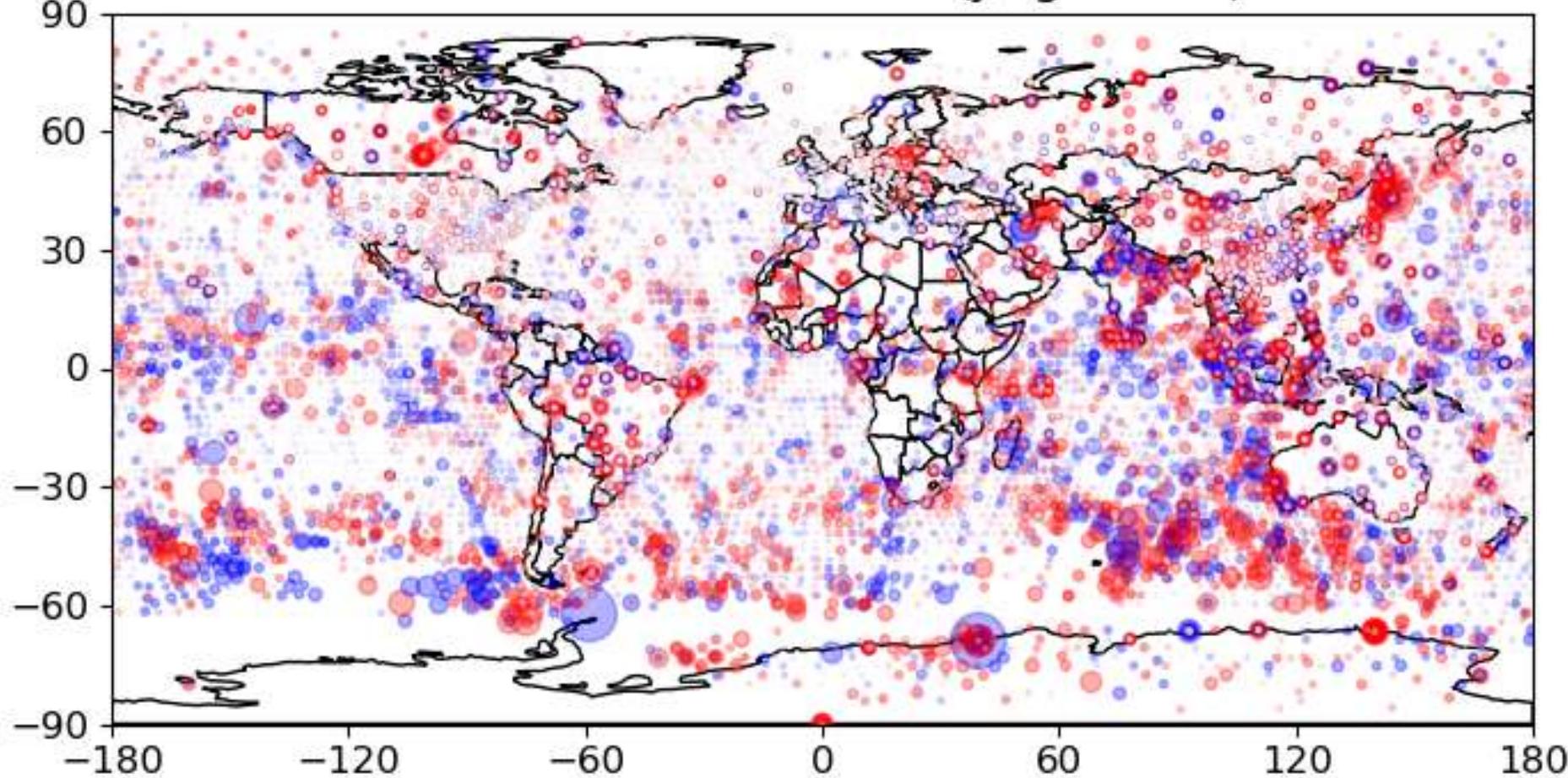
$$\approx \frac{1}{2} \frac{1}{m-1} \boxed{\delta \mathbf{y}_0^T} \mathbf{R}^{-1} \mathbf{Y}_0^a \mathbf{X}_{t|0}^{fT} C \left( \mathbf{e}_{t|0} + \mathbf{e}_{t|-6} \right)$$

$\mathbf{y}$ : 観測値,  $\mathbf{R}$ : 観測誤差共分散,  $\mathbf{Y}^a$ : 観測空間解析摂動,  $\mathbf{X}^f$ : 予報摂動,  $C$ : ノルム,  $\mathbf{e}$ : 誤差ベクトル

個々の観測が、予報を改善に貢献した寄与度を定量化する

# 定量化された「観測の価値」

EFSO :: Error Reduction in MTE ( $\text{J/kg} \times 1000$ ) vs. ERA



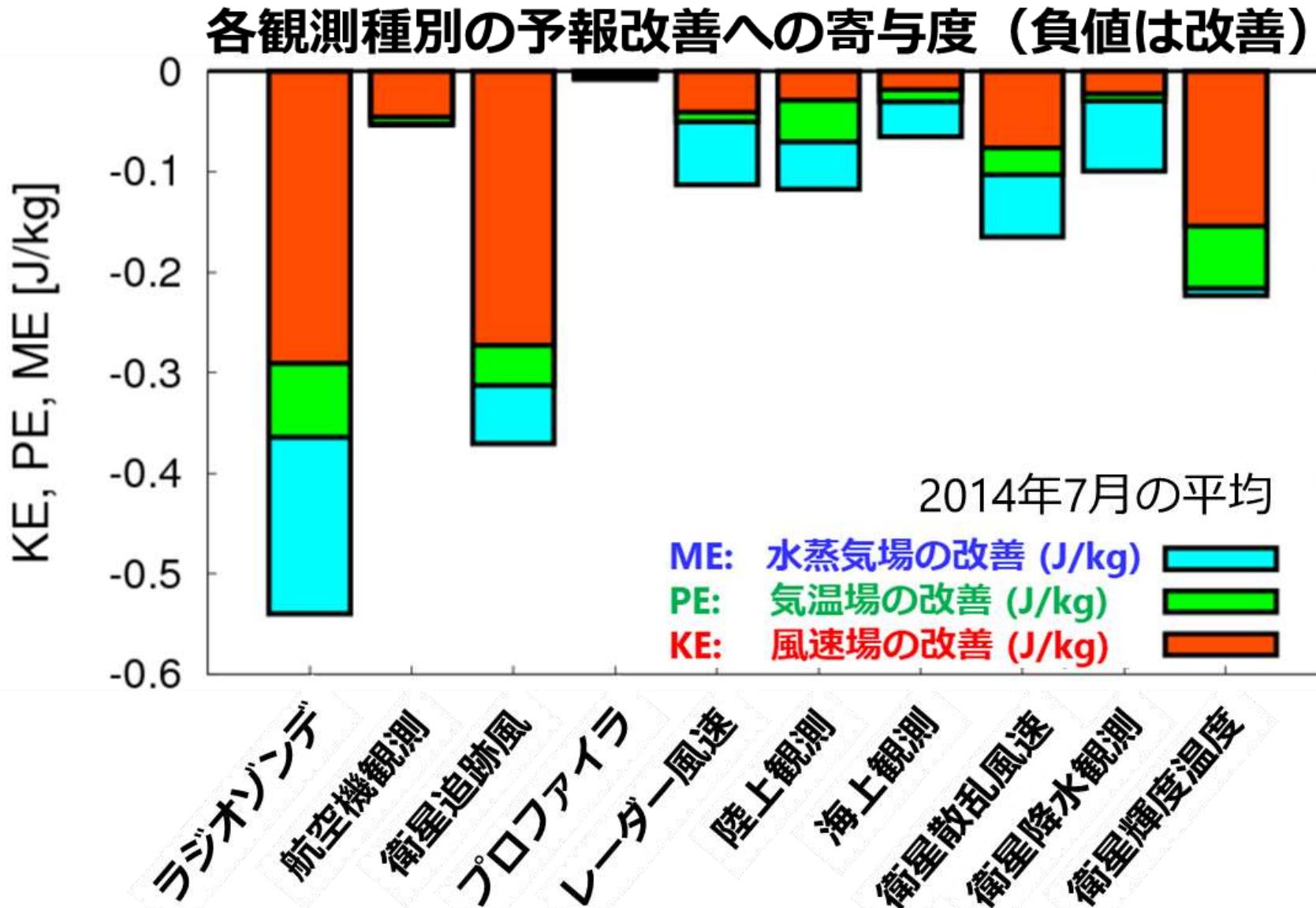
● ***Beneficial observations***

● ***Detrimental observations***

2014/07/11/00UTC; vs. ERA interim

Kotsuki et al. (2019; *QJRMS*)

# 観測の価値を評価: NICAM-LETKF



Kotsuki et al. (2019; *QJRMS*)

# EFSO Impact Estimates (vs. ERA Interim)

*detrimental*

AIRCFT

VADWND

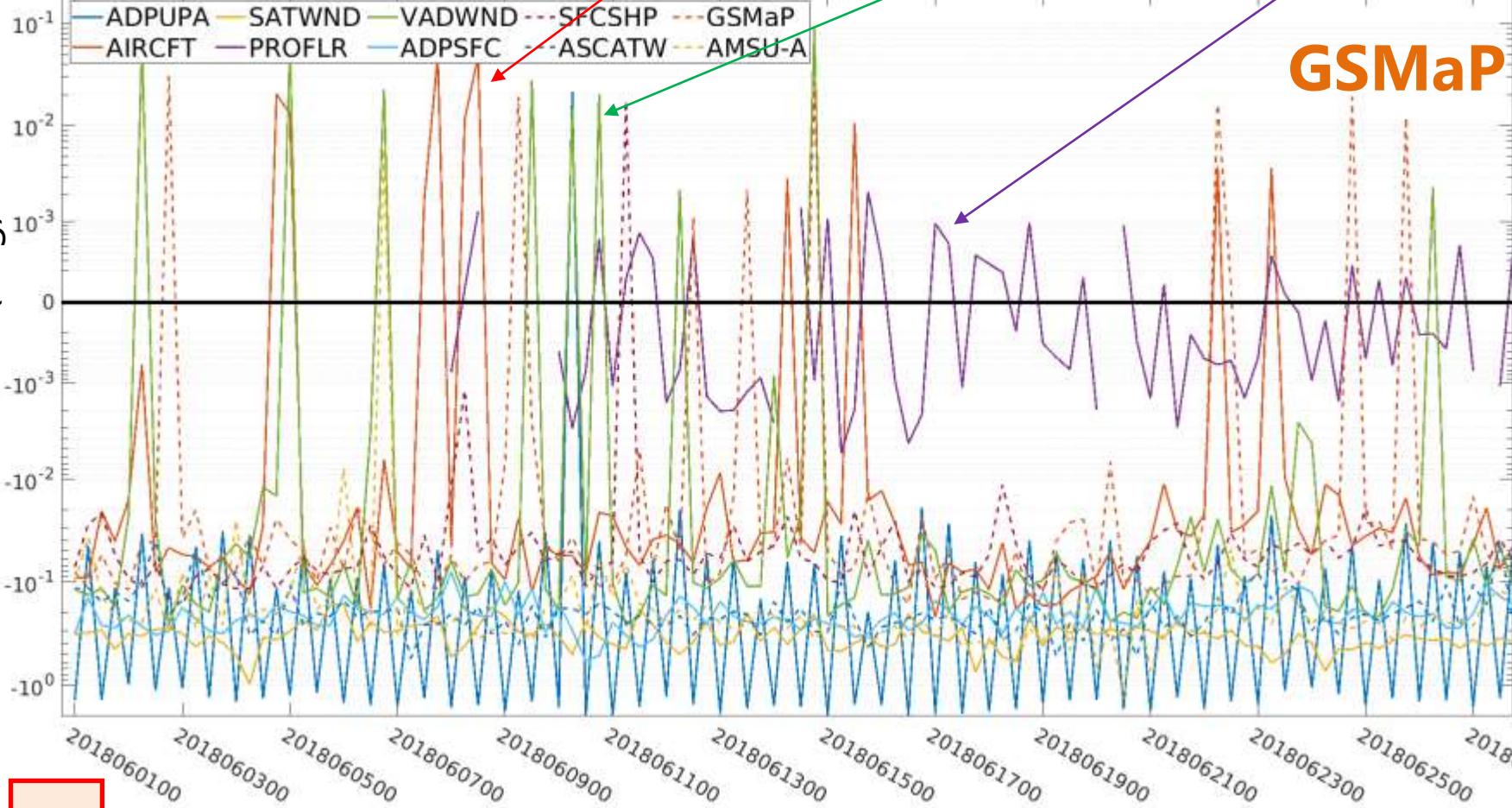
PROFLR



MTE (J/kg)

ADPUPA SATWND VADWND SFCSHP GSMaP  
AIRCFT PROFLR ADPSFC ASCATW AMSU-A

GSMaP

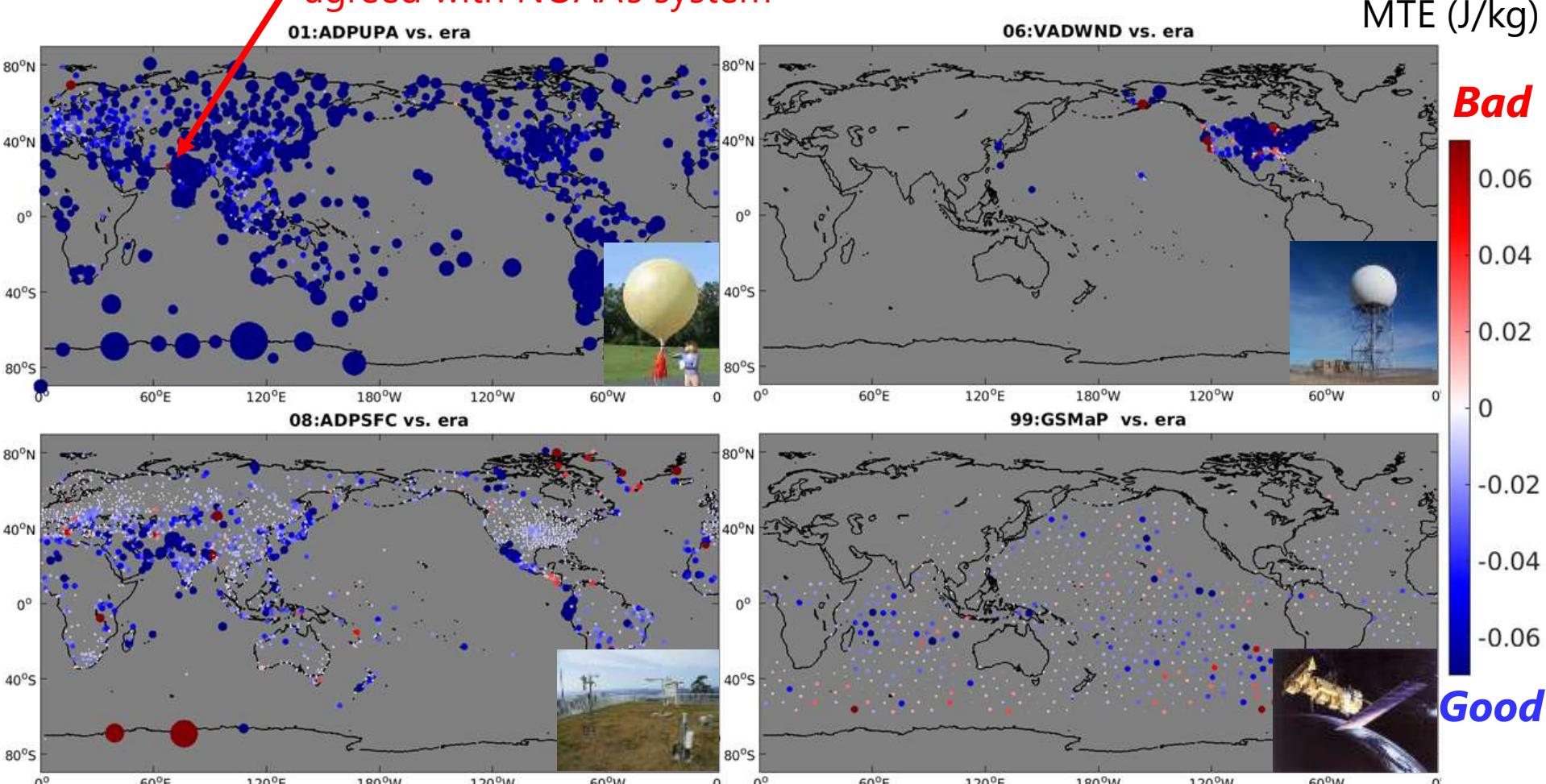


*beneficial*

FT: 24 hr

# EFSO Impact Estimates (vs. ERA Interim)

agreed with NOAA's system



**Can we detect statistically detrimental obs stations?**

FT: 24hr

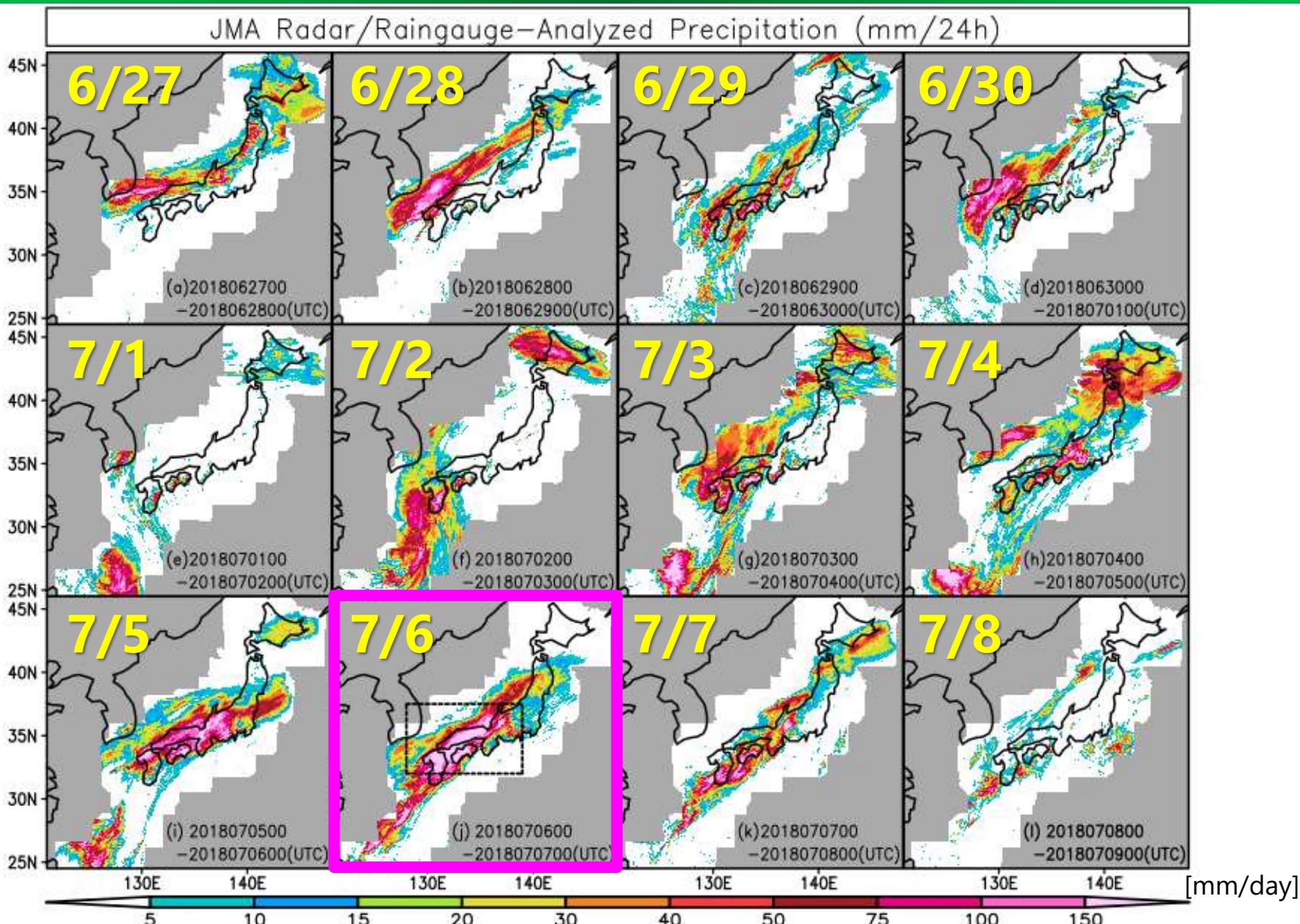
SAMPLE: 2018060100-2018063018

# **Impact of DA**

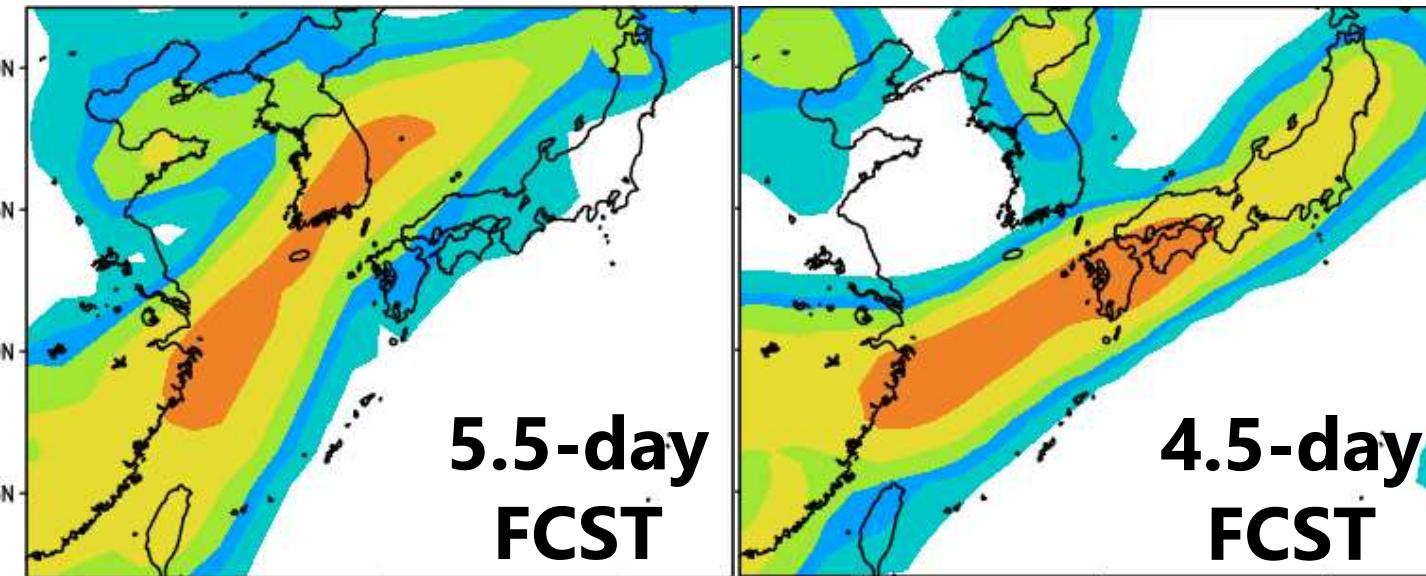
## **- A case of Record-breaking Rainfall in July 2018 -**

**Kotsuki et al. (2019; SOLA)**

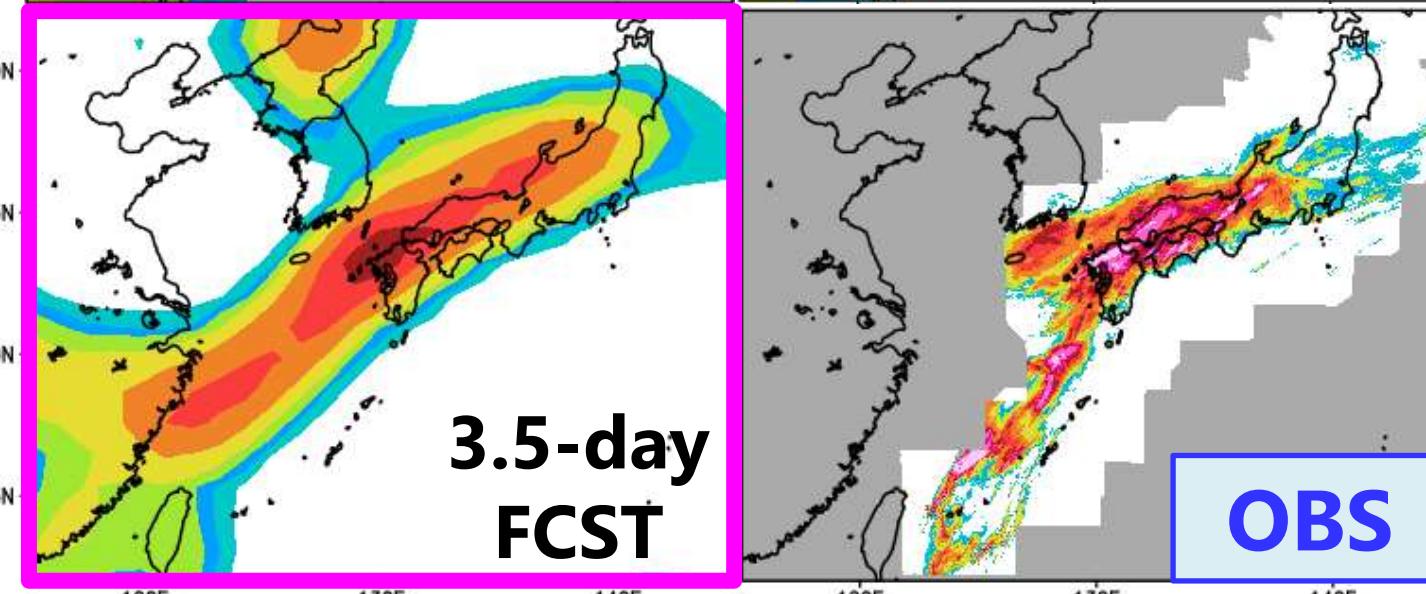
# Predictability of Record-breaking Rainfall in 2018



# RIKEN's Ensemble Rain Forecasts w/ NEXRA



7/1



7/2

7/3

7/4

7/5

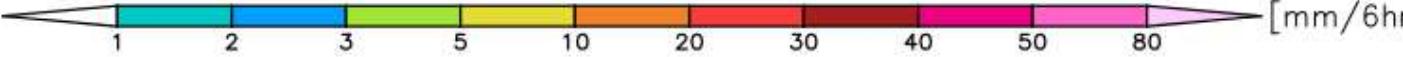
7/6

7/7

DA

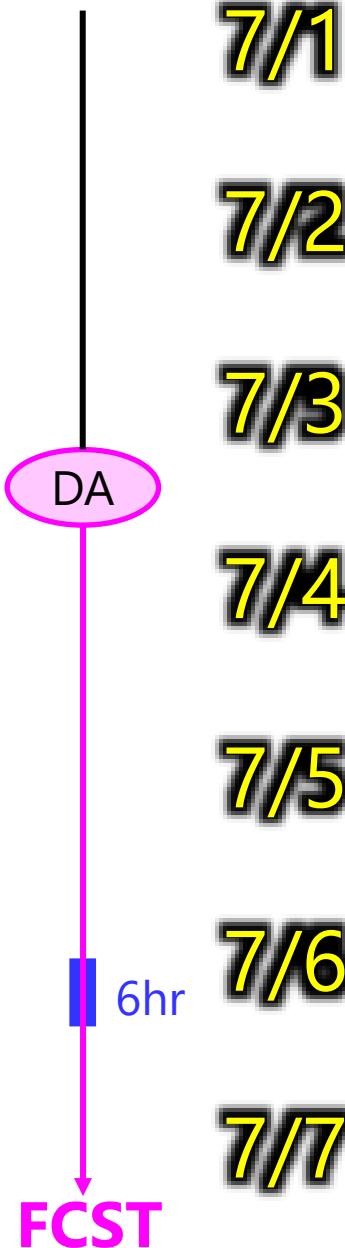
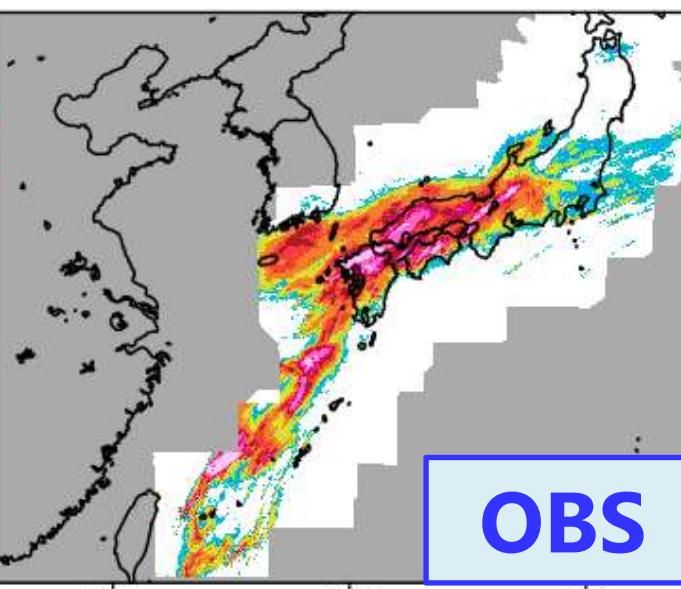
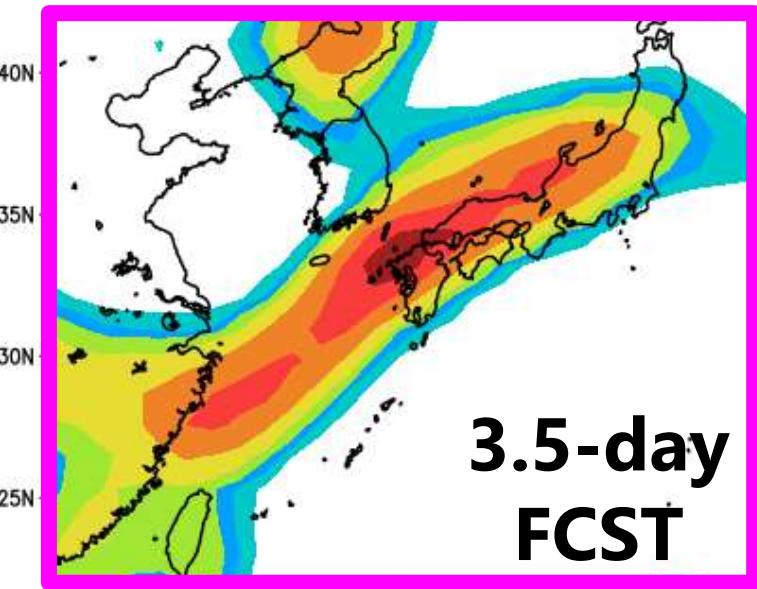
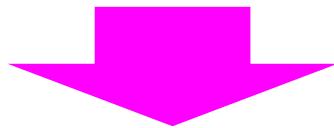
6hr

FCST

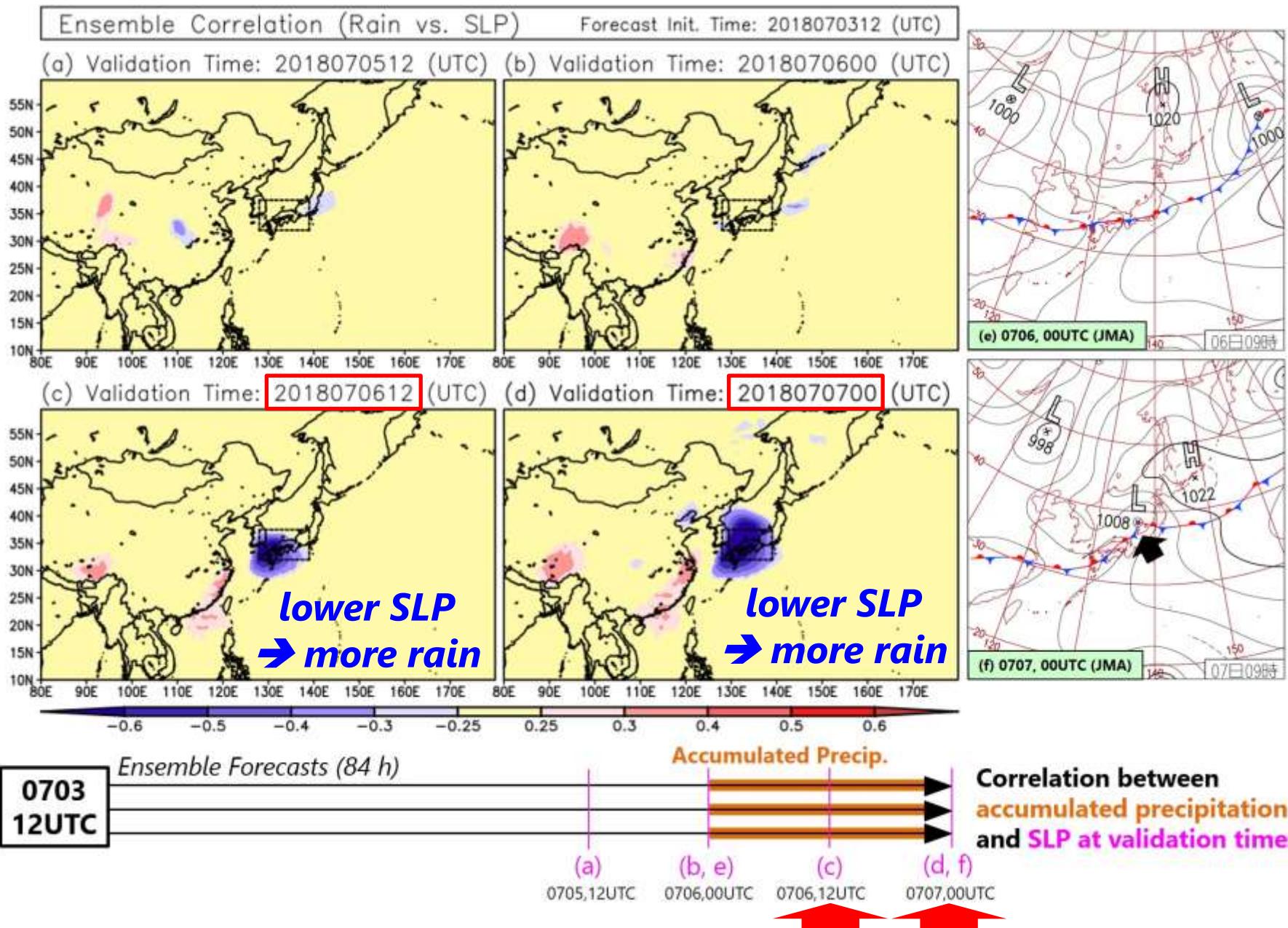


# Ensemble Forecasts w/ NEXRA

What happed?

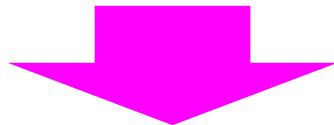


# Ensemble Correlation Analysis



# Ensemble Forecasts w/ NEXRA

How did DA improve state estimate?



7/1

7/2

7/3

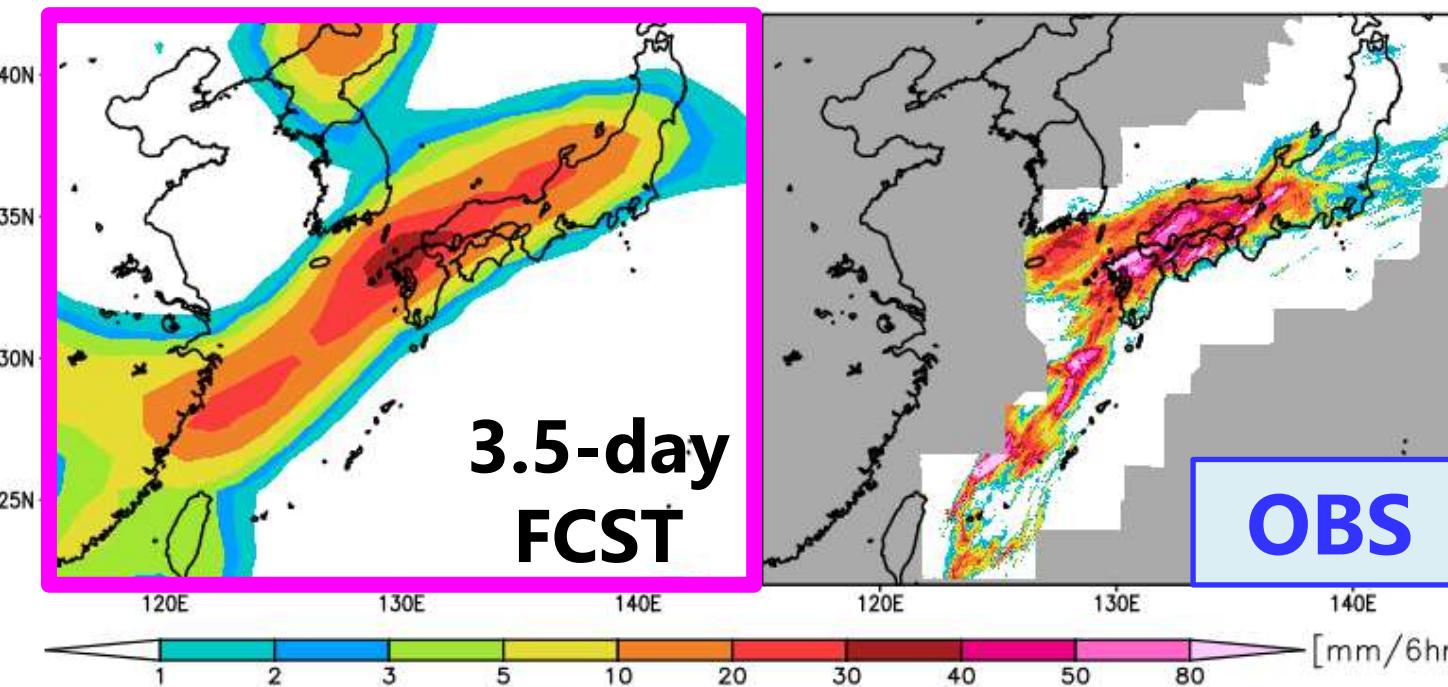
7/4

7/5

7/6

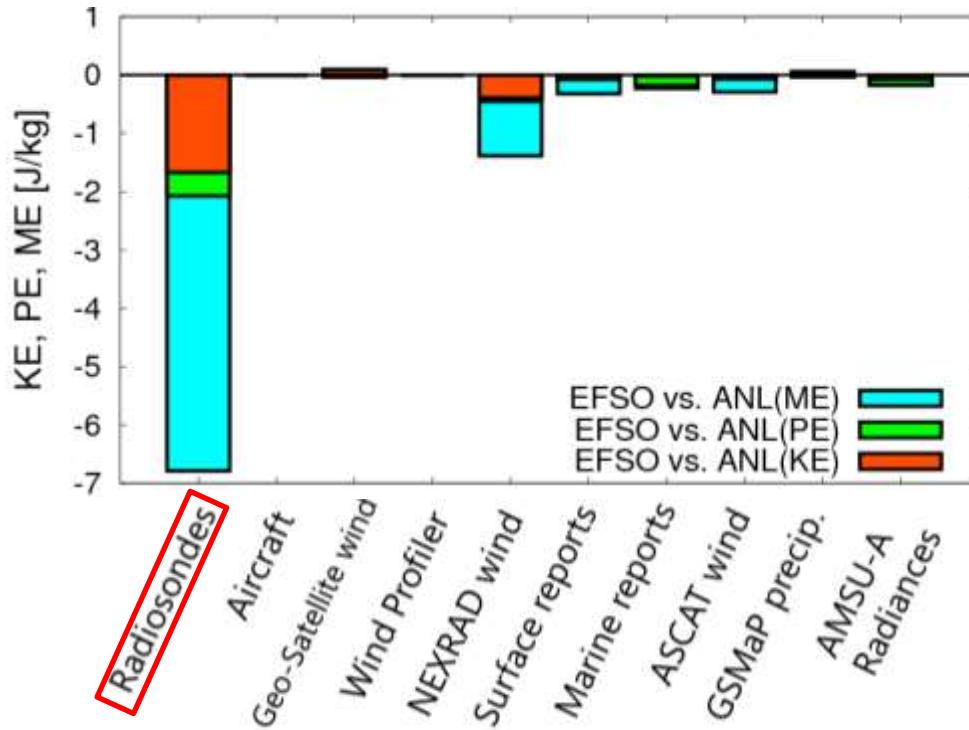
7/7

DA



# Evaluation of Observation Impacts w/ DA

**Obs @ 07/03 12UTC**



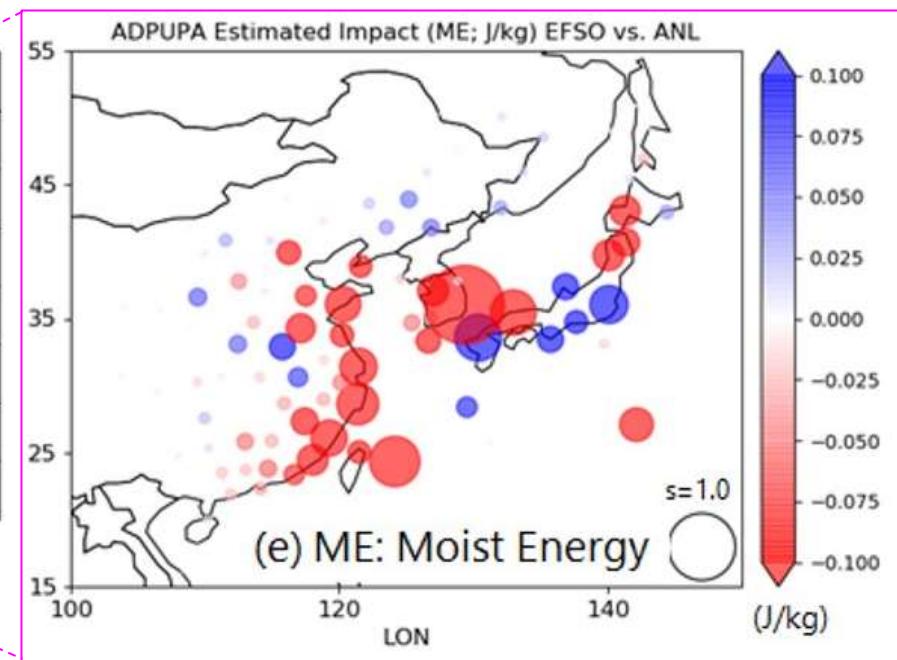
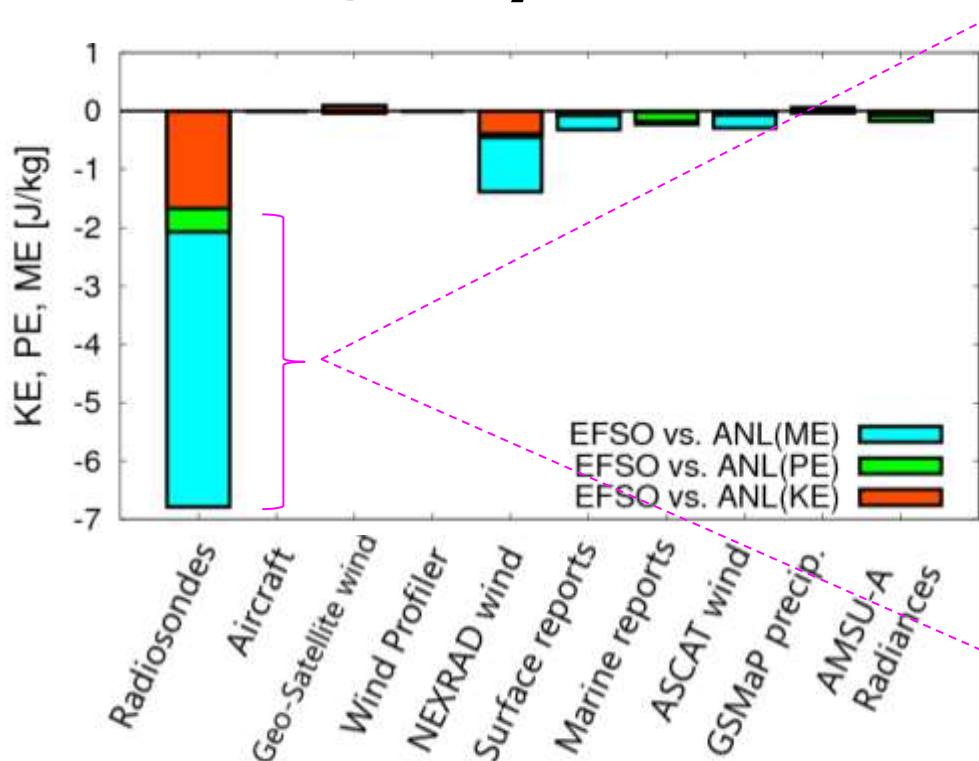
cf. Impact Estimates by EFSO  
Kotsuki et al. (2019; QJRMS)

- **ME : Moist Energy**      **(moisture field)**
- **PE : Potential Energy**      **(temperature & pressure)**
- **KE : Kinetic Energy**      **(wind field)**

# Evaluation of Observation Impacts w/ DA

**Obs @ 07/03 12UTC**

- : beneficial radiosondes
- : detrimental radiosondes

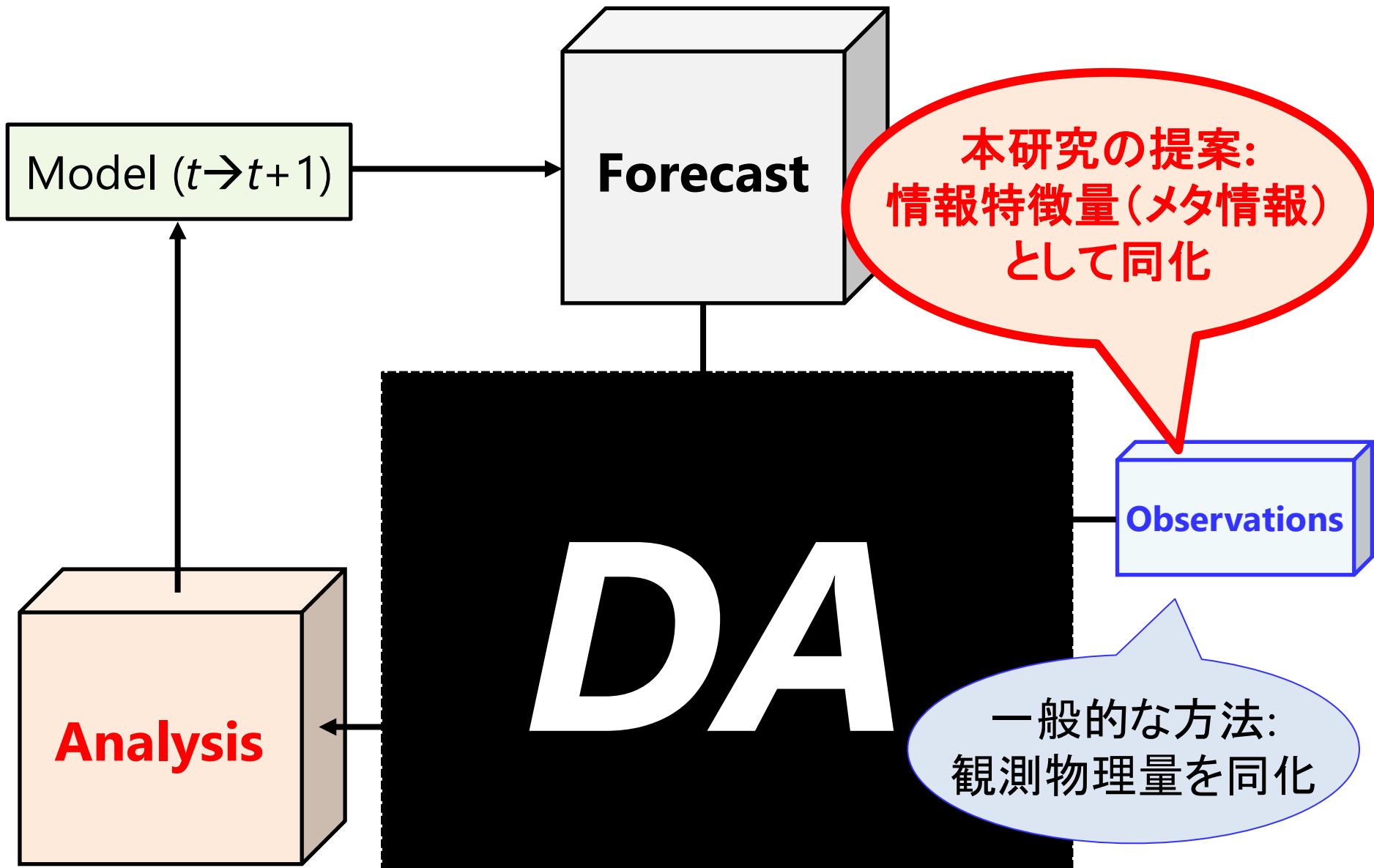


cf. Impact Estimates by EFSO  
Kotsuki et al. (2019; QJRMS)

- ME : Moist Energy (moisture field)
- PE : Potential Energy (temperature & pressure)
- KE : Kinetic Energy (wind field)

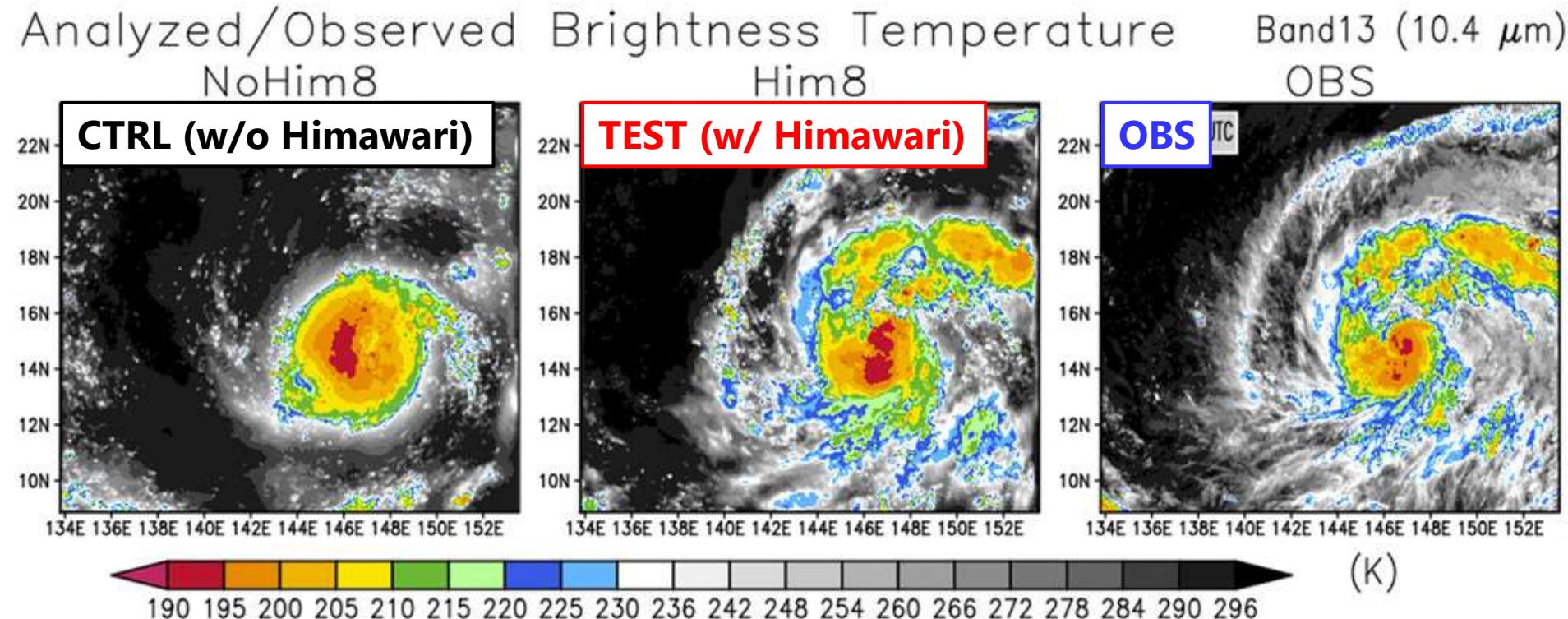
# **情報特徴量の データ同化**

# Workflow of Data Assimilation



# 「ひまわり」の輝度温度データ同化

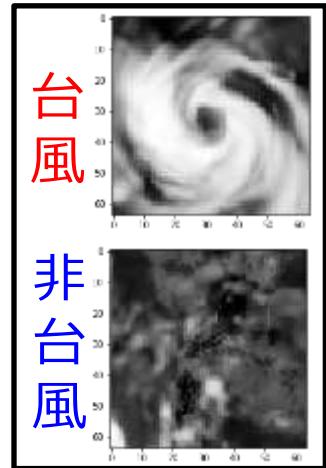
## 進展する全天候データ同化



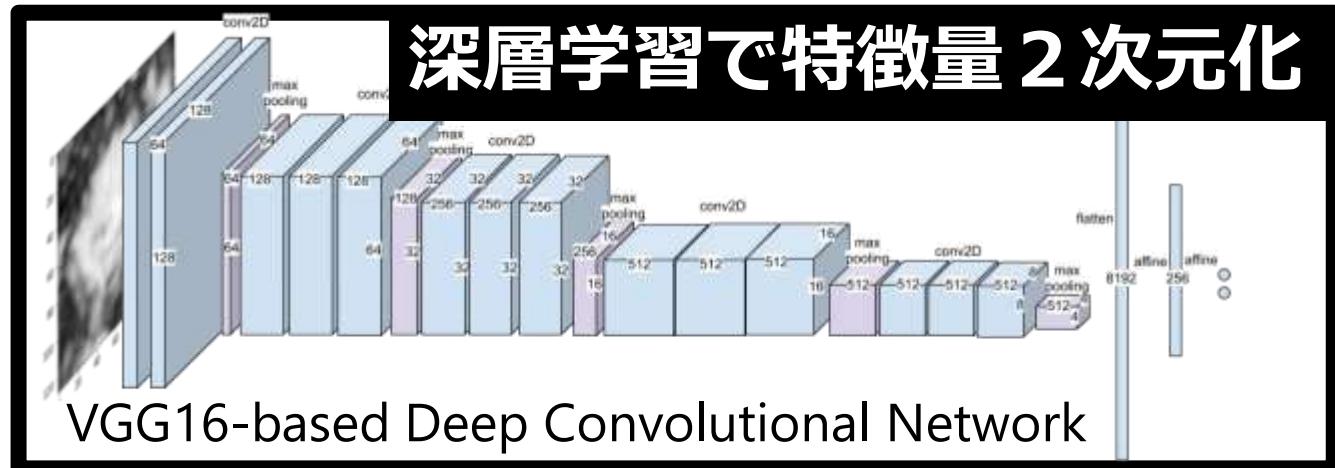
Honda et al. (2018; MWR) は  $0.20^\circ \times 0.20^\circ$  (約20km) に間引いて同化 (ひまわり自体は0.5~2km解像度)

→ 特徴量抽出でより観測情報を使えないか？

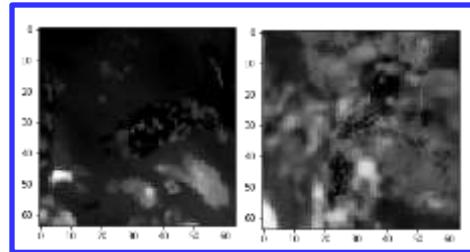
# VGG16 to extract typhoon features



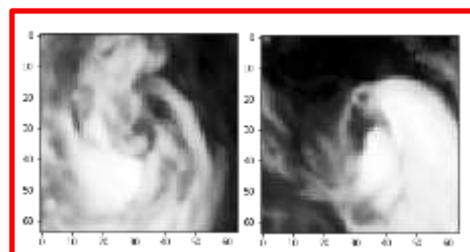
入力



明らかな非台風

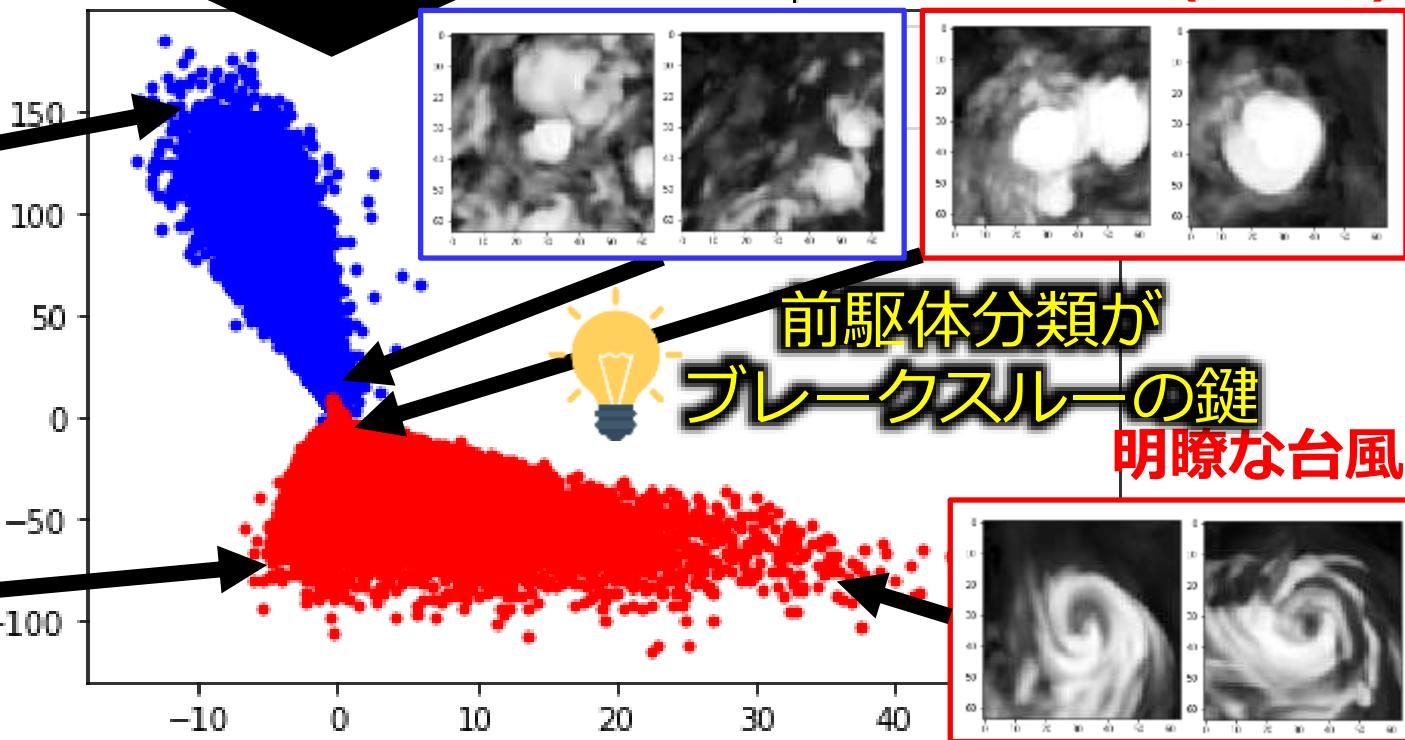


確かに台風

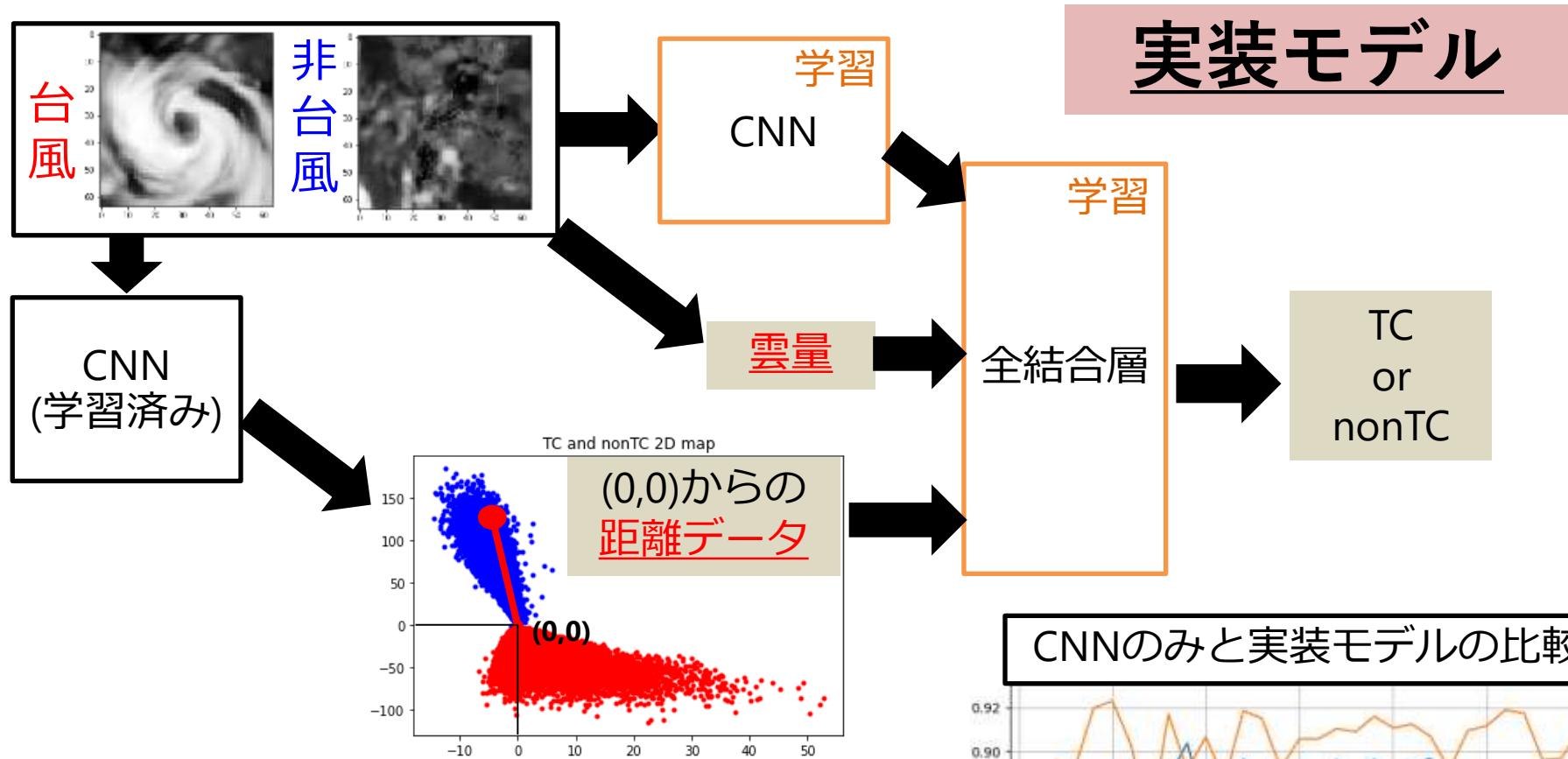


台風ではない

台風の卵(前駆体)



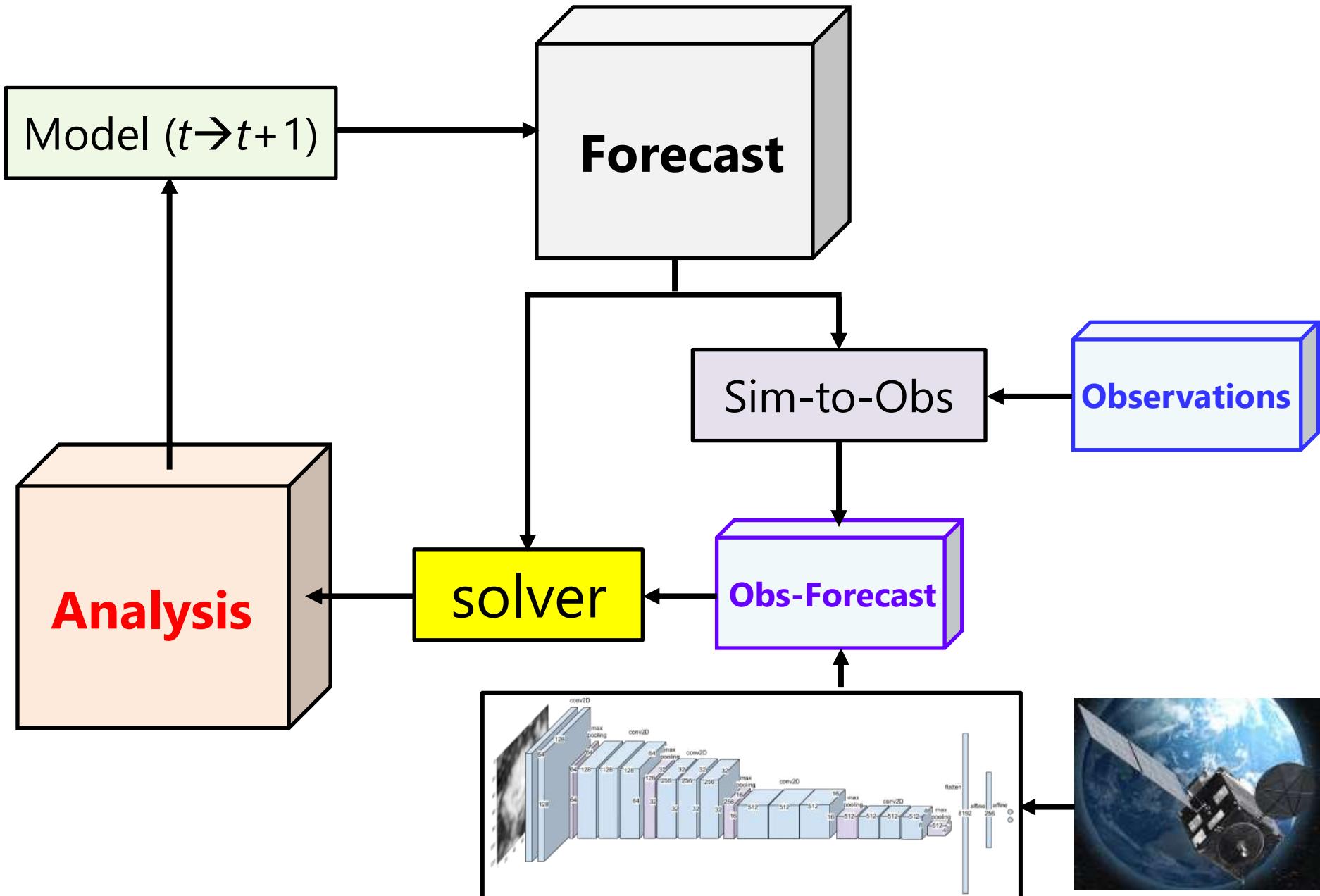
# Intermediate Step: Improving Typhoon Detection



特徴量(雲量, 距離)を明示的に与える事で、台風検知率が大きく向上。  
(本質的に同データだが学習効率化)



# 今後の計画



**設計に活かす  
データ同化**

# 設計に使うデータ同化: an example of JFE

ISIJ International, Vol. 57 (2017), No. 1, pp. 131–138

## Online Heat Pattern Control of a Shaft Furnace Based on a Real-time Visualization by Particle Filter

Yoshinari HASHIMOTO,<sup>1)\*</sup> Kazuro TSUDA,<sup>1)</sup> Takashi ANYASHIKI<sup>2)</sup> and Hidekazu FUJIMOTO<sup>2)</sup>

1) Instrument and Control Engineering Research Department, Steel Research Laboratory, JFE Steel Corp, 1 Kokan-cho, Fukuyama, Hiroshima, 721-8510 Japan. 2) Ironmaking Research Department, Steel Research Laboratory, JFE Steel Corp, 1 Kokan-cho, Fukuyama, Hiroshima, 721-8510 Japan.

(Received on August 10, 2016; accepted on September 26, 2016)

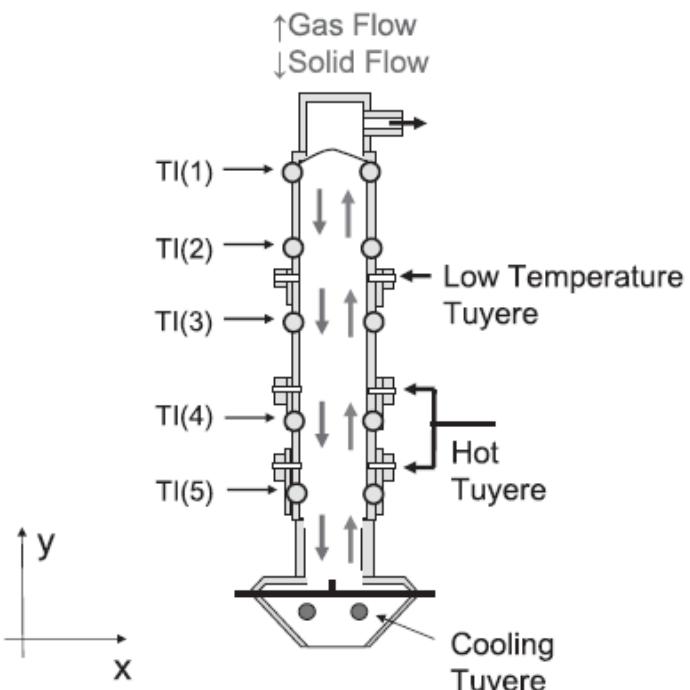


Fig. 1. Structure of the ferro-coke furnace.

Problem to be solved:  
to reproduce inner  
materials & conditions  
**from surface obs only.**

$$\frac{\partial(\rho_g C_g T_g)}{\partial t} + \frac{\partial(C_g u_g T_g)}{\partial x} + \frac{\partial(C_g v_g T_g)}{\partial y} = S\alpha(T_s - T_g) \dots (1)$$
$$+ R\Delta H_R \eta_1 + q$$

$$\frac{\partial(\rho_s C_s T_s)}{\partial t} + \frac{\partial(\rho_s C_s v_s T_s)}{\partial y} = S\alpha(T_g - T_s) + R\Delta H_R \eta_2 \dots (2)$$

$$q = -h(T_g - T_{out}) \dots (3)$$

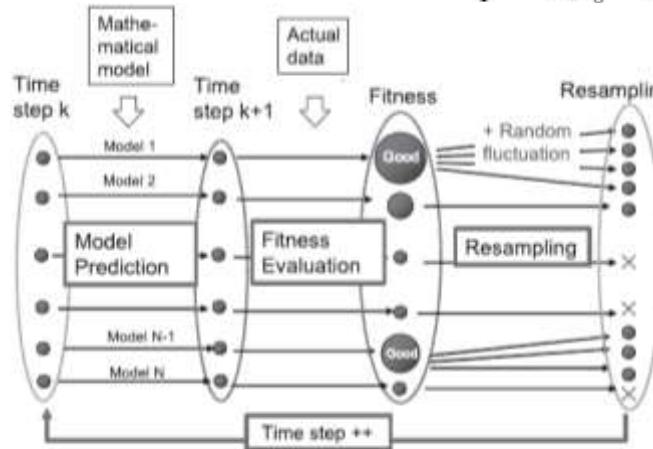
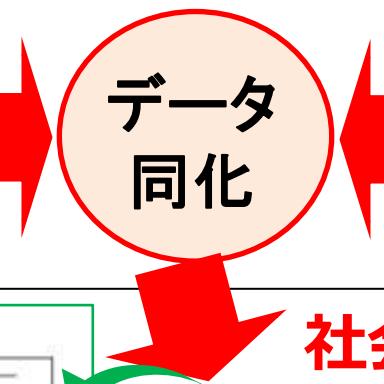
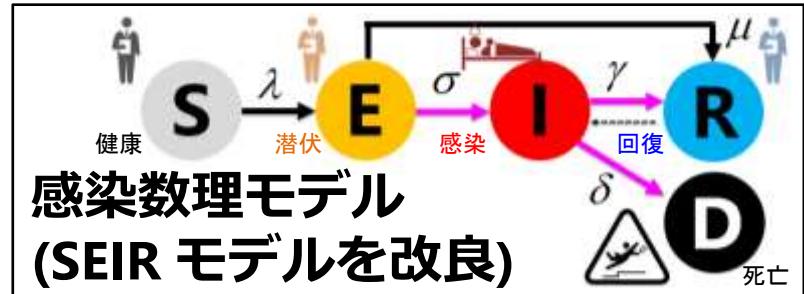


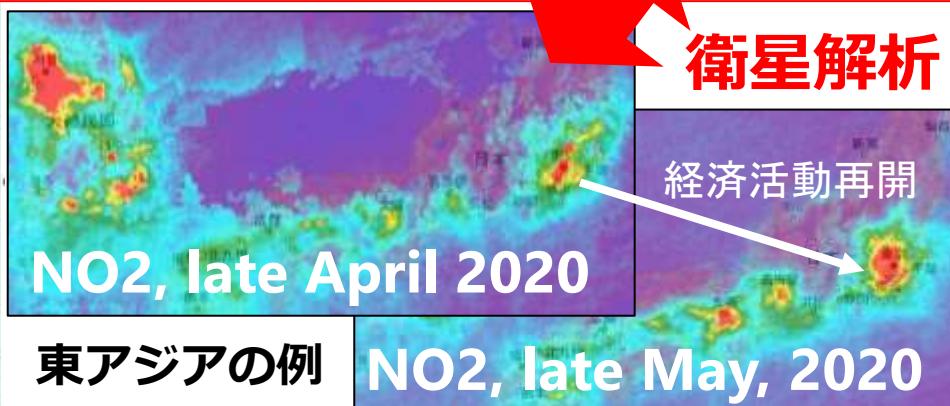
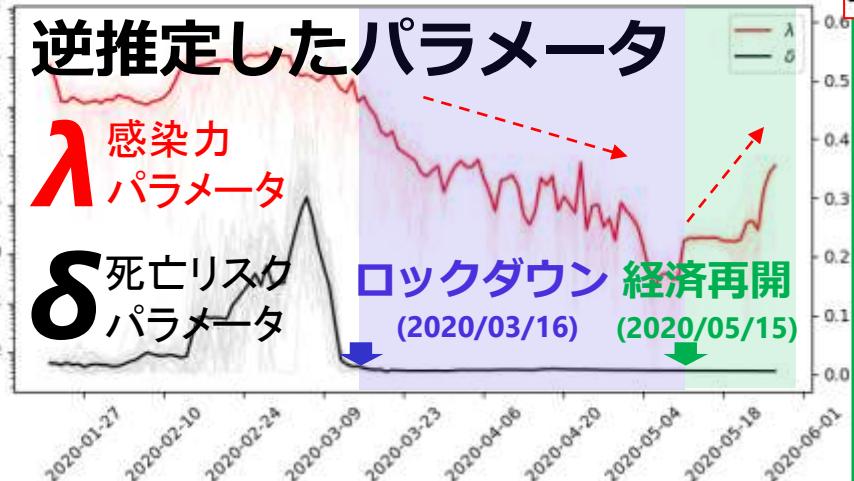
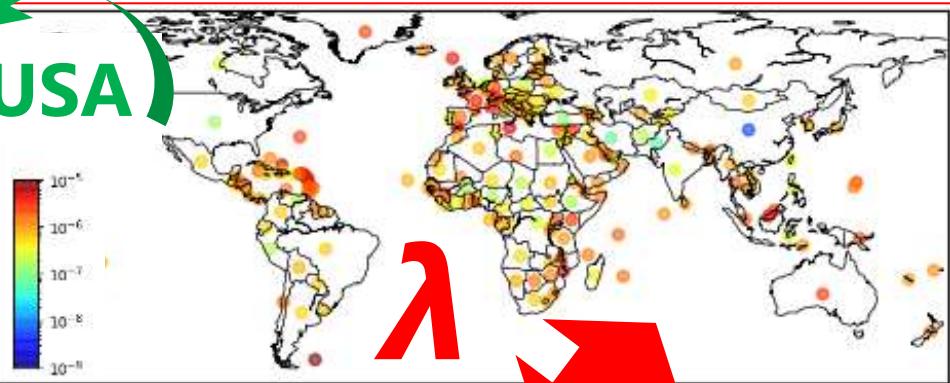
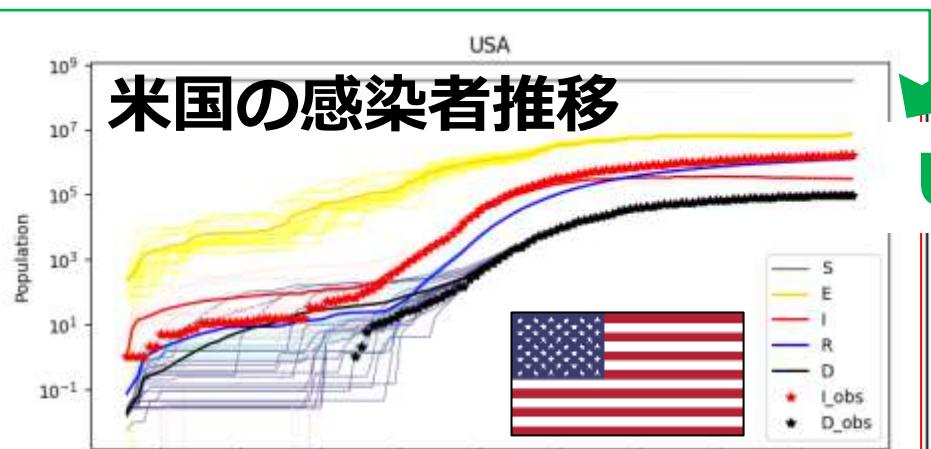
Fig. 2. The algorithm of particle filter.

Hashimoto et al. (2017)

# 数理×衛星: 新型コロナウィルス感染予測手法

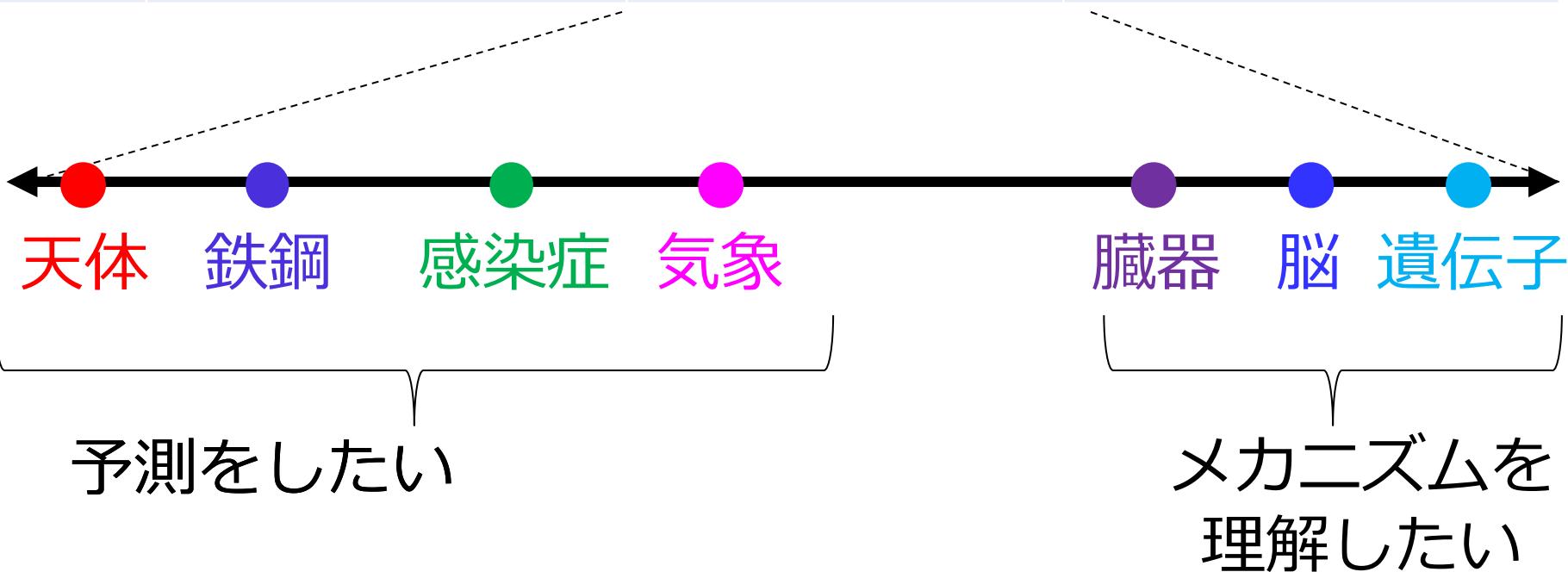


社会的感染力 " $\lambda$ " を逆推定



# DA combines Simulation & Data Science

	シミュレーション (第3の科学)	データ同化 (結び付け)	機械学習 (第4の科学)
性質	プロセス駆動型		データ駆動型
観測	少ない		大きい
数理	既知（硬い）		未知（柔らかい）
モデル	大自由度		低自由度



# インタビューからの実感

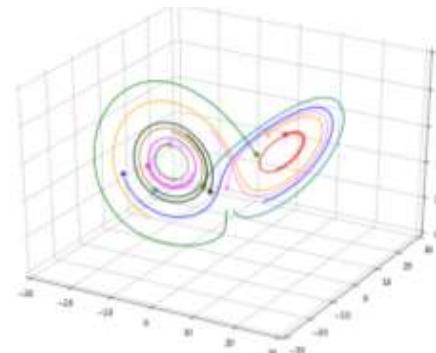
- ・ 総じて**現場のDA同化研究は想定以上に進展**
  - 社会との繋がりも薄かろうと考えていたデータ同化研究は、想定以上に実分野への拡がりを見せていた。
  - 基本的に、機械学習で出来る事をデータ同化でやる必要はない。基本的に解きたい問題は、「限られた(多くの場合表層的な)観測から、如何に内部状態を推定するか」であり、ここに物理プロセスを用いるデータ同化の優位性がある。
    - e.g. 鉄鋼、橋梁などの構造、酒造り
  - 模索すると良さそうなのは、機械とDAの併用か。
- ・ 一方で、**現場のボトルネックの解決には、現場に踏み込んだ共同研究が必要とも感じた。**
  - 特に、企業の研究の場合、情報の秘匿性もあるため、他者の技術者との意見交換・議論が難しい

# **データ同化の学習方法**

# Research Strategy

## 1. math & toy models

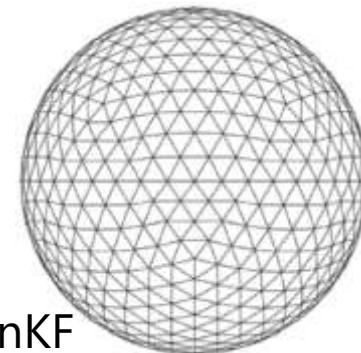
(e.g. Lorenz 96,  $n \sim O(10^2)$ ,  $p \sim O(10^2)$ )



## 2. intermediate models

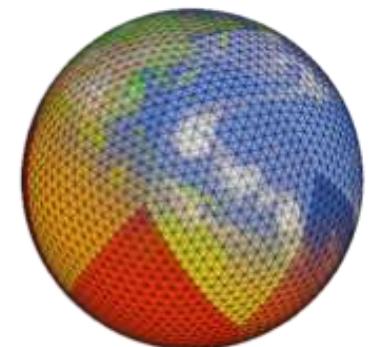
(e.g. SPEEDY,  $n \sim O(10^6)$ ,  $p \sim O(10^4)$ )

having  $n \times n \mathbf{P}^f$  is unaffordable ( $> 100$  Gb) → EnKF



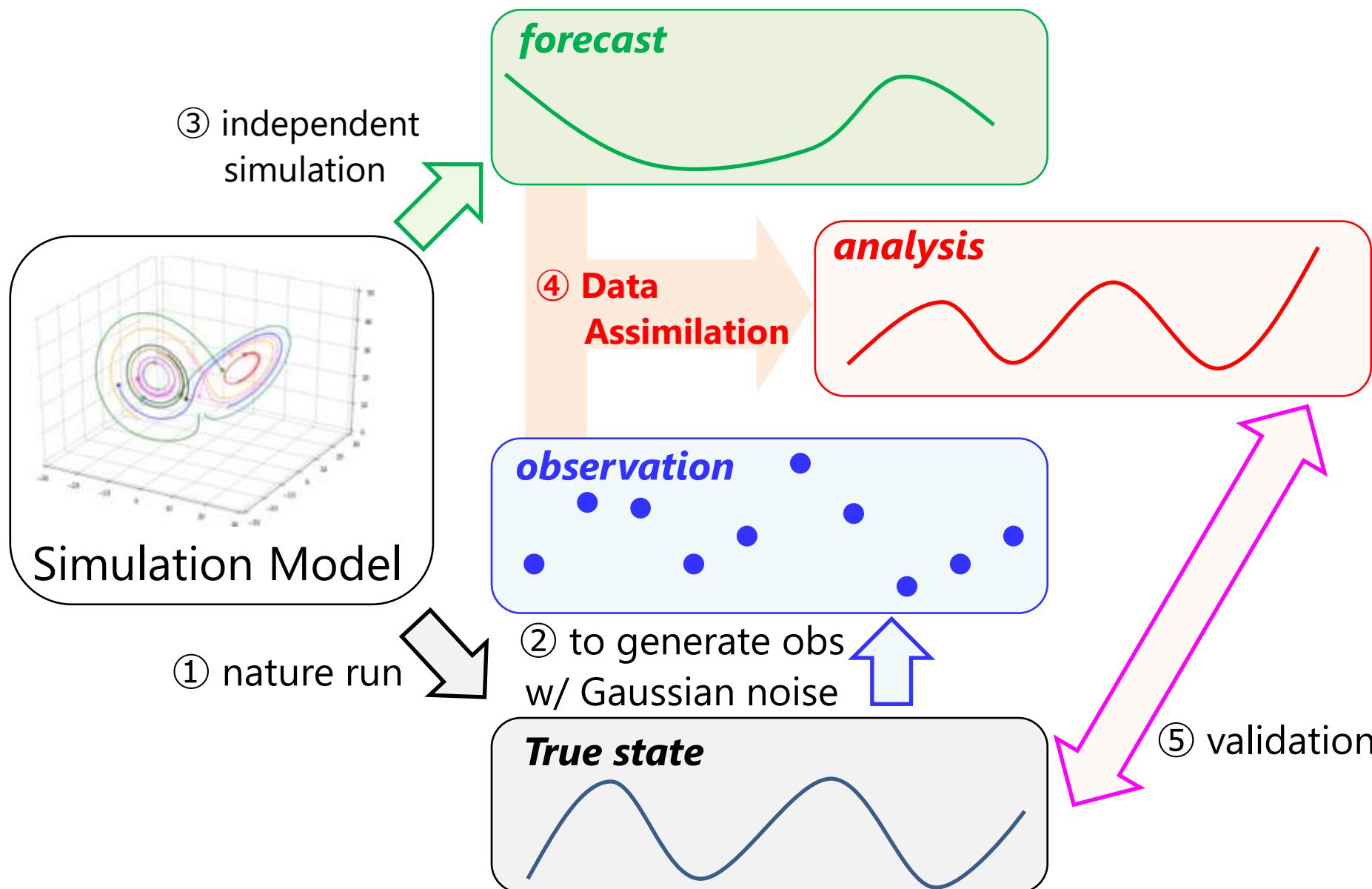
## 3. realistic models

(e.g. NWP,  $n > O(10^8)$ ,  $p > O(10^6)$ )



# Observing System Simulation Experiment (OSSE)

*also known as Idealized Twin Experiment*



# Data Assimilation Study w/ 40-variable Lorenz-96

## Lorenz-96 model (Lorenz 1996)

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

For  $j=1, \dots, J$ ,  $X_j = X_{j+J}$

### 力学系モデル・データ同化基礎技術の練習コース

2020年01月31日 小堀峻司  
updated 2020/03/19

#### 目的：

簡単力学モデル Lorenz の 40 变数モデル（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、自らコーディングすることで、力学モデルリングやデータ同化に関する実践的な「使える」基礎技術を得得する。

#### 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、専門で適切受け付ける。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、單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。

#### 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 はタイプライタリを用いずに自分でコーディングすること。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。

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 回と定義することの妥当性を確認する。

セント）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。

3. L96 を 2 年分積分し、最初の 10 年分をスピアアップとして捨てる。後半 10 年分を 6 時間毎に保存し、これを直角座標系上にプロットする。初期条件の自己循環性に対する影響を調べること。

5. 3 次元变分法と KF の比較実験を行う。この際、観測分布・觀測密度への依存性を調べる。

6. EnKF を実装し、KF と比較する。Whitaker and Hamill (2002)による Serial EnSRF, Bishop et al. (2001)による ETKF, Hunt et al. (2007)による LETKF, PO 法などの解法がある。2つ以上実装すること。

ヒント）気象分野の EnKF では、上述の手法が良く用いられている。カナダでは PO 法、米国気象局では Serial EnSRF。ドイツ・日本では LETKF など。小堀研で研究を進める場合、LETKF を用いた研究をしていくことが想定されるため、LETKF の実装には取り組んで欲しい。

#### コメント：

三好(2005,2006)を読めば、多くの事は理解できる。また追試を行う際に、どの程度の精度を期待できるのか、参考になる図が掲載されている。実装は式だけ分かれば良いが、KF、EnKF、3 次元变分法どちらは、自分でノートに式を書いて学ぶことを薦める。最初の基礎が無いと、後で深い研究は出来ない。

#### 発展課題：

1. 難易度 C, 研究発展性 B: KF と EnKF の重要なパラメータに、共分散脚張のファクターがある。これまでに手でチューニングしてきたが、観測空間統計を用いることで、この値を動的に推定することが出来る。このアルゴリズムを実装する。Miyoshi (2011)の動的共分散膨張法が実装しやすい。

2. 難易度 B, 研究発展性 A: データ同化により、同化される観測のインパクトを評価する手法（観測インパクト推定）がある。このうち、LETKF との親和性が高い、Ensemble Forecast Sensitivity to Observation (EFSO; Kalnay et al. 2012; Kotsuki et al 2018) を実装してみる。実際に、EFSO で解析値を改悪すると判定された観測を取り除くことで、解析値の RMSE が低下することを確認してみよう。

3. 難易度 S, 研究発展性 C: 4 次元变分法を実装し、EnKF と比較する。4 次元变分法には、アジョイントモデルを構築する他、近似的に  $40 \times 40$  行列の線形モデルを生成する方法もある。もしアジョントモデルを構築すれば、近似的な線形モデル行列との違いを調べてみるのも面白いかもしれない。

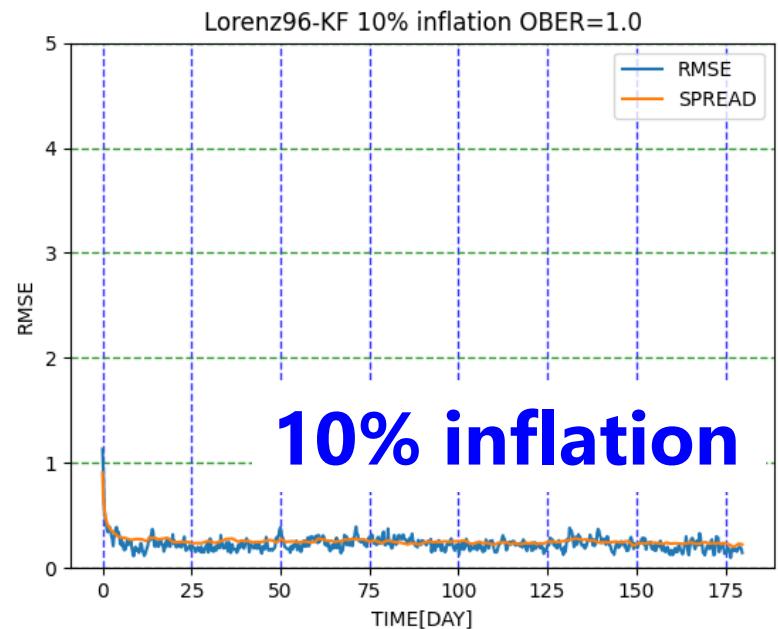
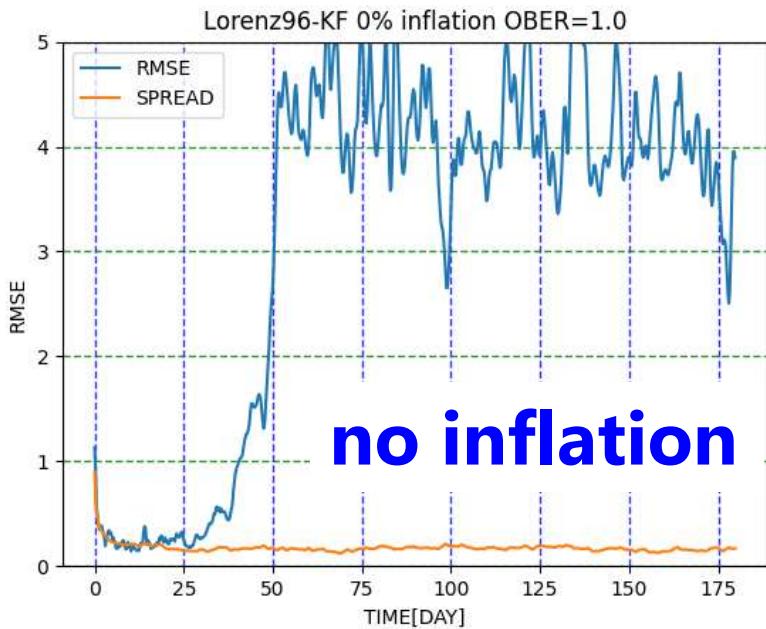
# (1) variance inflation (KF, EnKF)

**Empirical treatment for variance underestimation due to**

- (1) limited ensemble size
- (2) model nonlinearity
- (3) model imperfection

$$\mathbf{P}_{inf}^f = \alpha \times \mathbf{P}^f$$

*inflation factor (a tuning parameter)*



$$RMSE = \sqrt{\sum (x - x^{tru})^2 / n}$$

$$Spread = \sqrt{tr(\mathbf{P}^f) / n} = \sqrt{\sum \langle (x - x^{tru})^2 \rangle / n}$$

## (2) covariance localization (EnKF)

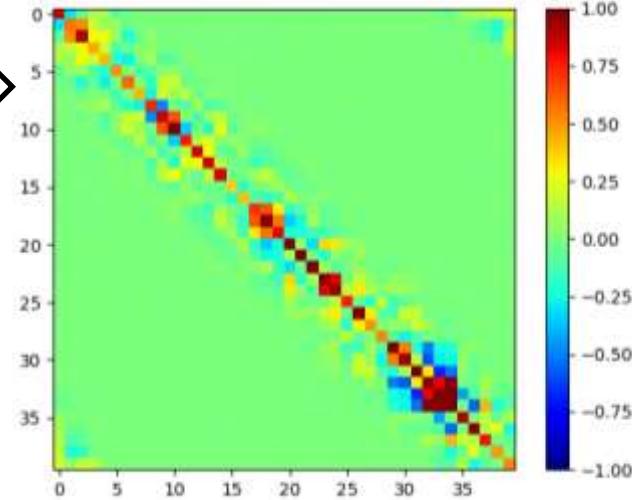
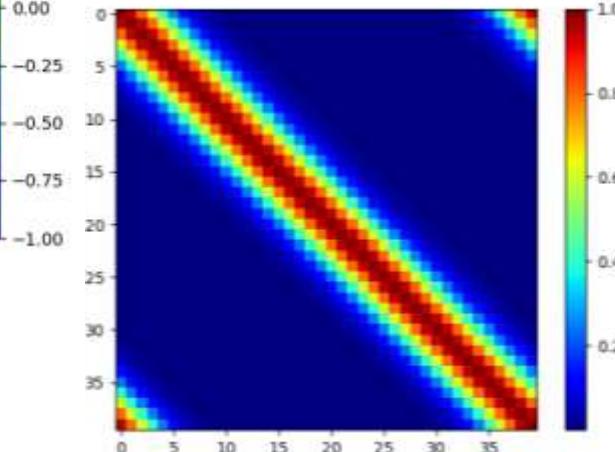
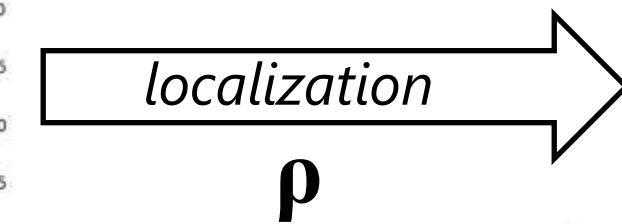
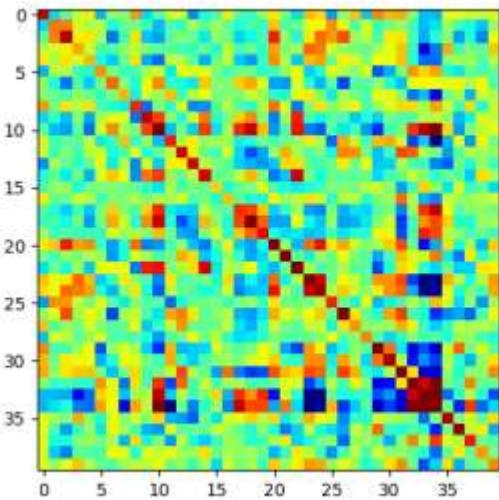
### Empirical treatment for

- (1) reducing sampling noise
- (2) increasing the rank

$$\mathbf{P}^f \rightarrow \rho \circ \mathbf{P}^f$$

$\circ$  : Schur product

$$\mathbf{P}^f \approx \frac{1}{m-1} \delta \mathbf{X}^f (\delta \mathbf{X}^f)^T$$



# まとめ

- データ同化について概念的説明
- 観測インパクト推定
- データ同化と機械学習の併用
- 設計に活かすデータ同化(私見)
- データ同化の学習方法



**Thank you for your attention!**

**Presented by Shunji Kotsuki**  
[\(shunji.kotsuki@chiba-u.jp\)](mailto:shunji.kotsuki@chiba-u.jp)

**Further information is available at**

Lab: <https://kotsuki-lab.com/>

Personal: <http://www.kotsuki-shunji.com/>