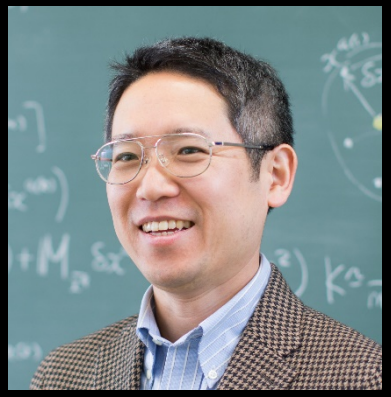
Big Data, Big Computation, and Machine Learning in Numerical Weather Prediction



Takemasa Miyoshi

Ph.D. (Meteorology)
Data Assimilation Scientist

Data Assimilation Research Team

RIKEN

Who am I?

B.S. from Kyoto U
↓
JMA administration (2y)
↓
JMA NWP (1.25y)
↓
UMD (2y, M.S. and Ph.D.)
↓
JMA NWP (3.5y)
↓
UMD (4y)
↓
RIKEN (8y+)

http://tedxsannomiya.com/en/speakers/takemasa-miyoshi/

http://data-assimilation.riken.jp/~miyoshi/

Takemasa Miyoshi, Ph.D.
Team Leader
Data Assimilation Research Team
RIKEN Center for Computational Science

Deputy Director
RIKEN interdisciplinary Theoretical and Mathematical Sciences (iTHEMS) Program

Chief Scientist
Prediction Science Laboratory
RIKEN Cluster for Pioneering Research

Visiting Professor
University of Maryland, College Park

Affiliate Professor
Graduate School of Science, Kyoto University

Visiting Principal Scientist
Application Laboratory, JAMSTEC

Research Counselor
Servicio Meteorológico Nacional (National Meteorological Service), Argentina

Education

- **2005** Ph.D. in Meteorology, University of Maryland, College Park, Maryland, USA (Dissertation PDF)
- **2004** M.S. in Meteorology, University of Maryland, College Park, Maryland, USA (Scholarly Paper PDF)
- **2000** B.S. in Physics, Faculty of Science, Kyoto University, Kyoto, Japan

RIKEN R-CCS

Data Assimilation Team

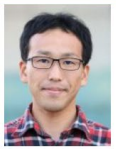
http://www.data-assimilation.riken.jp/

Data Assimilation Research Team

Data Assimilation Research Team was launched in October, 2012, in RIKEN Advanced Institute for Computational Science (AICS), conveniently located in the beautiful and historic city of Kobe. RIKEN is known as the flagship research institution in Japan. On April 1, 2018, RIKEN AICS was renamed RIKEN Center for Computational Science (R-CCS). R-CCS is operating the world's leading supercomputer "Fugaku", and also has a strong Research Division. R-CCS takes the lead in advancing the computational science and aims to be an international center of excellence for computational science in collaboration with a wide range of research organizations. R-CCS integrates the computer science and computational science to conduct most advanced research and development of a wide range of applied scientific computation, as well as of high performance computing technologies.

Data assimilation (DA) is a cross-disciplinary science to synergize computer simulations and real-world data, using statistical methods and applied mathematics. As computers become more powerful and enable more precise simulations, it will become more important to compare the simulations with actual observations. DA Team performs cutting-edge research and development on advanced DA methods and their wide applications, aiming to integrate computer simulations and real-world data in the wisest way. Particularly, DA Team tackles challenging problems of developing efficient and accurate DA systems for "big simulations" with real-world "big data" from various sources including advanced sensors. The specific foci include 1) theoretical and algorithmic developments for efficient and accurate DA, 2) DA methods and applications by taking advantage of the world-leading supercomputer and "big data" from new advanced sensors, and 3) exploratory new applications of DA in wider simulation fields. These advanced DA studies will enhance simulation capabilities and lead to a better use of high-performance computers.

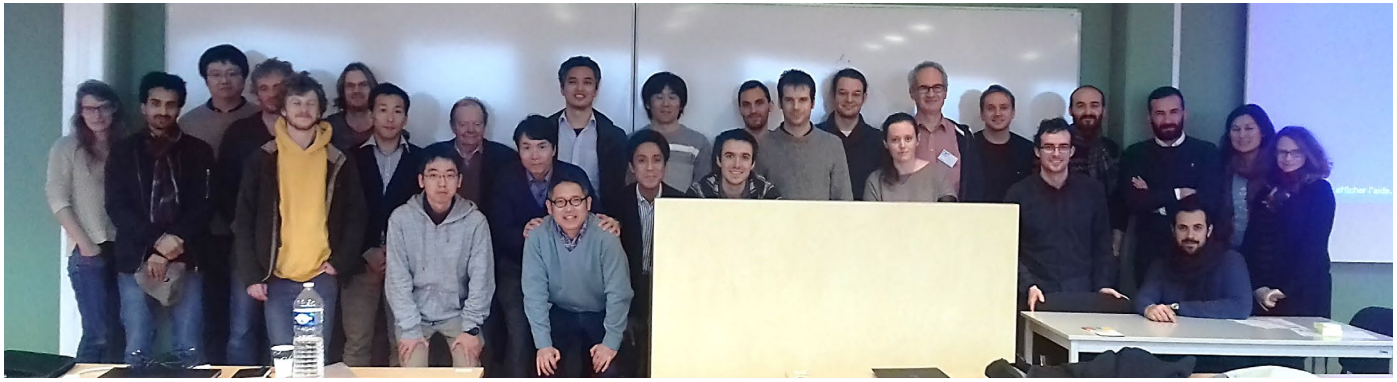
Team Leader	Research Scientist	Research Scientist	Special Postdoctoral Researcher
			
Takemasa Miyoshi	Koji Terasaki	Shigenori Otsuka	Takumi Honda

Postdoctoral Researcher	Postdoctoral Researcher	Postdoctoral Researcher	Postdoctoral Researcher
			
Kohei Takatama	James Taylor	Arata Amemiya	Maha Mdini

Postdoctoral Researcher	Postdoctoral Researcher	Postdoctoral Researcher	Technical Staff	Junior Research Associate
				
Shun Ohishi	Yasumitsu Maejima	Jianyu (Richard) Liang	Hideyuki Sakamoto	Qiwen Sun

The Second IMT-Atlantique & RIKEN Joint Workshop: "Statistical Modeling and Machine Learning in Meteorology and Oceanography"

- **Date: Feb. 10-13, 2020 (Mon-Thu)**
- **Place: IMT Atlantique, Brest, France (Room: B1-114)**
- **Language: English**



Program

Day 1: Feb. 10

Time	Speaker	Title
9:30-9:45	Takemasa Miyoshi & Pierre Tandeo	Opening (Perspective toward DA-AI fusion)
9:45-10:45	Michele Alessandro Bucci	Keynote
10:45-11:00	-	Break
11:00-11:30	Naonori Ueda	AI approach for advanced weather forecasting
11:30-12:00	Pierre Tandeo	Selection of dynamic model using analog data assimilation
12:00-13:30	-	Lunch break
13:30-14:00	Paul Platzer	Analog forecasting errors from a dynamical systems point of view
14:00-14:30	Arata Amemiya	Model bias correction by ML
14:30-15:00	Shigenori Otsuka	Toward hybrid NWP-AI system for precipitation nowcasting
15:00-15:15	-	Break
15:15-15:45	Maha Mdini	Toward model acceleration by ML
15:45-16:15	Maxime Beauchamp	A geostatistical journey through data and modeling in air quality
16:15-16:30	-	Introduction to breakout discussion
16:30-17:00	-	Breakout discussion

Day 2: Feb. 11

-

Time	Speaker	Title
9:30-10:00	Chen Wang	Classification of global ocean SAR images for broader applications
10:00-10:30	Tsuyoshi Yamaura	The parameter estimation system in SCALE for reduced-precision floating-point numbers
10:30-10:45	-	Break
10:45-11:15	Kenta Sueki	Estimation of key parameters in cloud microphysics using ensemble Kalman filter
11:15-11:45	Koji Terasaki	Accounting for the horizontal observation error correlation of satellite radiances in data assimilation
11:45-13:15	-	Lunch break
13:15-13:45	Marie Boutigny	Using precipitation radar for urban hydrology: motion interpolation and merging with rain gauges
13:45-14:15	Zhen Yicun & Jean-Marie Vent	Application of analog data assimilation to the spatial-temporal interpolation of sea-surface sediment concentration and sea-surface height
14:15-14:30	-	Break
14:30-15:00	Jules Guillot	Data-Model Coupling for SST-DA
15:00-15:30	Said Ouala	Data-driven identification of geophysical dynamics: incorporating stability constraints in neural networks models
15:30-16:00	-	Breakout discussion
16:00-17:00	-	Plenary discussion

IMT-Atlantique & RIKEN Online Joint Seminar Series (Jointly with Data Assimilation Seminar Series)

- **Dates: Feb.17 - Apr.14, 2021**
- **Language: English**
- **Place: Zoom**

Online seminar series

To join the video meeting, please contact the following address in advance. da-seminar@riken.jp

Theme: Statistical Modeling and Machine Learning in Meteorology and Oceanography

This online joint seminar series follows the first and second joint RIKEN-IMT Atlantique workshops in 2019 and 2020. The main topic is the fusion of Data Assimilation (DA) with Artificial Intelligence (AI) in the fields of Meteorology, Oceanography, and Climate. The use of AI and statistical techniques such as neural networks in geophysics has a potential to enhance our knowledge and to improve physical models' performance by exploiting more from available observations and by accelerating DA workflow for real time response. This seminar series aims to exchange ideas about potential future research on the fusion of DA and AI with HPC in the research fields of meteorology and oceanography for enhancing future collaborations between RIKEN and IMT-Atlantique based on the international agreement signed in 2019.

Seminar 1

● Feb. 17: Jointly with OceaniX project of IMT-Atlantique

Time	Speaker	Title
JST 21:30-23:00 **CET** 13:30-15:00	Dr. Takemasa Miyoshi	Big Data, Big Computation, and Machine Learning in Numerical Weather Prediction

Seminar 2: Application of deep-learning methods to environmental data

● Mar. 3

Time	Speaker	Title
JST 17:30-18:00 **CET** 9:30-10:00	Dr. Aurélien Colin (IMT Atlantique & CLS)	Semantic Segmentation of Metocean Processes and Estimation of Ancillary Data

Seminar 2: Application of deep-learning methods to environmental data

Mar. 3

Time	Speaker	Title
JST 17:30-18:00 **CET** 9:30-10:00	Dr. Aurélien Colin (IMT Atlantique & CLS)	Semantic Segmentation of Metocean Processes and Estimation of Ancillary Data
JST 18:00-18:30 **CET** 10:00-10:30	Dr. Hirotaka Hachiya	-
JST 18:30-19:00 **CET** 10:30-11:00	-	Discussion

Seminar 3: Model acceleration and emulation using ML

Mar. 18

Time	Speaker	Title
JST 17:30-18:00 **CET** 9:30-10:00	Dr. Maha Mdini	Accelerating Climate Model Computation by Neural Networks
JST 18:00-18:30 **CET** 10:00-10:30	Dr. Simon Benaïchouche (IMT Atlantique & e-odyn)	Variational learning of sea surface current reconstructions from AIS data streams
JST 18:30-19:00 **CET** 10:30-11:00	-	Discussion

Seminar 4: Characterization of model errors using ML

Mar. 31

Time	Speaker	Title
JST 16:30-17:00 **CET** 9:30-10:00	Dr. Yicun Zhen (Ifremer & IMT Atlantique)	-
JST 17:00-17:30 **CET** 10:00-10:30	Dr. Arata Amemiya	Connecting Data Assimilation and Neural ODEs
JST 17:30-18:00 **CET** 10:30-11:00	-	Discussion

Seminar 5

Apr. 14

Time	Speaker	Title
JST 16:30-17:30 **CET** 9:30-10:30	Dr. Pierre Tandeo	-
JST 17:30-18:00 **CET** 10:30-11:00	-	Discussion

Only in 10 minutes!!

(JMA brochure)

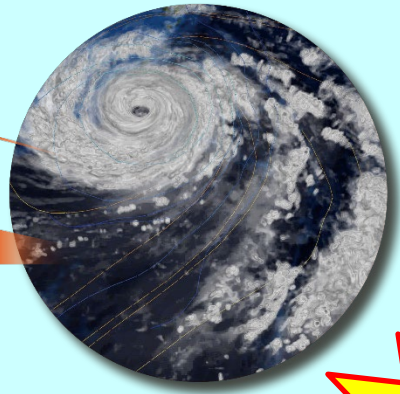
1.34 m in 10 minutes!!

5 people died in River Toga in Kobe on July 28, 2008

Big Data Assimilation

Observations

Simulations



Data Assimilation

100x

Big Data

New sensors, IoT

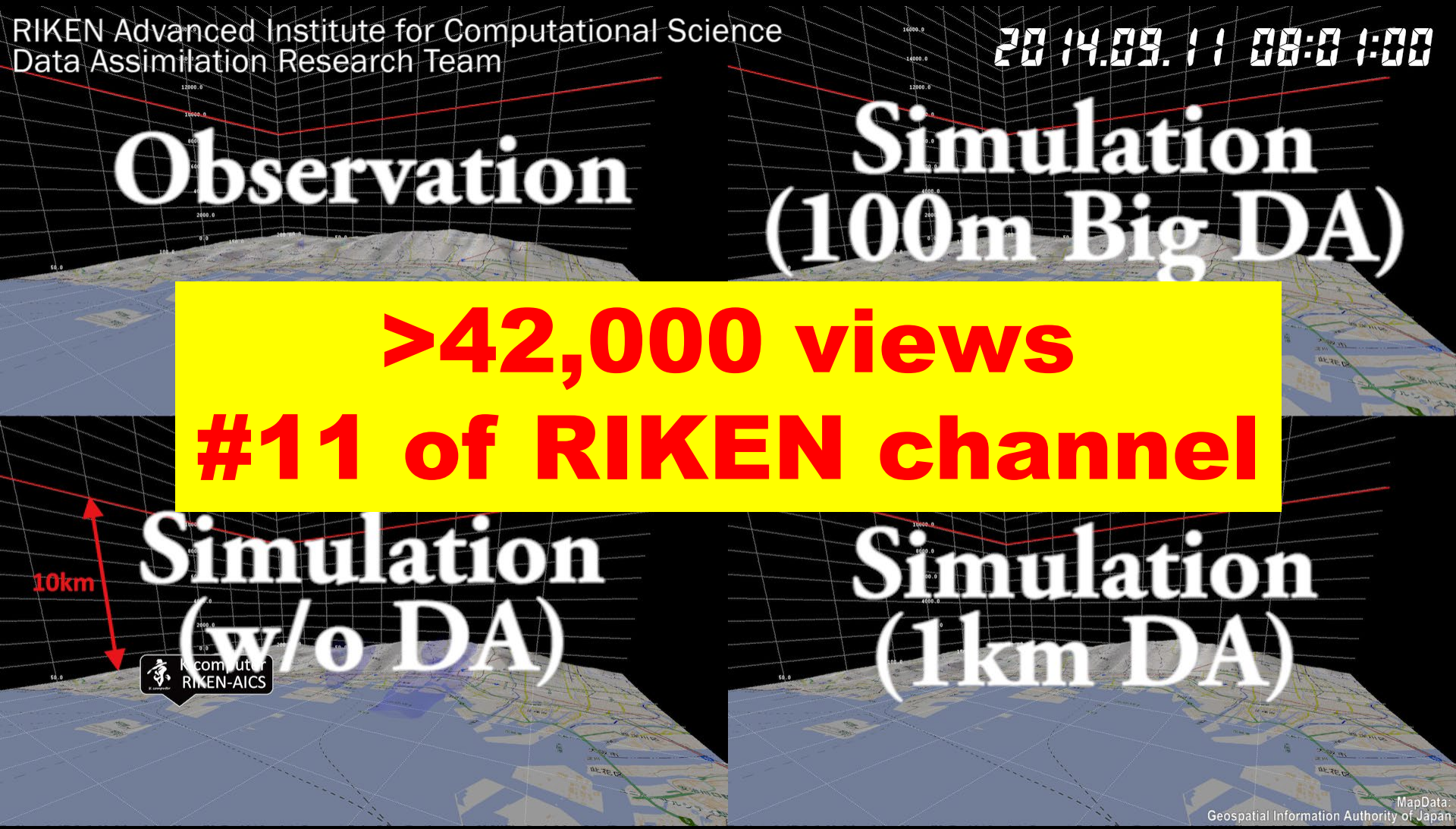
Big Data

100x

Powerful supercomputer

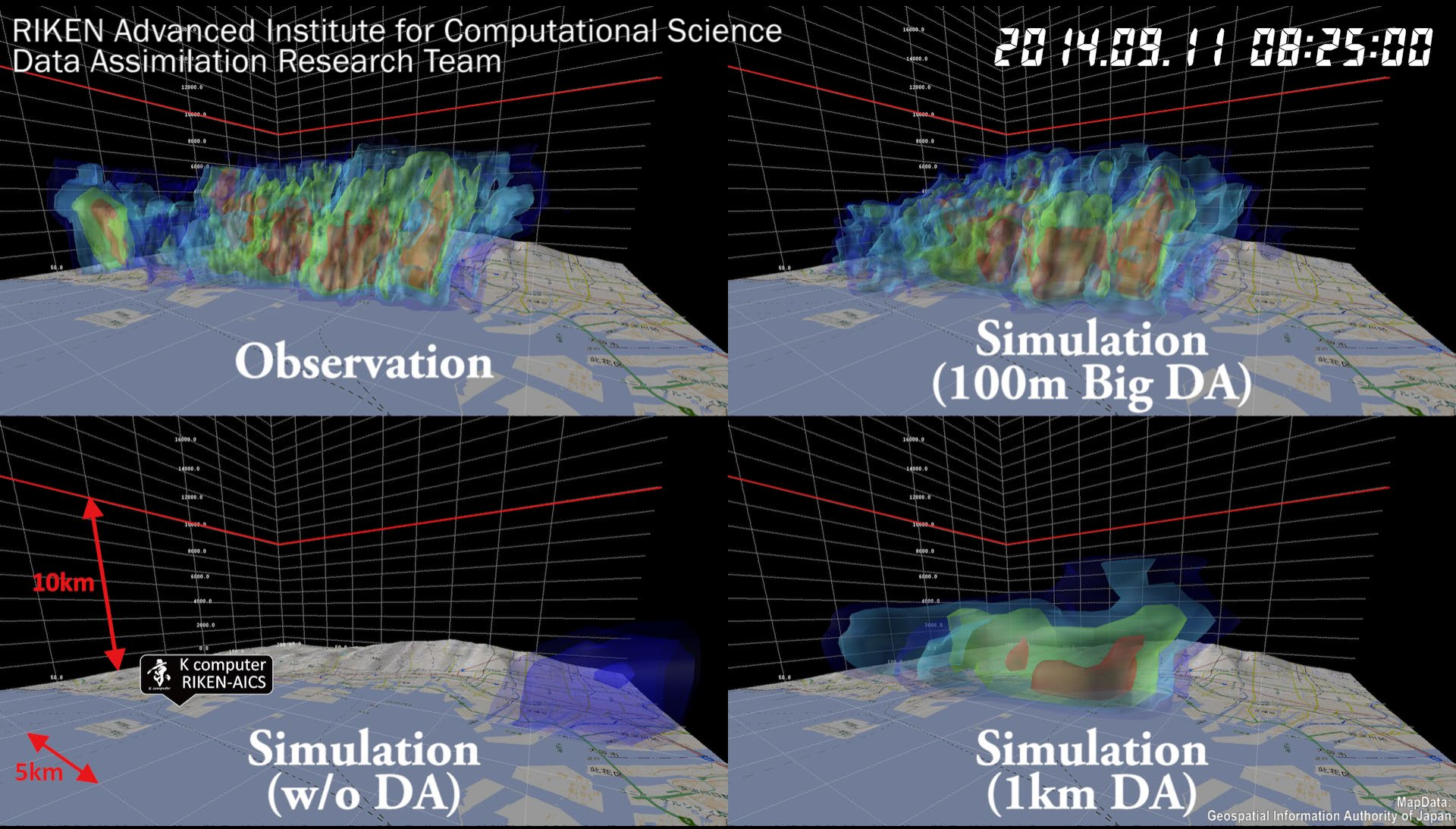
©RIKEN

9/11/2014, sudden local rain



Youtube https://www.youtube.com/watch?v=42NZTGdp1Js

9/11/2014, sudden local rain



Youtube https://www.youtube.com/watch?v=42NZTGdp1Js

Real-time forecast with 30-sec refresh during August 25-September 5, 2020

← 前の記事 ↑ 一覧へ戻る

Press announcement on August 21, 2020

2020年8月21日
理化学研究所
情報通信研究機構
大阪大学
株式会社エムティーアイ
筑波大学
東京大学
科学技術振興機構

30秒ごとに更新するゲリラ豪雨予報

－首都圏でのリアルタイム実証実験を開始－

理化学研究所（理研） 計算科学研究センターデータ同化研究チームの三好建正チームリーダー、情報通信研究機構 電磁波研究所リモートセンシング研究室の佐藤晋介研究マネージャー、大阪大学 大学院工学研究科の牛尾知雄教授、株式会社エムティーアイ ライフ事業部気象サービス部の小池佳奈部長、筑波大学 計算科学研究センターの朴泰祐教授、東京大学 情報基盤センターの中島研吾教授らの共同研究グループは、2020年8月25日から9月5日まで、首都圏において30秒ごとに更新する30分後までの超高速降水予報のリアルタイム実証実験を行います。

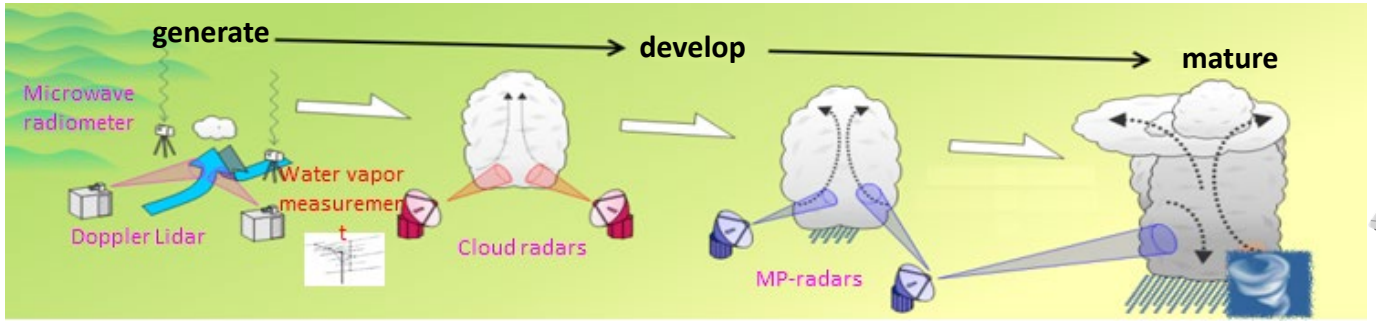
本研究成果は、近年増大する突発的なゲリラ豪雨[1]などの降水リスクに対して、コンピュータ上の仮想世界と現実世界をリンクさせることで、超スマート社会Society 5.0[2]の実現に貢献すると期待できます。

共同研究グループは、2016年にスーパーコンピュータ「京」[3]とフェーズドアレイ気象レーダ（PAWR）[4]を生かした「ゲリラ豪雨予測手法」を開発しました注1）。今回、この手法を高度化し、さいたま市に設置されている情報通信研究機構が運用する最新鋭のマルチパラメータ・フェーズドアレイ気象レーダ（MP-PAWR）[5]による30秒ごとの雨雲の詳細な観測データと、筑波大学と東京大学が共同で運営する最先端共同HPC基盤施設（JCAHPC）のスーパーコンピュータOakforest-PACS[6]を用いて、リアルタイムで30秒ごとに新しいデータを取り込んで更新し、30分後まで予測する超高速降水予報システムを開発しました。この予測データを、理研の天気予報研究のウェブページでは30秒ごとに分割して連続的に表示します。これまでの天気予報と比べて桁違いに速い速度で更新することにより、わずか数分の間に急激に発達するゲリラ豪雨を予測できます。このリアルタイム予報は世界初かつ唯一の取り組みで、研究開発に着手した2013年10月から継続してきたさまざまな成果の集大成です。

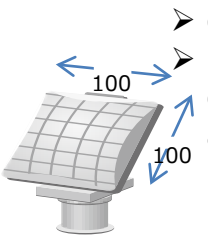
実証実験で得る予報データは、気象業務法に基づく予報業務許可のもと、理研の天気予報研究のウェブページ（理研天気予報研究）および株式会社エムティーアイのスマートフォンアプリ「3D雨雲ウォッチ」で8月25日午後2時から公開します。

Development of MP-PAWR

Multi-parameter phased array weather radar (MP-PAWR) was developed by SIP (Cross-ministerial Strategic Innovation Promotion Program) in 2014-2018 as a research subject of “torrential rainfall and tornadoes prediction.”



Early forecasting by water vapor, cloud, and precipitation observation

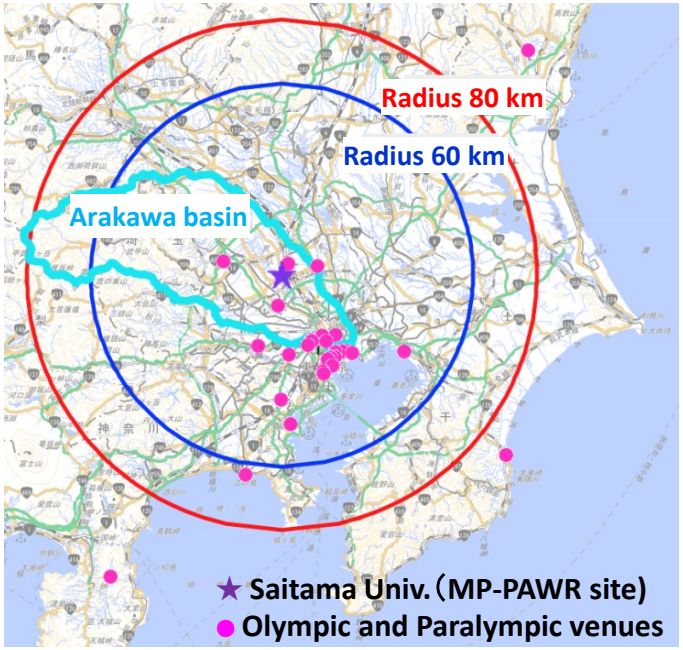


- dual polarization
- 100×100 elements array antenna

MP-PAWR features

MP-PAWR installed at Saitama Univ. on Nov 21, 2017, and observation began in July 2018.

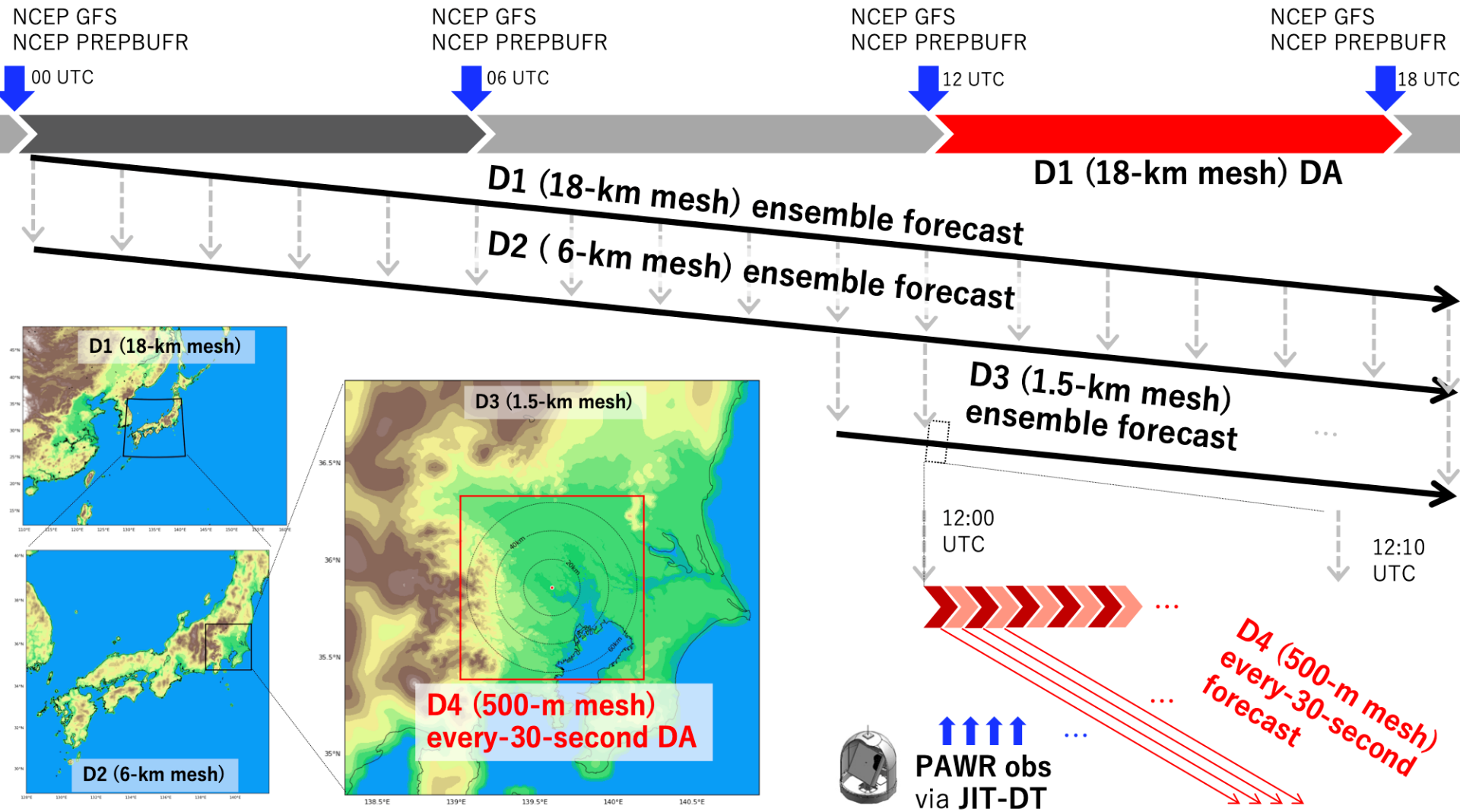
MP-PAWR antenna



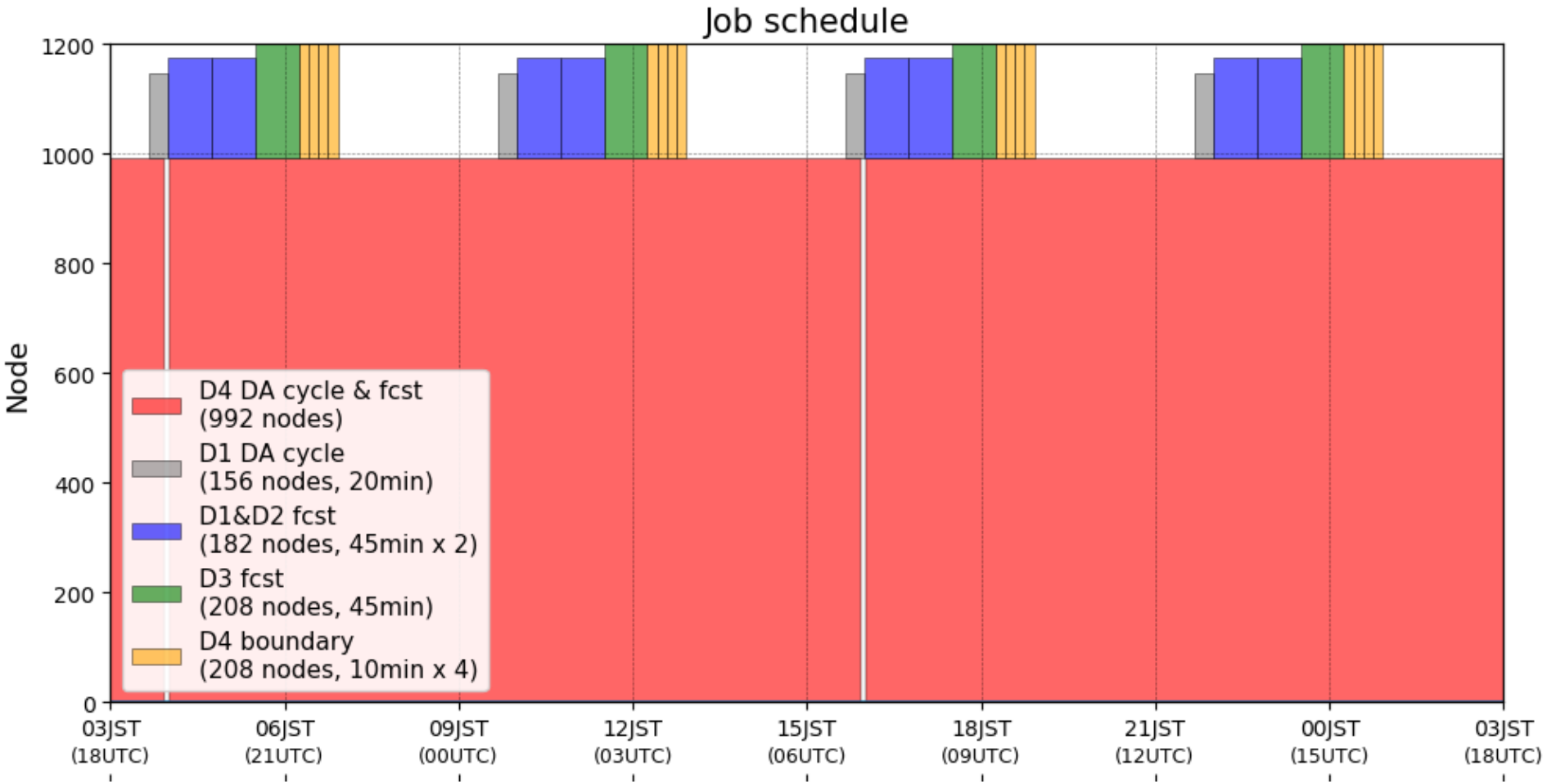
MP-PAWR observation area


東京大学情報基盤センター
スーパーコンピューティング部門
Supercomputing Division,
Information Technology Center,
The University of Tokyo
ENGLISH
サイト内検索
お問い合わせ • リンク • サイトマップ
ホーム
スパコン
利用案内
サポート
FAQ
研究会・イベント
広報・刊行物
研究成果
研究部門について
ホーム 〉 スパコン 〉 Oakforest-PACS 〉 サービス 〉 2020年度 Oakforest-PACS の運用について
スパコン
Oakbridge-CX
Oakforest-PACS
サービス
利用申込
運用スケジュール
2020年度 Oakforest-PACS の運用について
最先端共同HPC基盤施設（JCAHPC）と理化学研究所計算科学研究センター（理研R-CCS）は，2020年東京オリンピック・パラリンピック（※） 期間中に，関東地区に
「ゲリラ豪雨」リアルタイム予報と情報配信を協力して実施する予定です。ゲリラ豪雨は突発的で天気
る正確な予測が困難な局地的豪雨であり，今回は
Oakforest-PAC
上で，理研R-CCSの
E-LETKF」コードを使用して，埼玉大学に設置された世界初の
（MP-PAWR）の観測
基づき，リアルタ
日間， 動作確認のため
Special arrangement
for an exclusive use of Oakforest-PACS
of the U of Tokyo and Tsukuba U

Real-time workflow



Real-time job scheduling



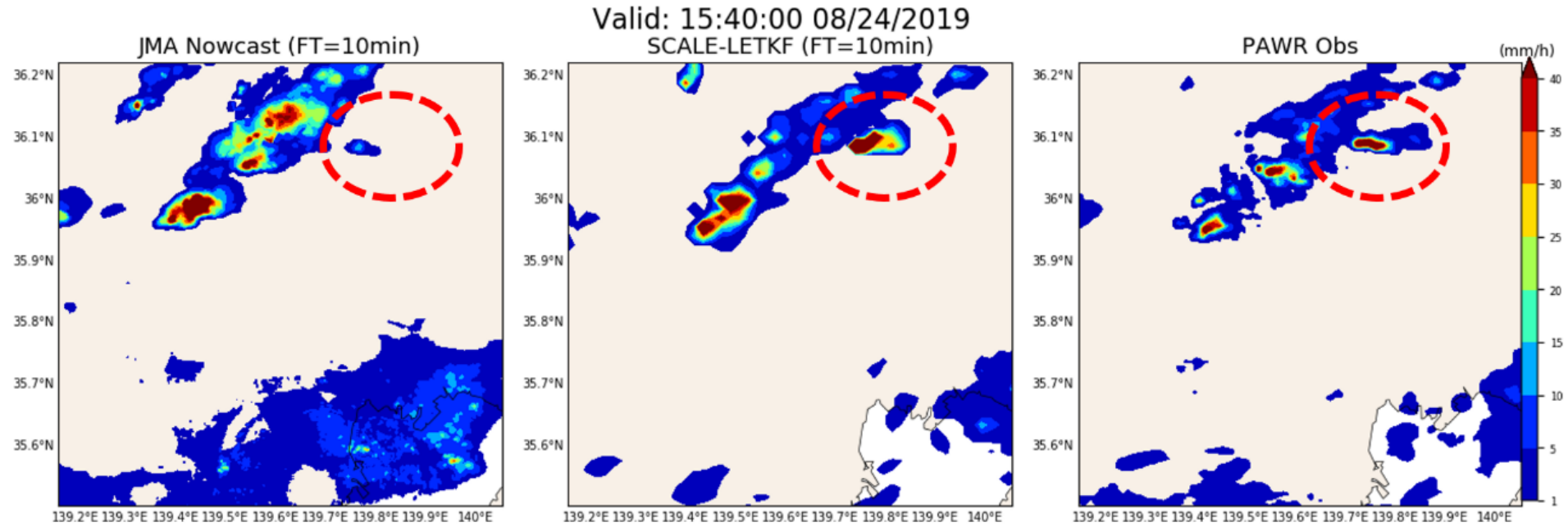
25 August 2019 00:40 JST

Past case study

JMA Nowcast 10-min lead

This study 10-min lead

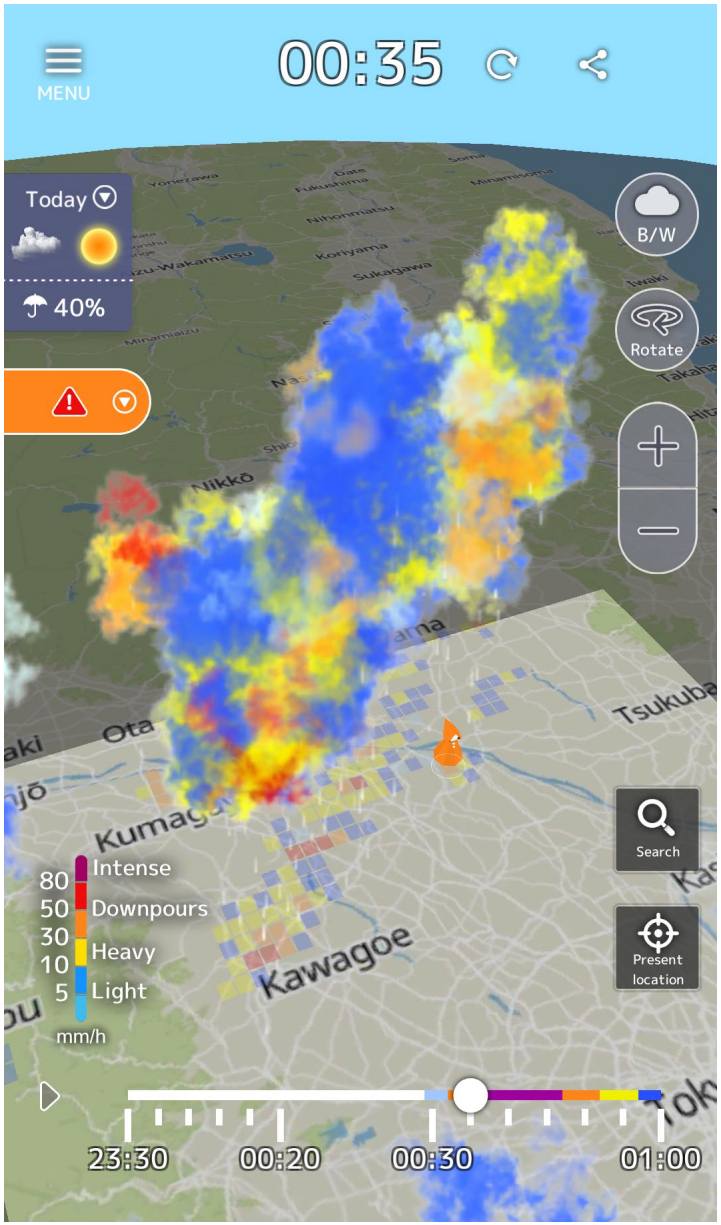
MP-PAWR observation

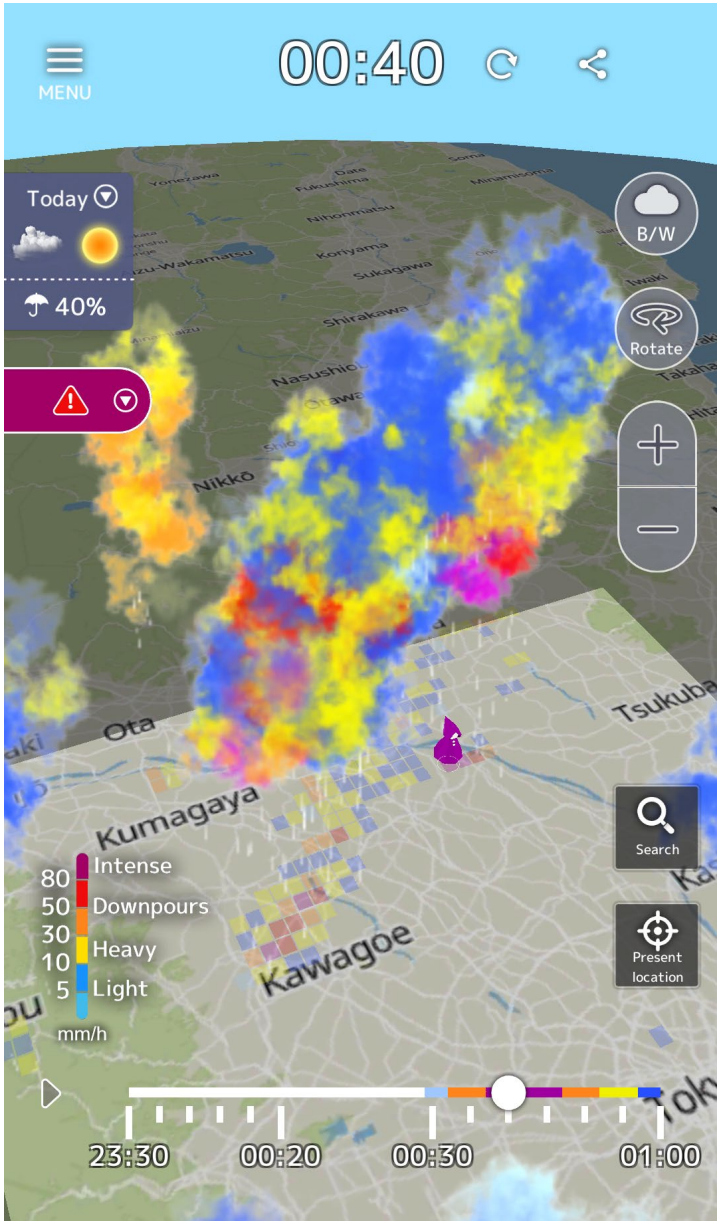


Process-driven model predicts rapid changes of rains

- Rapid development (red broken circles)
- Rapid weakening (left of red circles)

Smartphone app by MTI Co. Ltd.





データ同化と数値予報モデルによる関東地方の降水30分予報

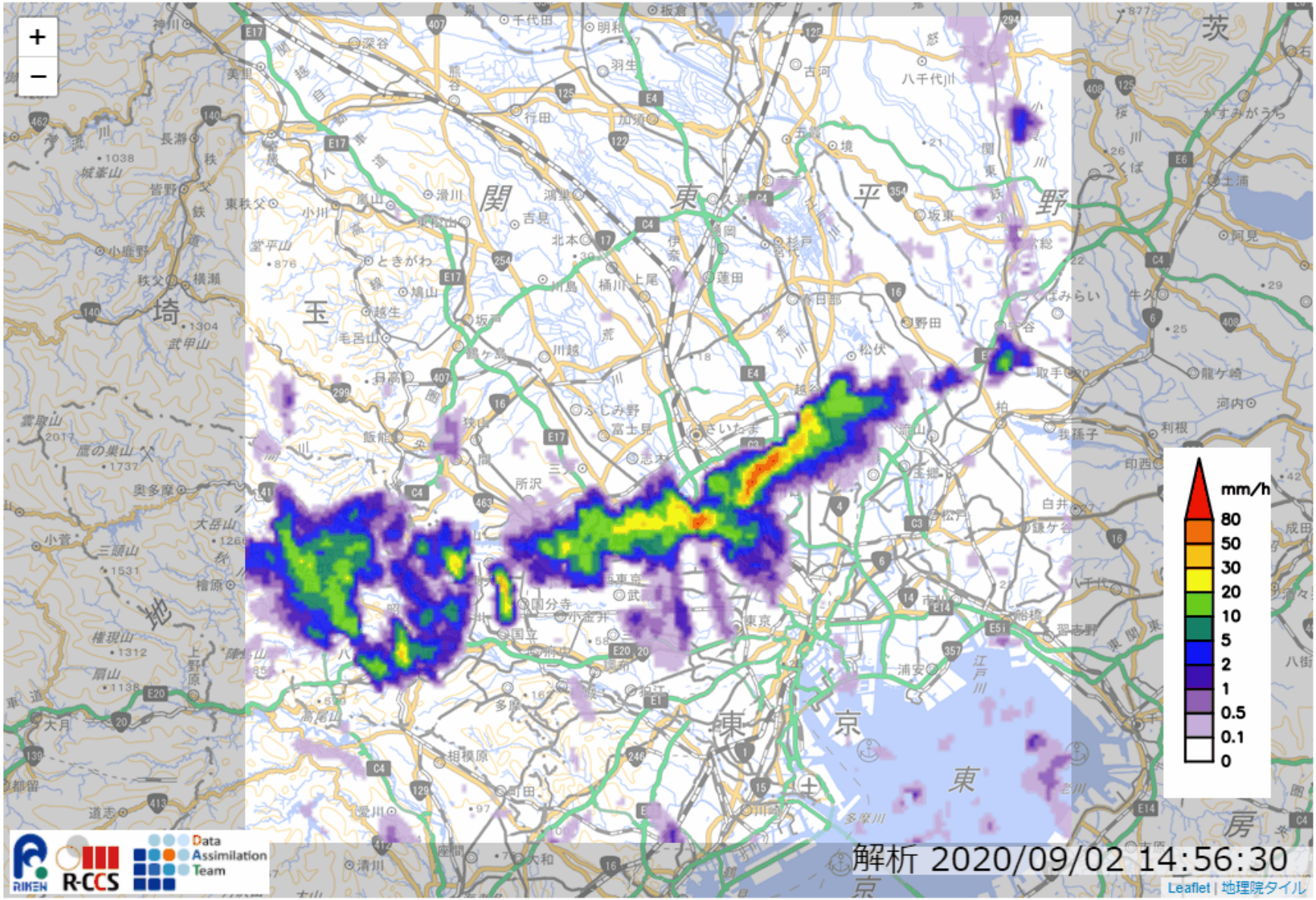
（30秒間隔で、30分先まで30秒ごとに予報、高度2km面）

予報開始時刻: 2020/09/02 15:06:00

<< 解析 2020/09/02 14:56:30 >>　観測/予測　解析/予測　?

☑ 30秒毎に新情報を自動ダウンロード(30分で自動OFF)　☑ アニメーション：　10分前から30分後まで

気象庁の警報・注意報を見る



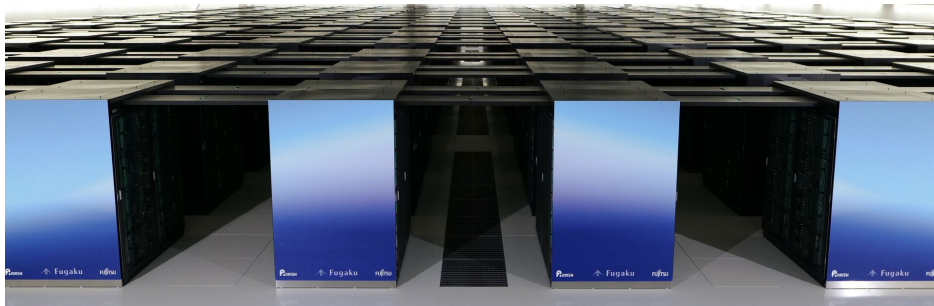
Plans

Verify & Publish

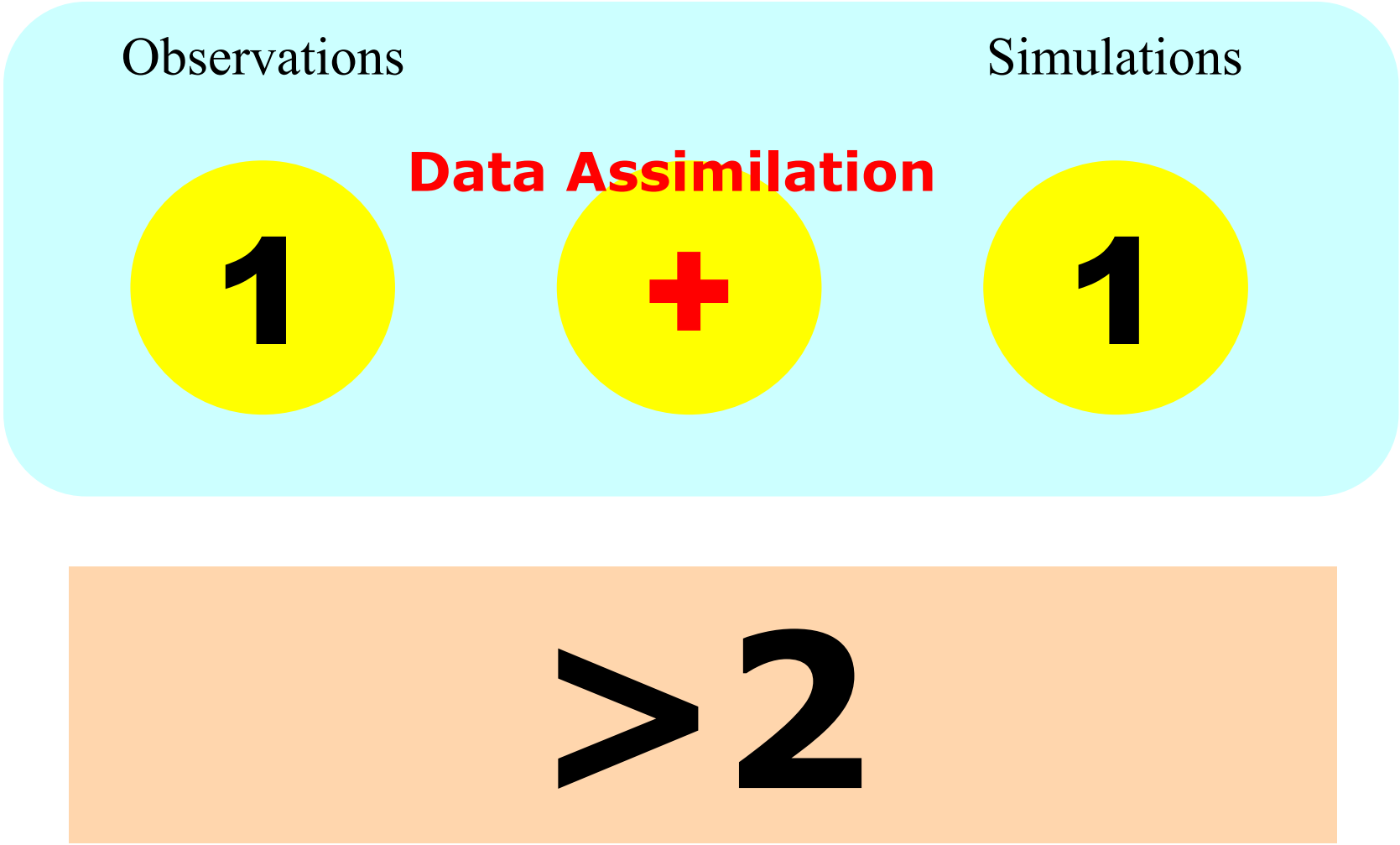
Improve

Test in 2021

New “Fugaku”



Data Assimilation (DA)



Data Assimilation (DA)

Data-driven
Induction
Real world

Process-driven
Deduction
Cyber world

Observations

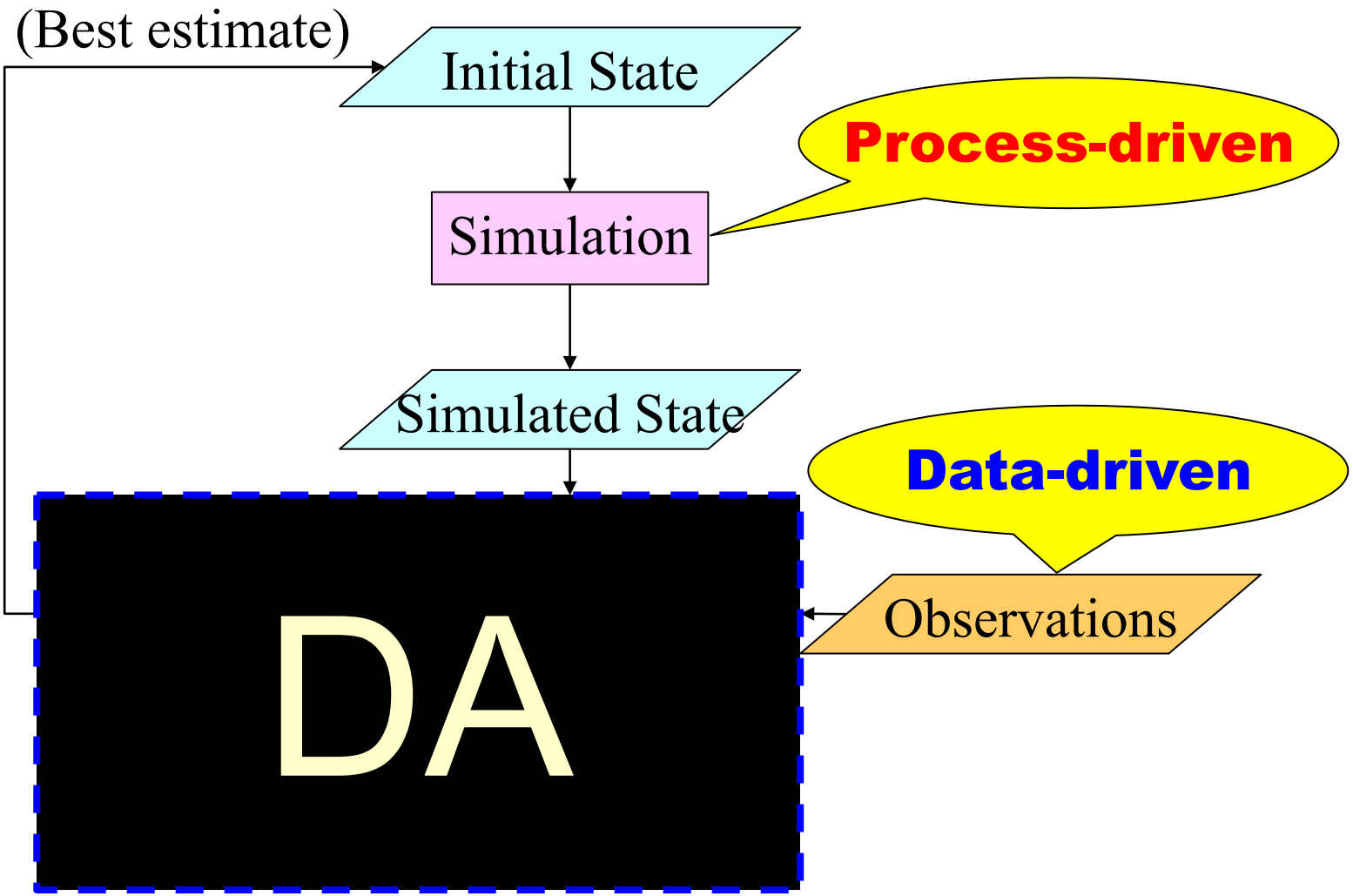
Simulations

Data Assimilation

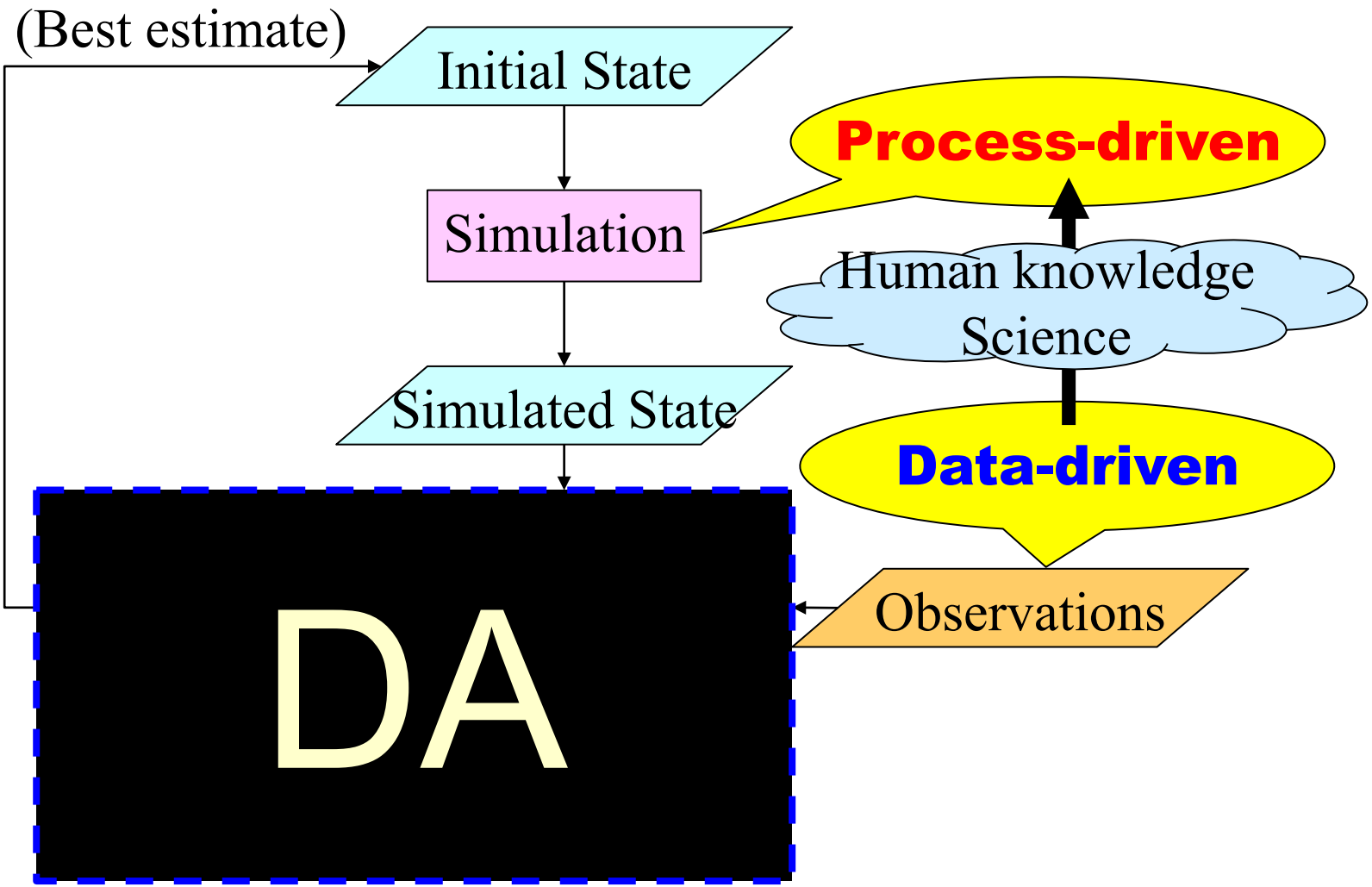
1 **+** **1**

>2

DA workflow


(Best estimate)
Initial State
Process-driven
Simulation
Simulated State
Data-driven
DA
Observations

DA workflow


(Best estimate)
Initial State
Simulation
Simulated State
DA
Process-driven
Human knowledge
Science
Data-driven
Observations

Scientific methods

Observations
Experiments

Noisy/missing data

Scientific methods

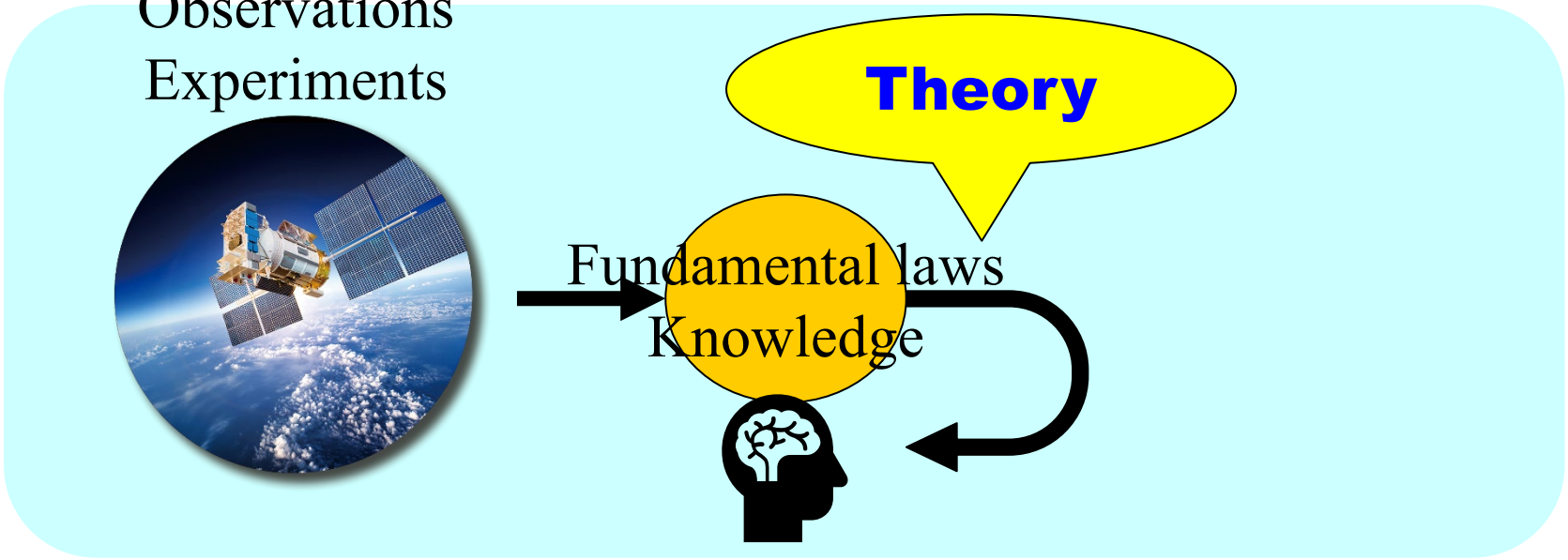
Observations
Experiments

Dealing with noise/miss

Fundamental laws
Knowledge

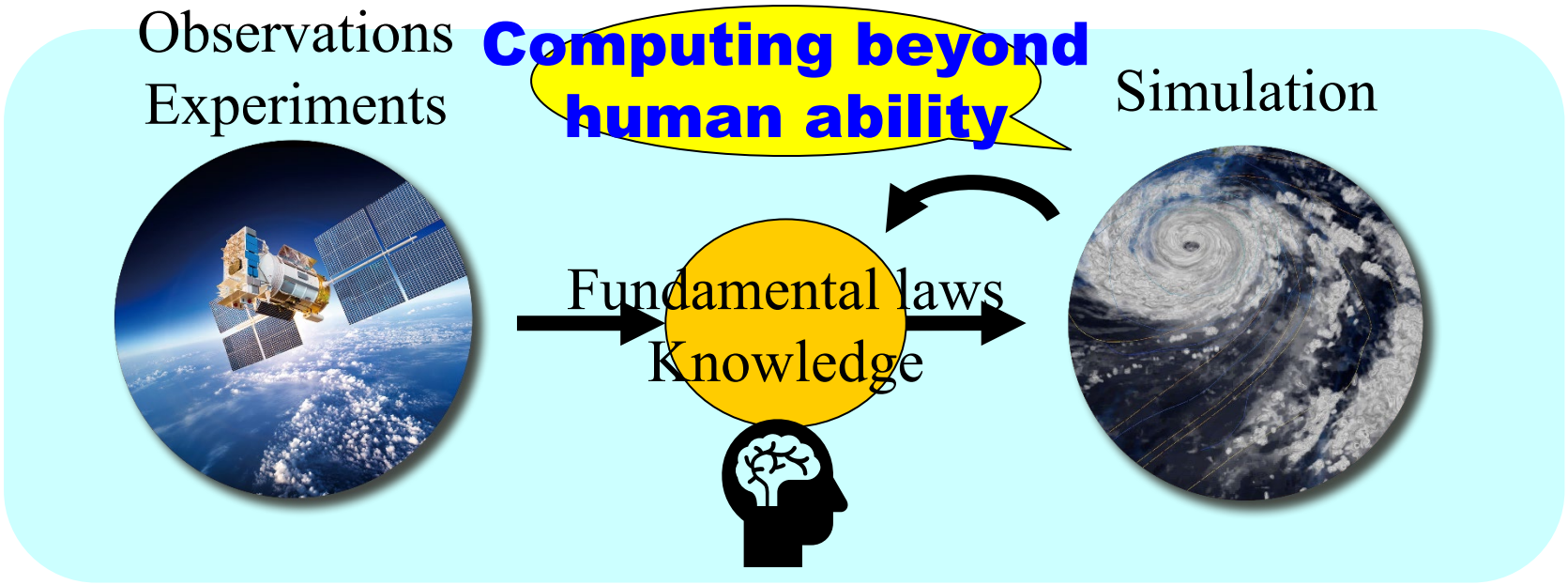
1st science (experimental)

Scientific methods



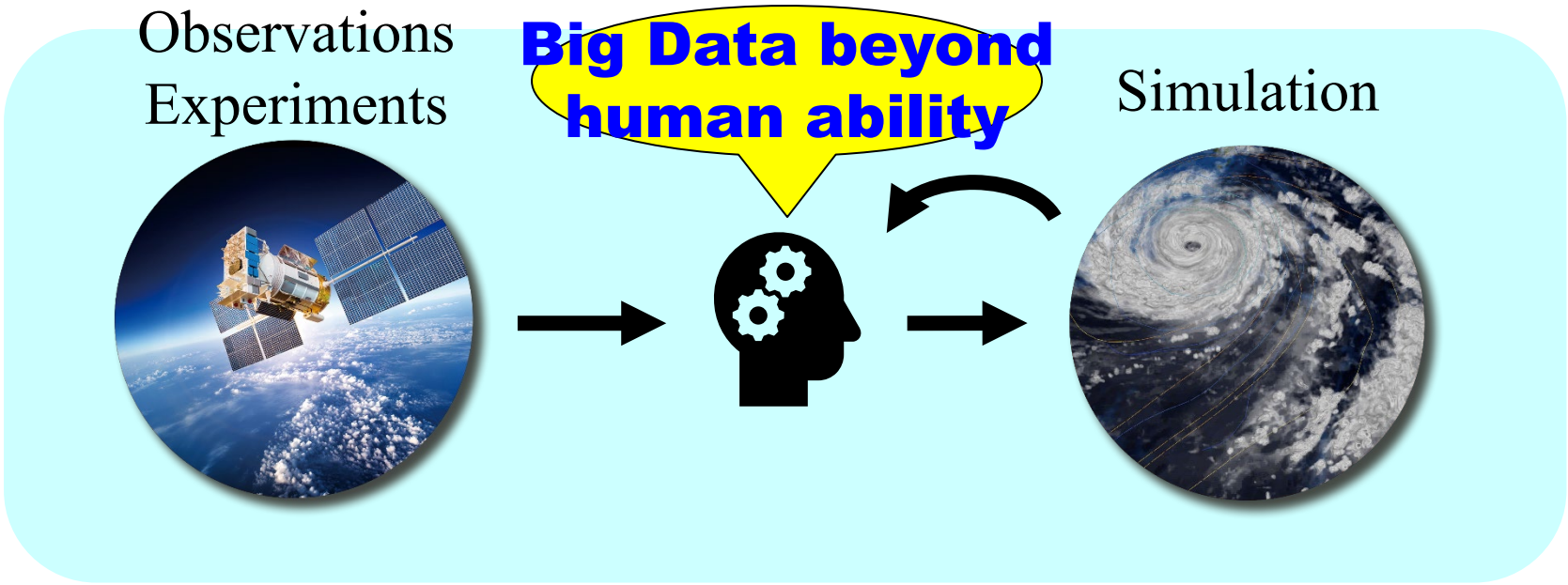
2nd science (theoretical)

Scientific methods



3rd science (computational)

Scientific methods

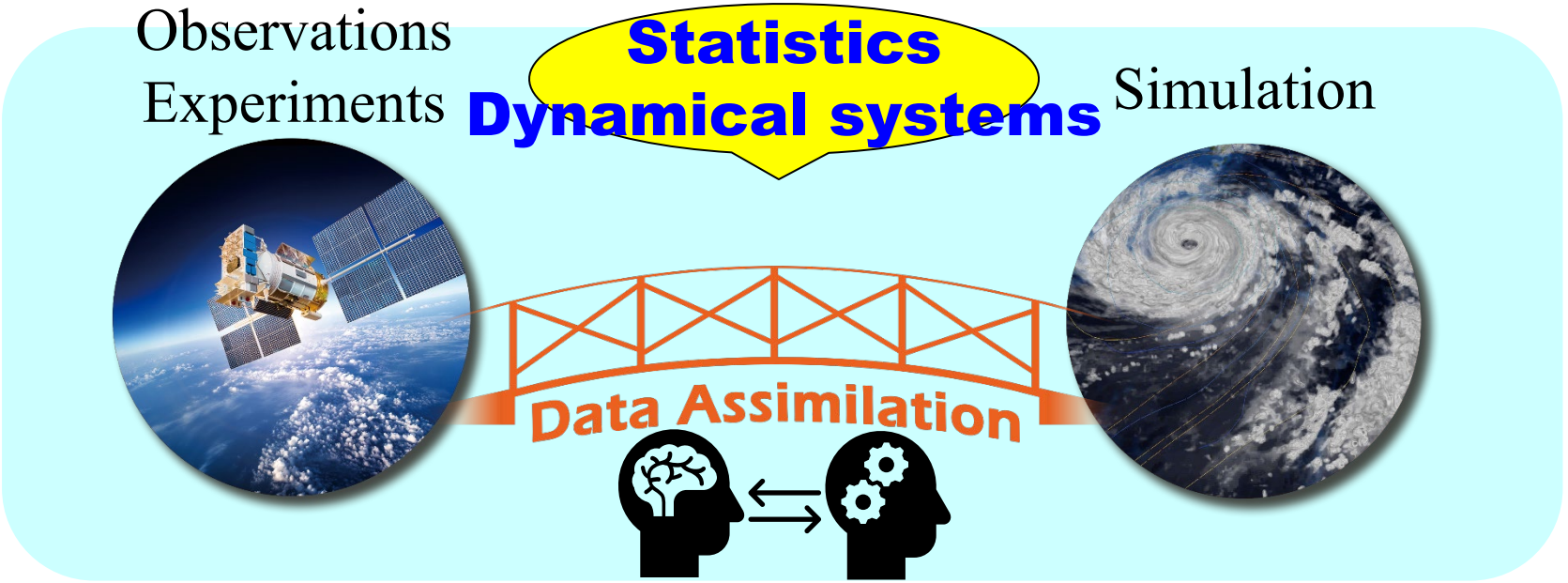


4th science (data-centric)

Data Assimilation

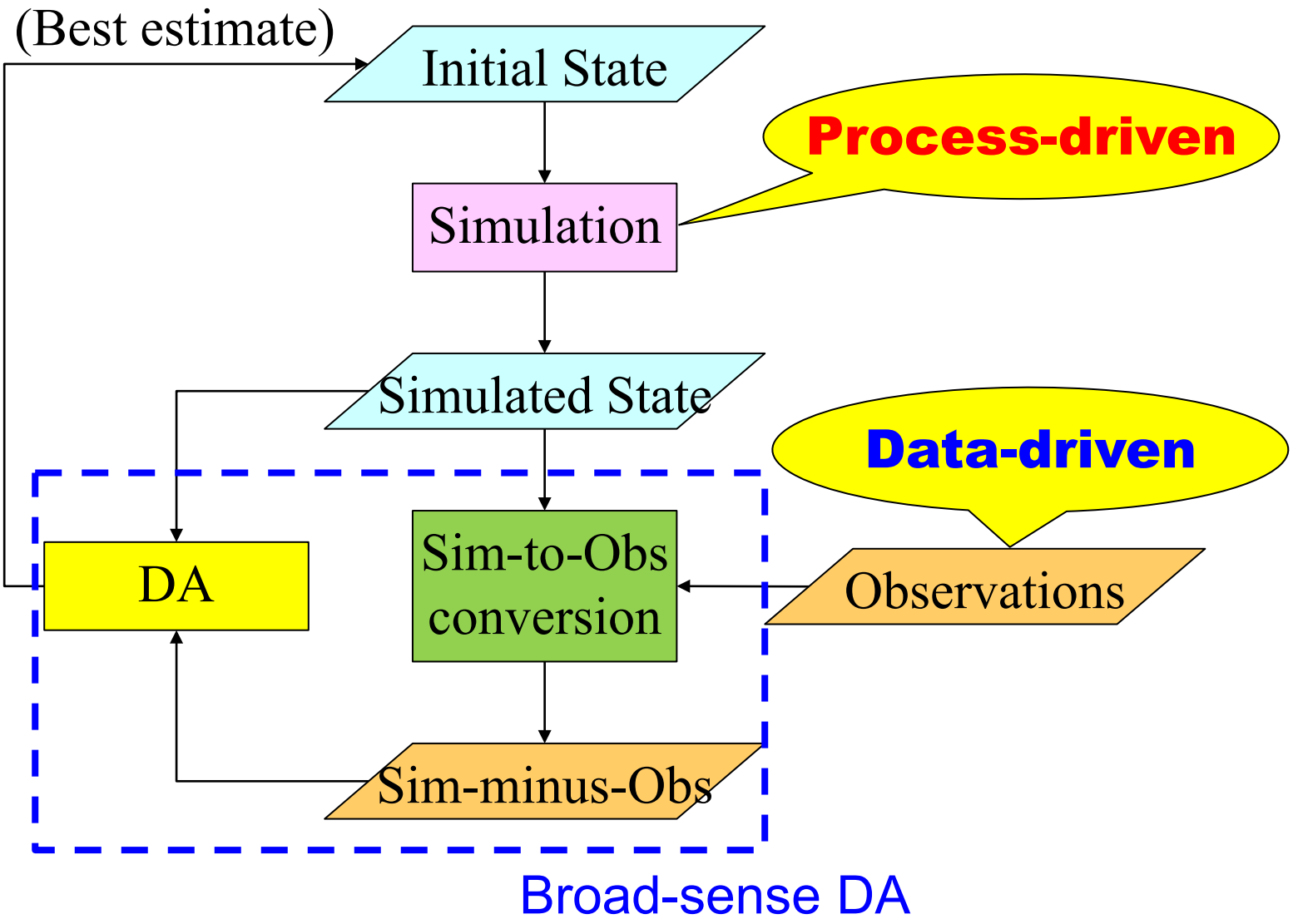
Data Assimilation connects data and simulation and brings synergy

The 5th paradigm?

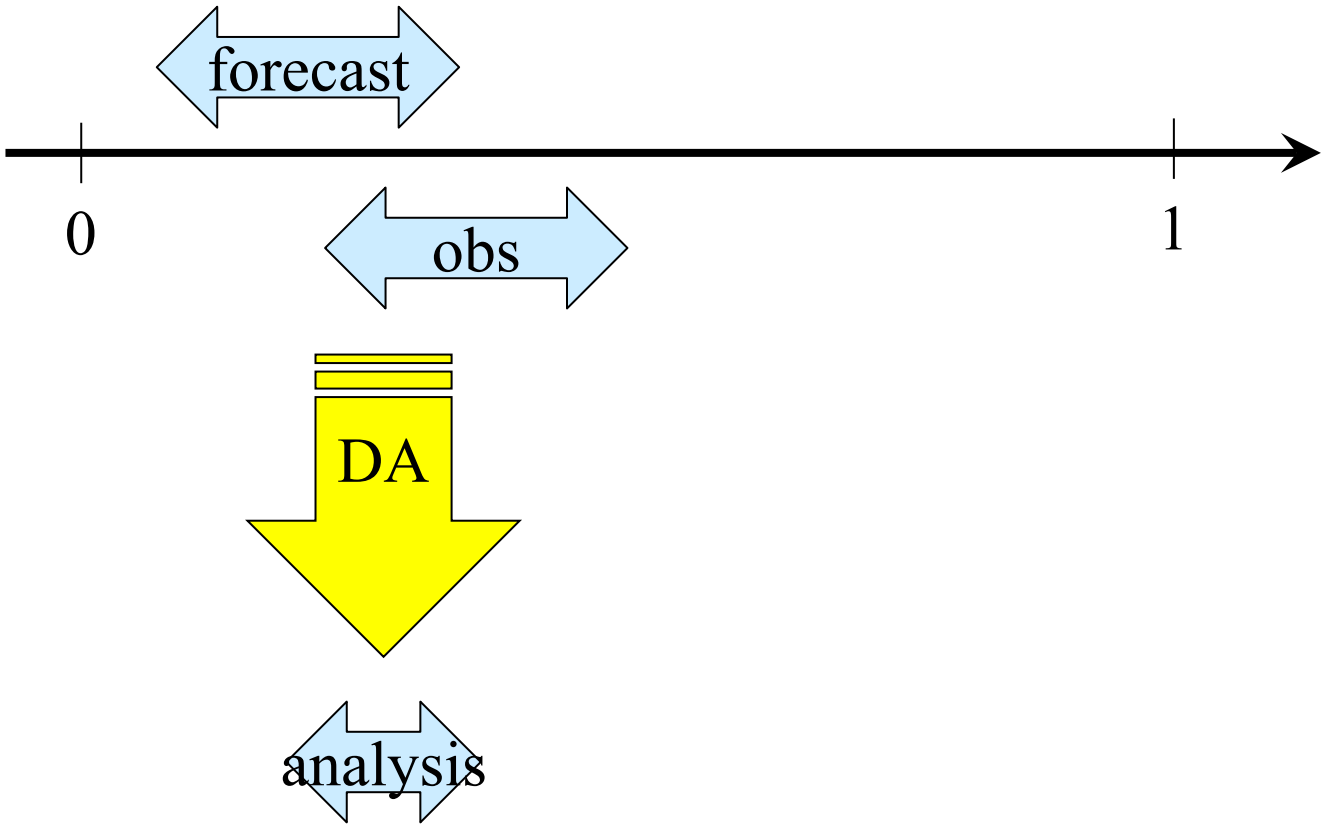


5th science ??
(data × computation)

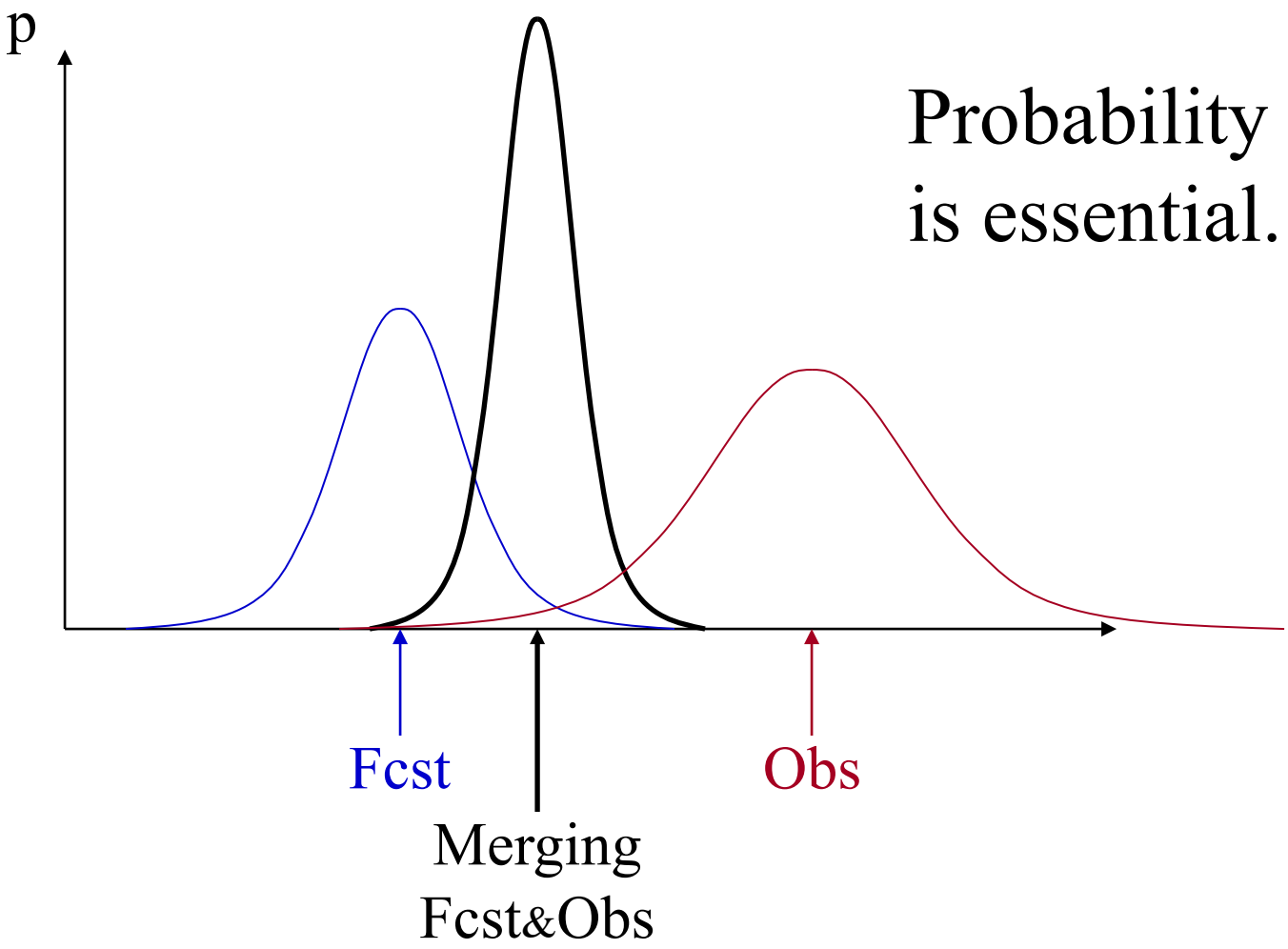
DA workflow


(Best estimate)
Initial State
Process-driven
Simulation
Simulated State
Data-driven
DA
Sim-to-Obs conversion
Observations
Sim-minus-Obs
Broad-sense DA

DA = math of errors

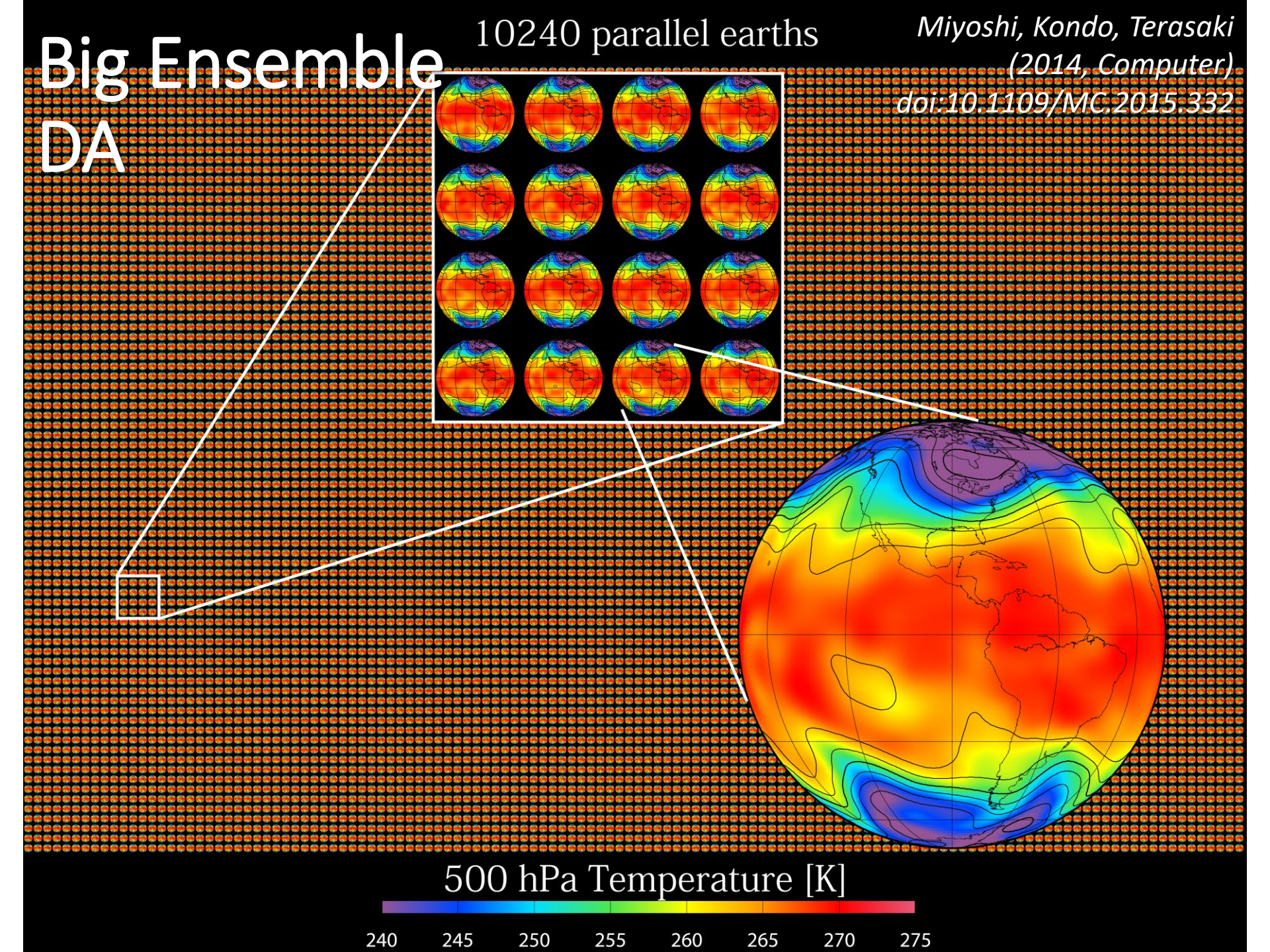


Merging 2 information (Bayesian estimation)

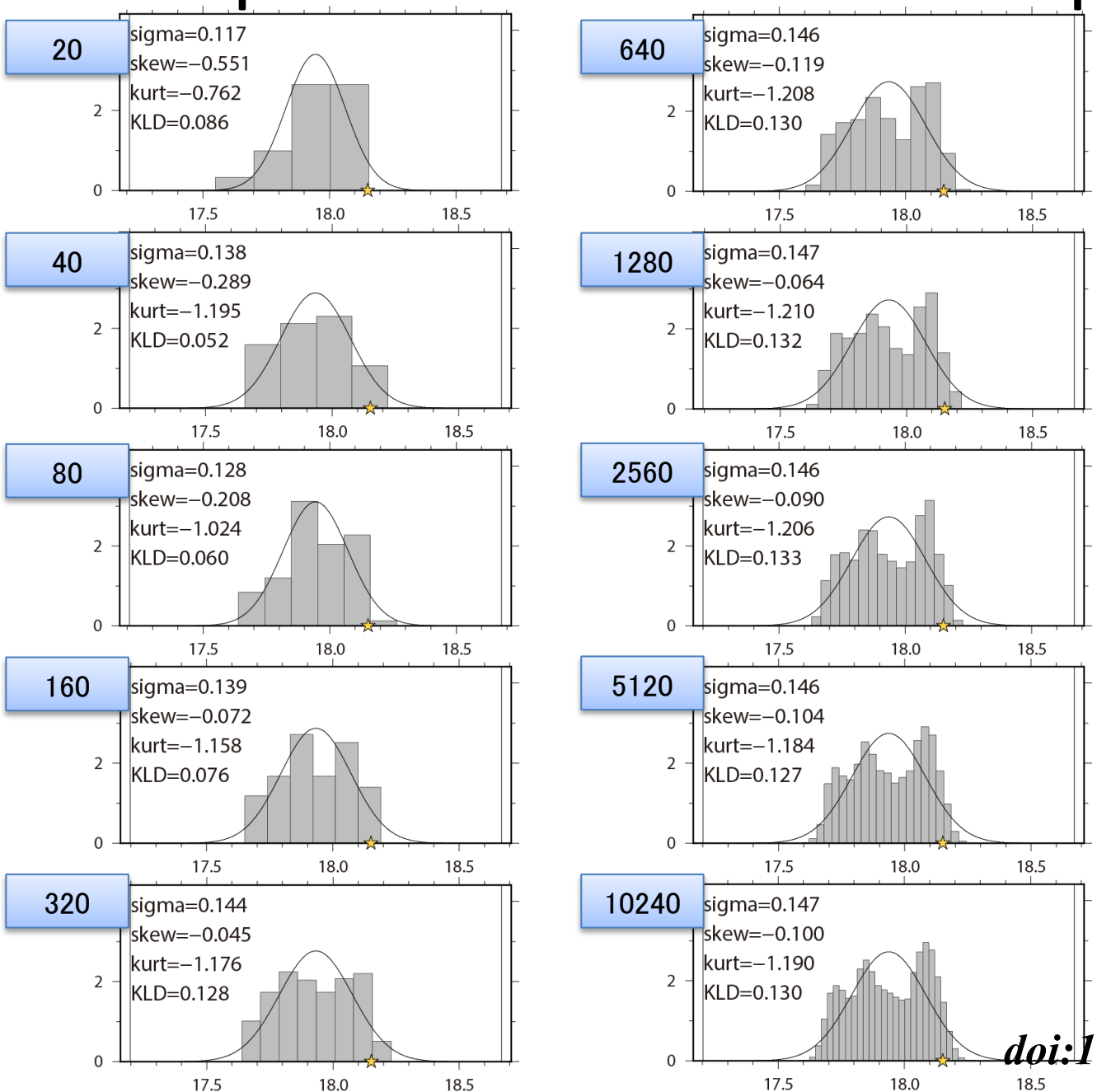


Probability distribution is essential.

Big Ensemble DA



Sample size = Resolution of probability

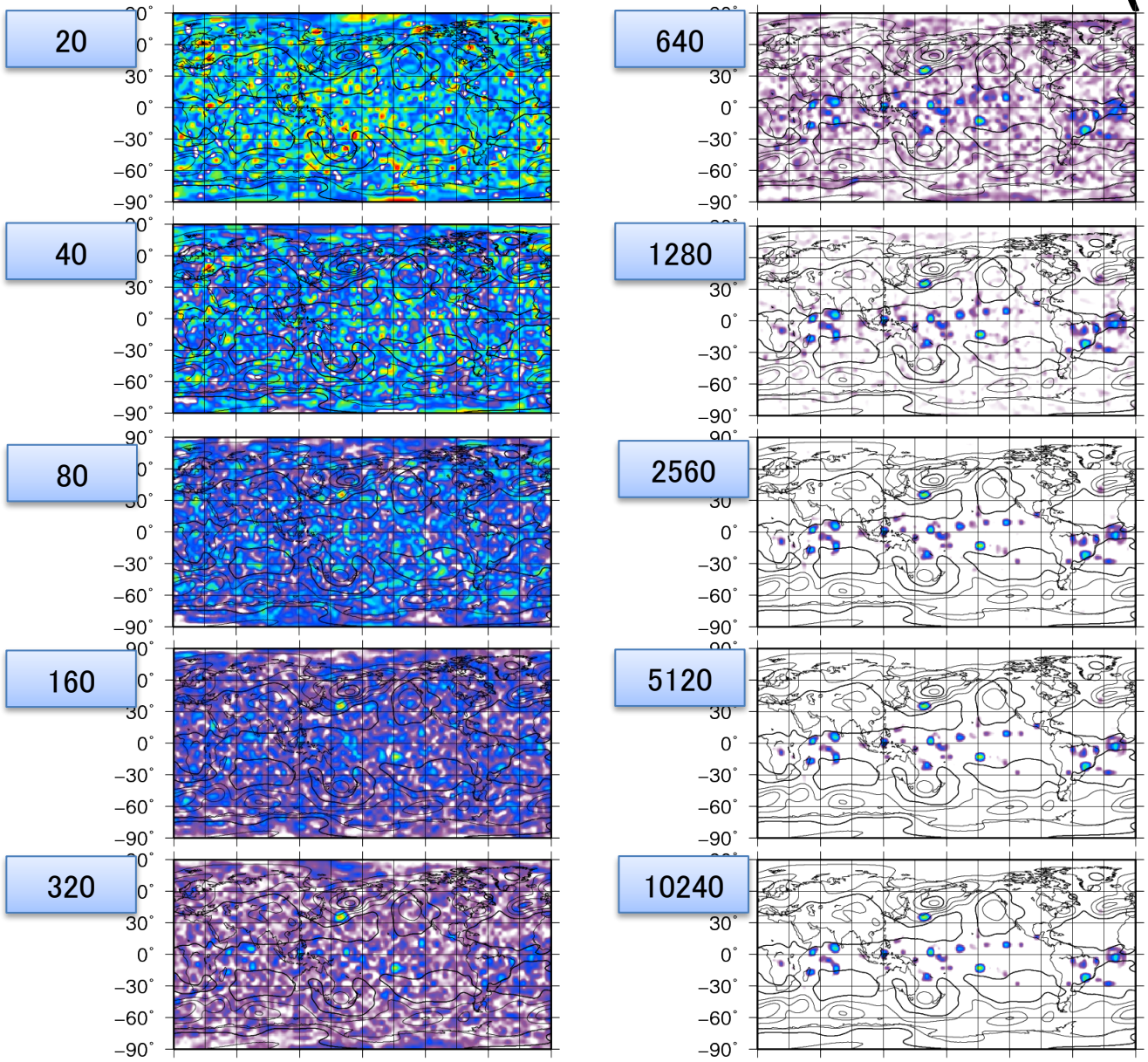


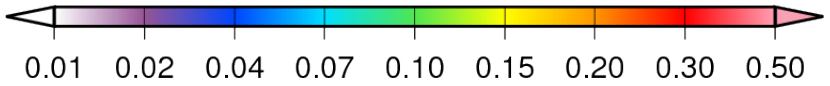
1.856N, 176.25E

Kondo&Miyoshi (2019, NPG)

doi:10.5194/npg-26-211-2019

Non-Gaussian metric (KLD)

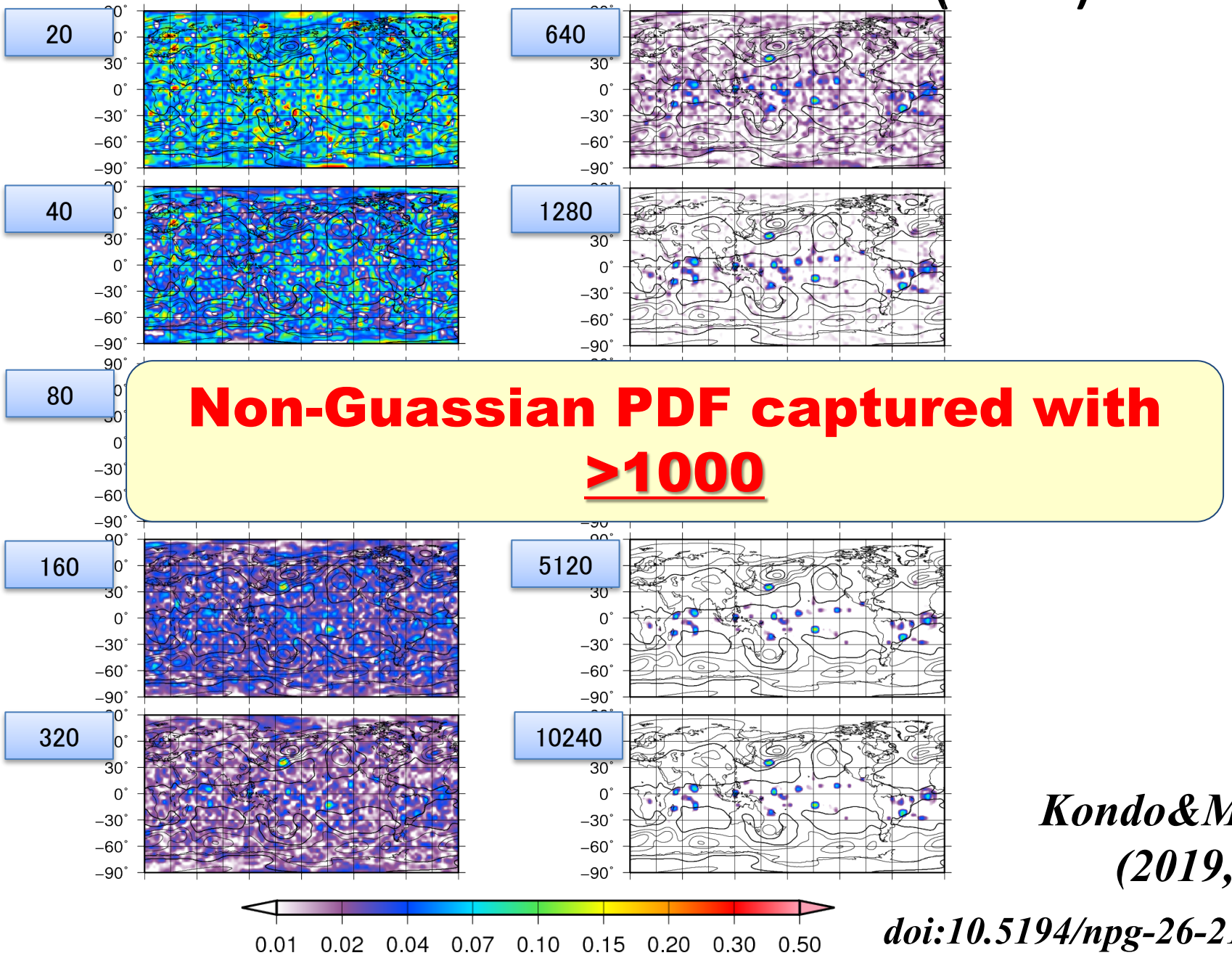




Kondo&Miyoshi (2019, NPG)

doi:10.5194/npg-26-211-2019

Non-Gaussian metric (KLD)



Non-Guassian PDF captured with >1000

Kondo&Miyoshi (2019, NPG)

doi:10.5194/npg-26-211-2019

Pushing the limits

Big Data × *Big Simulations*

Big ensemble (10240 ensemble members)

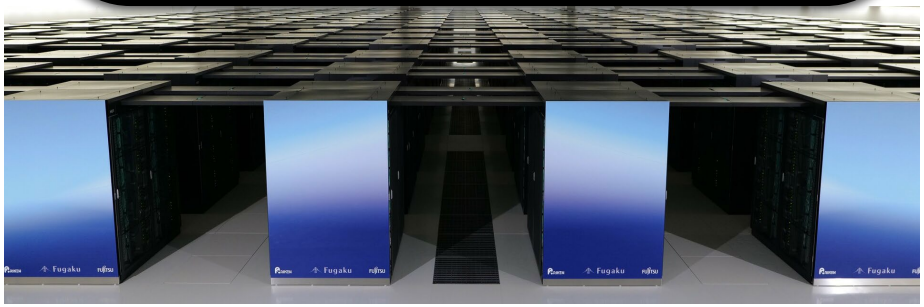
Rapid update (30-second update)

High resolution (100-m mesh)

➔ Future Numerical Weather Prediction

Fugaku

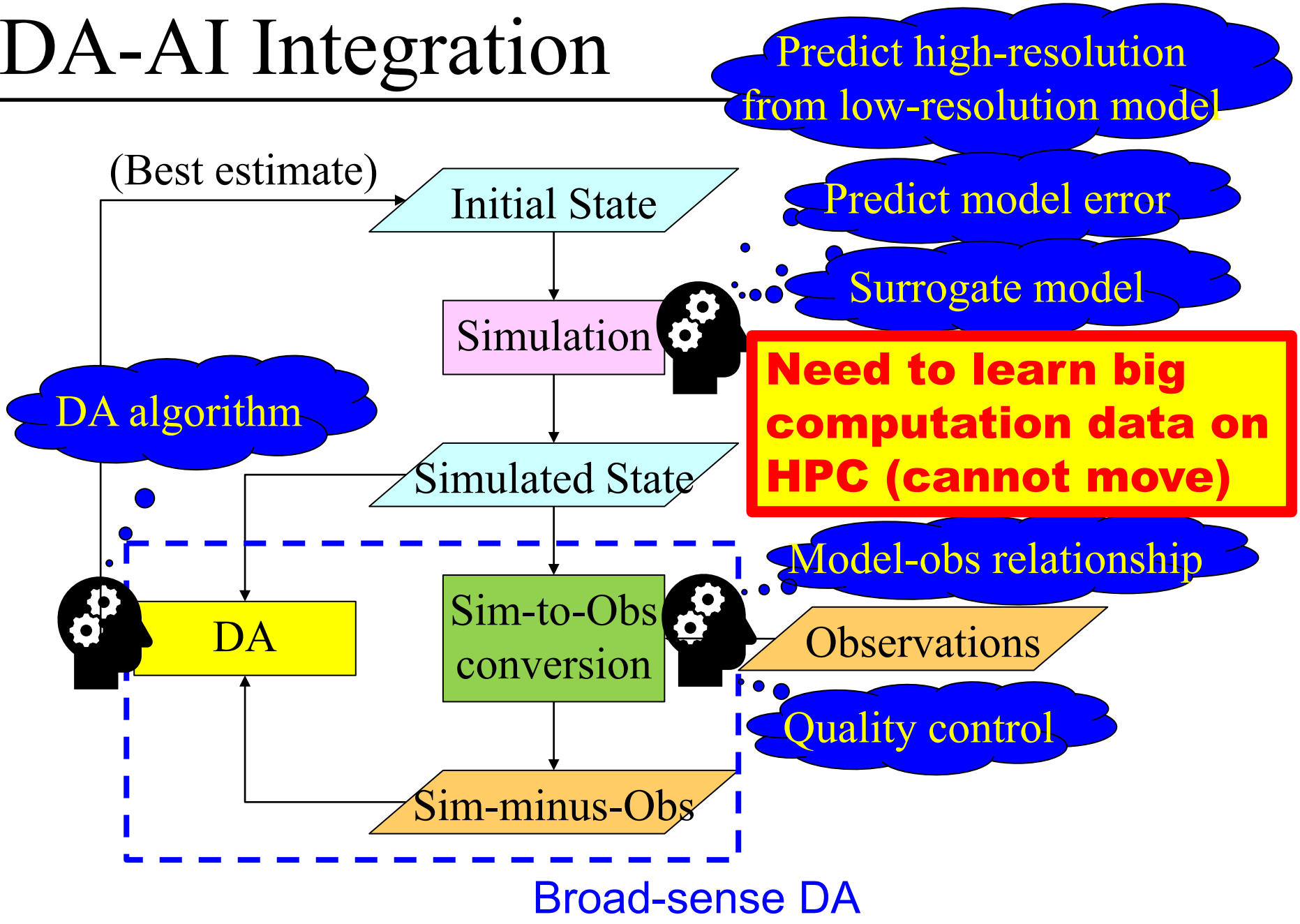
- Good for both Big DA and ML



K (and other HPCs)

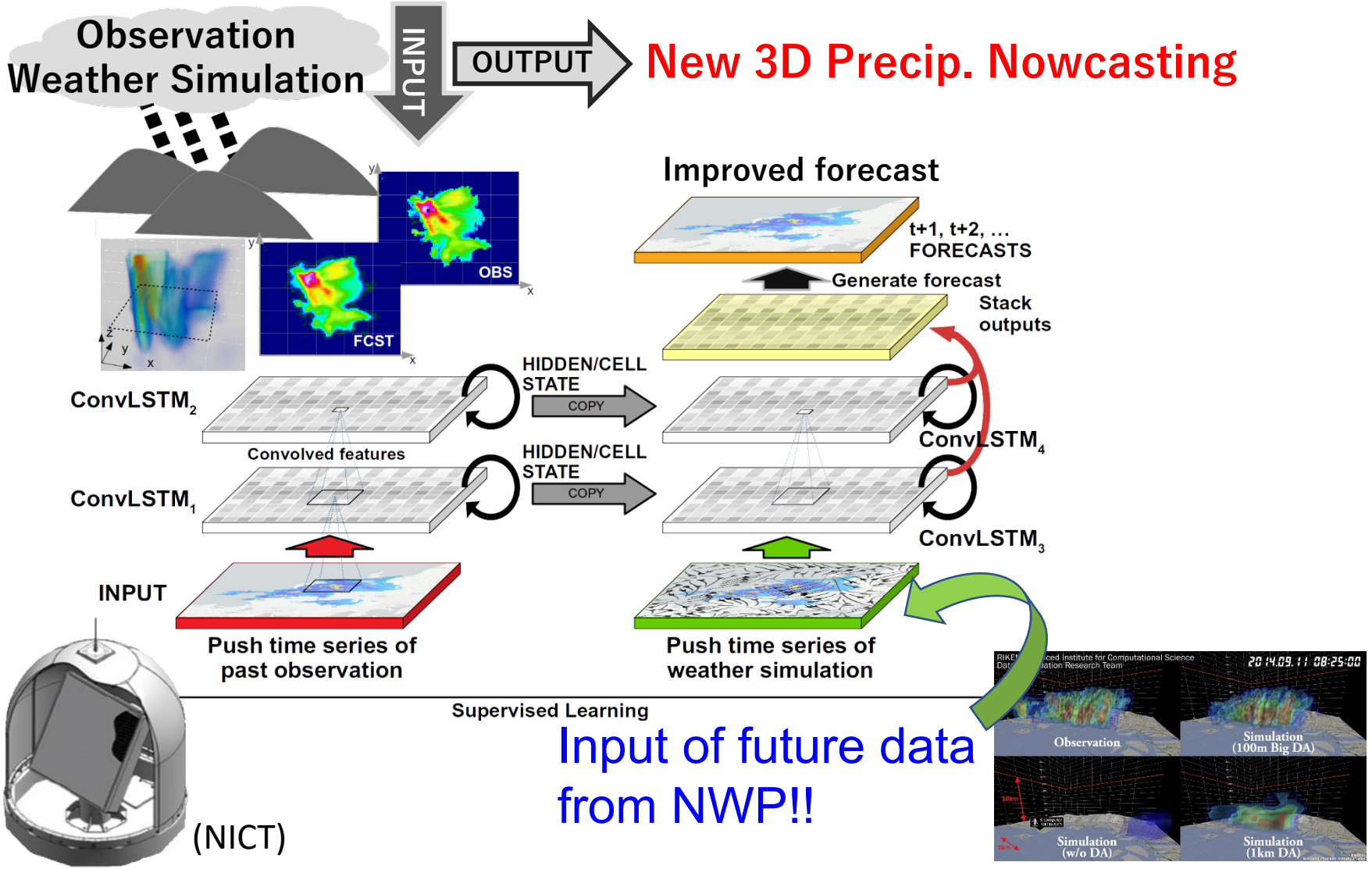
- Good for Big DA
- Not suitable for ML

DA-AI Integration



Otsuka, Miyoshi, et al.

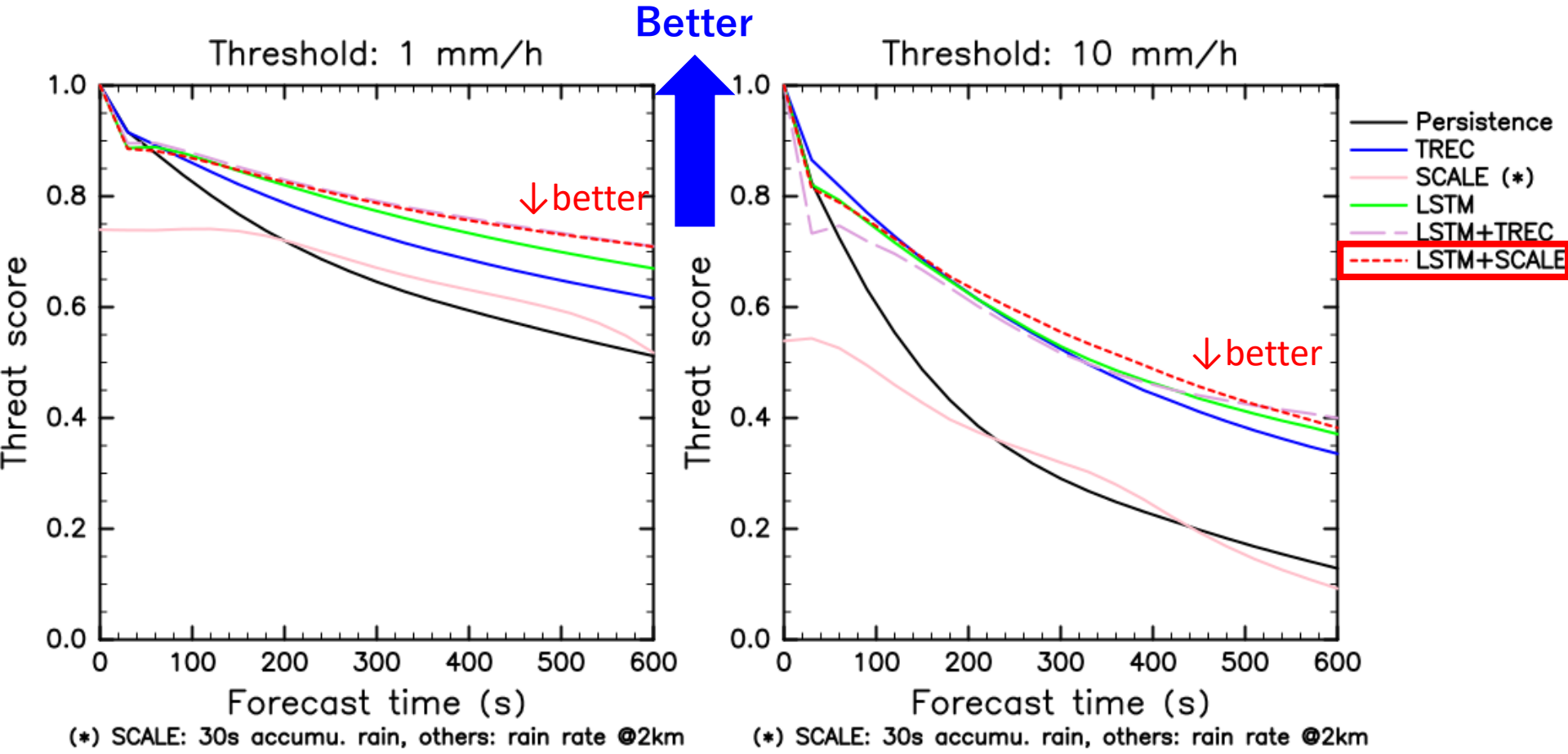
Fusing ML+DA+Simulation



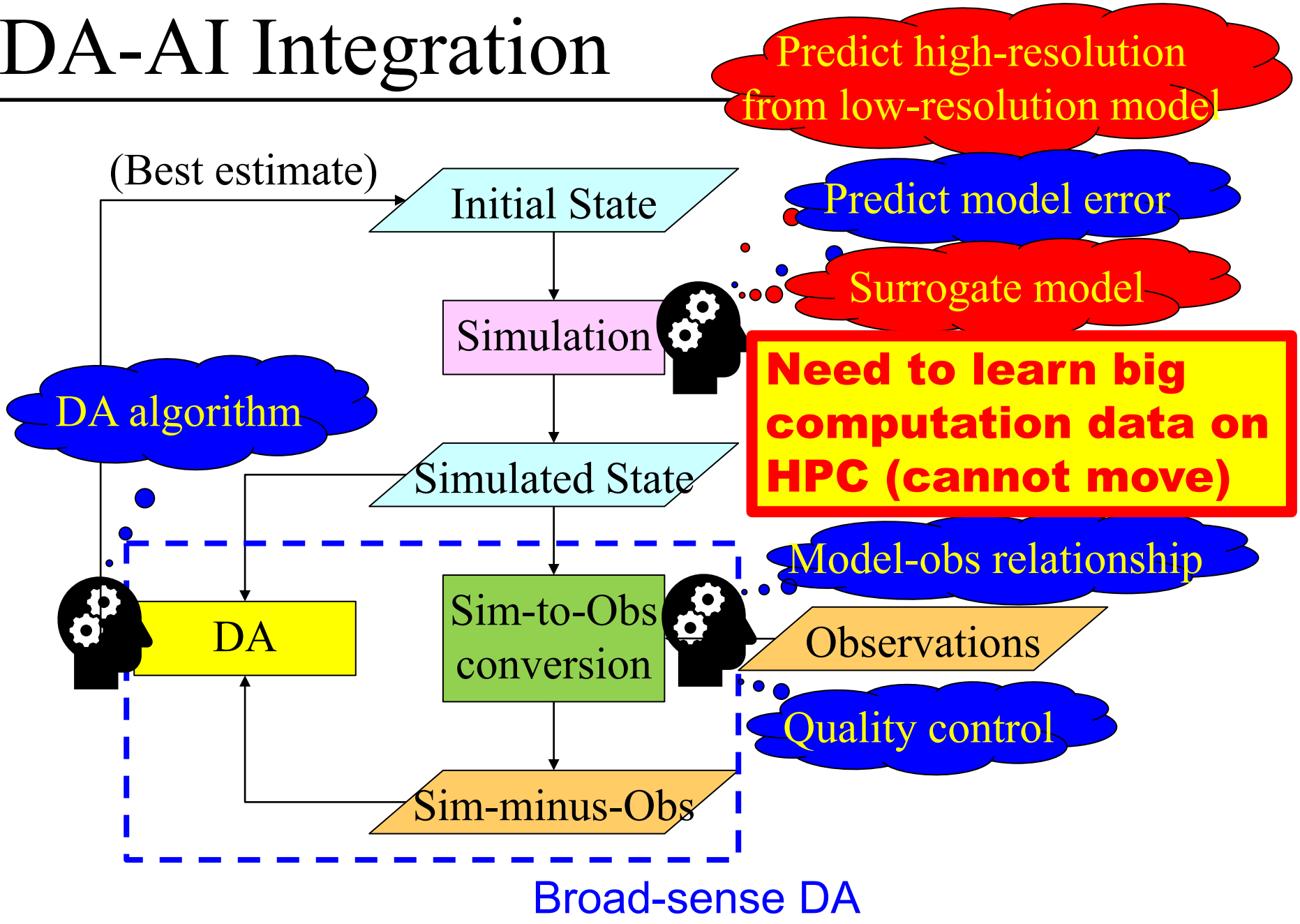
(NICT)

Otsuka, Miyoshi, et al.

Preliminary results:
Using future data in Conv-LSTM is effective.



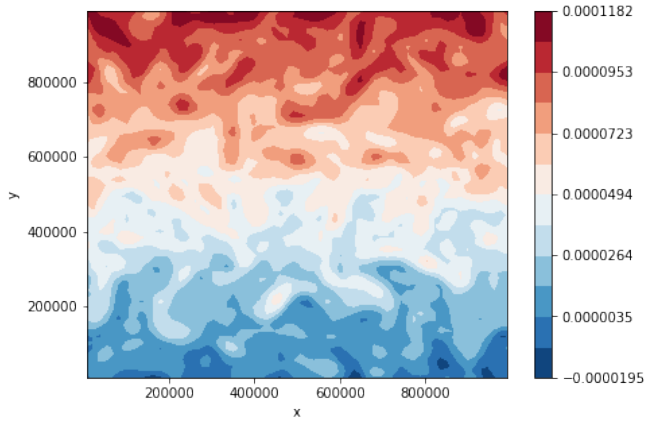
DA-AI Integration

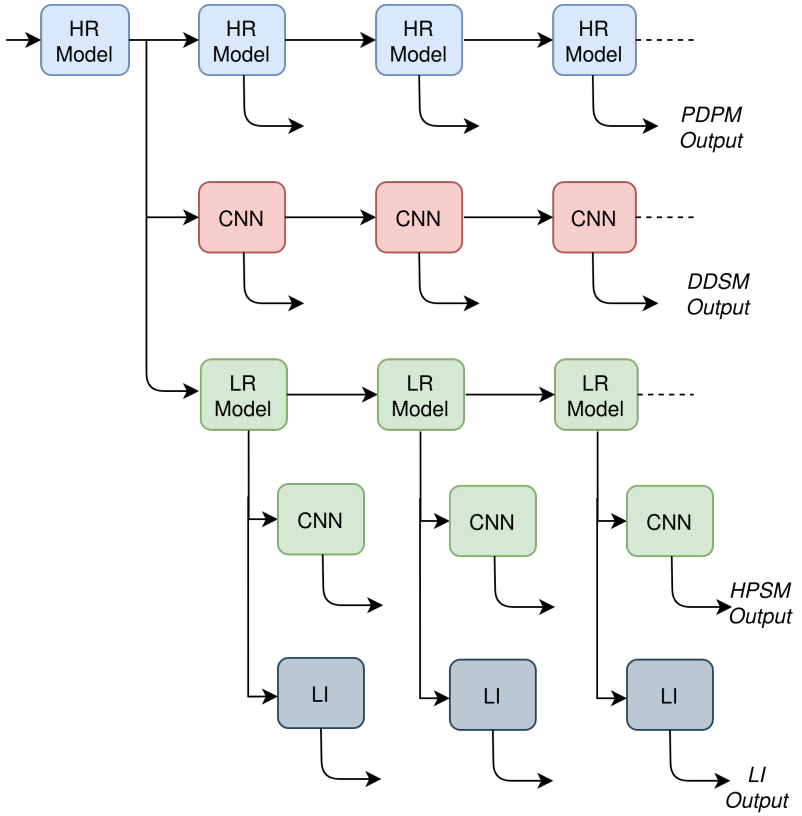


Climate Model Acceleration by Machine Learning

Experiment

- Quasi-Geostrophic model: Potential Vorticity
- Models:
  1. Process-Driven Physical Model *(PDPM)*
  2. Data-Driven Statistical Model *(DDSM)*
  3. Hybrid Physical-Statistical Model *(HPSM)*

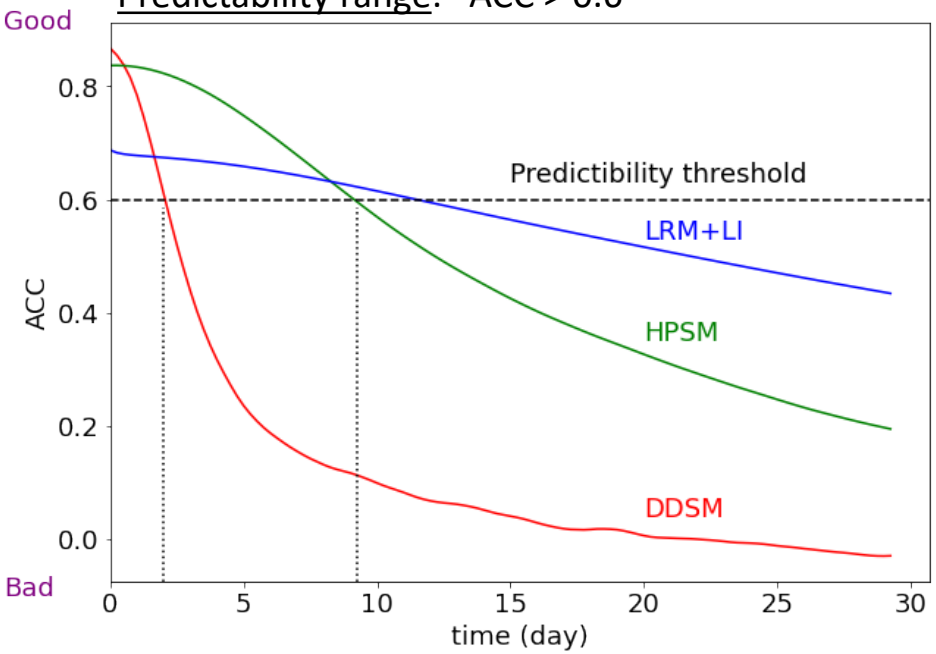




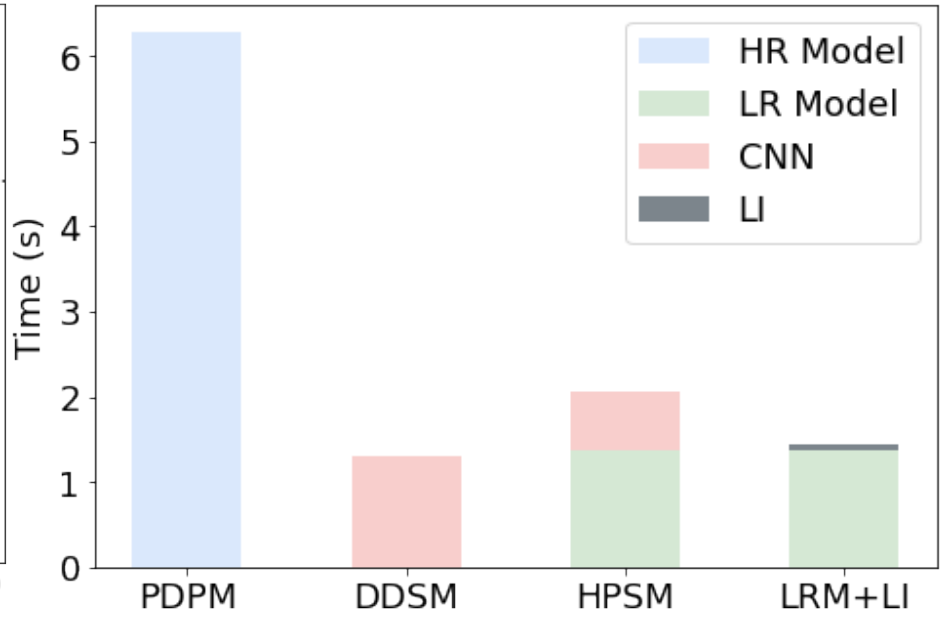
Model Acceleration: Results

Criterion: Anomaly Correlation Coefficient (ACC)

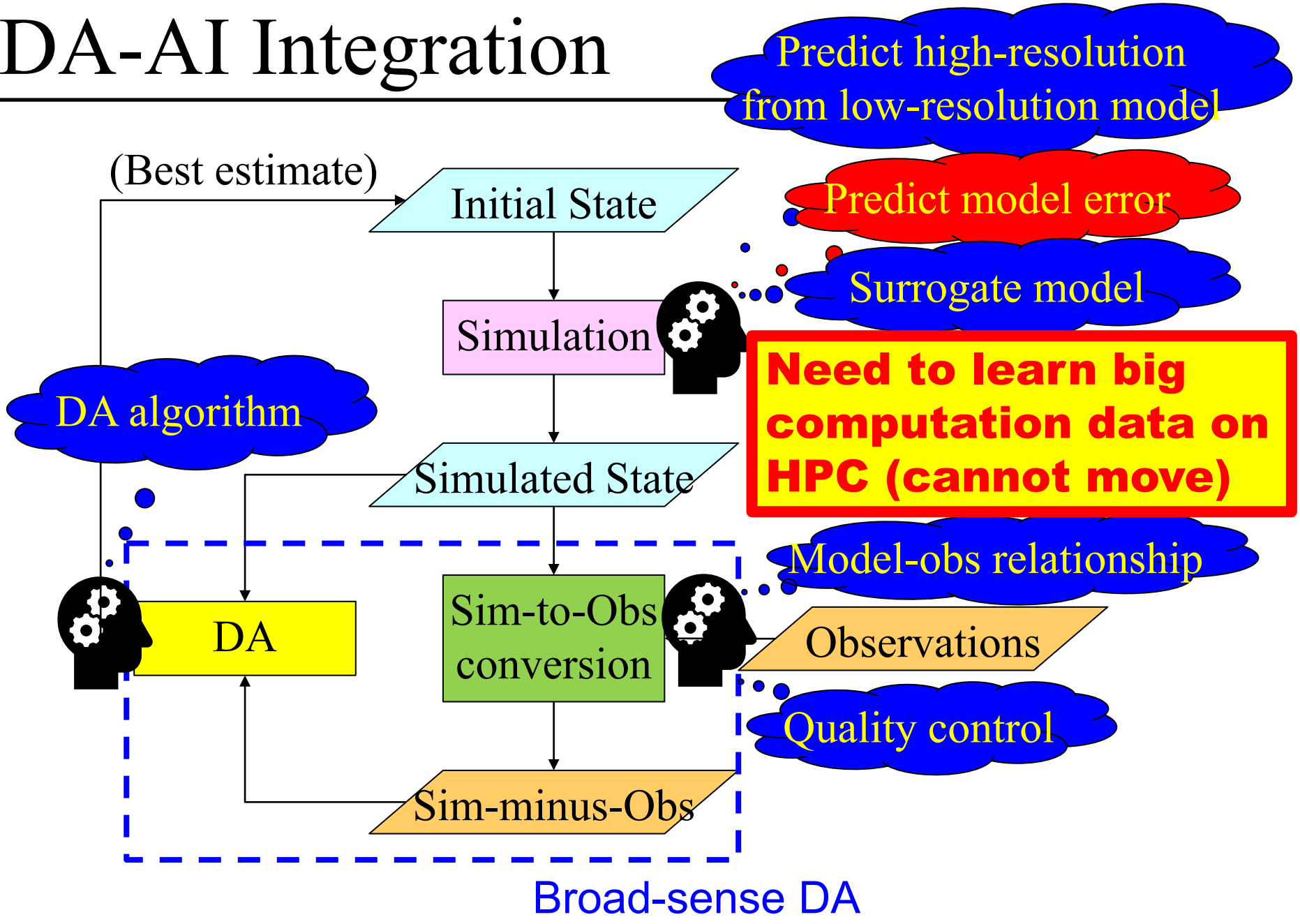
Predictability range: ACC > 0.6



Wall-clock computer time

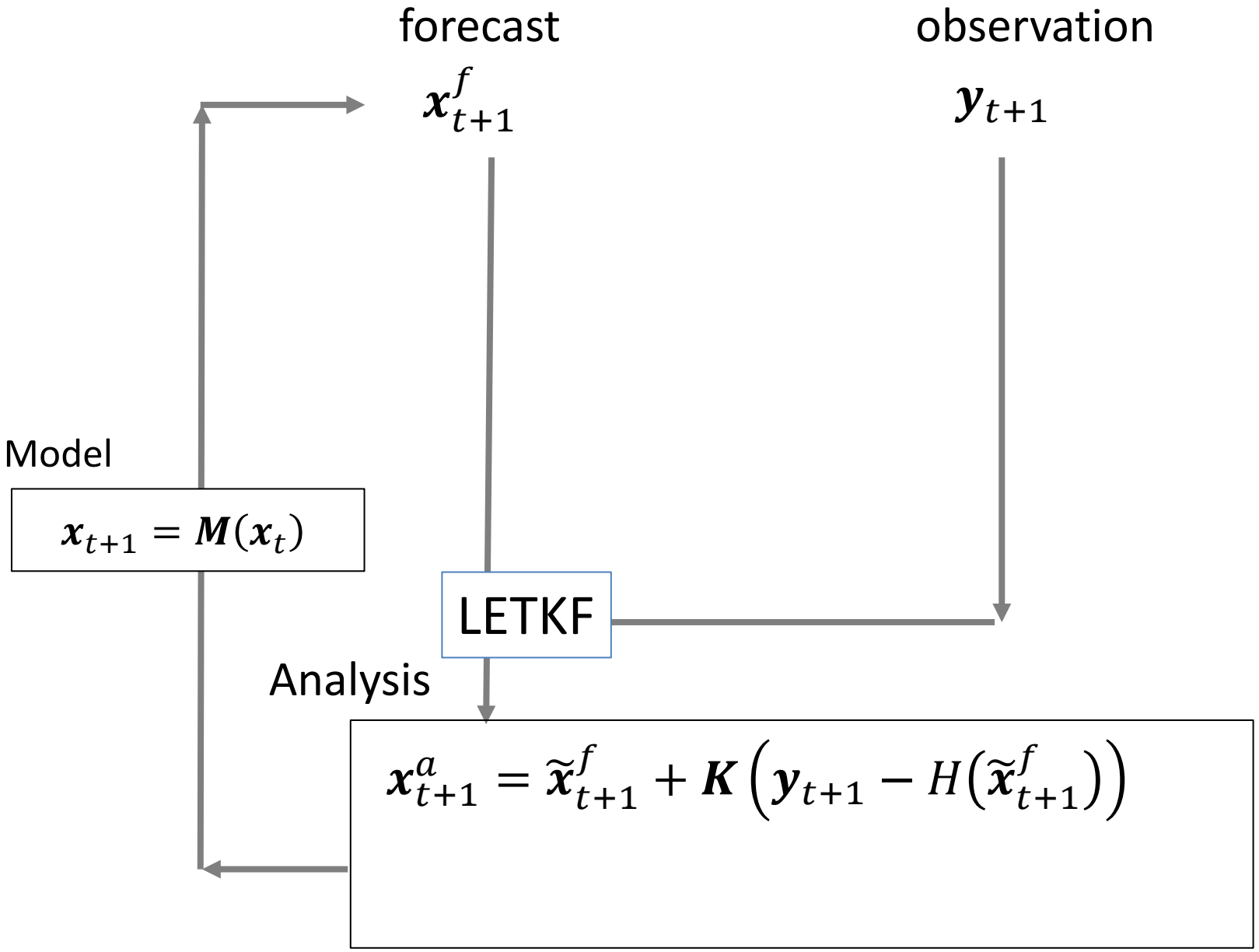


DA-AI Integration



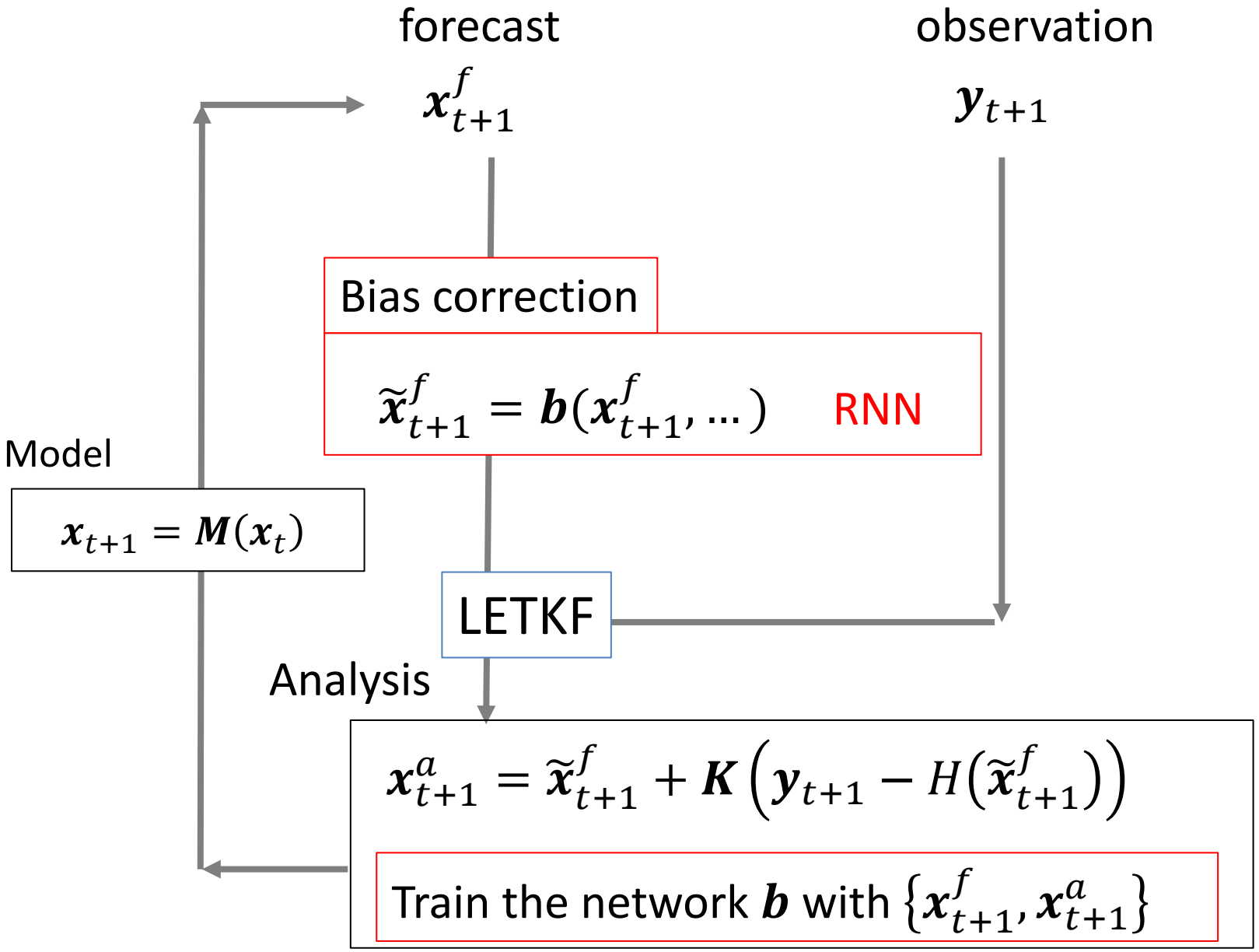
Nonlinear bias correction with ML

Amemiya, Mohta, Miyoshi



Nonlinear bias correction with ML

Amemiya, Mohta, Miyoshi



Idealized experiments: LSTM is effective.

Amemiya, Mohta, Miyoshi

"Nature run": Shear Lorenz96 model (Pulido et al., 2018)

$$\frac{d}{dt}x_k = x_{k-1}(x_{k+1} - x_{k-2}) - x_k + F - \frac{hc}{b} f_k(\boldsymbol{y})$$

$$\frac{d}{dt}y_j = cby_{j+1}(y_{j-1} - y_{j+2}) - cy_j + \frac{hc}{b} g_j(\boldsymbol{x})$$

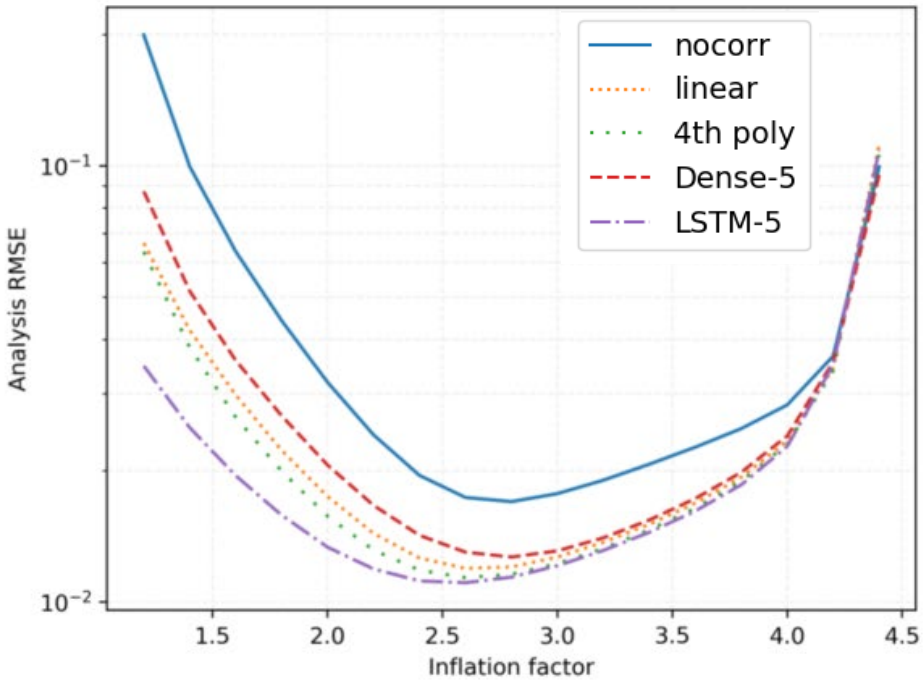
$$f_k(\boldsymbol{y}) = \sum_{j=(k-1)J/K+1}^{kJ/K} y_j$$

$$g_j(\boldsymbol{x}) = \alpha\left(x_{\mathrm{int}(j/K)+1} - x_{\mathrm{int}(j/K)-1}\right)$$

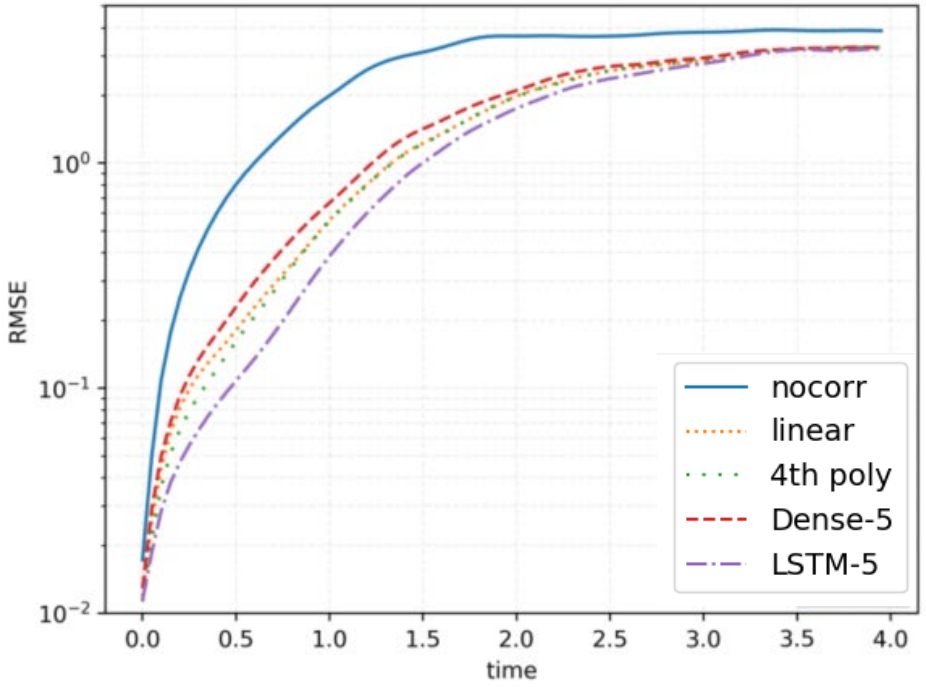
Forecast model

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

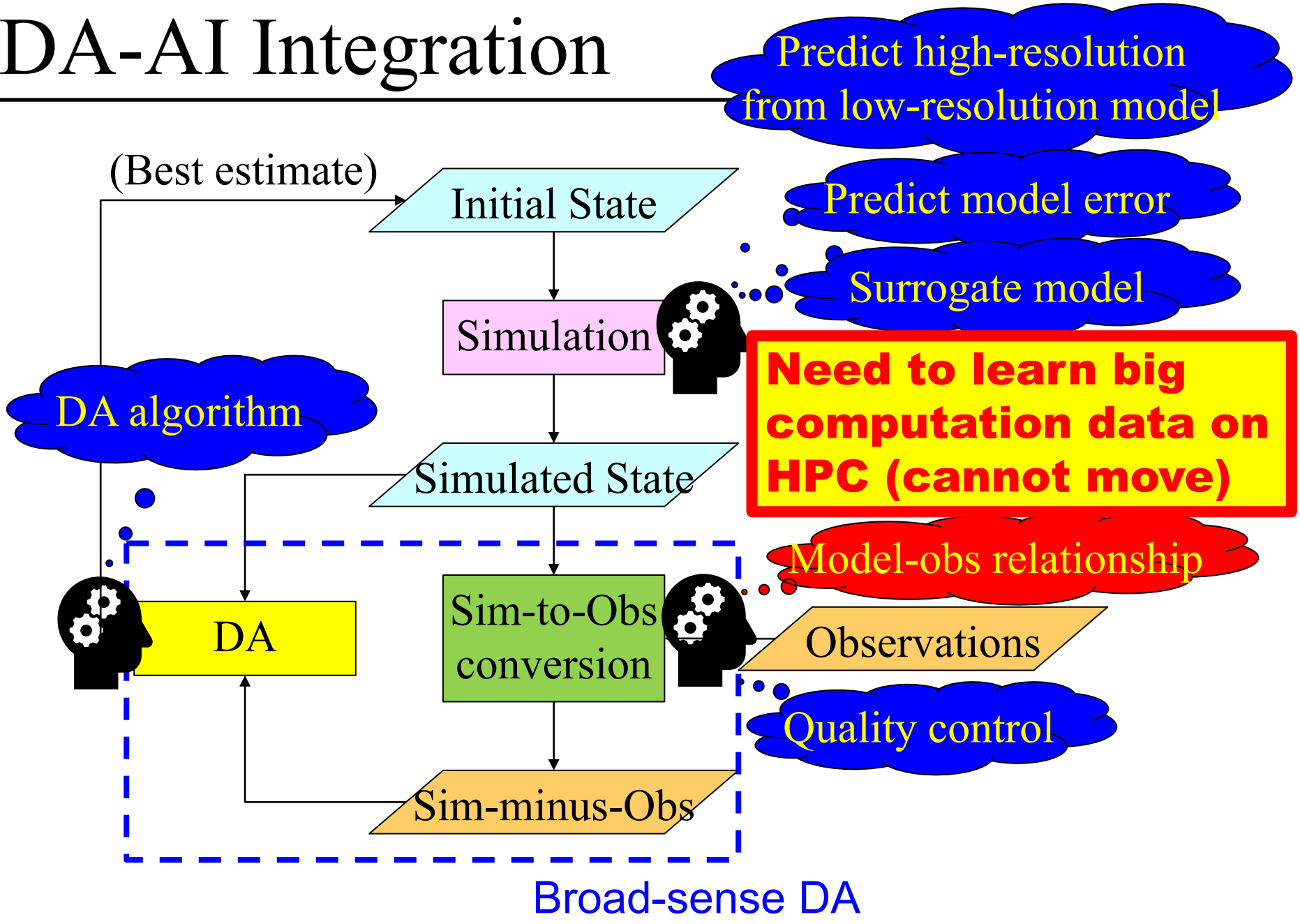
Analysis RMSE



Extended forecast RMSE



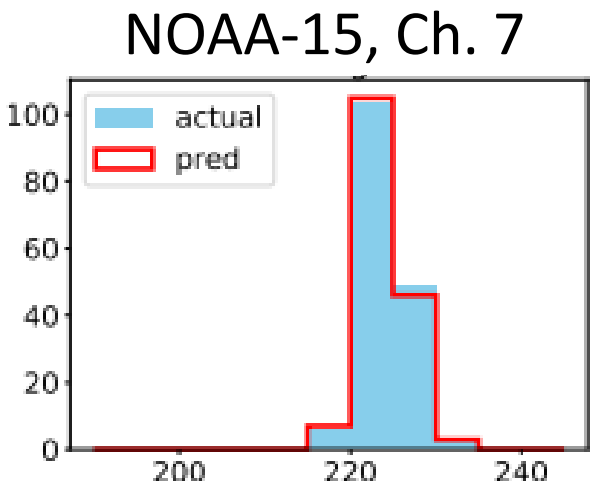
DA-AI Integration

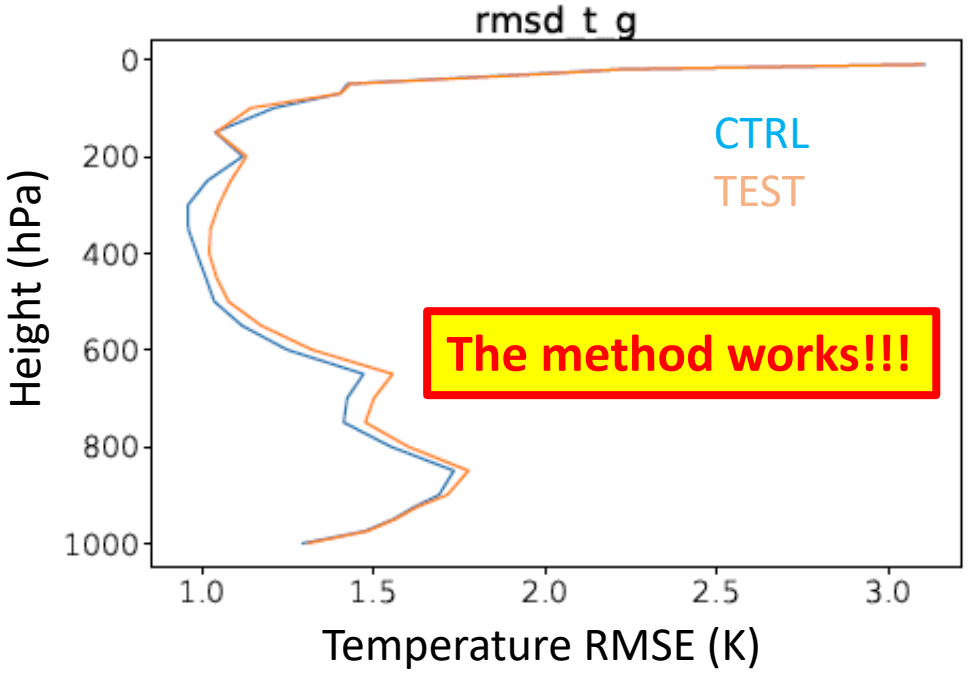


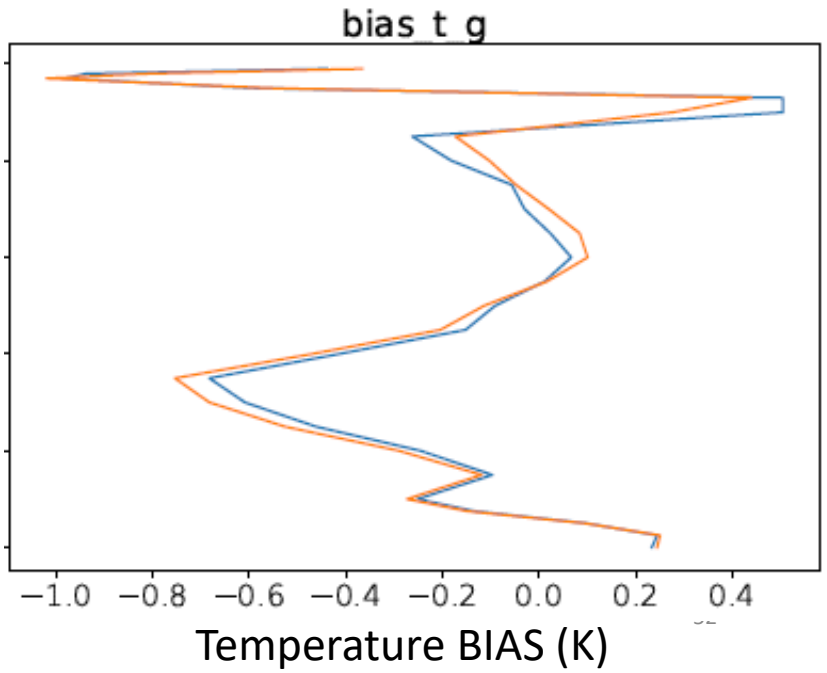
Satellite simulator with ML

Model Variables → Radiative transfer model → Satellite observations

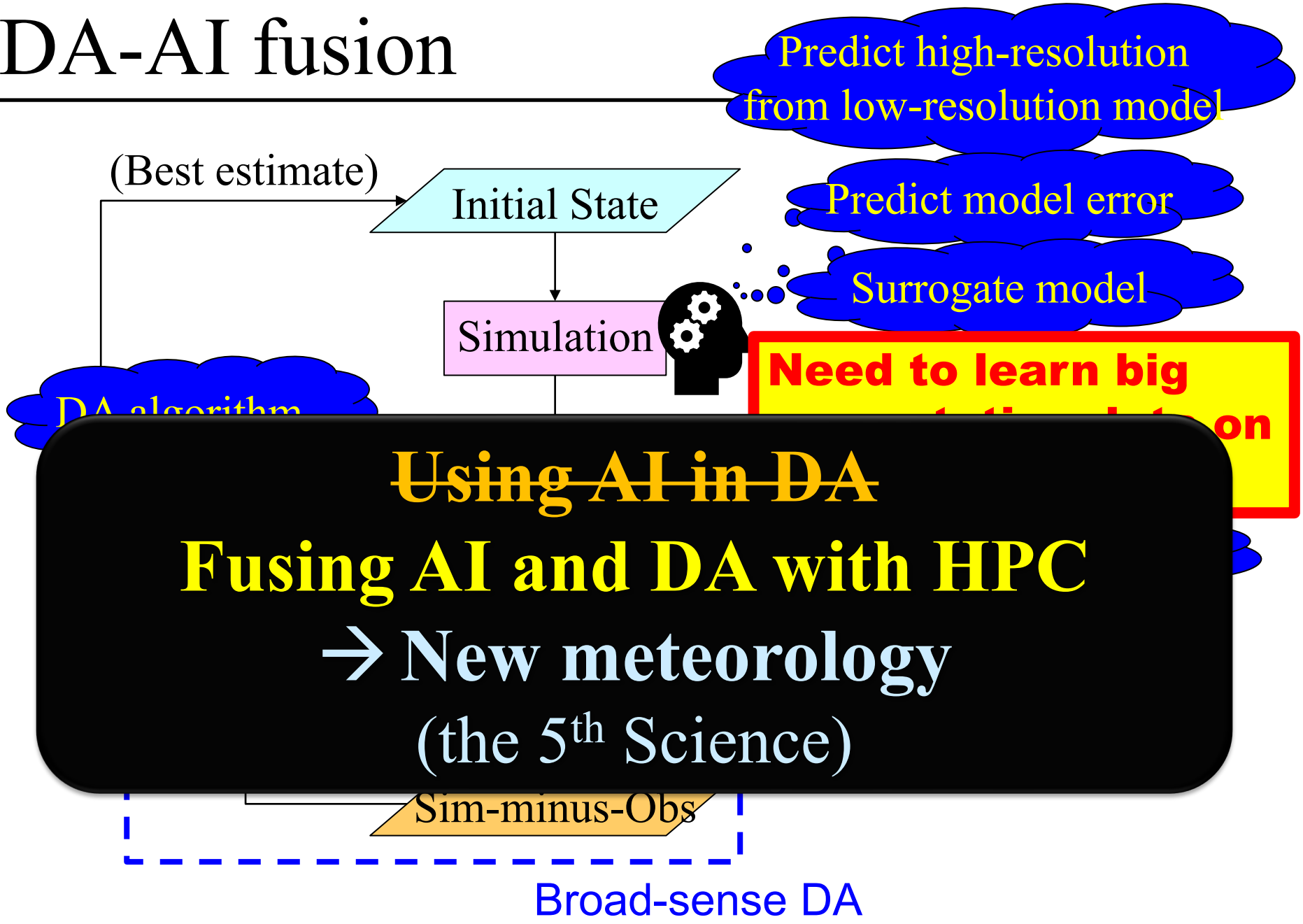
Model Variables → NN → Satellite observations







DA-AI fusion


Predict high-resolution from low-resolution model
(Best estimate)
Initial State
Predict model error
Surrogate model
Simulation
Need to learn big
on
Using AI in DA
Fusing AI and DA with HPC
→ New meteorology
(the 5th Science)
Sim-minus-Obs
Broad-sense DA