March 9, 2017, BDEC, Wuxi, China

"Big Data Assimilation"

for Extreme-scale Numerical Weather Prediction

Takemasa Miyoshi*

RIKEN Advanced Institute for Computational Science

*PI and presenting, Takemasa.Miyoshi@riken.jp

G.-Y. Lien, M. Kunii, J. Ruiz, Y. Maejima, S. Otsuka, K. Kondo, H. Seko, S. Satoh, T. Ushio, K. Bessho, K. Kamide, H. Tomita, S. Nishizawa, T. Yamaura, Y. Ishikawa

With many thanks to

JMA

UMD Weather-Chaos group

JST CREST "Big Data Assimilation" project

JAXA PMM "Ensemble Data Assimilation" project

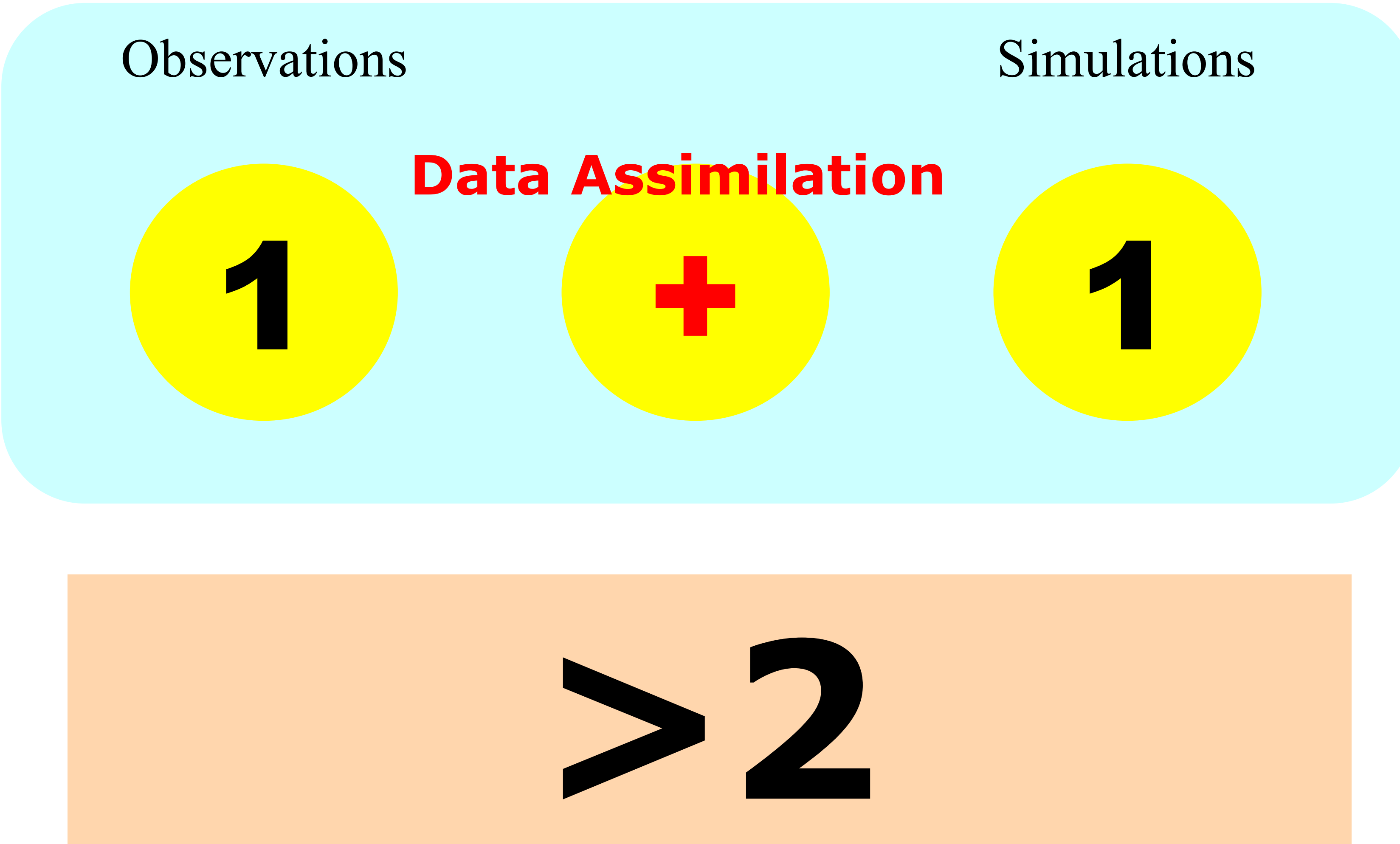
Japan's FLAGSHIP 2020 project

RIKEN Data Assimilation Research Team

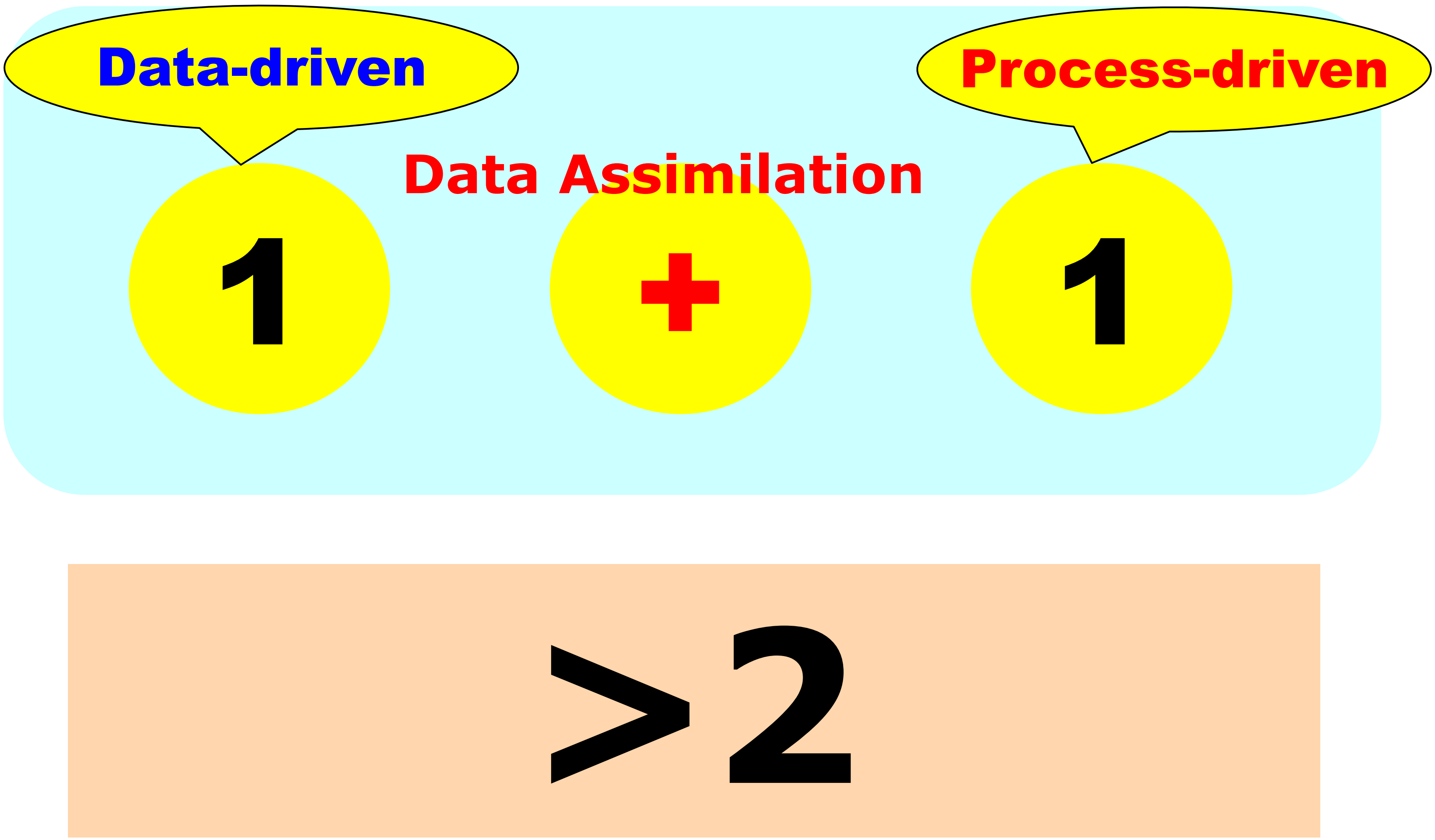
Data Assimilation (DA)

Data assimilation best combines observations and a model, and brings synergy.

Data Assimilation (DA)



Data Assimilation (DA)



“Big Data Assimilation” Revolutionizing Severe Weather Prediction

by

Takemasa Miyoshi, M. Kunii, J. Ruiz, G.-Y. Lien, S. Satoh, T. Ushio, K. Bessho, H. Seko, H. Tomita, and Y. Ishikawa

Bulletin of the American Meteorological Society

August 2016

doi:10.1175/BAMS-D-15-00144.1

国立研究開発法人
科学技術振興機構
Japan Science and Technology Agency
CREST

"Big Data Assimilation" Revolutionizing Severe Weather Prediction

BY Takemasa Miyoshi, Masaru Kunii, Juan Ruiz, Guo-Yuan Lien, Shinsuke Satoh, Tomoo Ushio, Kotaro Bessho, Hiromu Seko, Hirofumi Tomita, and Yutaka Ishikawa

Data assimilation (DA) integrates computer simulations and real-world observations based on statistical mathematics and dynamical systems theory, and plays a central role in numerical weather prediction (NWP). As computing and sensing technologies advance, DA will deal with "big simulations" and "big data." Here we focus on rapidly changing convective weather and explore a future direction of two orders of magnitude more rapid weather forecasting by innovating what we call "big data assimilation" (BDA) technology. Tremendous efforts have been devoted to convective-scale NWP and radar DA, including the U.S. effort on the "Warn-on-Forecast" project (Stensrud et al. 2009; 2013), which has been pioneering rapidly updated NWP to be used for warnings about convective-scale hazards. Sun et al. (2014) provided a comprehensive review on this subject with a rich body of literature. Extending a wealth of previous studies, this article presents the concept of BDA research and the first proof-of-concept results of a real high-impact weather case, exploring 30-min forecasts at 100-m grid spacing refreshed every 30 s—120 times more rapidly than hourly updated systems. This revolutionary NWP is only possible by taking advantage of the fortunate combination of Japan's most advanced technological developments: the 10-petaflops (floating-point operations per second) "K computer" and Phased Array Weather Radar (PAWR; Ushio et al. 2014; Yoshikawa et al. 2013). The science and analytics of big data, typically characterized by four "big V's" (volume, variety, velocity, and veracity), are growing rapidly, and BDA is one of the first two projects awarded by the Japanese government strategic funding program started in 2013 on general big data applications.[1]

In contemporary weather forecasting, radar observations and NWP play an essential role in real-time monitoring and short-term prediction of severe weather. The widely used parabolic-antenna radar observes rain intensity along a curvilinear beam track. The radar is rotated, and changes the azimuth and elevation angles to capture the whole sky typically in 5 min for 15 elevation angles. Also, typical convective-scale NWP updates forecasts every hour for the next $O(10)$ hours at $O(1)$-km grid spacing. However, convective weather systems evolve quickly in 5 min and undertake a nonlinear evolution. The current NWP systems that could possibly use all 5-min radar data at the highest frequency may still be far from sufficient to precisely represent individual convective activities.

Here we explore what the highest-end, next-generation supercomputing and sensing technologies can do at their full capacity, pioneering the future of weather forecasting for the next 10 years. The cutting-edge PAWR implemented in Osaka, Japan, in

[1] The other project is on pharmaceutical science, focusing on drug discovery and production.

AFFILIATIONS: Miyoshi—RIKEN Advanced Institute for Computational Science, Kobe, Japan, Department of Atmospheric and Oceanic Science, University of Maryland, College Park, Maryland, and Application Laboratory, Japan Agency for Marine-Earth Science and Technology, Yokohama, Japan; Kunii—RIKEN Advanced Institute for Computational Science, Kobe, Japan, and Meteorological Research Institute, Tsukuba, Japan; Ruiz—RIKEN Advanced Institute for Computational Science, Kobe, Japan, and CIMA, CONICET-University of Buenos Aires, Buenos Aires, Argentina; Lien, Tomita, and Ishikawa—RIKEN Advanced Institute for Computational Science, Kobe, Japan; Satoh—National Institute of Information and Communications Technology, Koganei, Japan; Ushio—Osaka University, Suita, Japan; Bessho—Meteorological Satellite Center, Kiyose, Japan; Seko—Meteorological Research Institute, Tsukuba, Japan

CORRESPONDING AUTHOR: Takemasa Miyoshi, RIKEN Advanced Institute for Computational Science, 7-1-26 Minatojima-minami-machi, Chuo-ku, Kobe 650-0047, Japan
E-mail: takemasa.miyoshi@riken.jp

The abstract for this article can be found in this issue, following the table of contents.
DOI:10.1175/BAMS-D-15-00144.1

A supplement to this article is available online (10.1175/BAMS-D-15-00144.2)

August 9, 2016

K computer and high-tech weather radar come together to predict sudden torrential rains

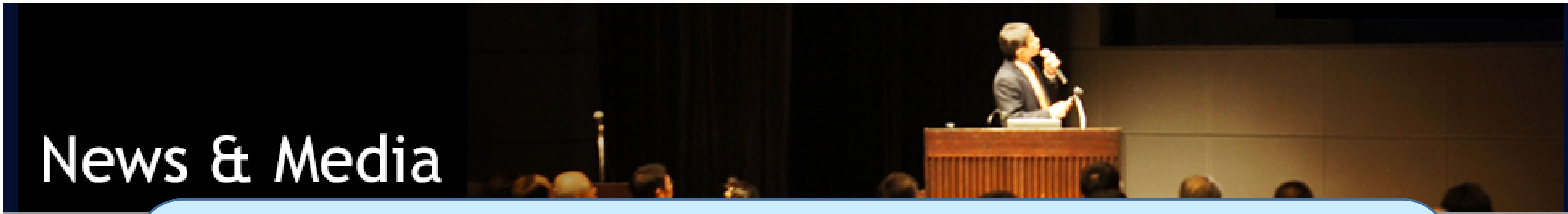
Today, supercomputer-based weather predictions are typically done with simulations that use grids spaced at least one kilometer apart, and incorporate new observational data every hour. However, due to the roughness of the calculations, these simulations cannot accurately predict the threat of torrential rains, which can develop within minutes when cumulonimbus clouds suddenly develop. Now, an international team led by Takemasa Miyoshi of the RIKEN Advanced Center for Computational Science (AICS) has used the powerful K computer and advanced radar observational data to accurately predict the occurrence of torrential rains in localized areas.

The key to the current work, to be published later this month in the August issue of the *Bulletin of the American Meteorological Society*, is “big data assimilation” using computational power to synchronize data between large-scale computer simulations and observational data.

Using the K computer, the researchers carried out 100 parallel simulations of a convective weather system, using the nonhydrostatic mesoscale model used by the Japan Meteorological Agency, but with 100-meter grid spacing rather than the typical 2-kilometer or 5-kilometer spacing, and assimilated data from a next-generation phased array

News & Media
Press Center
Press Releases
2016
2015
2014
2013
2012
2011
2010
2009
2008
2007
2006
2005
News
Events & Symposiums

RIKEN   About RIKEN | Research | News & Media | Careers | Outreach | Community



Home > News

Press Rele

August 9, 20

K comp
sudden

Today, supe
least one ki
roughness
which can
led by Take
powerful K
torrential r

The key to t
Meteorologic
computer simulations and observational data.

Using the K computer, the researchers carried out 100 parallel simulations of a convective weather system, using the nonhydrostatic mesoscale model used by the Japan Meteorological Agency, but with 100-meter grid spacing rather than the typical 2-kilometer or 5-kilometer spacing, and assimilated data from a next-generation phased array

Press conference on Monday, August 1, 11am.

Press release on Tuesday, August 9, 7am.

Covered by

- Newspapers (Asahi, Yomiuri, Nikkei, Nikkan-kogyo, Nikkei-sangyo, Kobe)
- TV broadcast (NHK, FNN)
- Web sources (HPC wire, mynavi, engadget, PC Watch, etc.)

2005

News

Events & Symposiums

日経産業新聞
2016年8月10日
朝刊8面

ゲリラ豪雨、正確に予測

理研など「京」と最新レーダーで

理化学研究所の三好建正チームリーダーらは、情報通信研究機構（NICT）、大阪大学などと共同で、スーパーコンピューター「京」と最新の気象レーダーのデータを組み合わせ、ゲリラ豪雨の発生を正確に予測する手法を開発した。ゲリラ豪雨の兆候をいち早くつかみ、被害の軽減につなげる狙いだ。

スパコンによる天気予報のシミュレーションは通常、1キロメートルより粗い解像度で、1時間ごとに観測データを取り込んで更新している。この空間・時間精度では、わずか数分間で積乱雲が急速に成長するゲリラ豪雨の発生を予測するのは難しかった。

開発した手法では理研の京とNICTや阪大が開発した気象レーダーのデータを利用し、100メートルの解像度で30秒ごとに更新する。シミュレーションと実測データを組み合わせる「データ同化」と呼ぶ手法を応用した。膨大なデータを素早く取り込んで計算することで、ゲリラ豪雨の予測が可能になるという。

ただ、現状ではデータ同化の作業に時間がかかり、すぐには実際の天気予報などに活用できない。実用化に向け、データの転送や計算の高速化などの課題克服に取り組む。

朝日新聞2016年8月9日夕刊2面

スパコン「京」ゲリラ豪雨再現

数分後まで成功 予測に光

短時間で急激に発達するゲリラ豪雨の雨雲の動きをスーパーコンピューター「京」で予測するシステムを、理化学研究所計算科学研究機構などのグループが開発した。過去のゲリラ豪雨を一部再現することに成功した段階だが、将来は、雨雲のもととなる水蒸気の塊が発生した時点で30分先の予測が可能になるかもしれない。

30秒ごとにゲリラ豪雨の分布を詳細にとらえる性能を持つ大阪大などの最新気象レーダーの観測データを京に取り込み、シミュレーションと組み合わせる手法。2013年に京都市を襲ったゲリラ豪雨の降り始め時点のデータをもとに予測したところ、数分後まではほぼ正確に再現できた。10分後以降は誤差が大きくなった。

将来は、観測データをリアルタイムで取り込み、「京」ではない普及型のスパコンで30分後までの雨雲を正確に予測することを目指す。実現には精度の向上に加え、現状では10分かかっている計算速度の大幅な短縮が必要となる。理研の三好建正チームリーダーは「10年程度で運用可能な段階にしたい」と話している。（野中良祐）

朝日新聞（大阪版）
2016年8月9日夕刊7面

「京」、ゲリラ豪雨予測へ一歩

理研などシステム開発

短時間で急激に発達するゲリラ豪雨の雨雲の動きをスーパーコンピューター「京」で予測するシステムを、理化学研究所計算科学研究機構などのグループが開発した。まだ過去のゲリラ豪雨を一部再現することに成功した段階だが、将来は、雨雲のもととなる水蒸気の塊が発生した時点で30分先の予測が可能になるかもしれない。

30秒ごとにゲリラ豪雨の分布を詳細にとらえる性能を持つ大阪大などの最新気象レーダーの観測データを京に取り込み、シミュレーションと組み合わせる手法。2013年に京都市を襲ったゲリラ豪雨の降り始め時点のデータをもとに、30分後までの変化を予測したところ、数分後まではほぼ正確に再現できた。10分後以降は誤差が大きくなった。

将来は、観測データをリアルタイムで取り込み、「京」ではない普及型のスパコンで30分後までの雨雲を正確に予測することを目指す。実現には精度の向上に加え、現状では10分かかっている計算速度の大幅な短縮が必要となる。理研の三好建正チームリーダーは「10年程度で運用可能な段階にしたい」と話している。（野中良祐）

神戸新聞2016年8月9日夕刊9面

ゲリラ豪雨30分前に予測

積乱雲の成長を再現

理研、スパコン「京」で

理化学研究所（理研）などは9日、30分後のゲリラ豪雨が予報可能となるシミュレーション手法を開発した、と発表した。最新鋭レーダーで観測された気象データを基に、雨雲の成長を神戸・ポートアイランドのスーパーコンピューター「京」で予測。研究グループは「ゲリラ豪雨が予測可能であることが示された。実用化に向けた大きな一歩だ」とする。（武藤邦生、佐伯竜一）

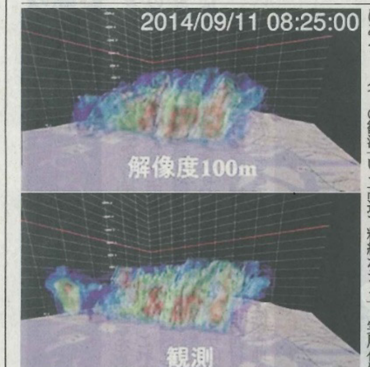
理研計算科学研究機構（神戸市中央区）の三好建正チームリーダーらによる成果で、近く米科学誌に掲載される。

ゲリラ豪雨を引き起こす積乱雲は直径が小さく、わずか10分ほどで急激に成長するため、現在の天気予報のシミュレーションの精度では、詳しい予測が困難だった。

グループは、雨雲の状況を短時間で立体的に観測できる新型レーダーのデータを使用。大阪大（大阪府吹田市）にあるレーダーの観測値と、京の計算値を組み合わせ、現在の天気予報の100倍以上という高速で精細なシミュレーションに成功した。

神戸・阪神間に局地的豪雨をもたらした2014年9月11日朝の神戸市上空の雨雲をシミュレーションしたところ、実際の観測とほぼ一致。雨雲の内部の構造まで再現できた。

実用化には10年程度かかるとみられ、三好リーダーは「計算時間の短縮や予測精度の向上に取り組みたい」としている。



シミュレーションによって再現された2014年9月11日午前8時25分の神戸市上空の雨雲（上）と実際の観測データ（理研計算科学研究機構・三好建正チームリーダー提供）

読売新聞2016年8月10日朝刊2面

局地的大雨をスパコンで再現

理研「将来的に予測」

短時間で急激な大雨が降る「局地的大雨」をより正確に予測するのにつながる手法を開発したと、理化学研究所などのチームが米科学誌に発表した。最新のレーダーで30秒ごとに集めたデータをスーパーコンピューターで計算し、過去に実際に起きた局地的大雨の再現に成功した。チームは「将来的に、局地的大雨の予測を可能にする技術」と見ている。

気象庁のスパコンとレーダーを使う現在の手法は、2キロ・メートル四方ごとの降水量を予測しているが、その計算には1時間ほどかかり、局地的大雨を予測するのが難しかった。チームは最新のレーダーを使い、100メートル四方ごとの雨雲の分布や風速のデータを30秒ごとに収集。理研のスパコン「京」を使い、実際に起きた局地的大雨の大気の状態で試算した結果、雨の状況が正確に再現できた。

日経新聞2016年8月9日夕刊16面

ゲリラ豪雨予測スパコンで正確

理研など手法開発

理化学研究所と情報通信研究機構（NICT）、大阪大などは9日、スーパーコンピューター「京」を使ってゲリラ豪雨の発生を正確に予測する手法を開発したと発表した。積乱雲の発達状況を瞬時に把握できる最新の気象レーダーを活用する。ゲリラ豪雨の兆候をいち早くつかむことができるという。

日刊工業新聞2016年8月10日朝刊25面

ゲリラ豪雨30分先予測

理研

解像度100m 30秒ごと更新

理化学研究所計算科学研究機構データ同化研究チームの三好建正チームリーダーらは、30分後までのゲリラ豪雨を予測する手法を開発した。スーパーコンピューター「京」と、フェーズドアレイ気象レーダーから得られるデータを組み合わせることで、解像度100メートルで30秒ごとに新しい観測データを取り込んで更新する天気予報シミュレーションを実現。実際のゲリラ豪雨の動きを詳細に再現することに成功した。

スーパーコンピューターを使った天気予報シミュレーションは一般的に、1キロメートルより粗い解像度で1時間ごとに新しい観測データを取り込んで更新する。しかし、ゲリラ豪雨の場合、数分の間に積乱雲が急激に発生・発達するため、1時間の更新間隔では予測が困難だった。

また、1キロメートルより粗い解像度では、ゲリラ豪雨を引き起こす積乱雲を十分に解像できなかった。超高速かつ超高精細な天気予報につながる可能性がある。

情報通信研究機構、大阪大学などとの共同研究。成果は、今月下旬に米科学誌ブリティン・オブ・ザ・アメリカン・メテオロジカル・ソサエティーに掲載される。

Proceedings of the IEEE

(Miyoshi et al., 2016)

November 2016

INVITED PAPER

"Big Data Assimilation" Toward Post-Petascale Severe Weather Prediction: An Overview and Progress

This article summarizes the activities and progress of the big data assimilation project for severe weather prediction and concludes with perspectives toward the post-petascale supercomputing era.

By Takemasa Miyoshi, Guo-Yuan Lien, Shinsuke Satoh, Tomoo Ushio, Kotaro Bessho, Hirofumi Tomita, Seiya Nishizawa, Ryuji Yoshida, Sachiho A. Adachi, Jianwei Liao, Balazs Gerofi, Yutaka Ishikawa, Masaru Kunii, Juan Ruiz, Yasumitsu Maejima, Shigenori Otsuka, Michiko Otsuka, Kozo Okamoto, and Hiromu Seko

ABSTRACT | Following the invention of the telegraph, electronic computer, and remote sensing, "big data" is bringing another revolution to weather prediction. As sensor and computer technologies advance, orders of magnitude bigger data are produced by new sensors and high-precision computer simulation or "big simulation." Data assimilation (DA) is a key to numerical weather prediction (NWP) by integrating the real-world sensor data into simulation. However, the current DA and NWP systems are not designed to handle the "big data" from next-generation sensors and big simulation. Therefore, we propose "big data assimilation" (BDA) innovation to fully utilize the big data. Since October 2013, the Japan's BDA project has been exploring revolutionary NWP at 100-m mesh refreshed every 30 s, orders of magnitude finer and faster than the current typical NWP systems, by taking advantage of the fortunate combination of next-generation technologies: the 10-petaflops K computer, phased array weather radar, and geostationary satellite Himawari-8. So far, a BDA prototype system was developed and tested with real-world retrospective local rainstorm cases. This paper summarizes the activities and progress of the BDA project, and concludes with perspectives toward the post-petascale supercomputing era.

KEYWORDS | Atmospheric measurements; computer applications; Kalman filtering; optimal control; phased array radar; remote sensing; simulation; supercomputers; weather forecasting

Manuscript received January 15, 2016; revised April 5, 2016 and August 7, 2016; accepted August 15, 2016. Date of publication September 26, 2016; date of current version October 18, 2016. This study was supported by CREST, Japan Science and Technology Agency (JST). *(Corresponding author: Takemasa Miysohi.)*

T. Miyoshi is with the RIKEN Advanced Institute for Computational Science, Chuo-ku, Kobe 650-0047, Japan, with the Department of Atmospheric and Oceanic Science, University of Maryland, College Park, MD 20740 USA, and also with the Application Laboratory, Japan Agency for Marine-Earth Science and Technology, Yokohama 236-0001, Japan (e-mail: takemasa.miyoshi@riken.jp).

G.-Y. Lien, **H. Tomita**, **S. Nishizawa**, **R. Yoshida**, **S. A. Adachi**, **B. Gerofi**, **Y. Ishikawa**, **Y. Maejima**, and **S. Otsuka** are with the RIKEN Advanced Institute for Computational Science, Chuo-ku, Kobe 650-0047, Japan (e-mail: guo-yuan.lien@riken.jp; htomita@riken.jp; s-nishizawa@riken.jp; ryoshida@riken.jp; sachiho.adachi@riken.jp; bgerofi@riken.jp; yutaka.ishikawa@riken.jp; yasumitsu.maejima@riken.jp; shigenori.otsuka@riken.jp).

S. Satoh is with National Institute of Information and Communications Technology, Koganei 184-8795, Japan (e-mail: satoh@nict.go.jp).

T. Ushio is with the Osaka University, Suita 565-0871, Japan (e-mail: ushio@comm.eng.osaka-u.ac.jp).

K. Bessho was with the Meteorological Satellite Center, Kiyose 204-0012, Japan. He is now with the Forecast Department, Japan Meteorological Agency, Tokyo 100-8122, Japan (e-mail: kbessho@met.kishou.go.jp).

J. Liao was with the RIKEN Advanced Institute for Computational Science, Chuo-ku, Kobe 650-0047, Japan. He is now with the Southwest University of China, Chongqing 400715, China (e-mail: jianwei.liao@riken.jp).

M. Kunii, **M. Otsuka**, and **K. Okamoto** are with the RIKEN Advanced Institute for Computational Science, Chuo-ku, Kobe 650-0047, Japan, and also with the Meteorological Research Institute, Tsukuba 305-0052, Japan (e-mail: mkunii@mri-jma.go.jp; motsuka@mri-jma.go.jp; kokamoto@mri-jma.go.jp).

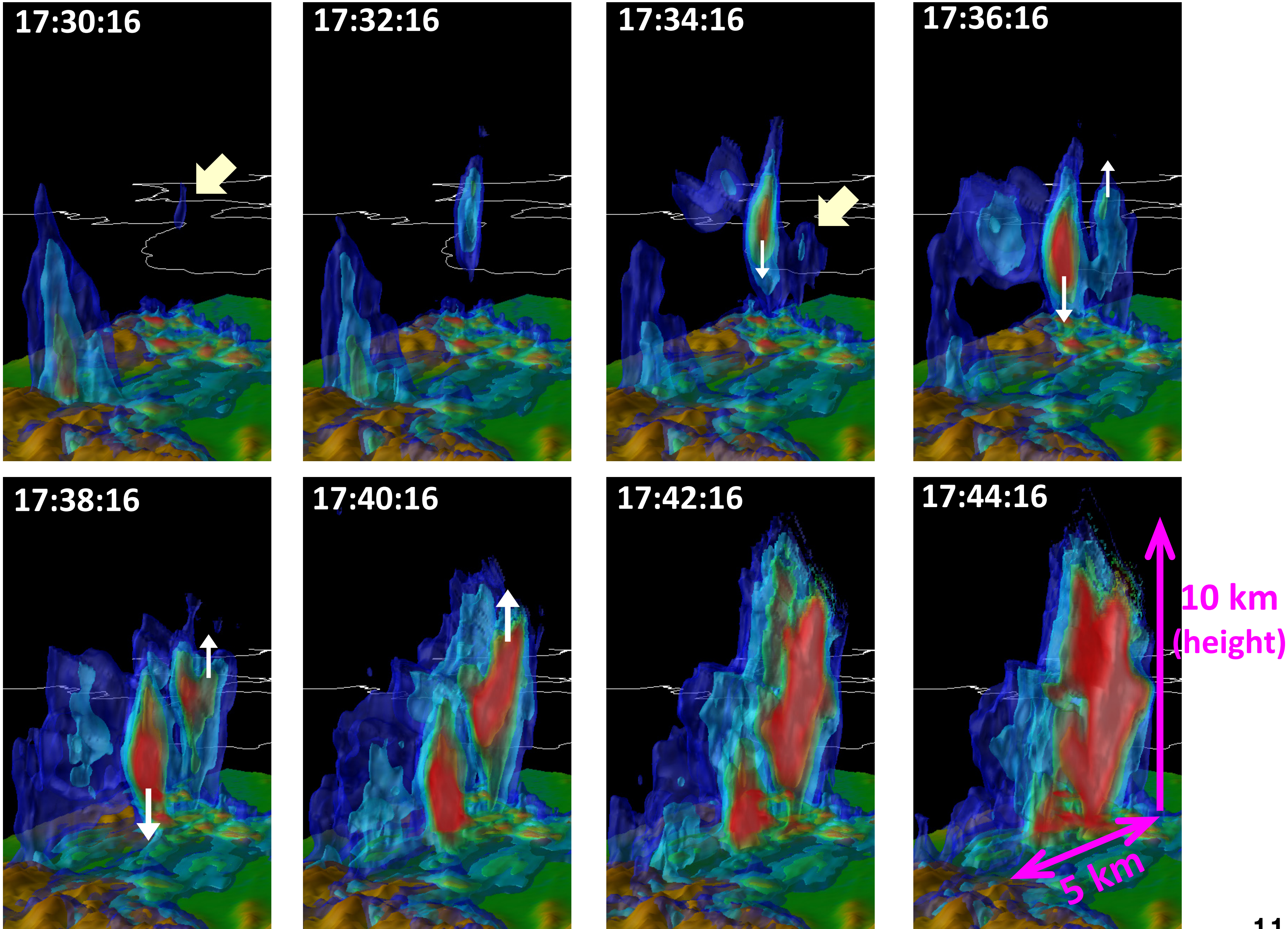
J. Ruiz is with the RIKEN Advanced Institute for Computational Science, Chuo-ku, Kobe 650-0047, Japan, and also with the CIMA, CONICET-University of Buenos Aires, Buenos Aires, Argentina (e-mail: jruiz@cima.fcen.uba.ar).

H. Seko is with the Meteorological Research Institute, Tsukuba, Japan (e-mail: hseko@mri-jma.go.jp).

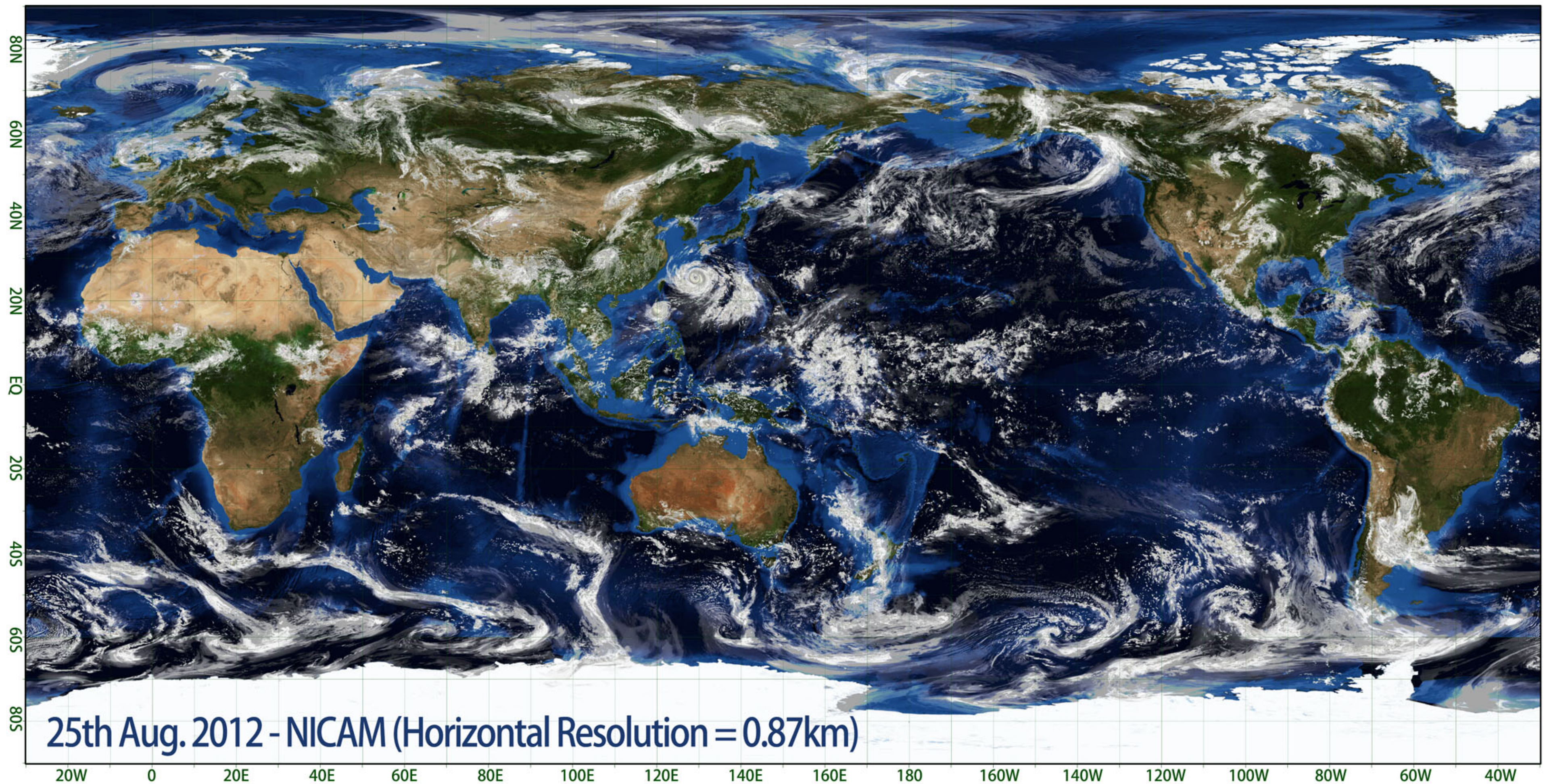
Digital Object Identifier: 10.1109/JPROC.2016.2602560

Only in 10 minutes!!

(Courtesy of NICT)



Global 870-m simulation (*Miyamoto et al. 2013*)



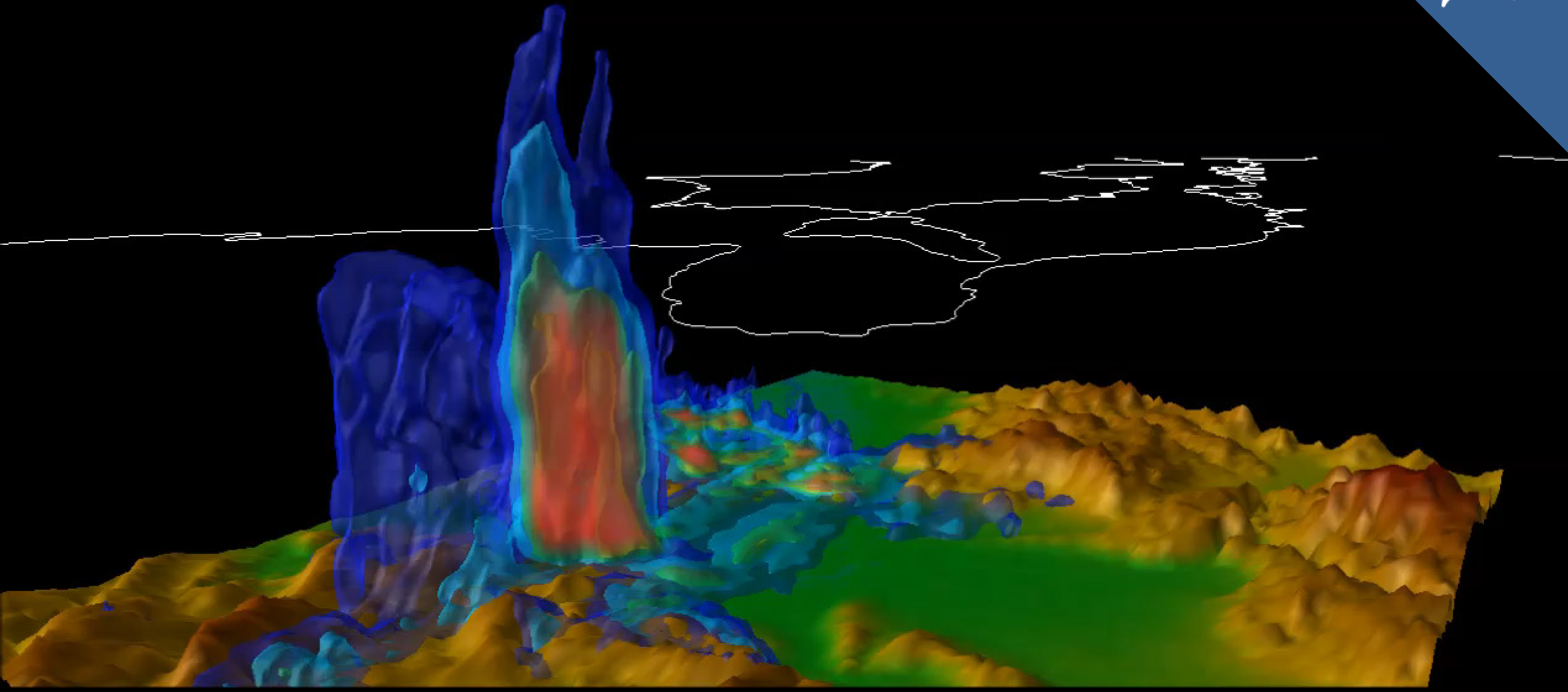
Visualized by Ryuji Yoshida

©JAMSTEC・AORI (SPIRE Field3), RIKEN/AICS
Visualized by Ryuji Yoshida

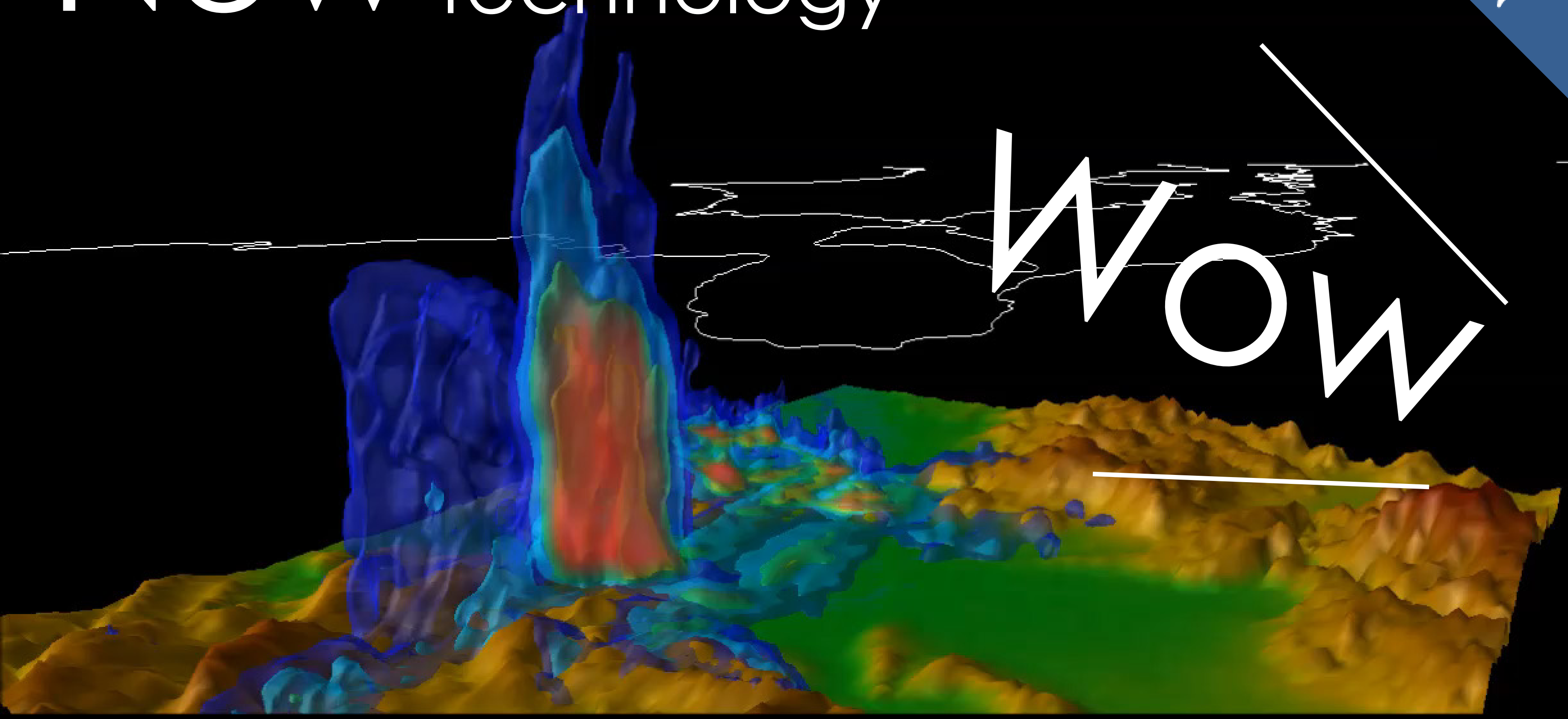
cf. TEDxSannomiya
http://tedxsannomiya.com/speakers/takemasa-miyoshi/

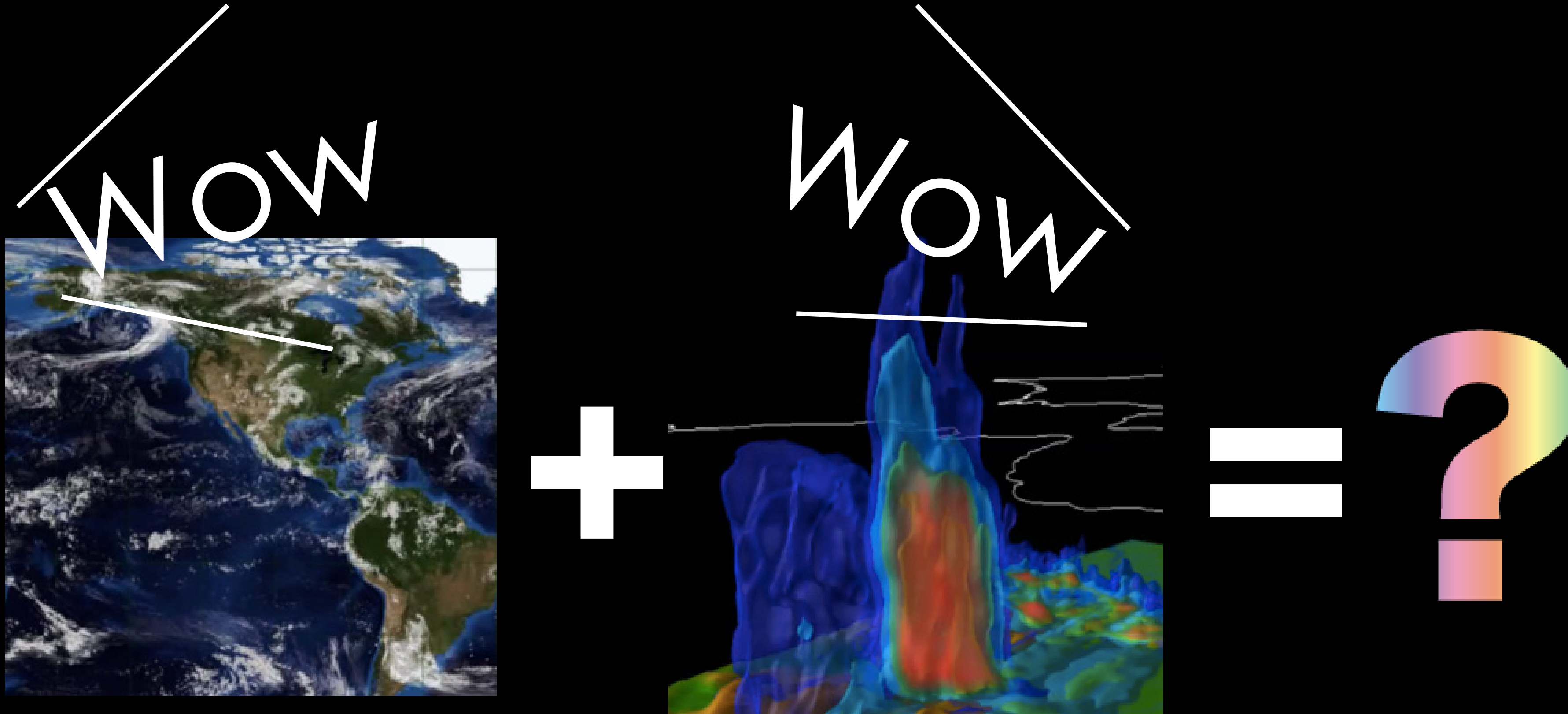

京
K computer
京
K computer
©RIKEN

Raindrops in the air

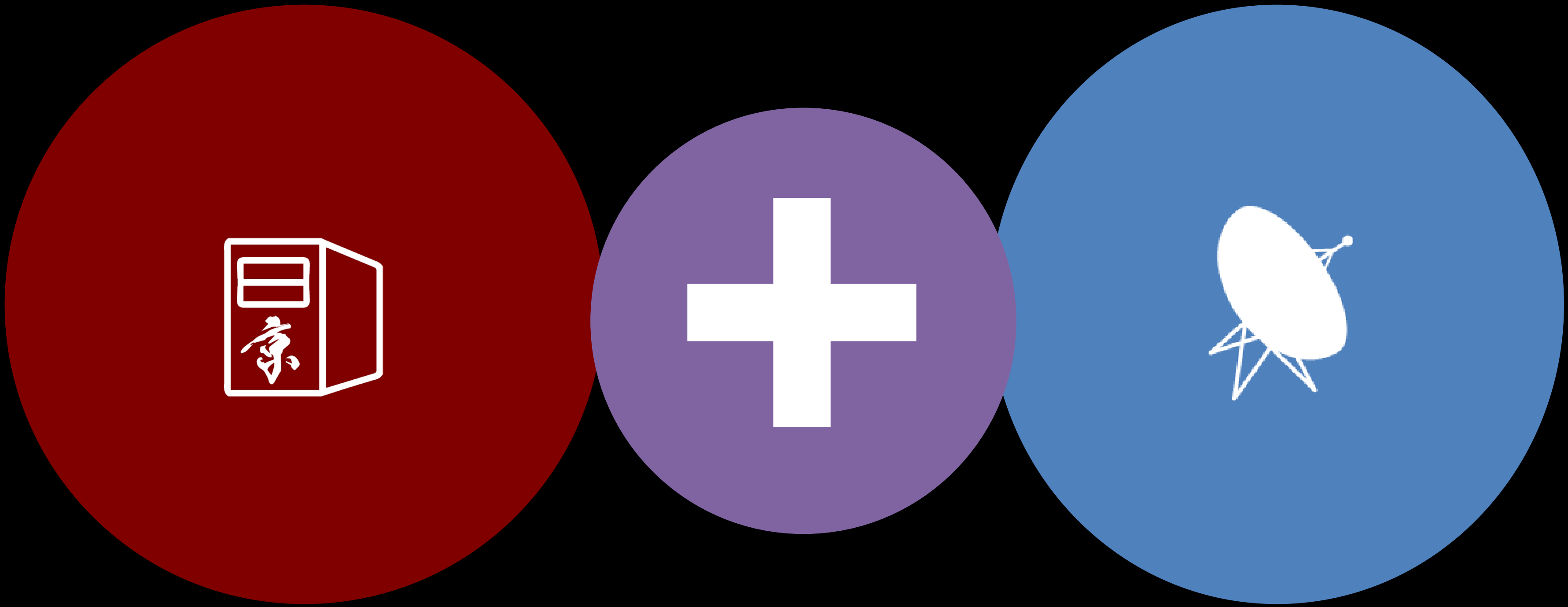


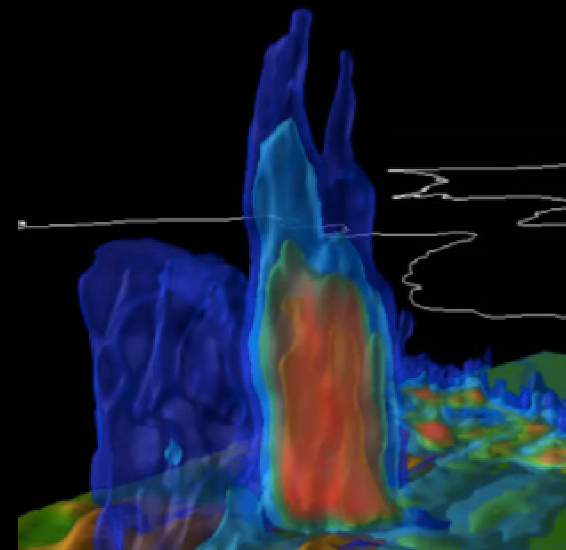
New radar technology




Wow
+
Wow
=
?

Data Assimilation





= ~~Sudden~~ heavy rain

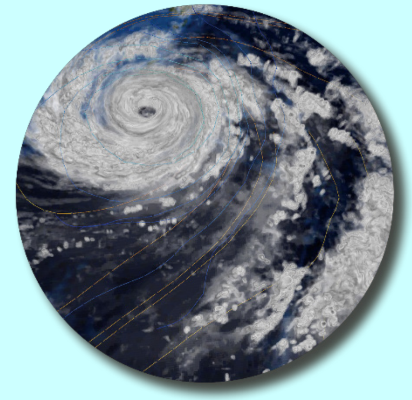
Sources of Big Data

Observations

Big Data

Advanced obs technology

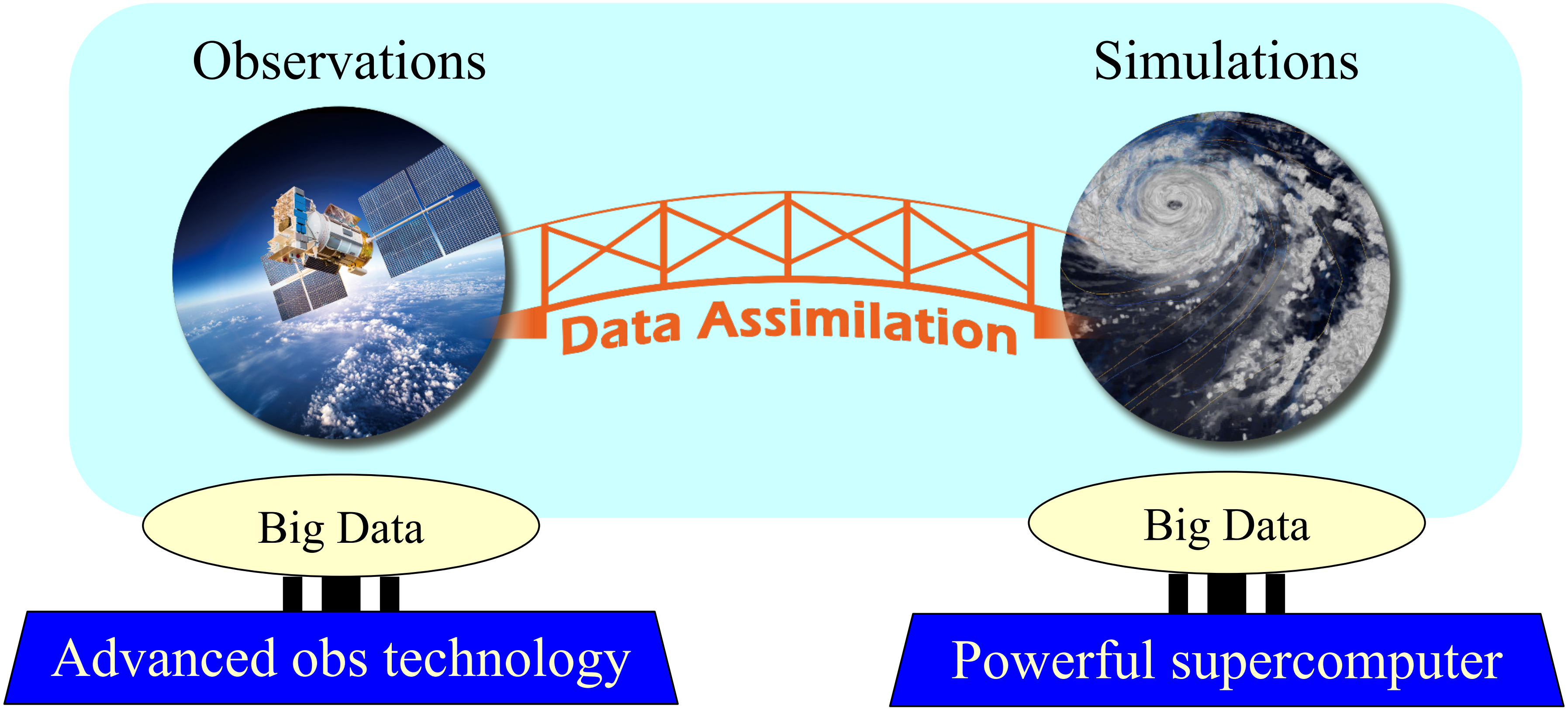
Simulations



Big Data

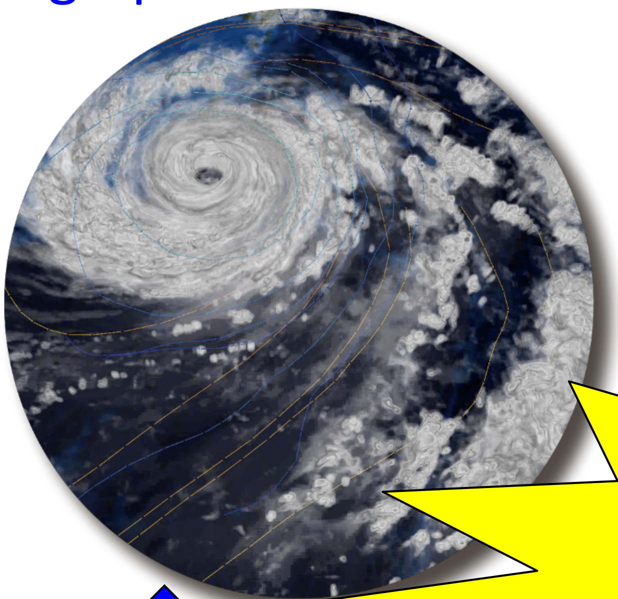
Powerful supercomputer

Big Data Assimilation



Pioneering "Big Data Assimilation" Era

High-precision Simulations

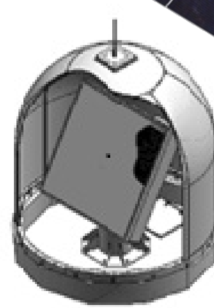


国立研究開発法人 科学技術振興機構 Japan Science and Technology Agency

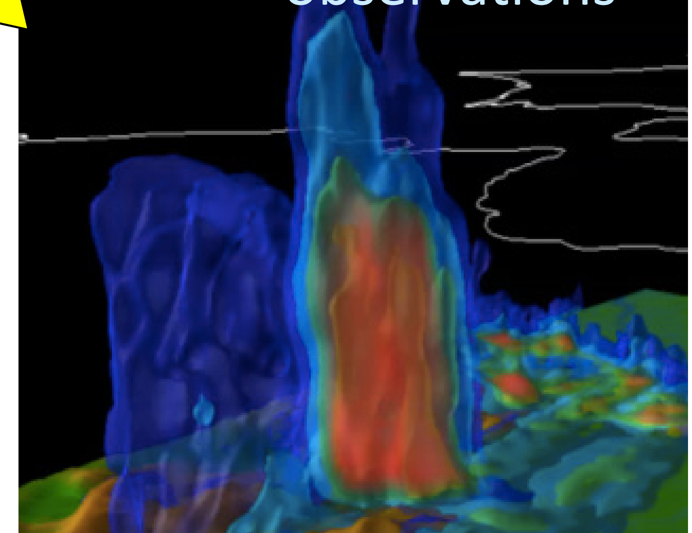
CREST

Future-generation technologies available 10 years in advance

BDA BIG DATA ASSIMILATION

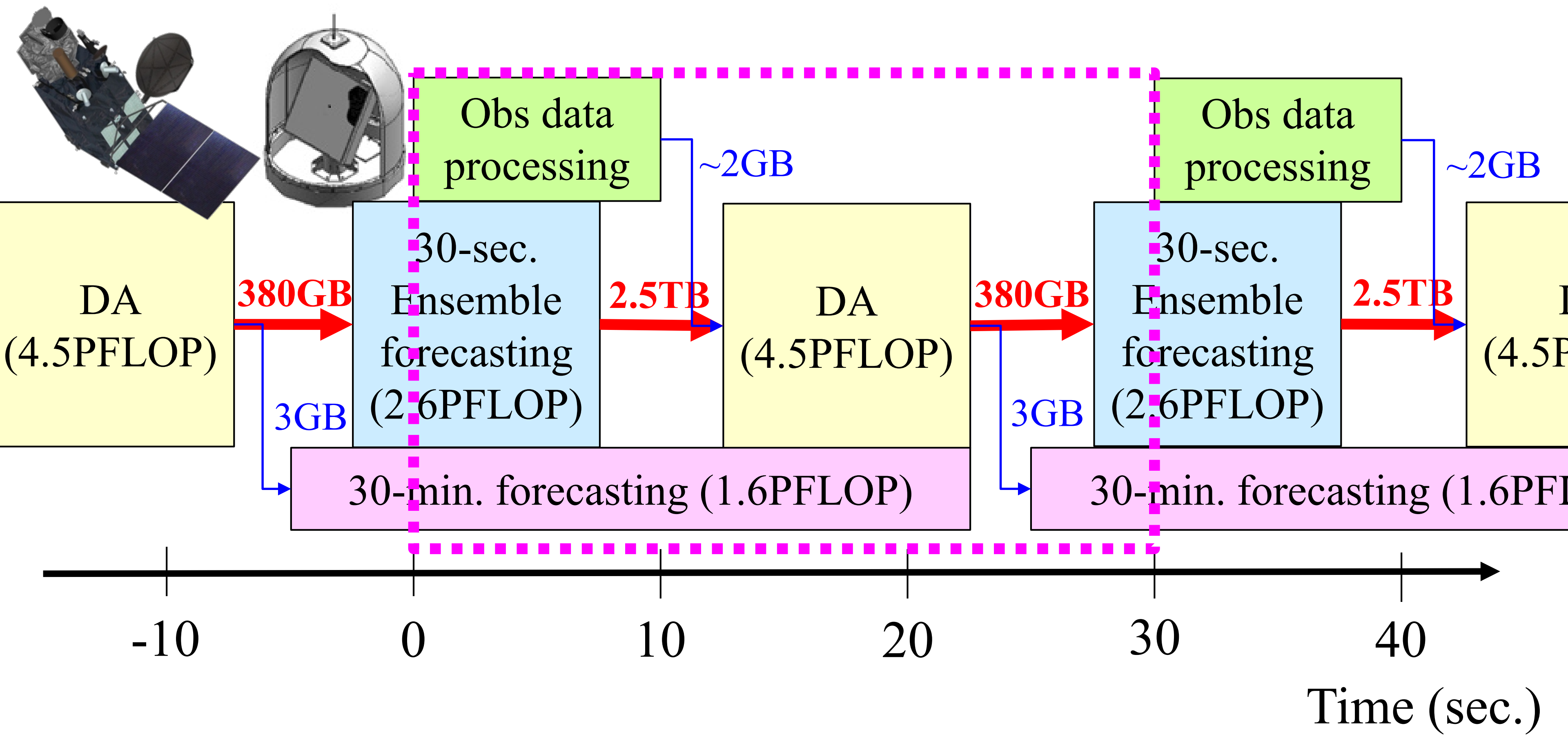


High-precision observations



Mutual feedback

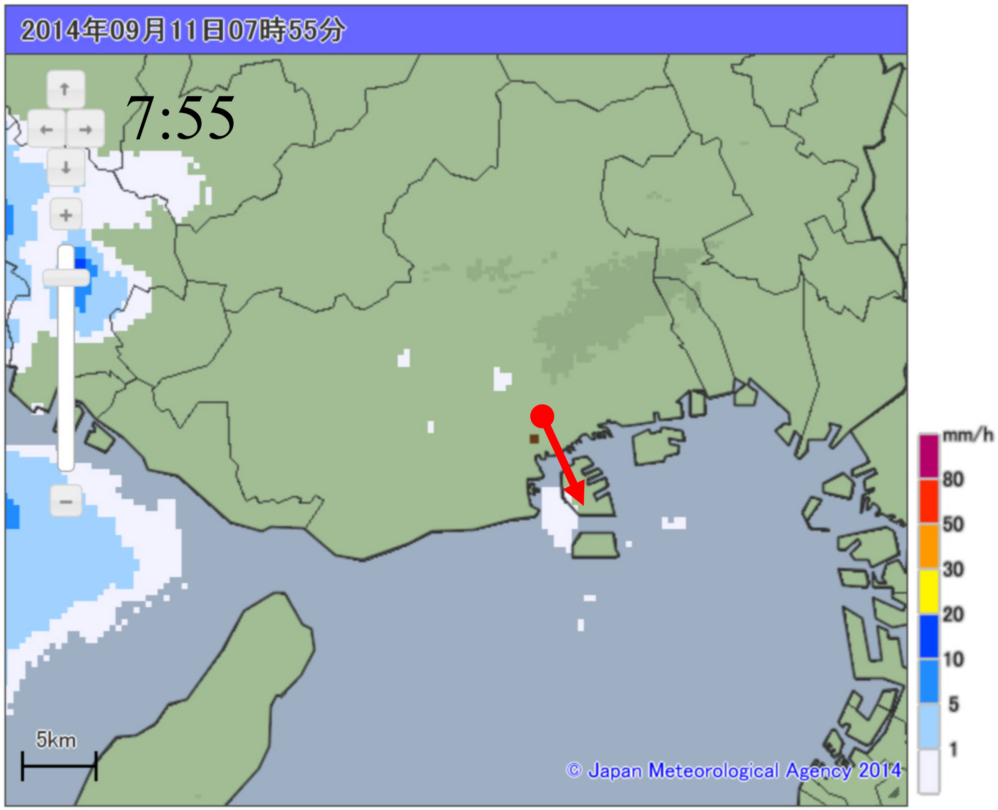
Revolutionary super-rapid 30-sec. cycle



<u>120 times more rapid</u> than hourly update cycles

9/11/2014 morning, sudden rain

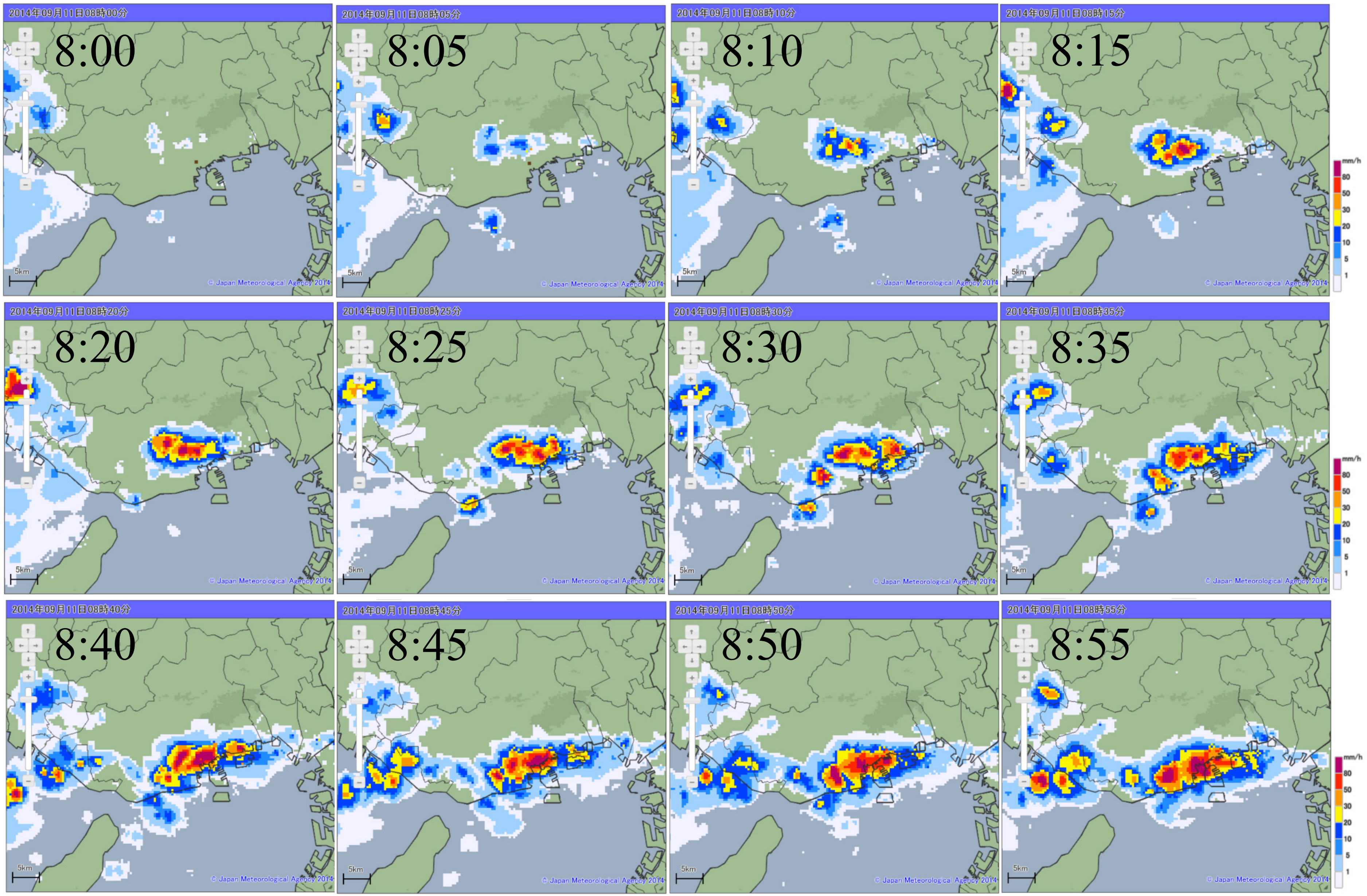
JMA observation



I looked at this obs at 8 am.

Decided to bike to office.
It takes about 30 min.

9/11/2014 morning, sudden rain



9/11/2014 morning, sudden rain

JMA observation

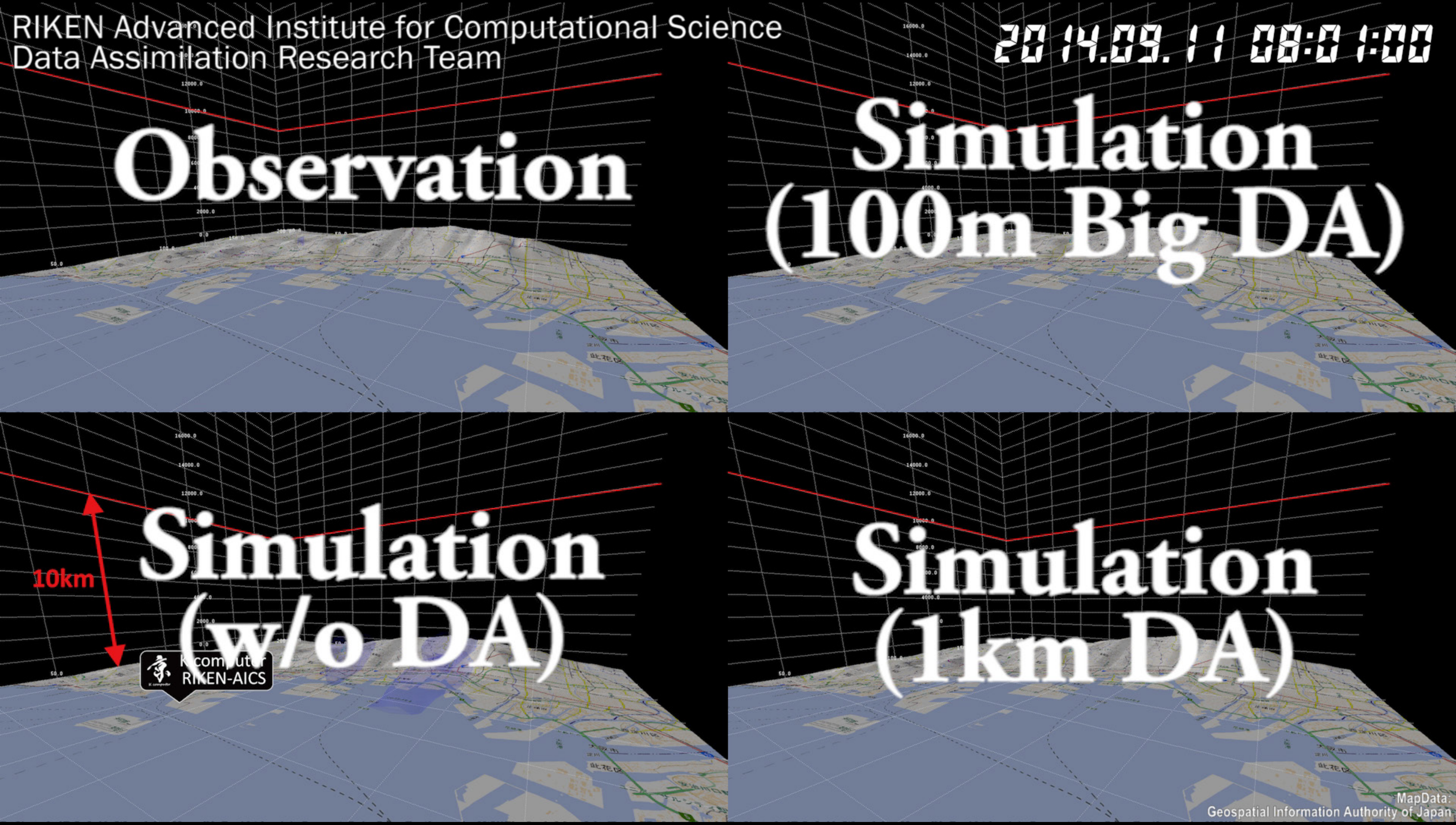
I looked at this obs at 8 am.

Decided to bike to office.
It takes about 30 min.

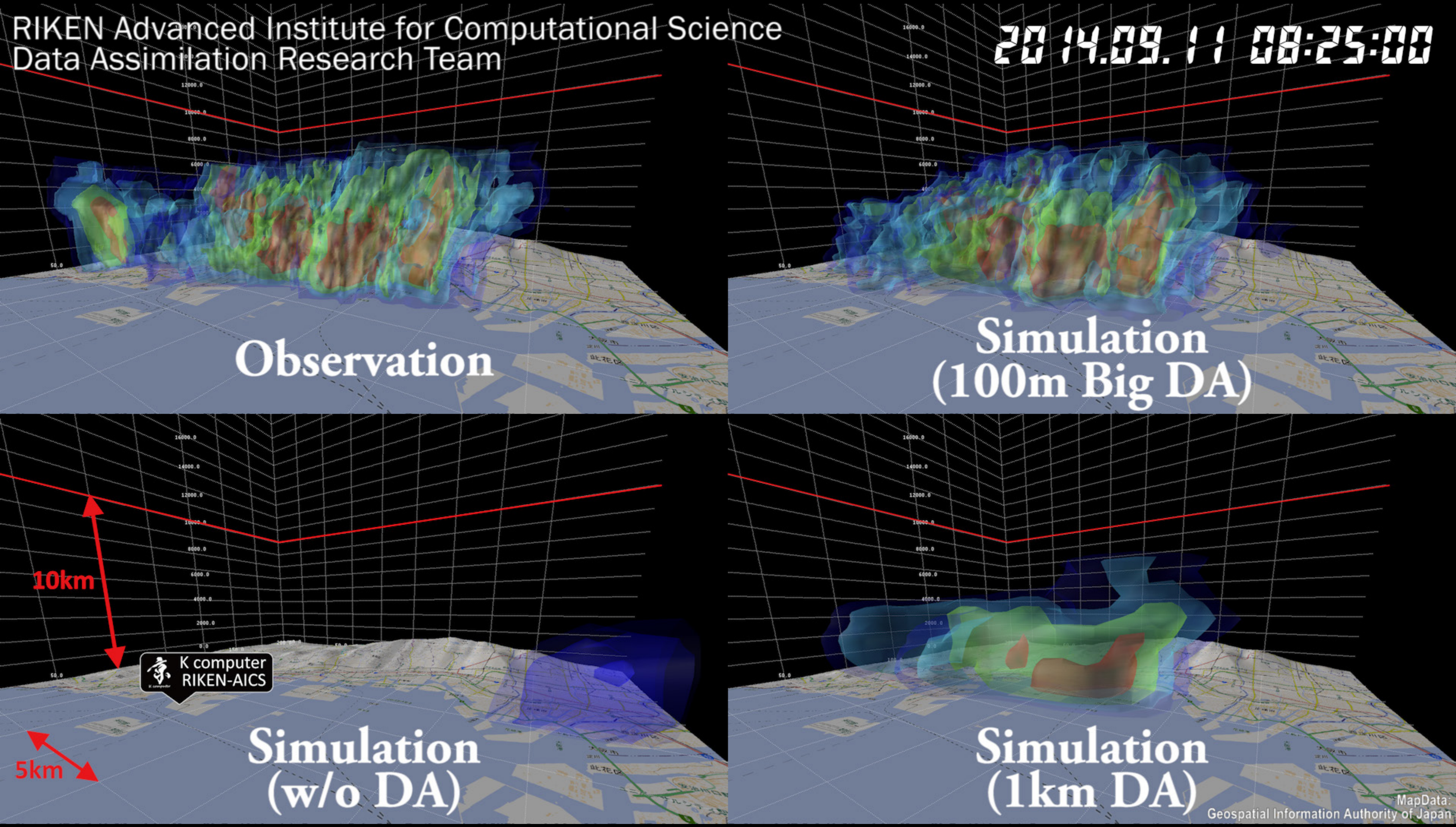
It is almost impossible to predict from this obs!

9/11/2014, sudden local rain

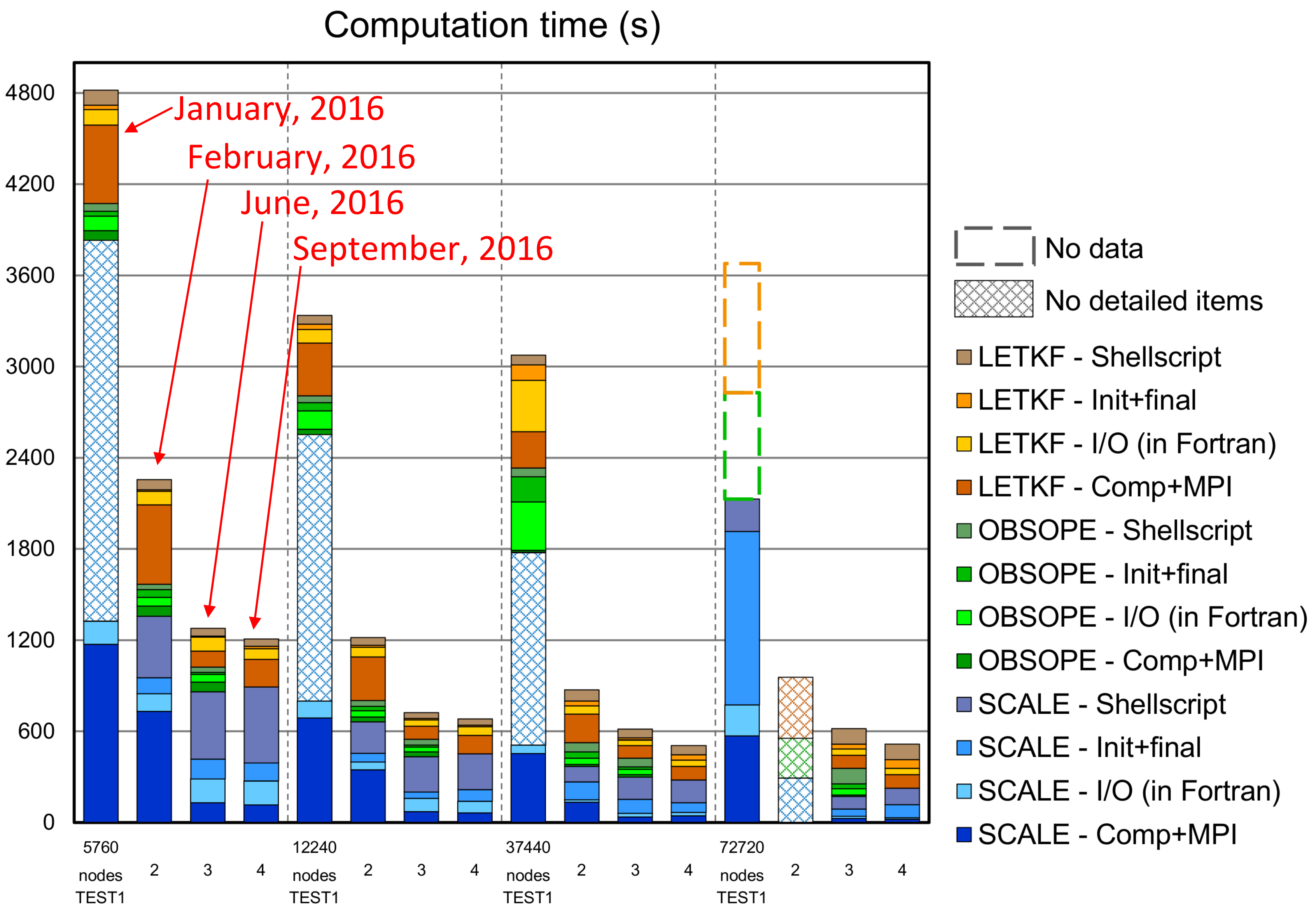
9/11/2014, sudden local rain



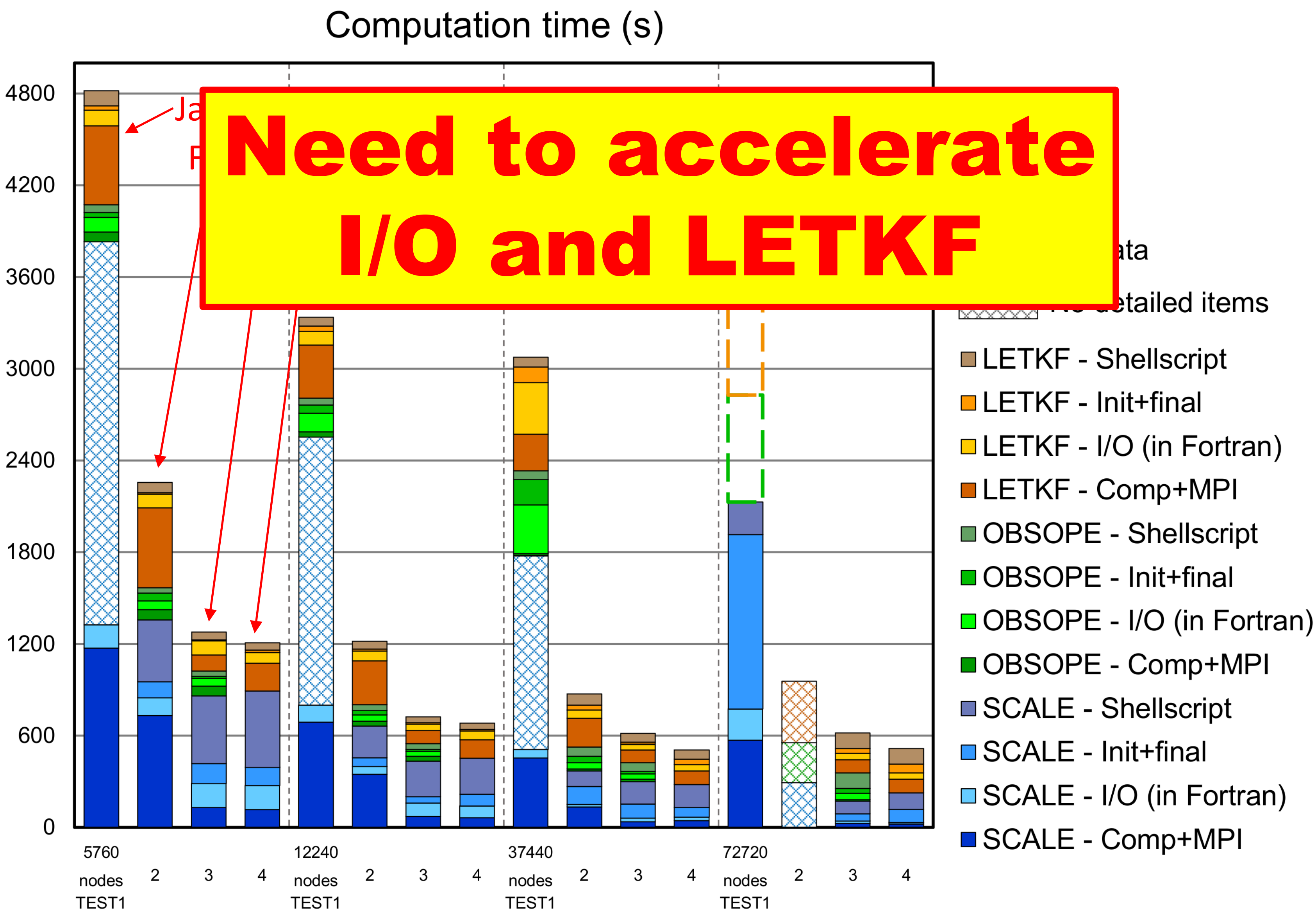
9/11/2014, sudden local rain



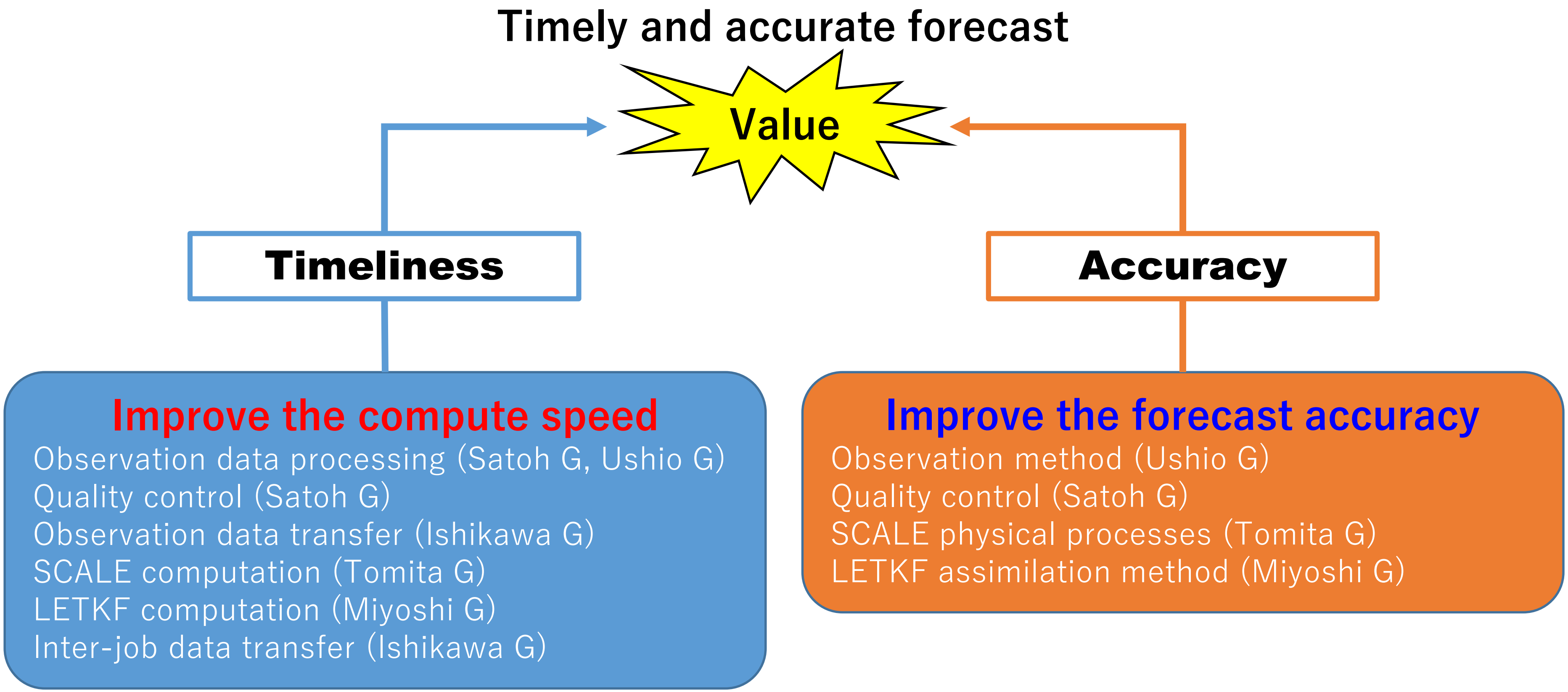
Huge-job test results – September, 2016



Huge-job test results – September, 2016



Future directions



RIKEN real-time weather service

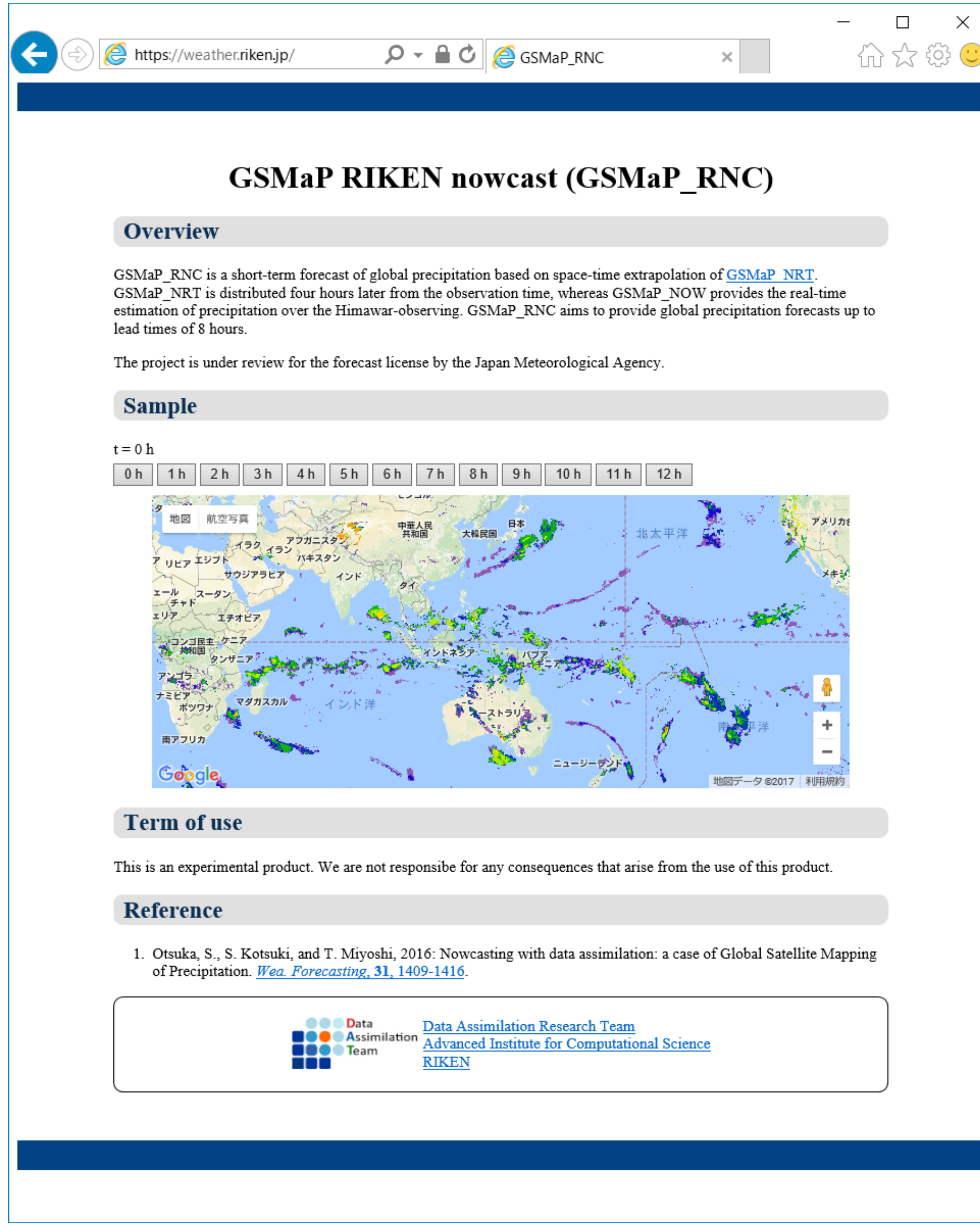
➔http://weather.riken.jp

Hourly-updated 12-h global precipitation prediction in real time

- ➢ JMA forecast license was issued. Final preparation stage.

Science → Service

- ➢ Engineering challenges
- ➢ Prediction algorithm
- ➢ Best use of computer



Computers keep advancing…

- With the "post-K" supercomputer (~2020), we can afford 100 samples of the global 870-m simulation.

With the Post-K, we aim to run 1000-sample global NICAM-LETKF at 3.5-km resolution

in close collaboration with the FLAGSHIP 2020 project

The Japanese 10-Peta-Flops K computer

Geophysical Research Letters
AN AGU JOURNAL
Explore this journal >

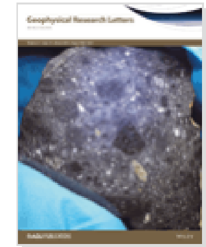
Research Letter

The 10,240-member ensemble Kalman filtering with an intermediate AGCM

Takemasa Miyoshi ✉, Keiichi Kondo, Toshiyuki Imamura

First published: 23 July 2014 Full publication history

DOI: 10.1002/2014GL060863 View/save citation

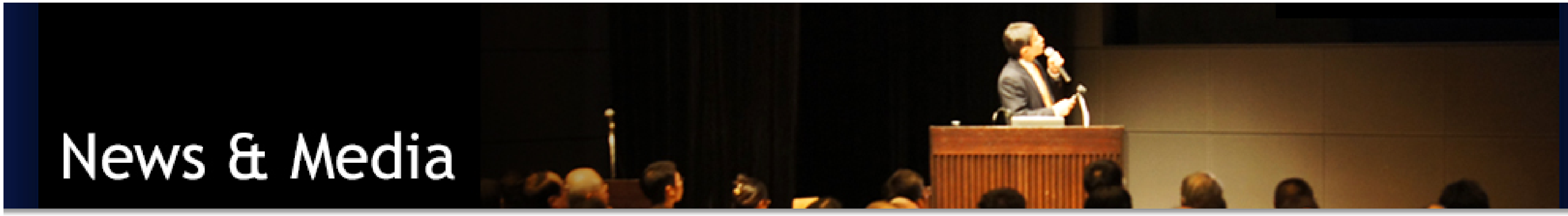


View issue TOC
Volume 41, Issue 14
28 July 2014
Pages 5264–5271

Abstract

The local ensemble transform Kalman filter (LETKF) with an intermediate atmospheric general circulation model (AGCM) is implemented with the Japanese 10 petaflops (floating point operations per second) “K computer” for large-ensemble simulations of 10,240 members, 2 orders of magnitude greater than the typical ensemble size of about 100. The computational challenge includes the eigenvalue decomposition of 10,240 × 10,240 dense covariance matrices at each grid point. Using the efficient eigenvalue solver for the K computer, the LETKF computations are accelerated by a factor of 8, allowing a 3 week experiment of 10,240-member LETKF with an intermediate AGCM for the first time. The flow-dependent 10,240-member ensemble revealed meaningful long-range error correlations at continental scales. The surface pressure error correlation shows teleconnection patterns like the Pacific North American pattern. Specific humidity error correlation shows continental scale wave trains. Investigations with different ensemble sizes suggest that at least several hundred members be necessary to capture these continental scale error correlations.

RIKEN



Press Release

Previous Index Next

July 23, 2014

K computer runs largest ever ensemble simulation of global weather

Ensemble forecasting is a key part of weather forecasting today. Computers typically run multiple simulations, called ensembles, using slightly different initial conditions or assumptions, and then analyze them together to try to improve forecasts. Now, in research published in *Geophysical Research Letters*, using Japan's flagship 10-petaFLOPS K computer, researchers from the RIKEN Advanced Institute for Computational Science (AICS) have succeeded in running 10,240 parallel simulations of global weather, the largest number ever performed, using data assimilation to reduce the range of uncertainties.

The assimilation of the 10,240 ensemble data sets was made possible by a cross-disciplinary collaboration of data assimilation experts and eigenvalue solver scientists at RIKEN AICS. The "Local Ensemble Transform Kalman Filter" (LETKF), an already efficient system, was further improved by a factor of eight using the "EigenExa" high-performance eigenvalue solver software, making possible a three-week computation of data from the 10,240 ensembles for simulated global weather. By analyzing the 10,240 equally probable estimates of atmospheric states, the team discovered that faraway observations, even going beyond 10,000 kilometers in distance, may have an immediate impact on eventual state of the estimation. This finding suggests the need for further research on advanced methods that can make better use of faraway observations, as this could potentially lead to an improvement of weather forecasts.

News & Media
Press Center
Press Releases
2014
2013
2012
2011
2010
2009
2008
2007
2006
2005
News
Events & Symposiums
Publications
Videos

RIKEN

About RIKEN | Research | News & Media | Careers | Outreach | Community

News & Media

Home > News & Media > Press Releases > Press Releases 2014 >

Press Release

Previous | Index | Next

July 23, 2014

Like 59 | Tweet 21

K computer runs largest ever ensemble simulation of global weather

Ensemble forecasti
called ensembles,
try to improve fore
10-petaFLOPS K co
have succeeded in
using data assimila

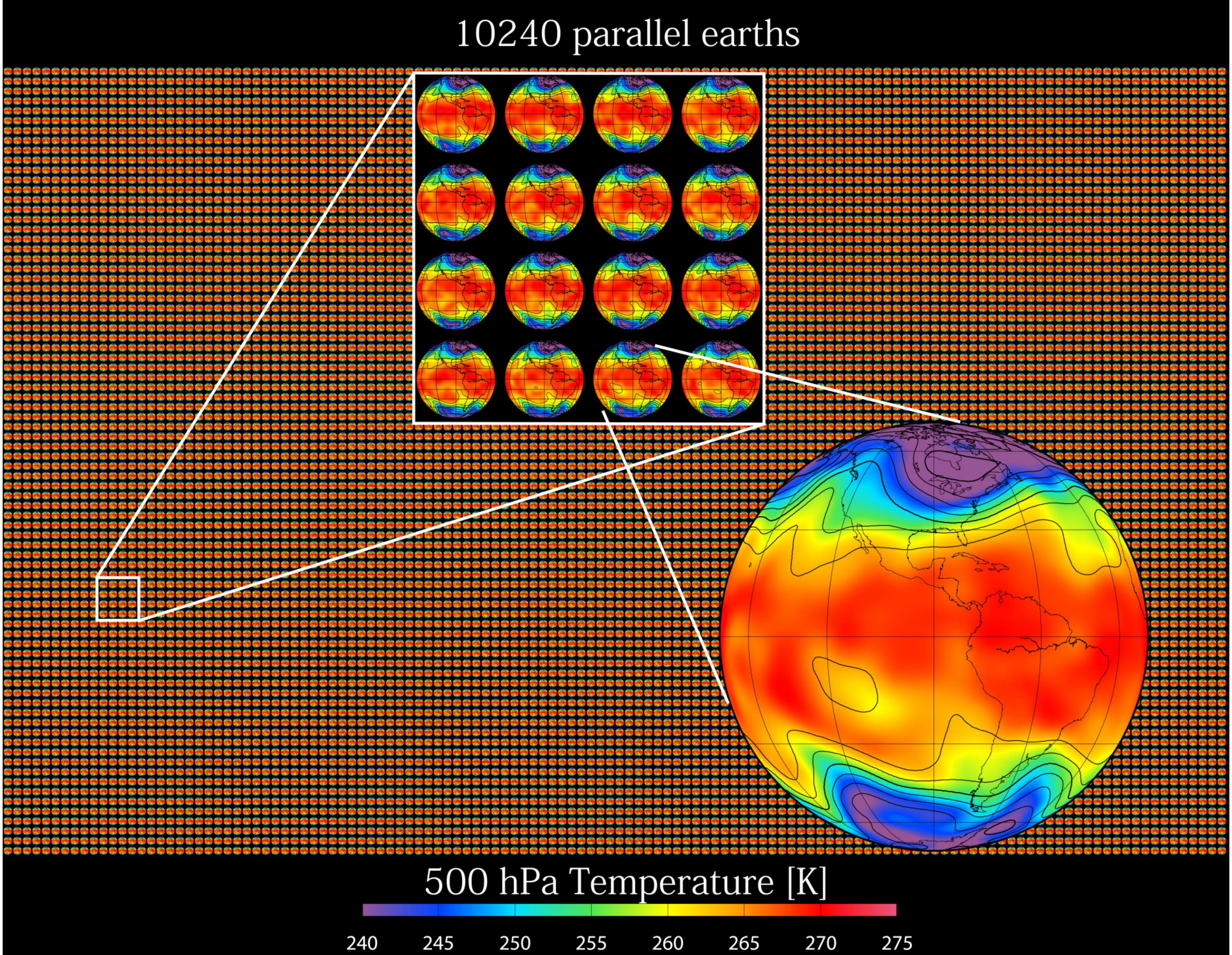
A simulated study using the T30/L7 SPEEDY AGCM

(Miyoshi, Kondo, Imamura 2014)

The assimilation of the 10,240 ensemble data sets was made possible by a cross-disciplinary collaboration of data assimilation experts and eigenvalue solver scientists at RIKEN AICS. The "Local Ensemble Transform Kalman Filter" (LETKF), an already efficient system, was further improved by a factor of eight using the "EigenExa" high-performance eigenvalue solver software, making possible a three-week computation of data from the 10,240 ensembles for simulated global weather. By analyzing the 10,240 equally probable estimates of atmospheric states, the team discovered that faraway observations, even going beyond 10,000 kilometers in distance, may have an immediate impact on eventual state of the estimation. This finding suggests the need for further research on advanced methods that can make better use of faraway observations, as this could potentially lead to an improvement of weather forecasts.

News & Media

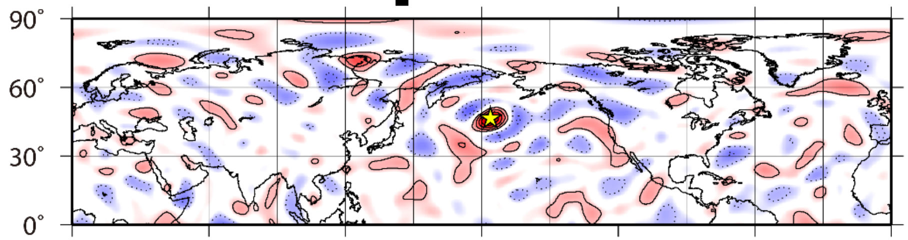
- Press Center
- Press Releases
  - 2014
  - 2007
  - 2006
  - 2005
- News
- Events & Symposiums
- Publications
- Videos

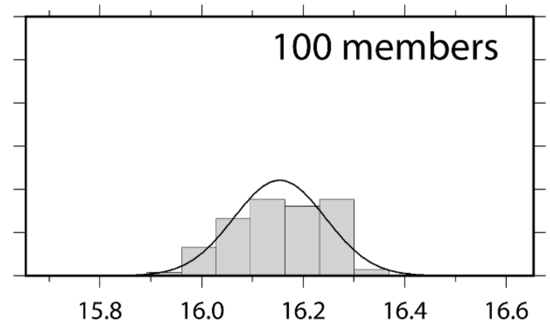

10240 parallel earths
500 hPa Temperature [K]
240
245
250
255
260
265
270
275

Advantage of large ensemble

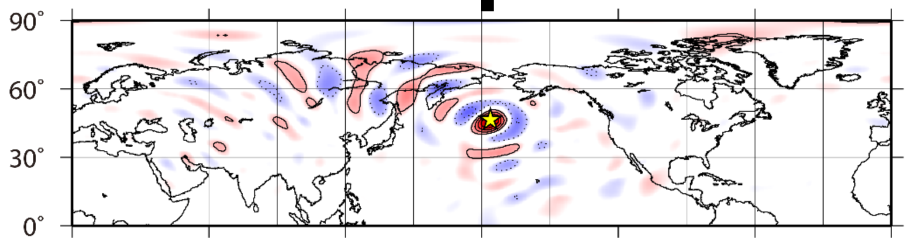
(Miyoshi, Kondo, Imamura 2014)

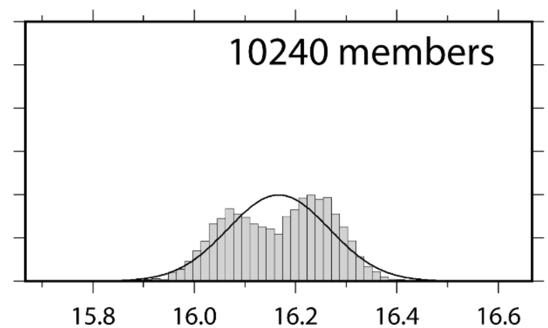
100 samples





10240 samples





Sampling noise reduced

High-precision probabilistic representation

RIKEN | About RIKEN | Research | News & Media | Careers | Outreach | Community

Home > News & Media > Press Releases > Press Releases 2015 >

Press Release

Previous | Index | Next

November 11, 2015

Like 13 | Tweet

Largest ensemble simulation of global weather using real-world data

When performing numerical weather predictions, it is important that the simulation itself be accurate, but it is also key for real-world data, based on observations, to be accurately entered into the model. Typically, weather simulations work by having the computer conduct a number of simulations based on the current state, and then entering observational data into the simulation to nudge it in a way that puts it closer to the actual state. The problem of incorporating data in the simulation—data assimilation—has become increasingly complex with the large number of types of available data, such as satellite observations and measurements taken from ground stations. Typically, supercomputers today spend an approximately equal amount of time running the simulations and incorporating the real-world data.

News & Media

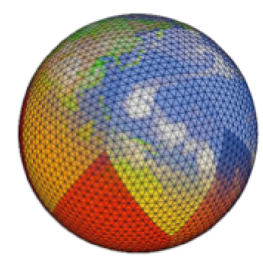
- Press Center
- Press Releases
  - 2015
  - 2014
  - 2013
  - 2012
  - 2011
  - 2010
  - 2009
  - 2008

A real-world study using the NICAM

(Miyoshi, Kondo, Terasaki 2015)

NICAM-LETKF

(Terasaki et al. 2015)



Computer

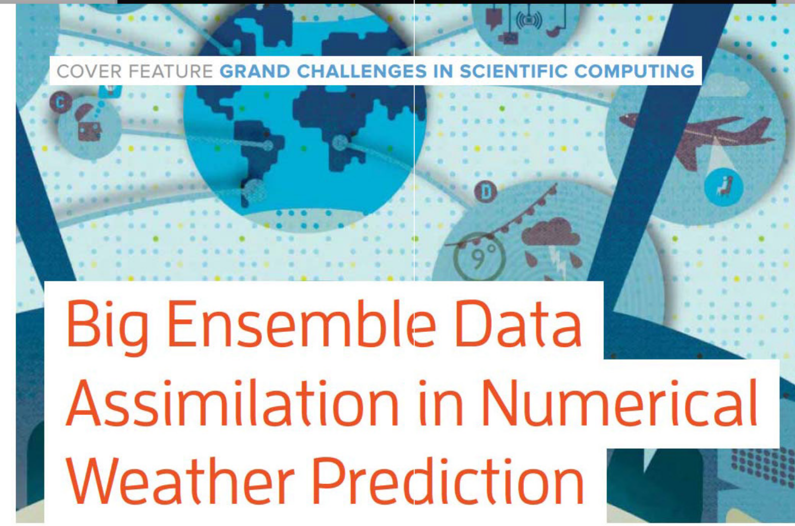
MULTIMEDIA

Cover feature!

Computer Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page

COVER FEATURE **GRAND CHALLENGES IN SCIENTIFIC COMPUTING**

VIDEO



Big Ensemble Data Assimilation in Numerical Weather Prediction

Takemasa Miyoshi, RIKEN Advanced Institute for Computational Science, University of Maryland, and Japan Agency for Marine-Earth Science and Technology

Keiichi Kondo and Koji Terasaki, RIKEN Advanced Institute for Computational Science

Powerful computers and advanced sensors enable precise simulations of the atmospheric state, requiring data assimilation to connect simulations to real-world sensor data using statistical mathematics and dynamical systems theory. Numerical weather prediction (NWP) thus enables simulations that more closely represent the real world. The authors explore the NWP-associated challenges in managing big data through supercomputing.

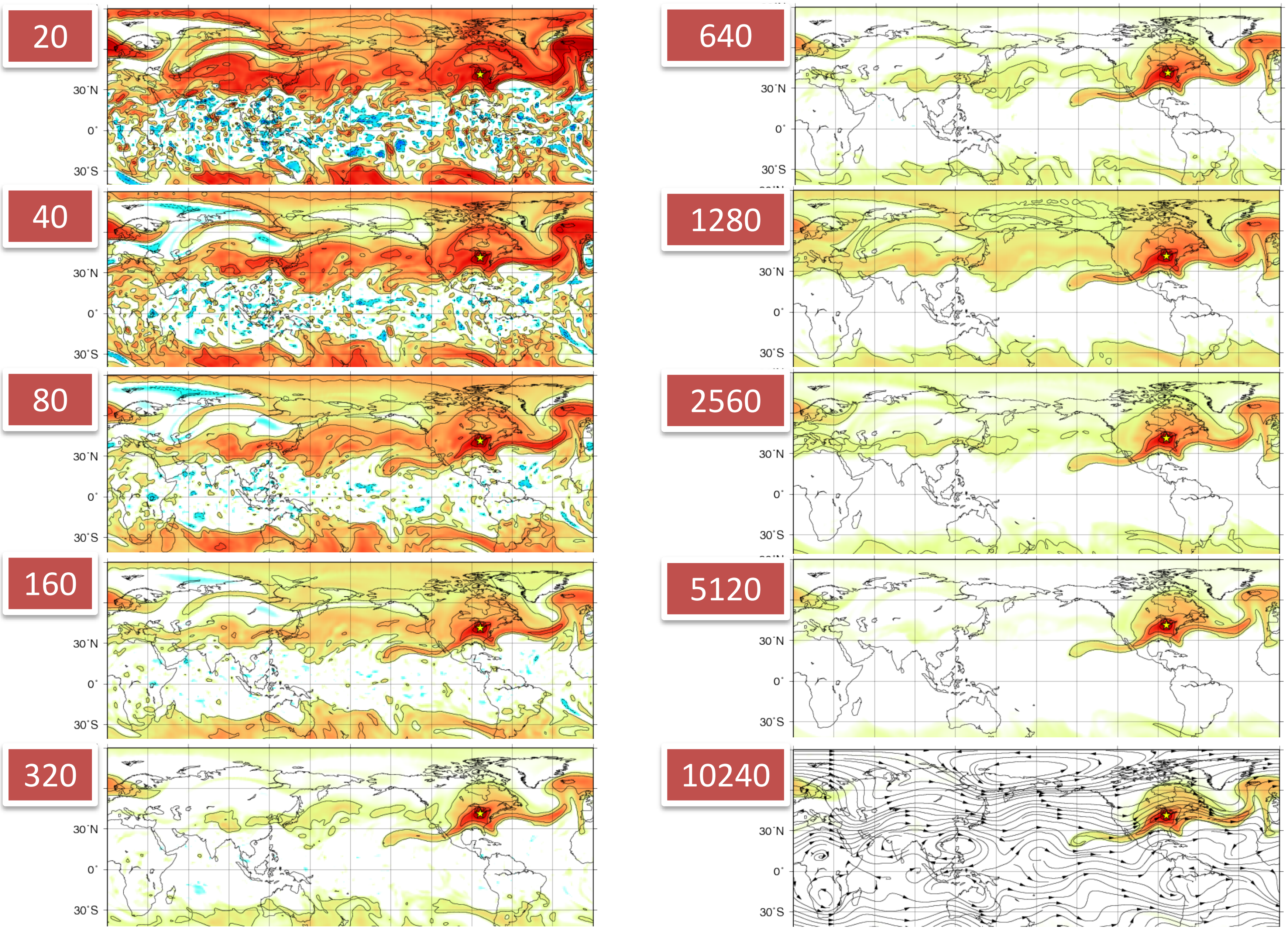
High-performance computing (HPC) is essential for *numerical weather prediction* (NWP), the method by which computer models of the atmosphere are used to predict the weather. Advances in computing power enable higher resolution and more complex physical representations of the atmosphere. Although these more advanced representations have led to more accurate weather forecasts from supercomputers than the first models from 1950, the technology is still far from ideal.[1]

In NWP, synchronizing the computer simulation with the real world is essential to accurately determine the atmosphere's current state and likely evolution. Although more precise simulations and more powerful computing are helpful in improving accuracy, data assimilation (DA) plays a key role in improving integration between the computer simulation and real-world observation data.[2,3] DA also employs HPC; in fact, global NWP systems devote equivalent computational resources to DA and 10-day forecast simulation.

To accurately represent the probability density function (PDF) in the ensemble Kalman filter (EnKF)—an advanced DA approach widely used in NWP—within the global atmosphere, we used a large sample size and the

With subsets of 10240 samples

Kondo&Miyoshi (2015)



Accelerating SPEEDY-LETKF

○ Improve the eigenvalue solver (67～84% of the LETKF)

- EigenExa (Imamura et al. 2011)
  - Timing [sec.] (48 nodes)

Ensemble size	40	80	160	320	640	1280	2560
Original	**1.5**	**3.7**	14.9	72.6	346.8	2035.3	-
EigenExa	3.1	6.6	**12.2**	**30.6**	**91.0**	**320.4**	**1595.0**

  - Timing [sec.] (4608 nodes)

Ensemble size	2560	5120	10240
Original	165.4	1017.5	7419.3
EigenExa	**22.3**	**118.0**	**802.1**

9x

○ Remove vertical localization, remove the vertical loop

- Timing with 10240 samples with 4608 nodes

	Timing [sec.]	RAM usage [GB]
Original	802.1	14.2
No vert. loc.	**150.6**	**8.8**

5x

Accelerating NICAM-LETKF

- ○ MPI_SCATTER, GATHER ➔ ALLTOALL
  - 20 samples, 804 nodes
    19.3 → **11.8** sec.

- ○ Optimized DGEMM for the K computer
  - 1809 → **808** sec.

- ○ Remove the vertical loop
  - 1280 samples, 804 nodes

	EISPACK	EigenExa	No vertical loc.
Timing for LETKF	12488 sec.	2025 sec.	**200 sec.**
FLOPS/Peak FLOPS	7.4 %	44.6 %	15.6 %

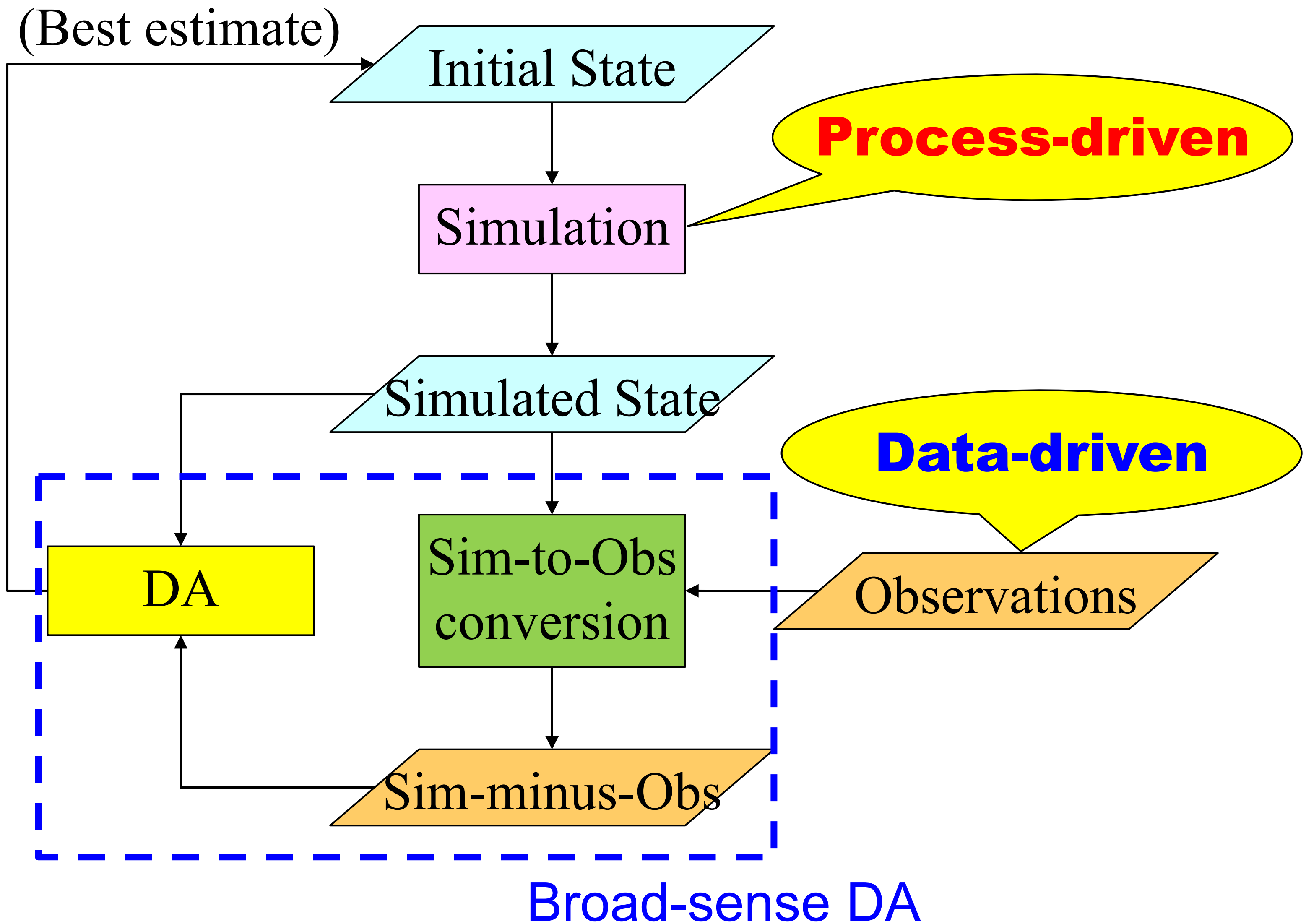
Data size in NICAM-LETKF

- 80 samples vs. 10240 samples

	80 samples	10240 samples
# of files	25,600	3,276,800
data size	30 GB	3,937 GB
timing	332 sec. (804 nodes)	140±30 min. (~6000 nodes)

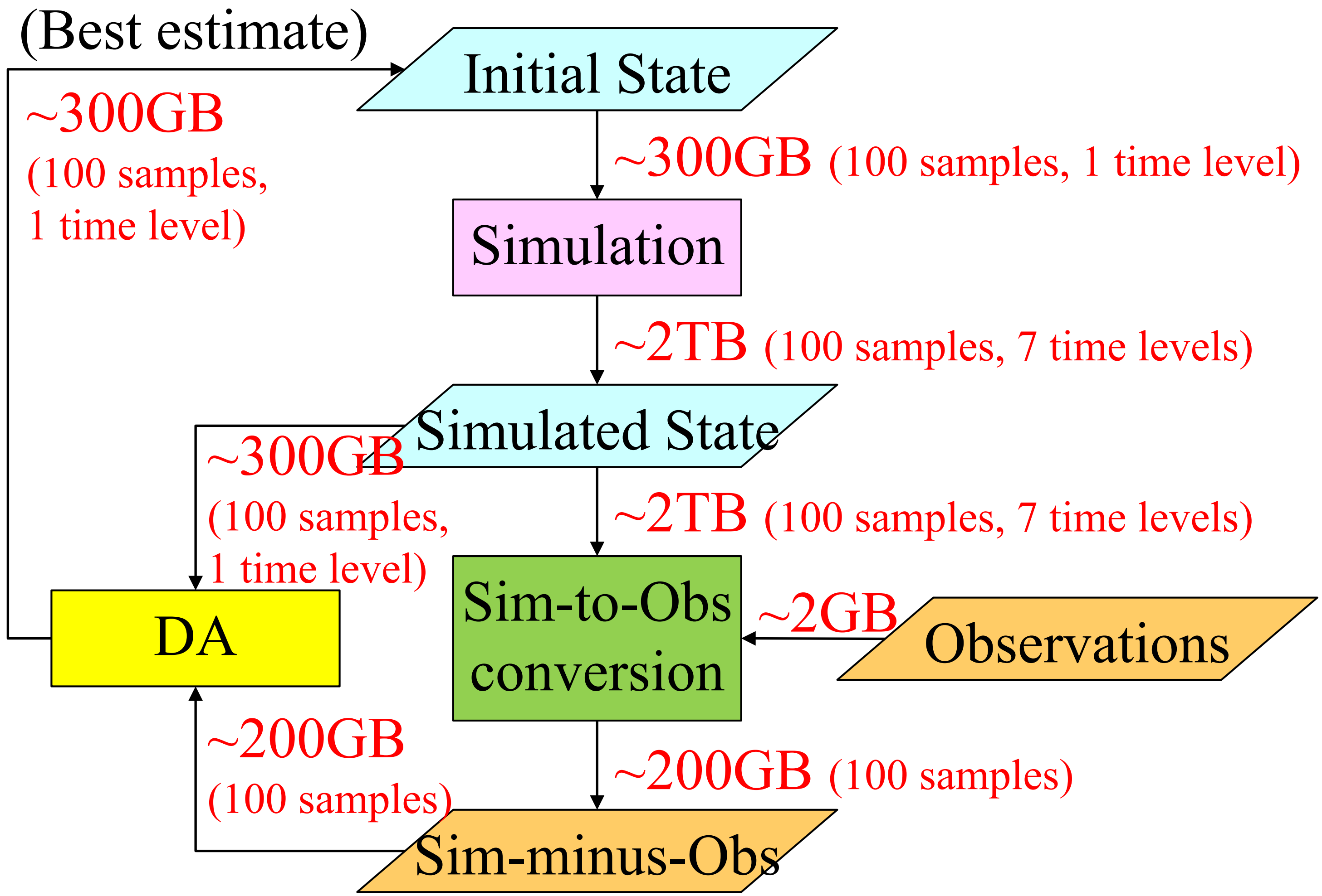
190x

DA workflow

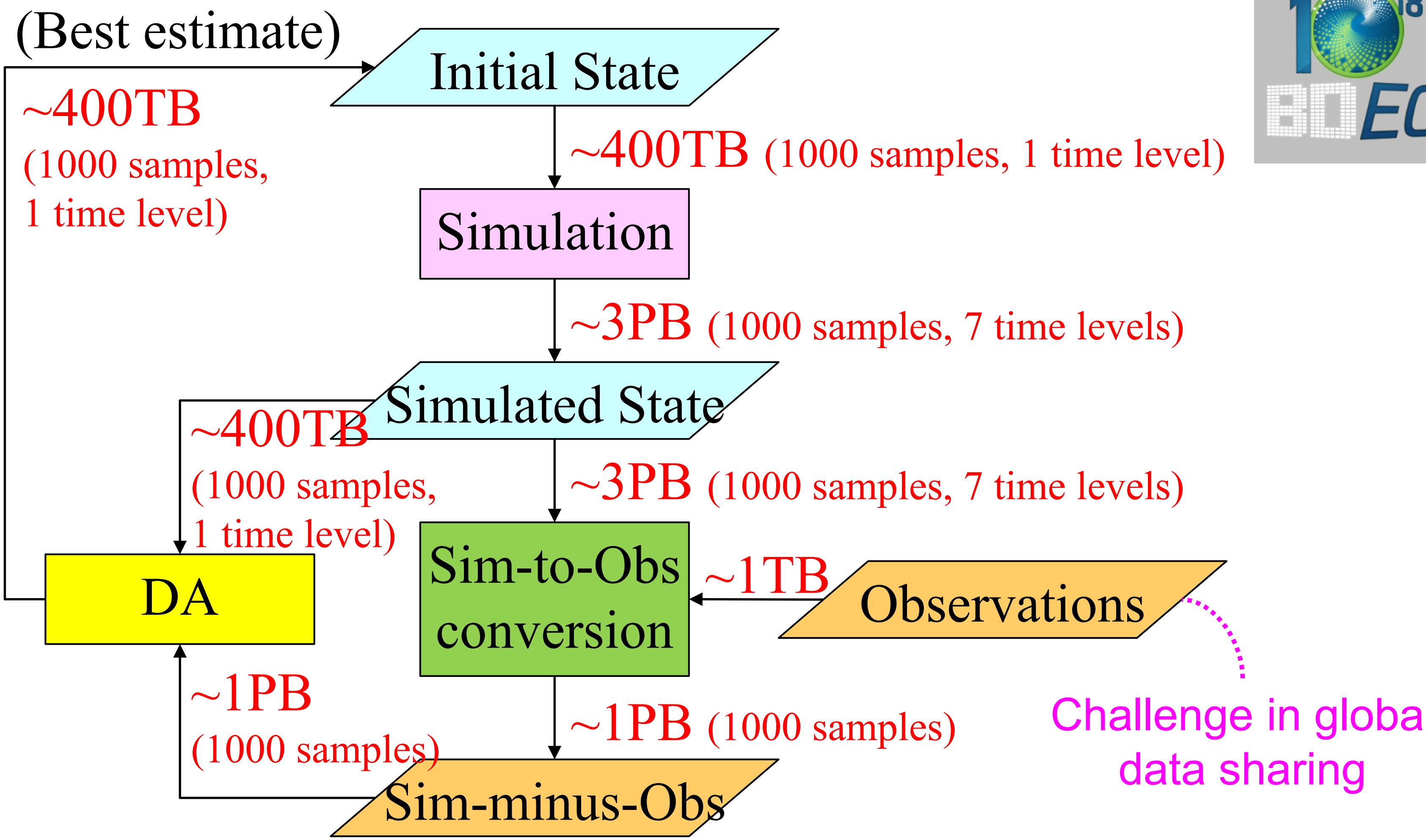


(Best estimate)
Initial State
Process-driven
Simulation
Simulated State
Data-driven
DA
Sim-to-Obs
conversion
Observations
Sim-minus-Obs
Broad-sense DA

Workflow with current data size



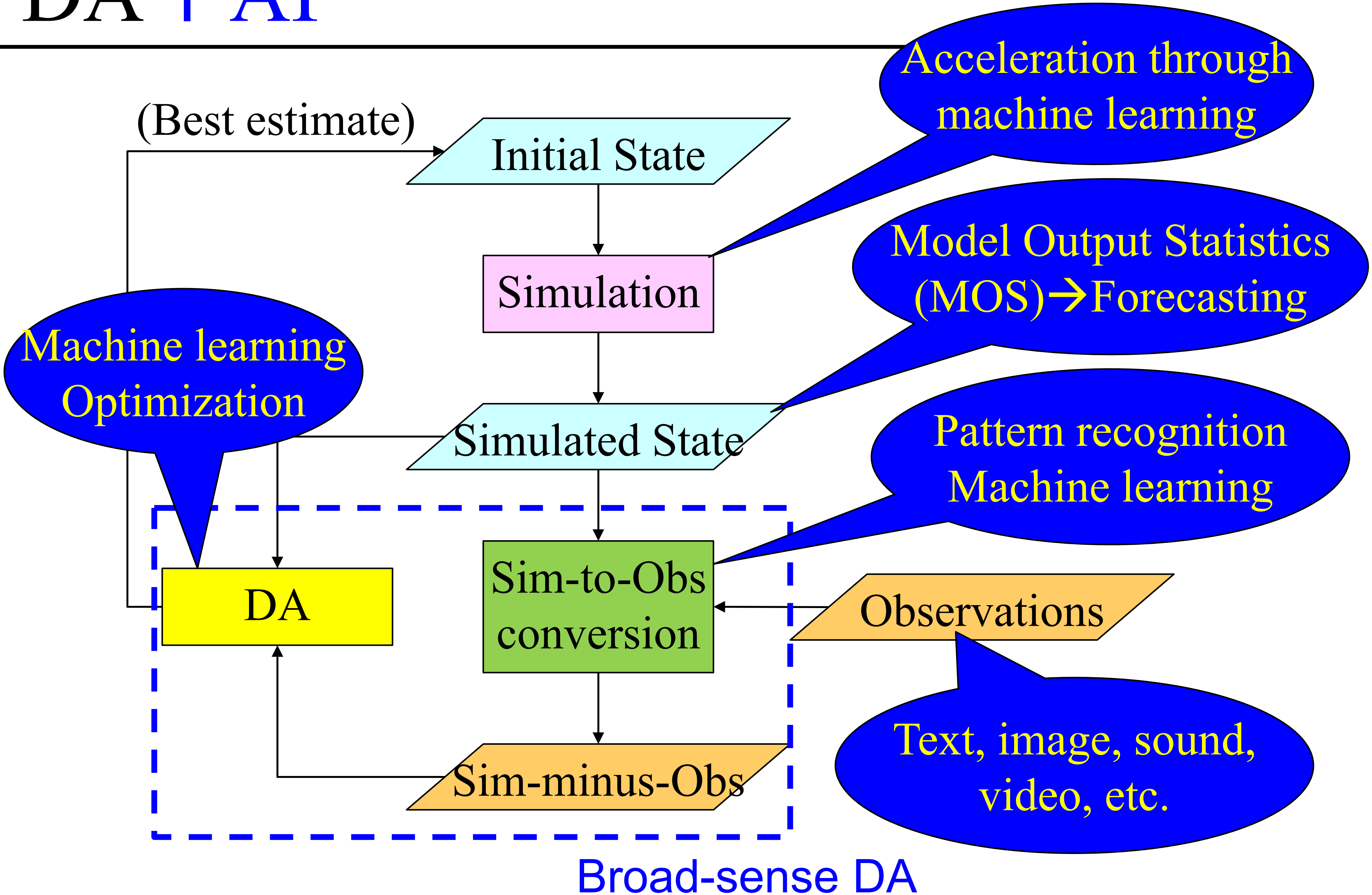
Workflow with extreme-scale



I/O intensive!

Repetitions of I/O between separate programs

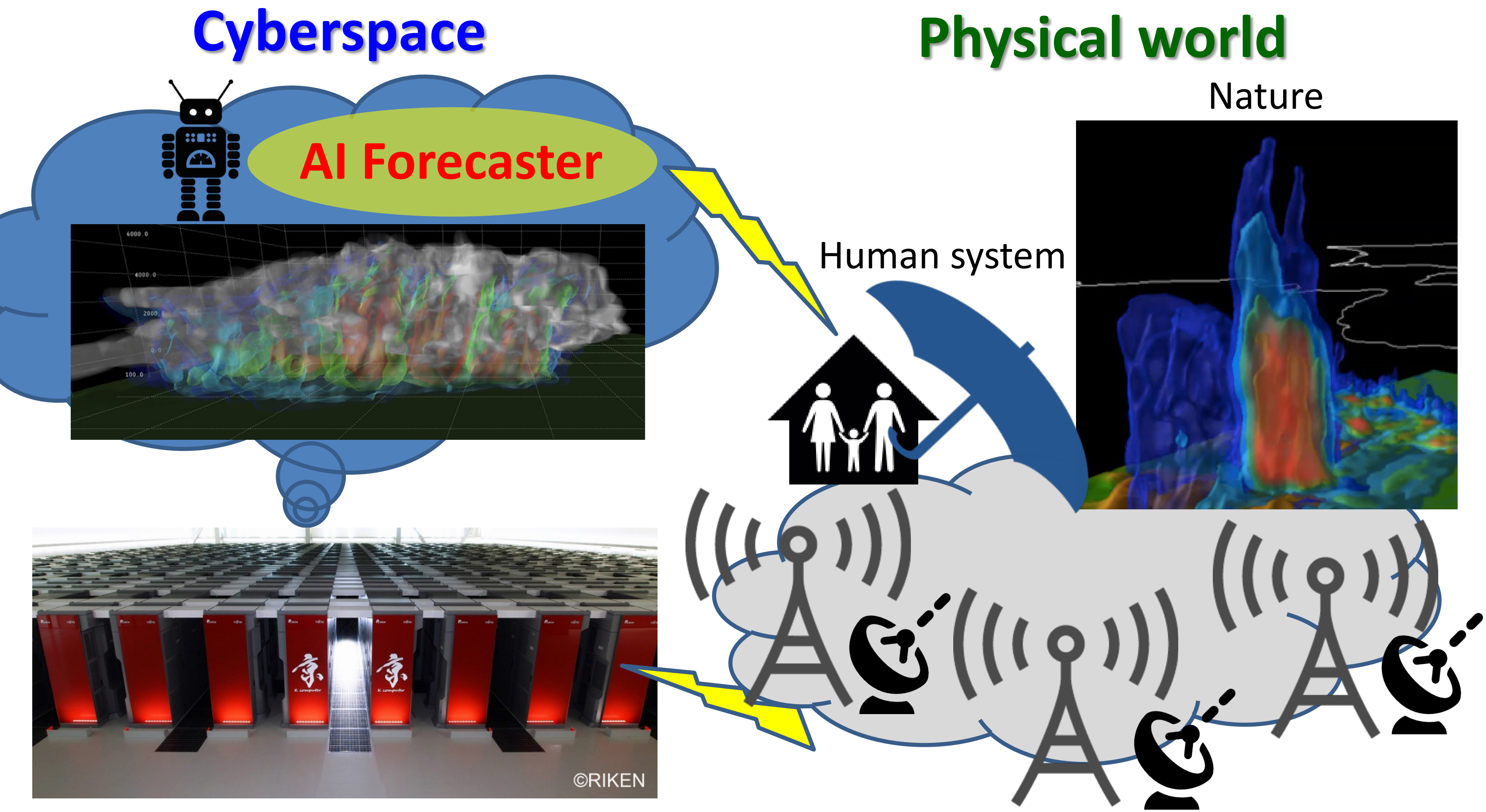
DA + AI



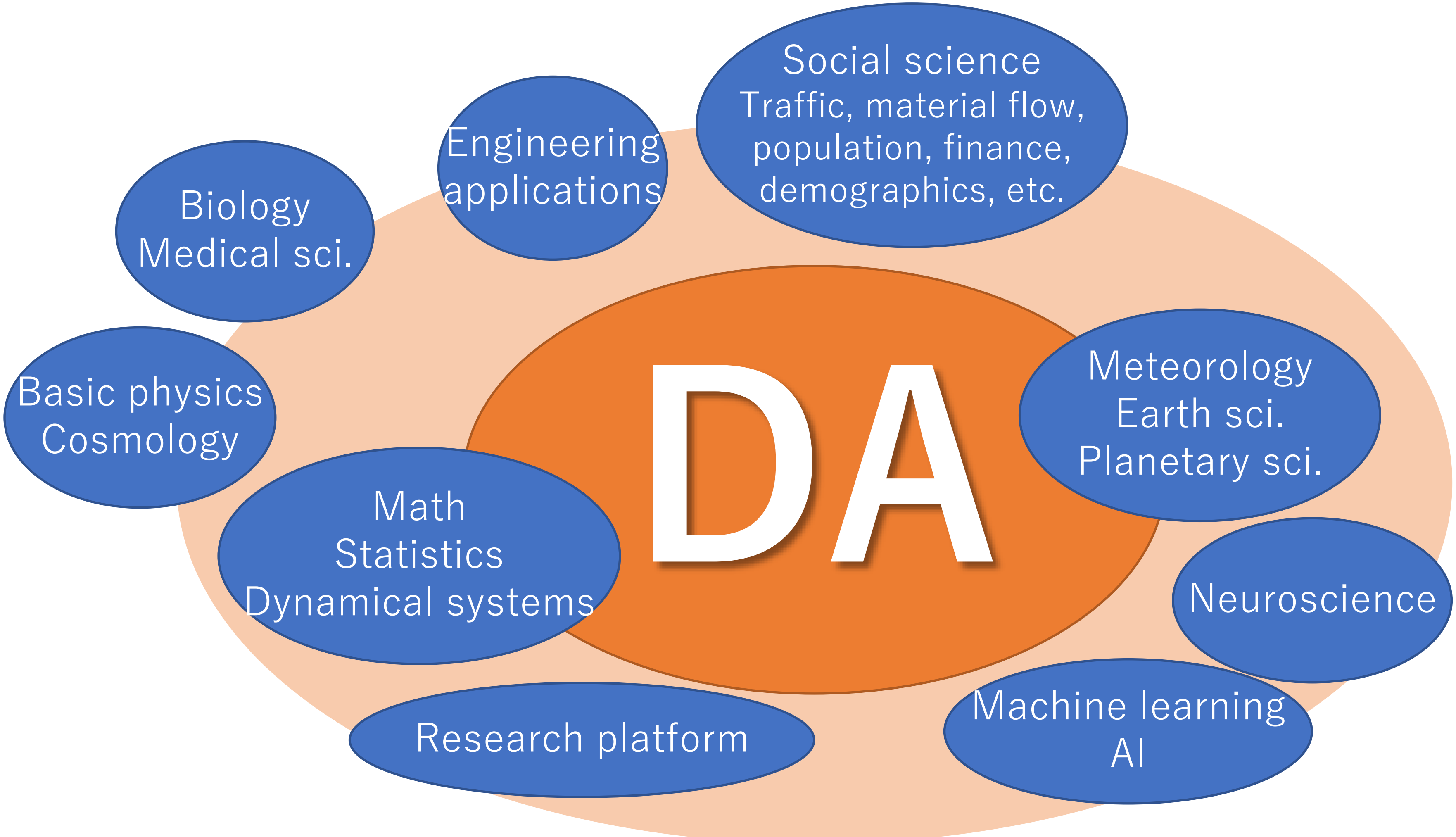
(Best estimate)
Initial State
Acceleration through machine learning
Simulation
Model Output Statistics (MOS)→Forecasting
Machine learning Optimization
Simulated State
Pattern recognition Machine learning
DA
Sim-to-Obs conversion
Observations
Text, image, sound, video, etc.
Sim-minus-Obs
Broad-sense DA

Cyber-Physical framework for weather prediction

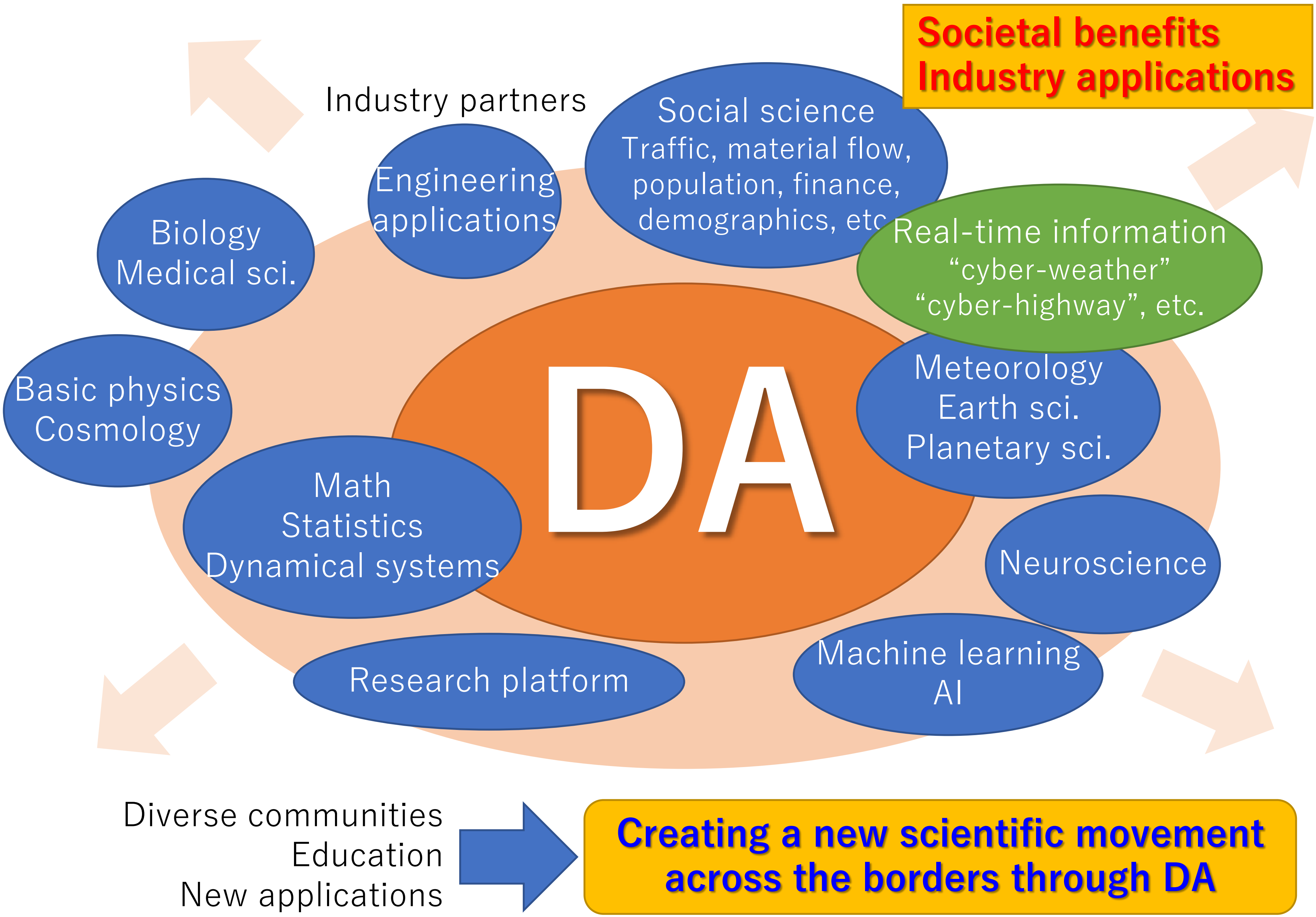
Data Assimilation is the key




DA
Math
Statistics
Dynamical systems
Meteorology
Earth sci.
Planetary sci.



Social science
Traffic, material flow,
population, finance,
demographics, etc.
Engineering
applications
Biology
Medical sci.
Basic physics
Cosmology
DA
Meteorology
Earth sci.
Planetary sci.
Math
Statistics
Dynamical systems
Neuroscience
Research platform
Machine learning
AI


Societal benefits
Industry applications
Industry partners
Social science
Traffic, material flow, population, finance, demographics, etc.
Engineering applications
Biology
Medical sci.
Real-time information
"cyber-weather"
"cyber-highway", etc.
Basic physics
Cosmology
DA
Meteorology
Earth sci.
Planetary sci.
Math
Statistics
Dynamical systems
Neuroscience
Research platform
Machine learning
AI
Diverse communities
Education
New applications
Creating a new scientific movement across the borders through DA