

*“A 1024-Member Ensemble Data Assimilation
with 3.5-Km Mesh Global Weather Simulations”*

「3.5kmメッシュ全球気象シミュレーションを用いた
1024メンバーアンサンブルデータ同化」

(2020年ACM ゴードン・ベル賞ファイナリスト論文)

スーパーコンピュータ「富岳」を用いた史上最大規模の気象計算

八代 尚 (国立環境研/理研R-CCS)、寺崎 康児 (理研R-CCS)、河合 佑太 (理研R-CCS)、工藤 周平 (理研R-CCS)、
三好 建正 (理研R-CCS)、今村 俊幸 (理研R-CCS)、南 一生 (理研R-CCS)、井上 晃 (富士通(株))、
錦 龍生 (富士通(株))、佐治 隆行 ((株)メトロ)、佐藤 正樹 (東大大気海洋研)、富田 浩文 (理研R-CCS)

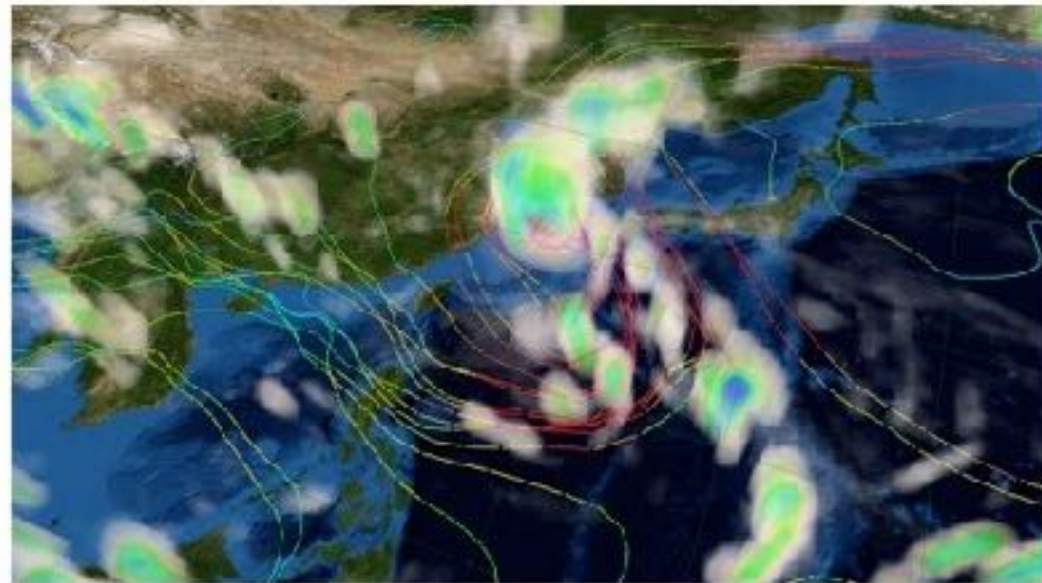
本研究のポイント

- 世界最大規模とは？
→空間解像度 × アンサンブル数の組み合わせが最大
- 気象シミュレーション × データ同化の複合計算
- 演算性能だけではない「富岳」の総合力を実証する結果
→「京」に対して100倍の性能
- 6年に渡る「富岳」コデザインの成果

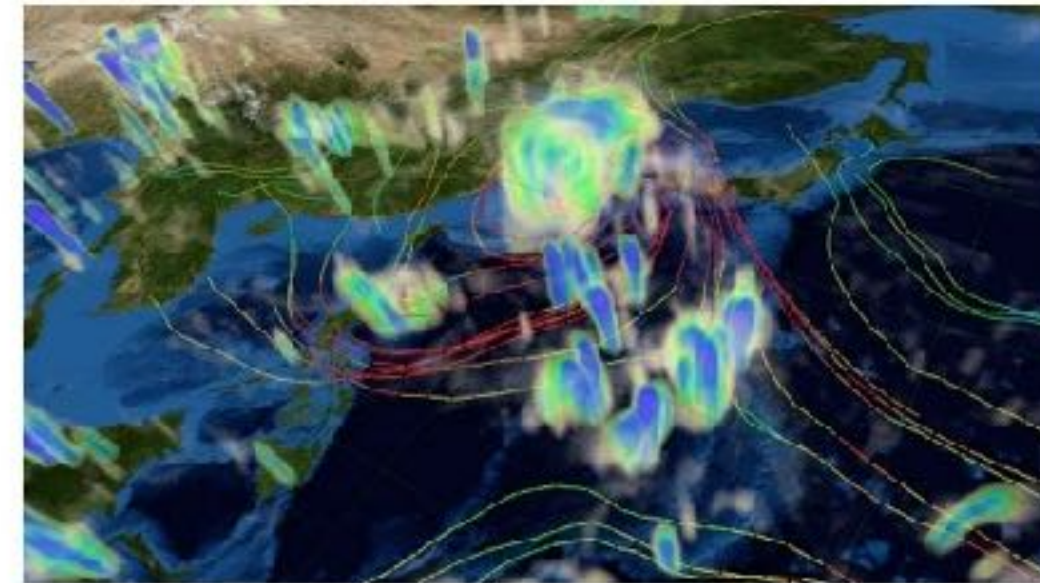
気象シミュレーションと空間解像度

Typhoon Bolaven 2012

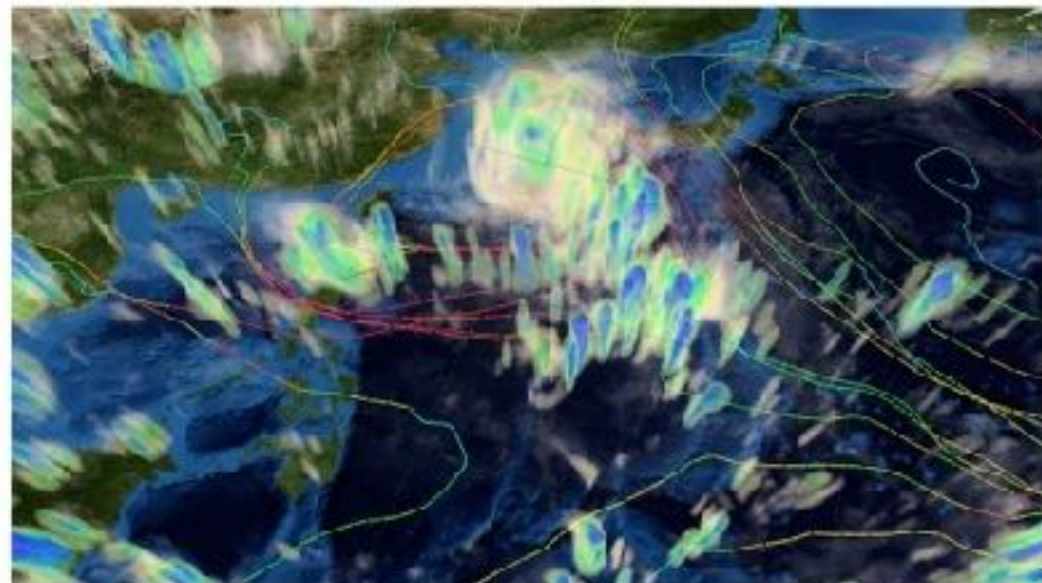
28kmメッシュ



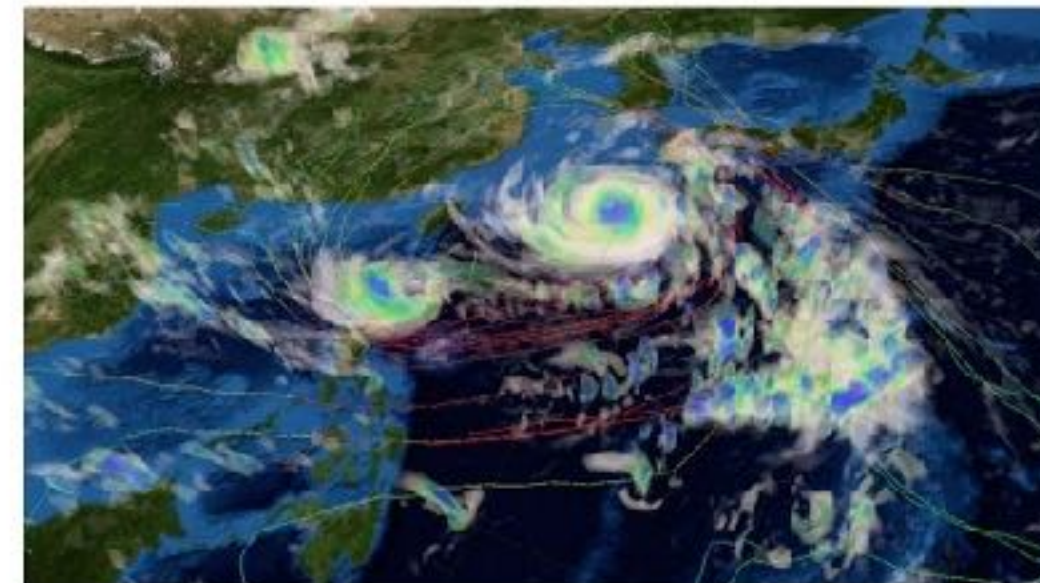
14kmメッシュ



7kmメッシュ



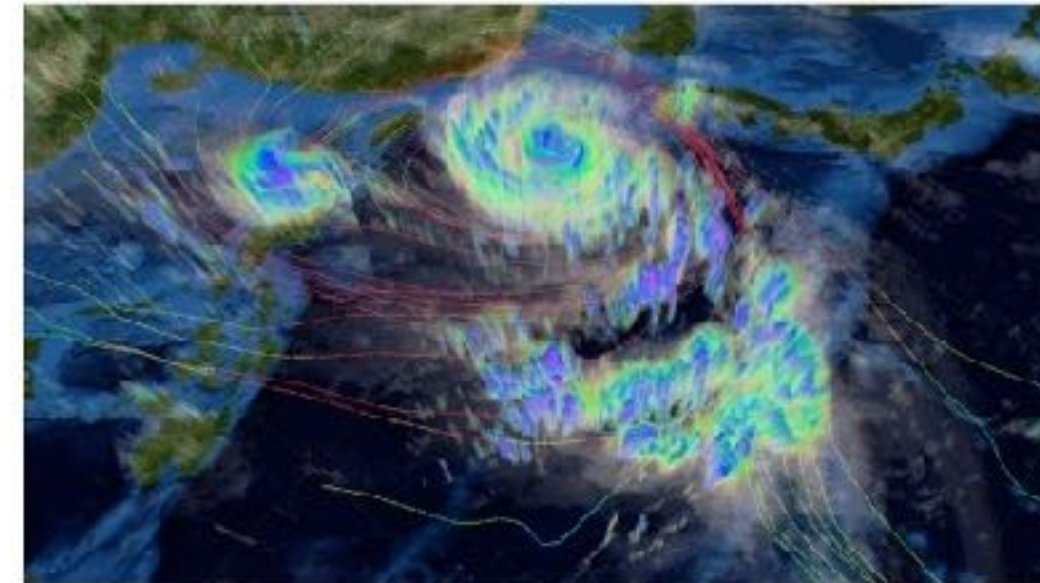
3.5kmメッシュ



1.7kmメッシュ



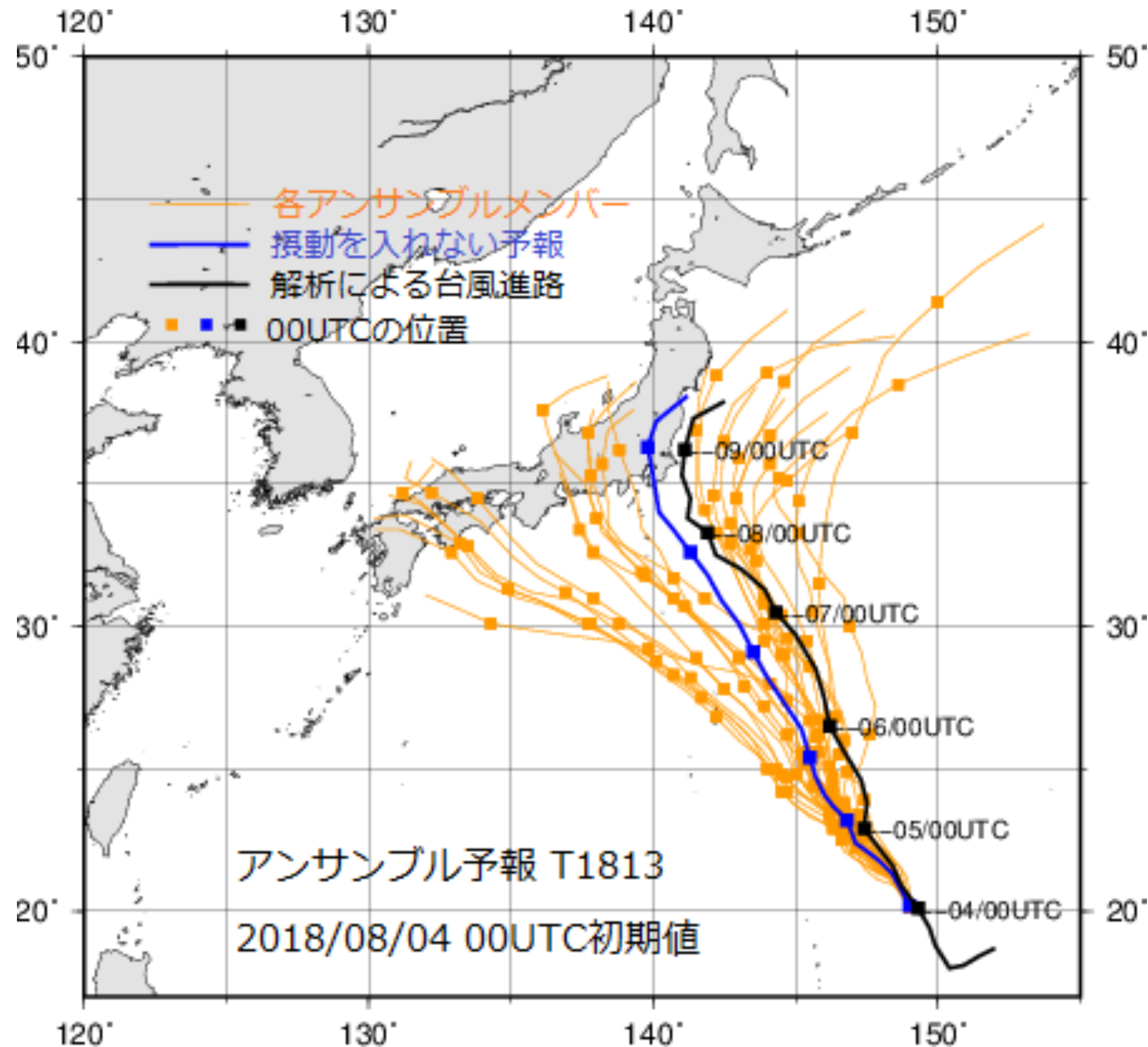
870mメッシュ



- 気象シミュレーションは地球の大気を格子等で分割して計算している
- 地球の大気は大きいいため、全てをカバーしようとする点の間隔が粗くなり、再現・予測精度が落ちる
- 空間解像度が2倍→格子点の数が4倍→計算量は8倍

「京」コンピュータでの2012年台風15号のシミュレーション

アンサンブル計算



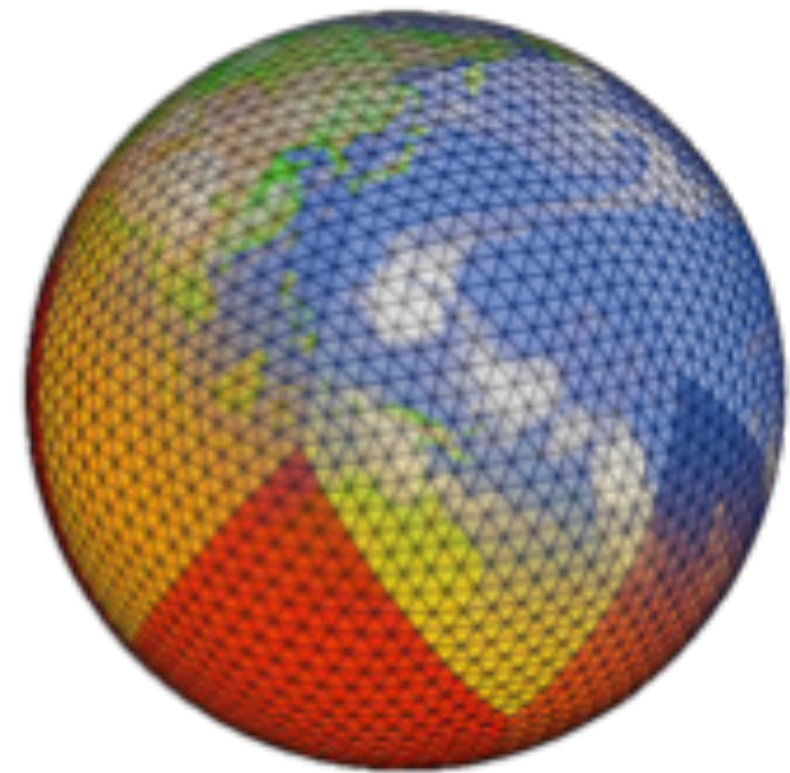
台風進路のアンサンブル計算の例、気象庁HPより

- 初期値や境界条件を少しずらした計算
- シミュレーションを進めると結果のズレがだんだん大きくなっていく
- 予測の信頼度・確率情報を得られる
- アンサンブル数を増やせば統計的により精度の高い情報が得られるが、計算量は増える

「富岳」の能力をもってすれば、どこまで解像度を高めて、アンサンブル数を増やすことが出来るのか？
気象計算の「グランドチャレンジ」

気象のデータ同化

現実の大気の情報
教えてもらわないとわからない



モデル
シミュレーション

高性能計算
(HPC)

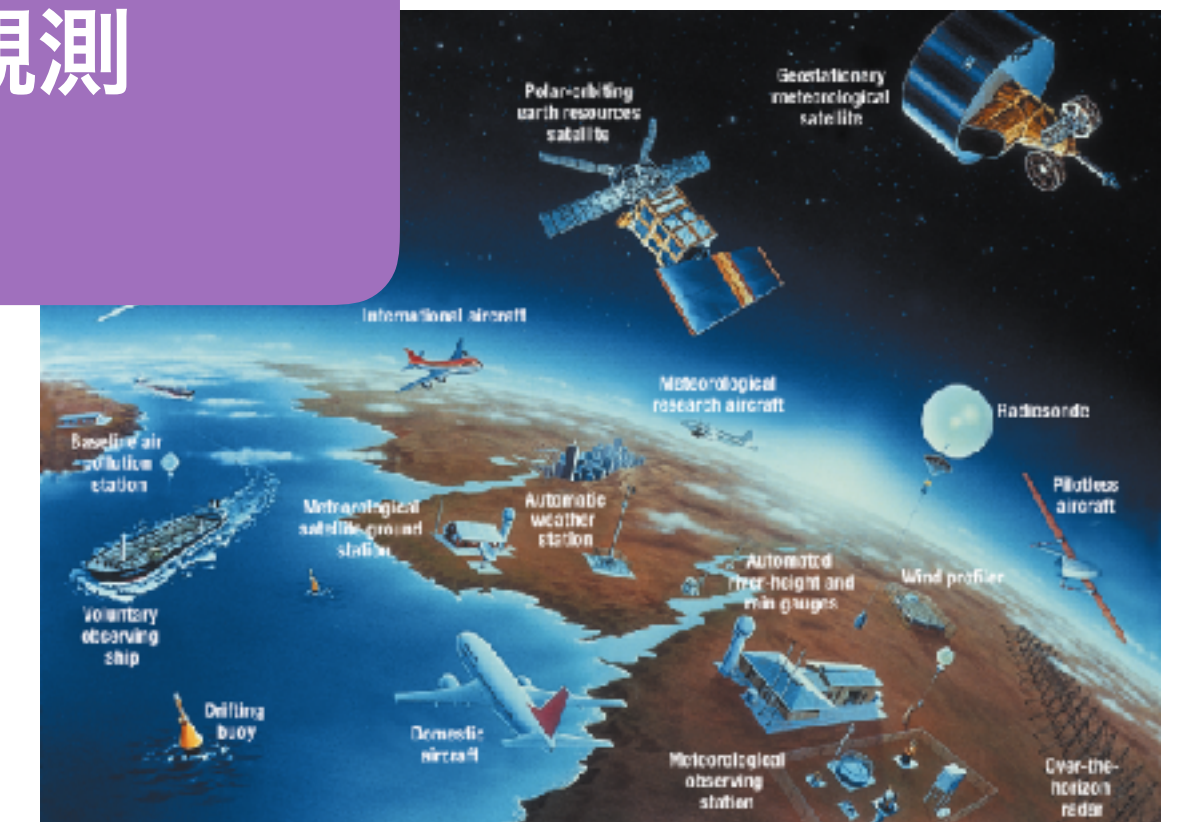
データ同化

高性能データ解析
(HPDA)

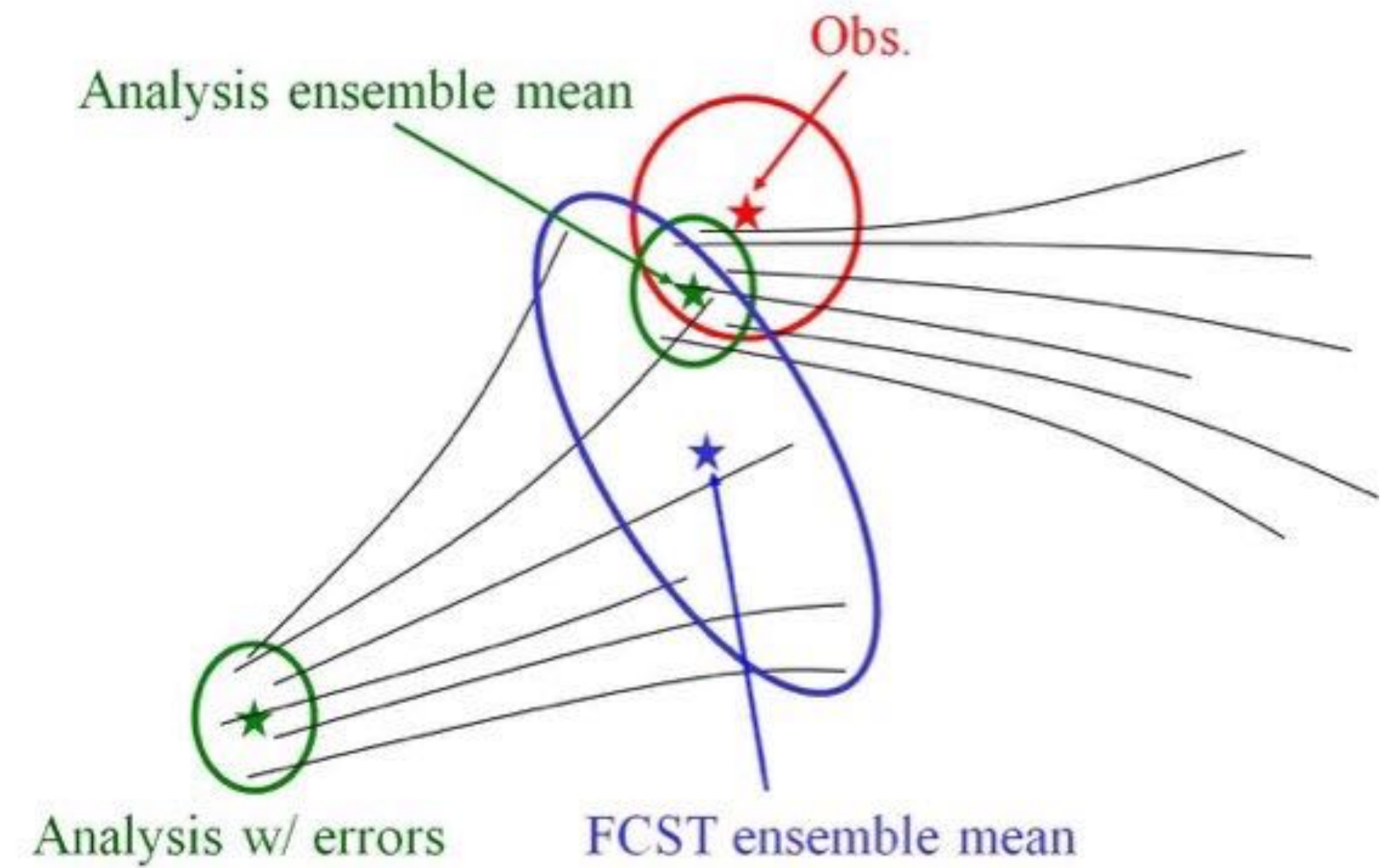
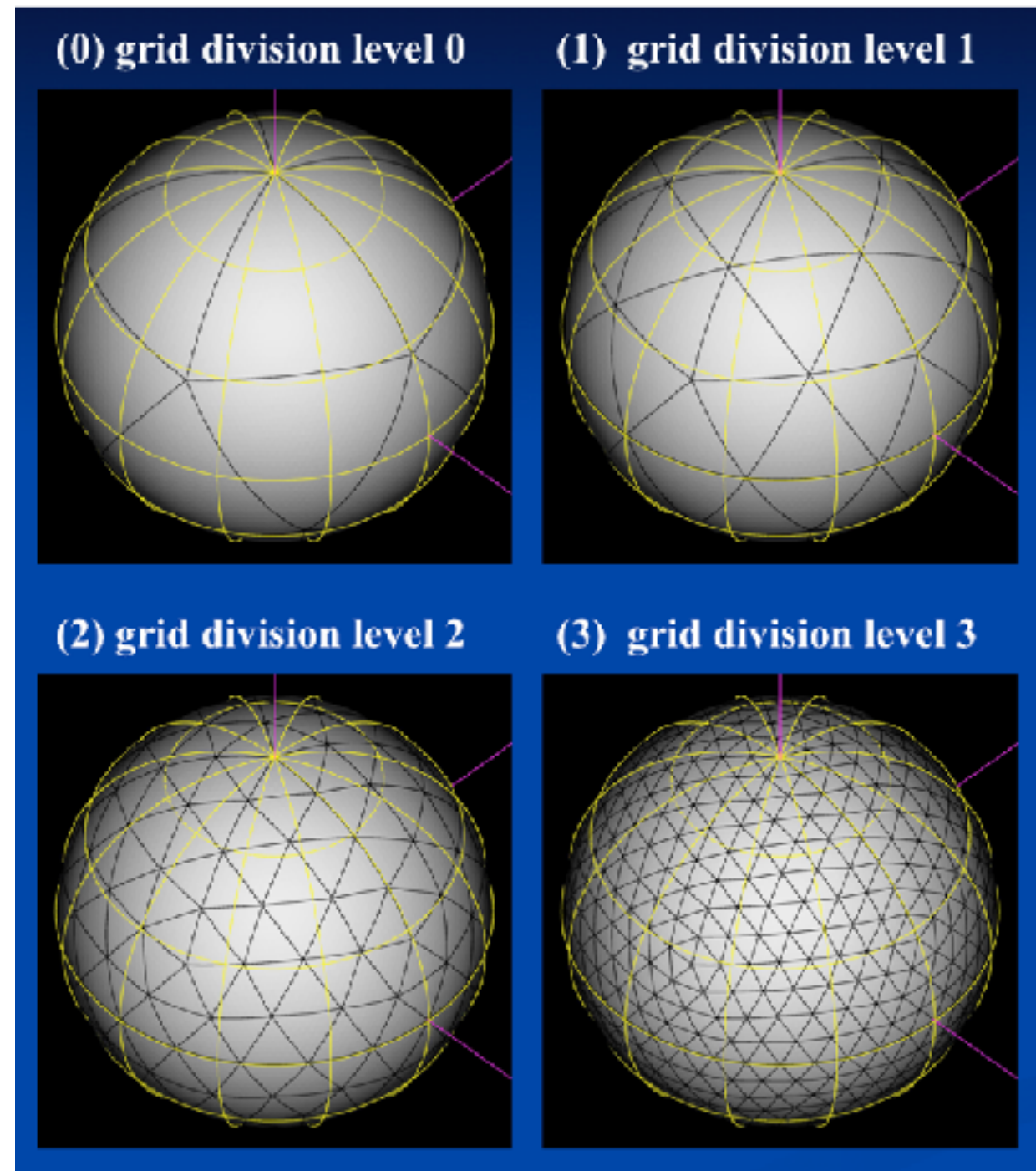
数学的な手法を用いて
観測とシミュレーションを融合！

地球上をくまなく
観測しているわけではない

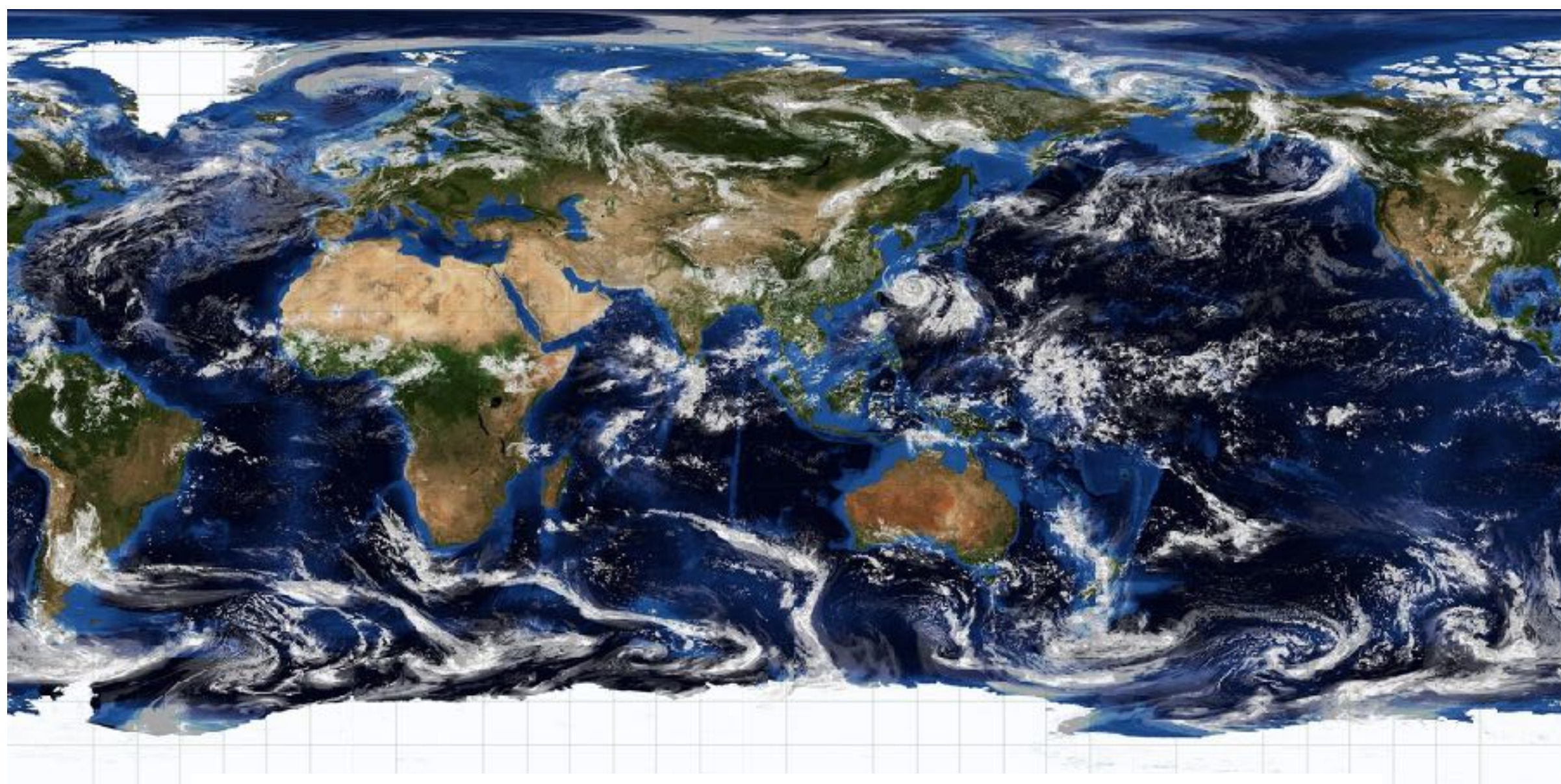
地球観測



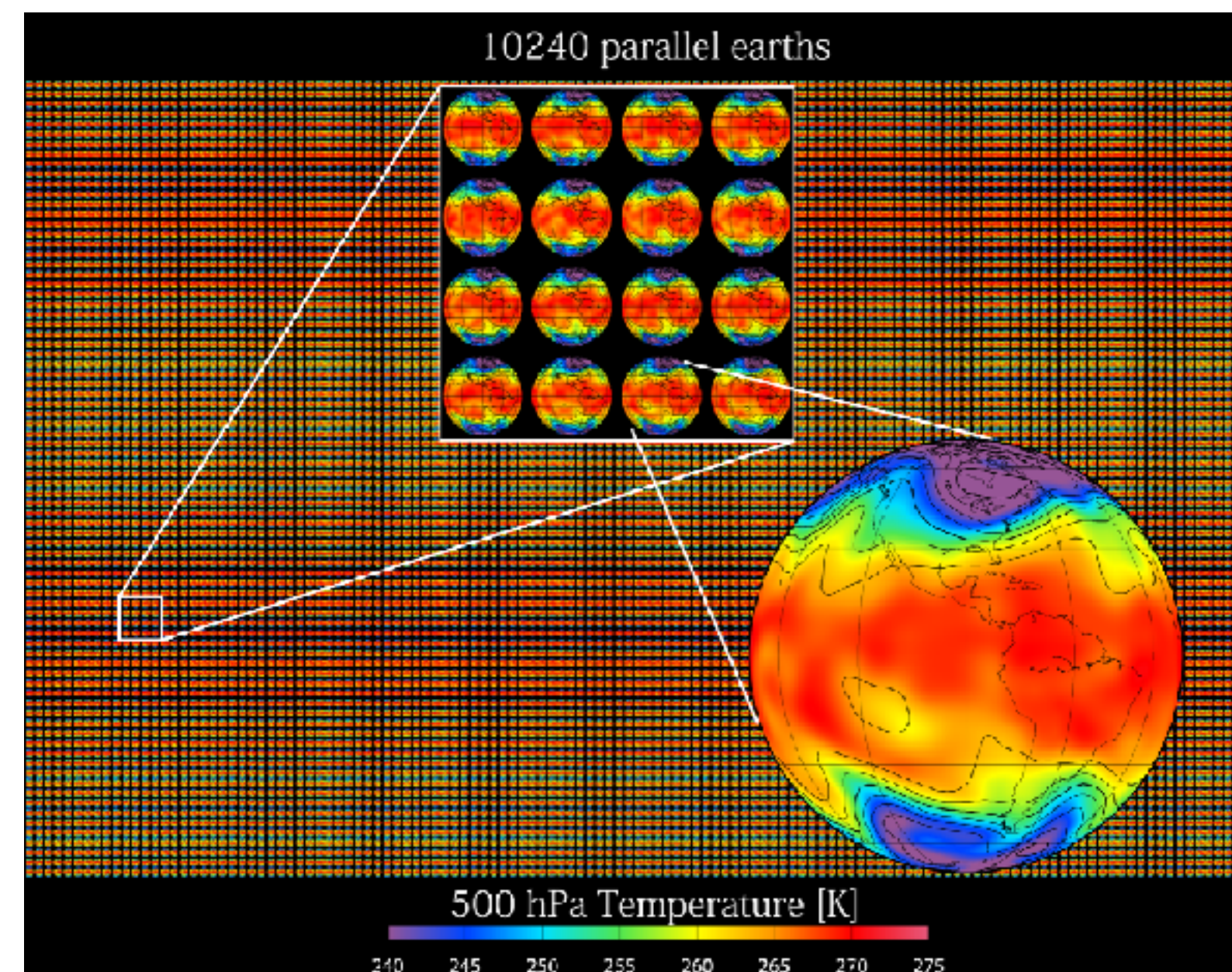
- 本研究では、気象予報計算の前段階として行なわれるデータ同化計算の性能を検証した
- アンサンブルデータ同化は、いくつかあるデータ同化手法の一つ
- 世界の気象機関が日々行っている同種の計算と比べて500倍以上の規模



- 正20面体非静力大気モデルNICAM
- 入道雲ひとつひとつまで計算する超高解像度計算を実現する次世代大気モデル
- 局所アンサンブル変換カルマンフィルタLETKF
- 世界の様々な天気予報システムで使われている優れた手法



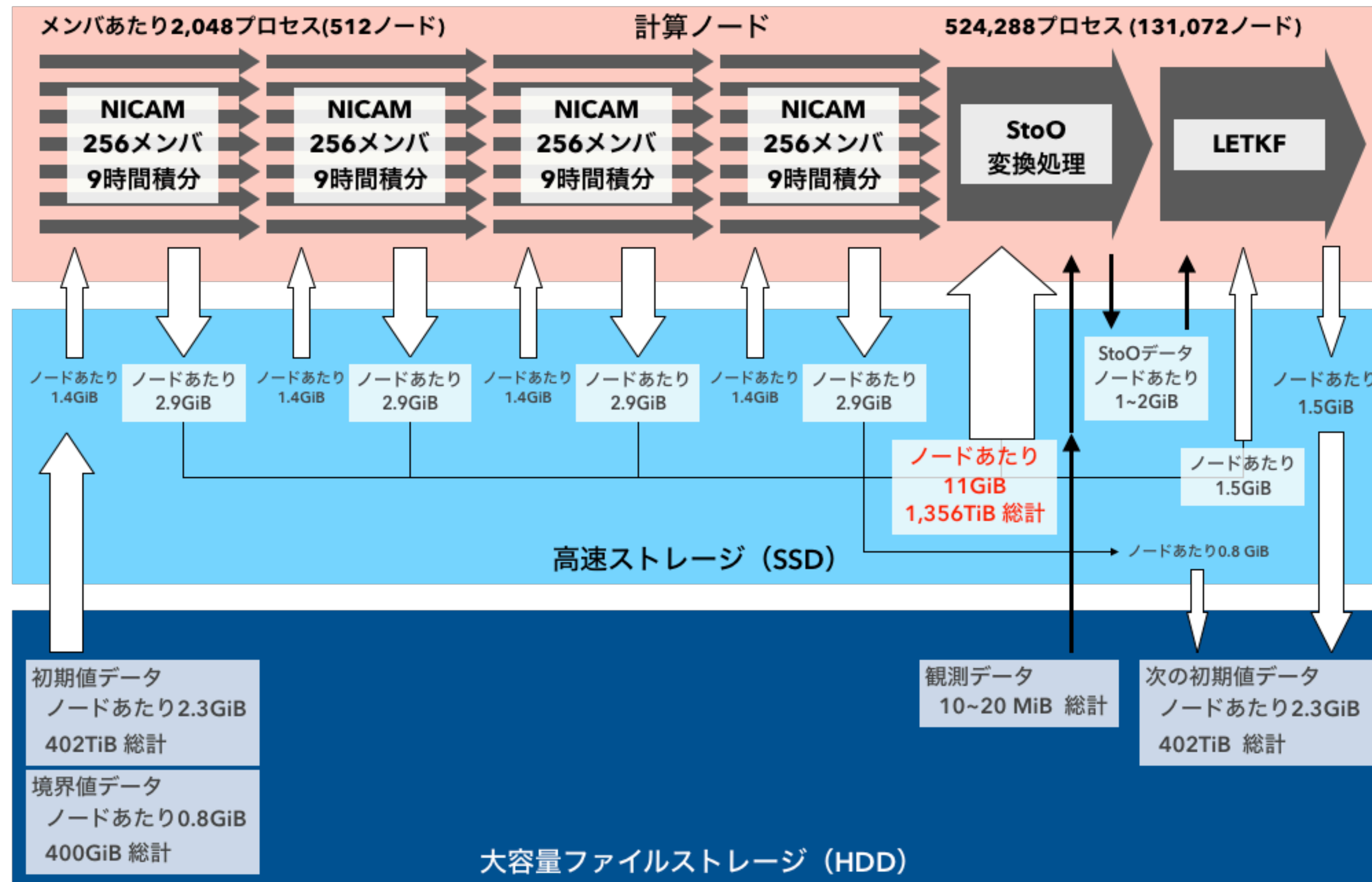
「京」での全球870mシミュレーション



「京」での1万アンサンブルデータ同化実験

それぞれスーパーコンピュータ「京」でこれまでにない規模の計算を行ってきた
2つのアプリケーションを組み合わせた実験

計算の流れ

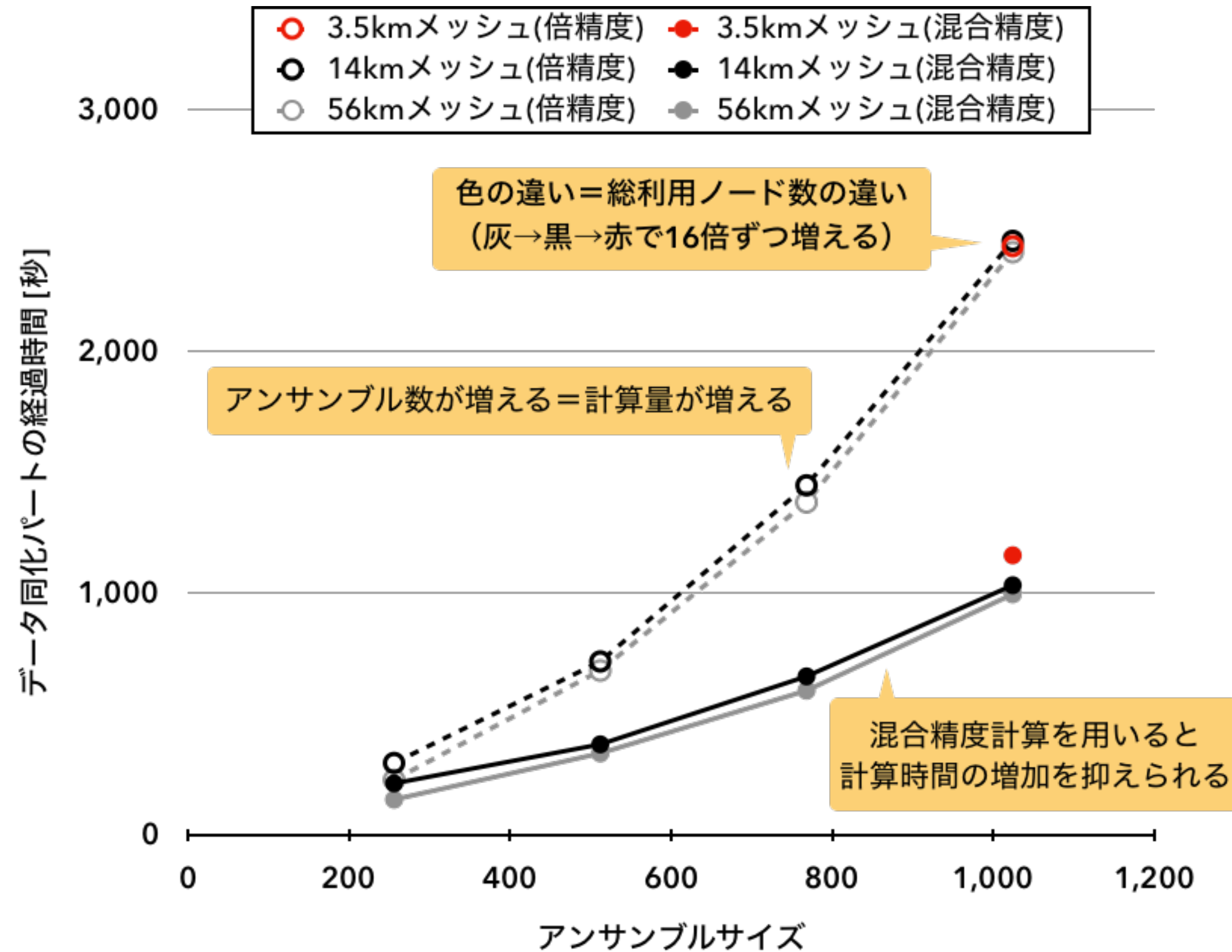


- 気象シミュレーション (NICAM) の計算→データ同化 (LETKF) の計算→繰り返し
- シミュレーションが計算した1.3ペタバイトのデータを利用してデータ同化を行う
- ファイル入出力にかかる時間を最短にするような設計の変更→ふたつのソフトウェア間のコデザイン

様々な改良とコデザイン

- キーワードは「**データの移動**」
- 富岳のハードウェア、ソフトウェアとのコデザイン
 - ：強化された**データ転送性能**は首位を獲ったベンチマークの結果にも表れている
- 浮動小数点精度を落とした**混合精度計算**を積極的に利用
 - 各データのバイト数が減ると、移動も演算も高速に
 - 計算結果が劣化しないか？→科学的要請と計算性能の間のコデザイン
- 数十万行のプログラムコードから、データ移動や待ち時間ばかりかかっている部分を効率よく見つけ出す手法の開発

データ同化部分の 実験結果



- 512ノード (5TB) →8,192ノード (80TB) →131,072ノード (1.3PB) と増やしていっても、計算時間をほとんど同じに抑えられた：ファイル入出力にかかる時間がほぼ同じ！
- 最大規模のシミュレーション+データ同化では、総格子点数は4兆点
- 131,072ノード (6,291,456計算コア、富岳の82%) を用いて、4時間弱で計算が終わる→「京」では2週間

まとめ



- 「富岳」で史上最大の気象計算を実現
- 単純な計算の繰り返しではない、実用的で複雑な科学計算を、富岳全体を使って実現した
→京で2週間かかる計算を4時間に短縮し、「アプリケーション性能で100倍」
- 1.3ペタバイトものデータを気象モデルとデータ同化システムがやりとりする
- 将来の気象予報の精度向上のためには、計算の大規模化は避けて通れない
 - データの移動量、転送性能を重視したスパコンとソフトウェアの開発を行えば、現在の**500倍以上**の大規模化が可能であることを示した
 - スパコン、シミュレーション科学、データ科学の各分野が密接に連携して開発する = **コデザイン**が、性能向上にとっても重要であることを示した

• 発表論文

- H. Yashiro, K. Terasaki, Y. Kawai, S. Kudo, T. Miyoshi, T. Imamura, K. Minami, H. Inoue, T. Nishiki, T. Saji, M. Satoh, and H. Tomita, "A 1024-Member Ensemble Data Assimilation with 3.5-Km Mesh Global Weather Simulations," in SC20: International Conference for High Performance Computing, Networking, Storage and Analysis (SC), Atlanta, GA, US, 2020 pp. 1-10. doi: 10.1109/SC41405.2020.00005

• 研究グループ概要

- NICAM、データ同化システムの開発、実験の実施と全体のとりまとめ：八代 尚（国立環境研/理研R-CCS）
- NICAMの開発と最適化：佐藤 正樹（東大大気海洋研）、富田 浩文（理研R-CCS）、河合 佑太（理研R-CCS）
- データ同化システムの開発と最適化：三好 建正（理研R-CCS）、寺崎 康児（理研R-CCS）
- 数値計算ライブラリの開発と最適化：今村 俊幸（理研R-CCS）、工藤 周平（理研R-CCS）
- NICAM、データ同化システムの富岳への最適化：南 一生（理研R-CCS）、井上 晃（富士通）、錦 龍生（富士通）、佐治 隆行（メトロ）

• 謝辞

- 中野満寿男 様（JAMSTEC）、小玉知央 様（JAMSTEC）、NICAM開発者の皆様、LETKFを用いたデータ同化システム開発者の皆様
- 理研フラッグシップ2020プロジェクトの皆様、富士通と関係各社の皆様
- 本研究は、文部科学省スーパーコンピュータ「富岳」成果創出加速プログラム「防災・減災に資する新時代の大アンサンブル気象・大気環境予測」の一環として実施したものです。また、本研究の一部は、スーパーコンピュータ「富岳」の計算資源の提供を受け、実施しました。