

超重量級計算はこの先どうなるのか？：
気象・気候シミュレーション研究が期待するもの

八代 尚 (国立環境研究所)

第22回PCクラスタシンポジウム

2022/12/6

自己紹介

- 八代 尚 (Hisashi Yashiro)

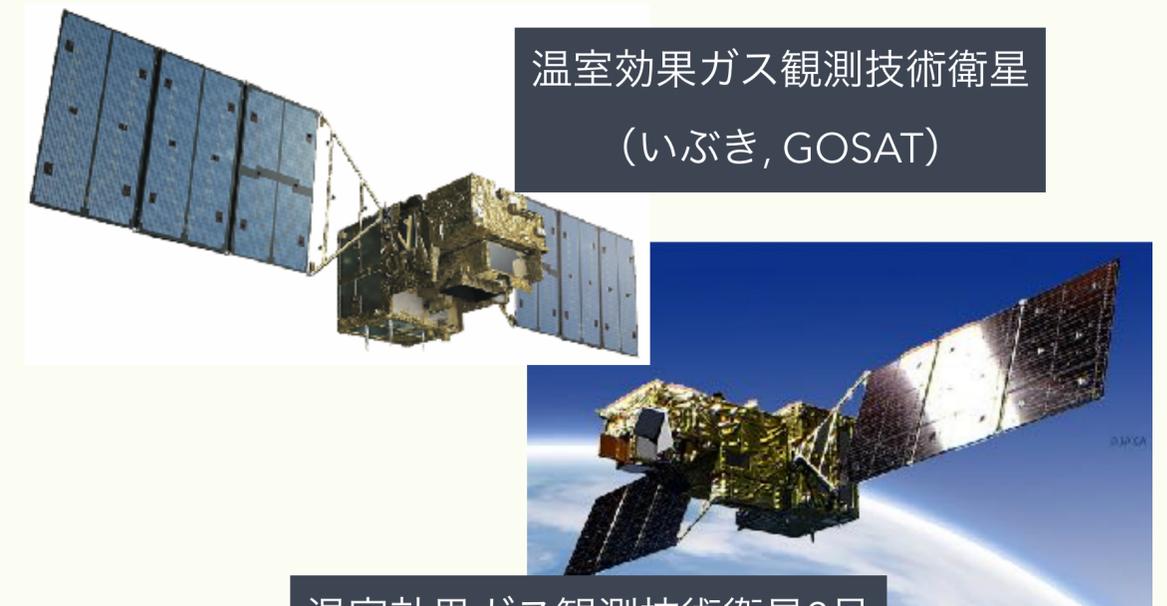
- 1978年生：青森県弘前市
- 専門：大気化学、高性能計算、大気科学
- 海洋研究開発機構→理研R-CCS(AICS)→国立環境研究所
- 温室効果ガス観測衛星「GOSAT」プロジェクトが現在の本務



スバルバル島ニーオルスン基地(北緯79度)



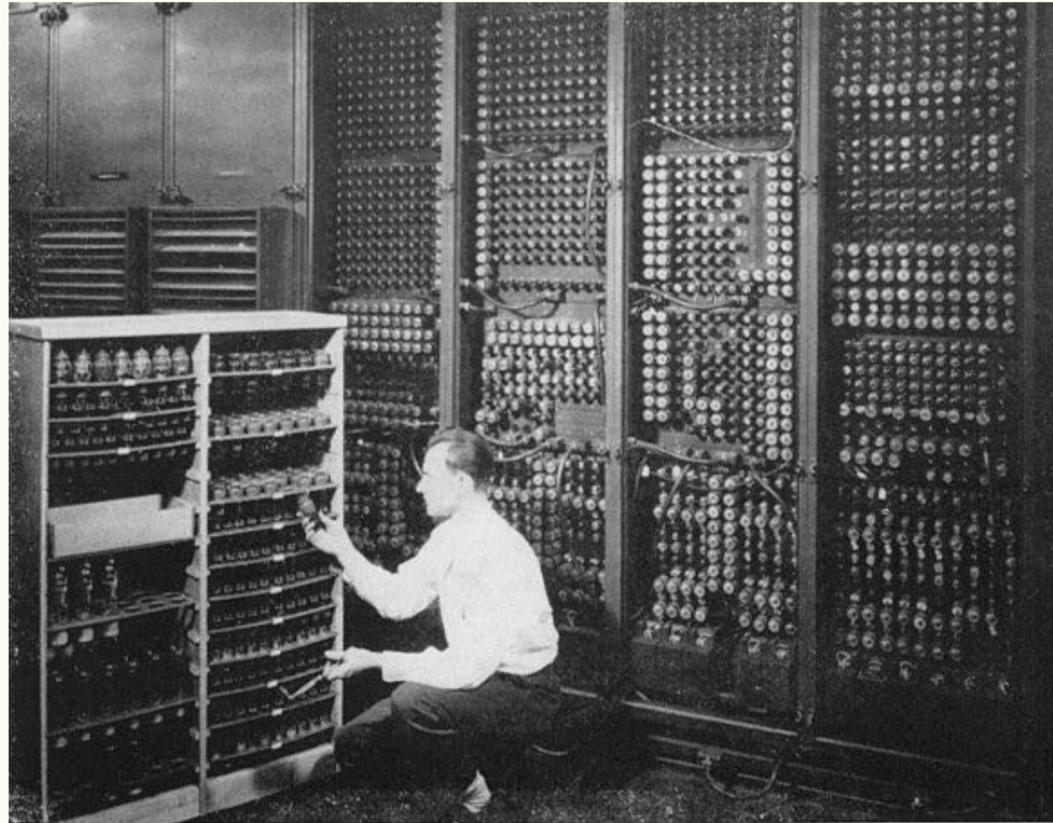
JAXA三陸大気球観測所 (岩手県大船渡市)



温室効果ガス観測技術衛星
(いぶき, GOSAT)

温室効果ガス観測技術衛星2号
(いぶき2号, GOSAT-2)

気象の数値シミュレーションと計算機



Replacing a bad tube meant checking among ENIAC's 19,000 possibilities.

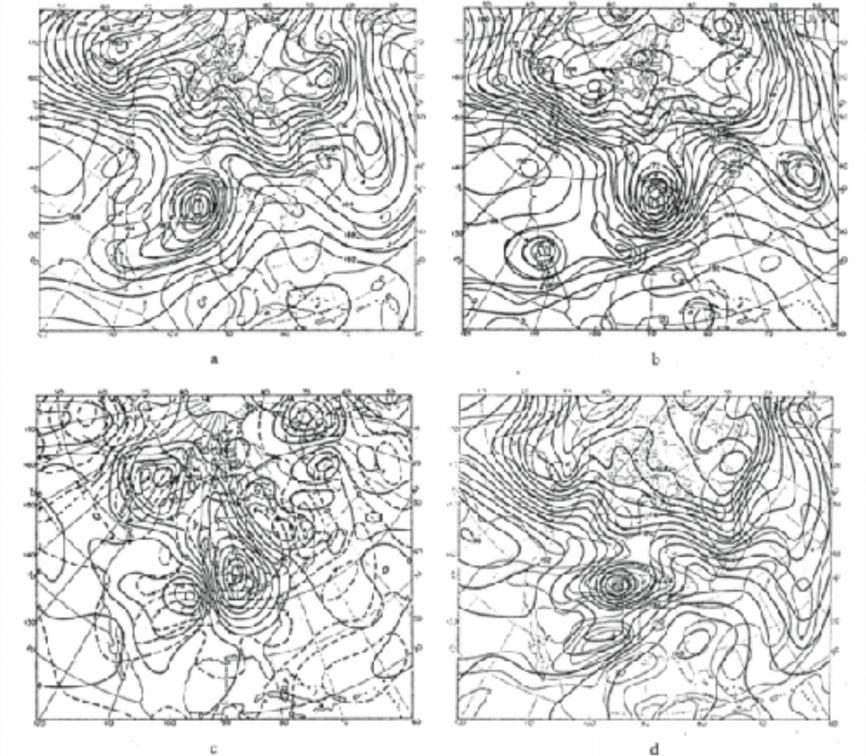
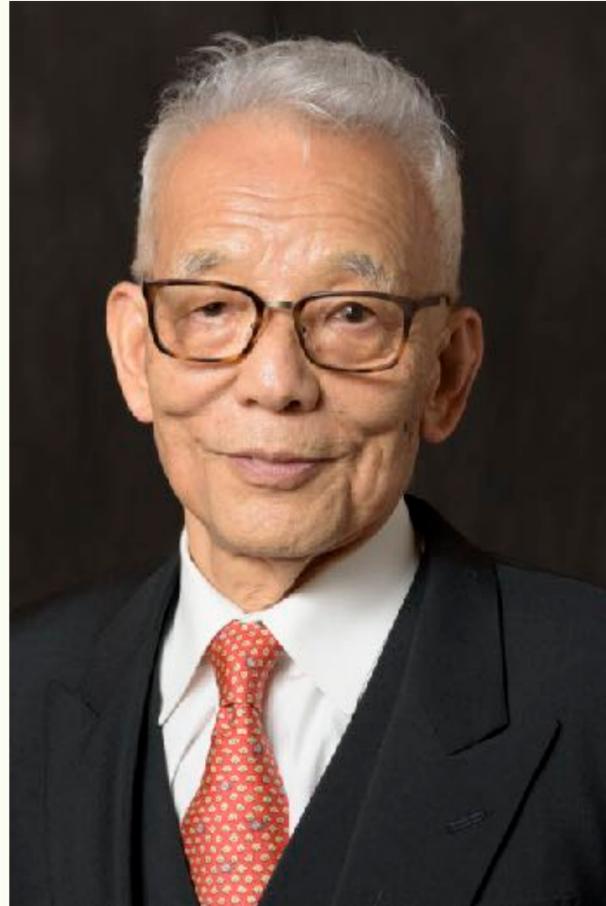


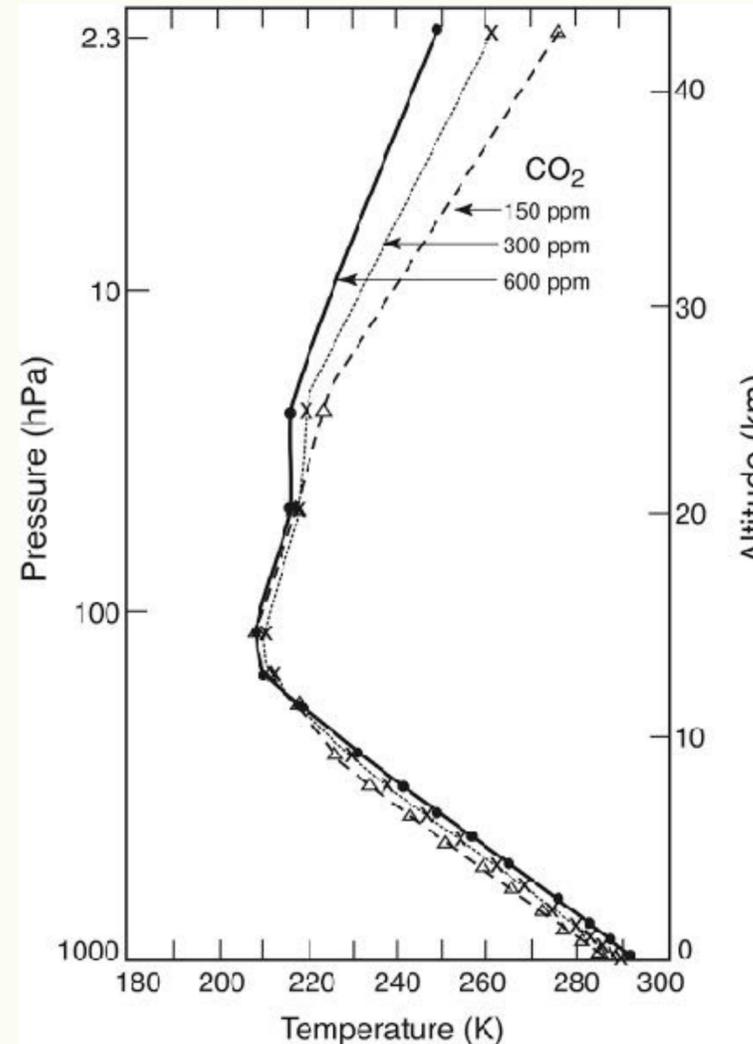
FIG. 3. Forecast of CFvN from 5 Jan 1949. (a) Analysis of 500-hPa geopotential height (thick lines) and absolute vorticity (thin lines) for 0300 UTC 5 Jan. (b) Analysis for 0300 UTC 5 Jan. (c) Observed change (solid) and forecast change (dashed). (d) Forecast height and vorticity valid at 0300 UTC 6 Jan (from CFvN). Height units are hundreds of feet, contour interval is 200 ft. Vorticity units and contour interval ($10^{-4} s^{-1}$).

ENIACによる数値気象シミュレーション（1950）から連綿と続く歴史

気候の数値シミュレーションと計算機



©Novel Prize Outreach



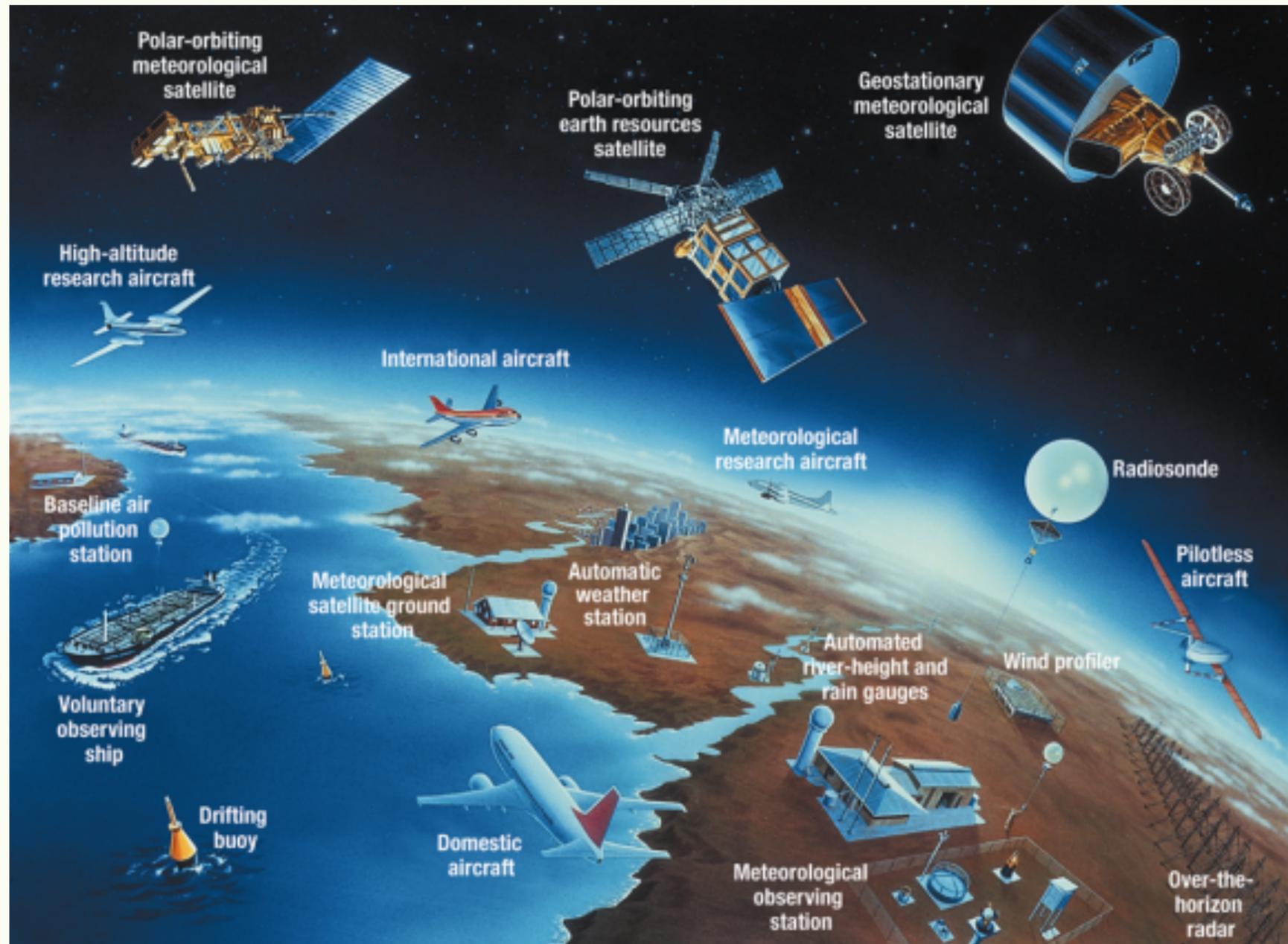
from Manabe and Wetherald (1967)



IBM 7030 stretch (from Wikimedia commons)

真鍋先生のノーベル物理学賞受賞は、地球物理学分野だけでなく、シミュレーションを用いた計算科学の受賞という意味でも大きな意義を持つ

気象分野のビッグデータ利用



from WMO

- 気象学者はビッグデータのユーザとしてもかなりの古株
- 時々刻々、様々なセンサによる観測結果が収集され、気象予報に役立てられている
- センシング→シミュレーション・解析→社会への情報提供という流れは、Society5.0のやろうとしていることそのもの

気象・気候アプリケーションとは？

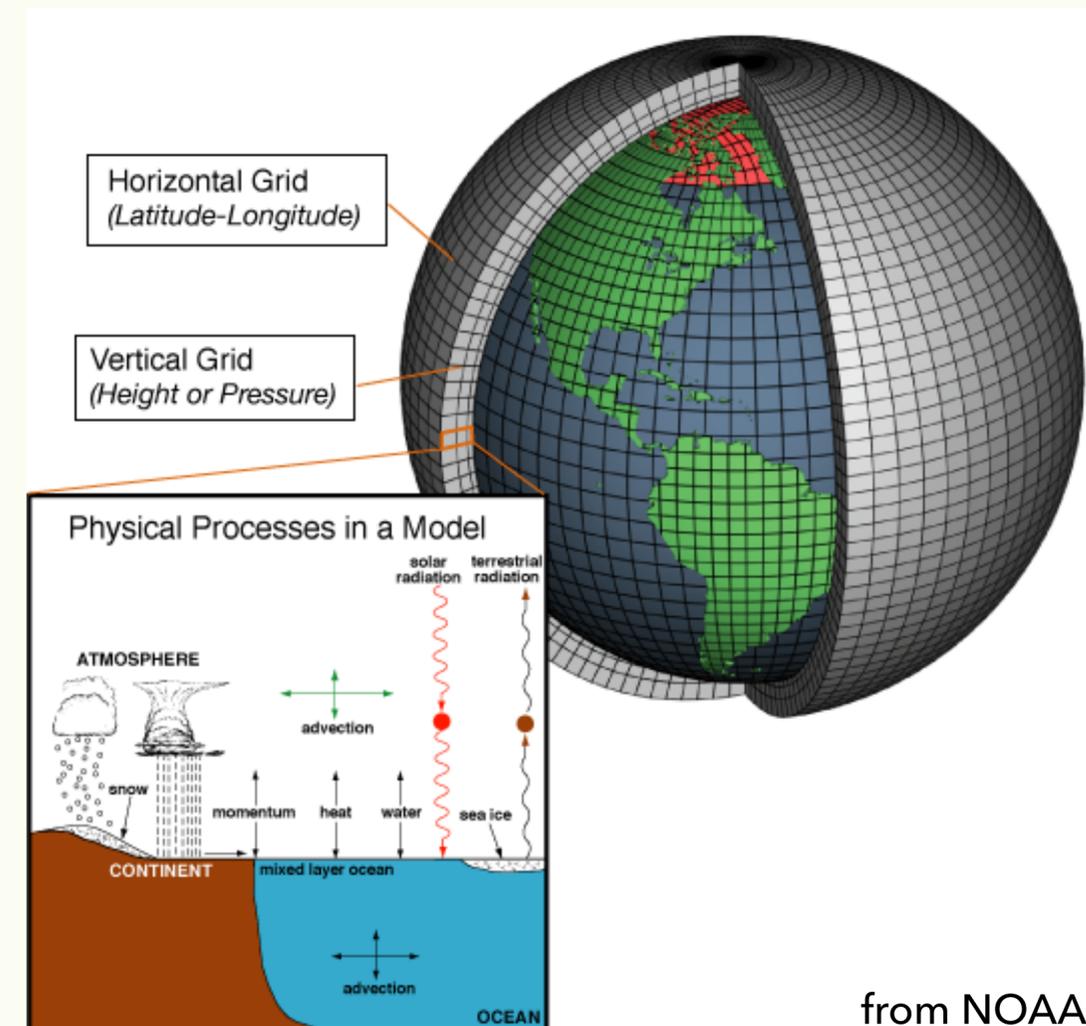
気象シミュレーション：数時間～数ヶ月

- 大気の状態を予測するのが主
- 初期値問題、データ同化の利用
- 解像度重視、リアルタイム性重視

気候シミュレーション：数年～数百年

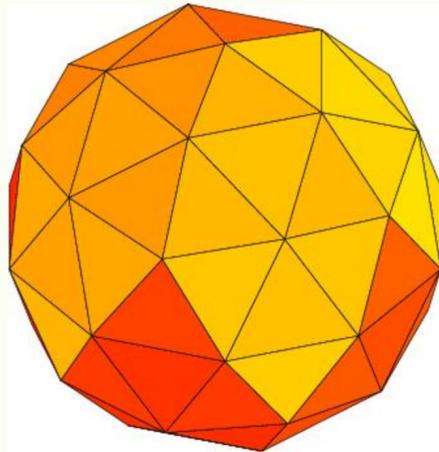
- 大気＋海洋＋海氷＋陸上生態系を含めた再現と予測
- 境界値問題
- 計算速度重視

- 気象計算は第一原理計算とは程遠い空間スケールで解いている：解くべき系の大きさが大きすぎるため
- 流体計算は一部に過ぎない：様々な物理的過程を解いている

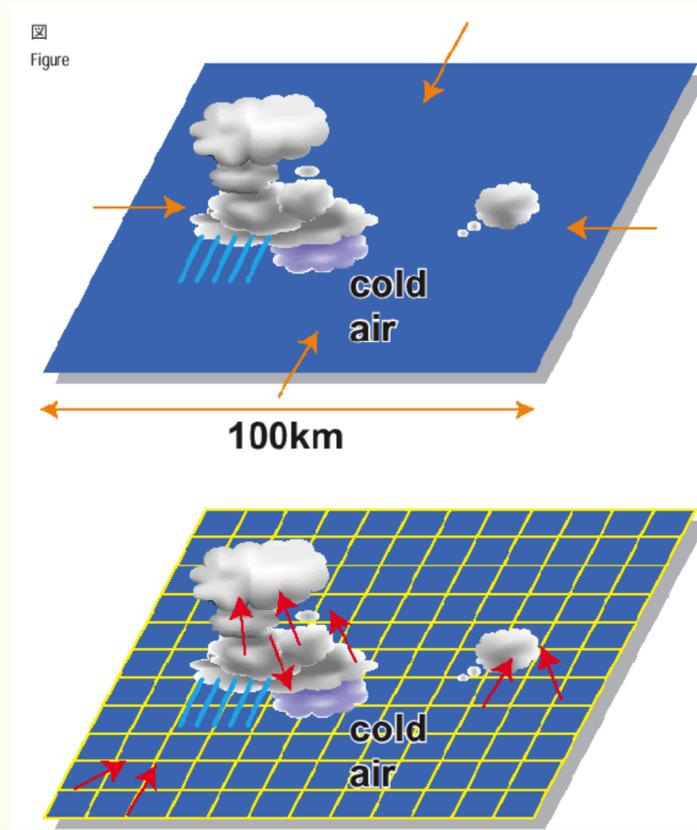
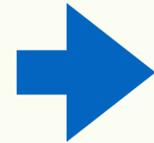


空間解像度によって変わるソルバー

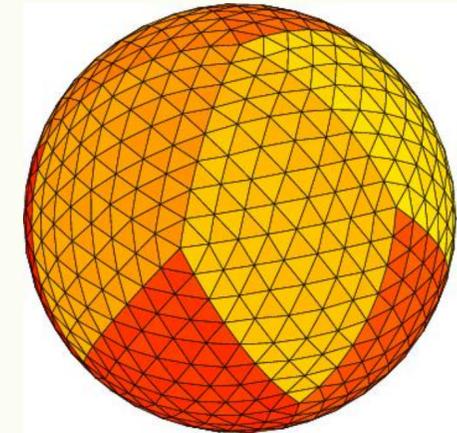
低解像度大気モデル



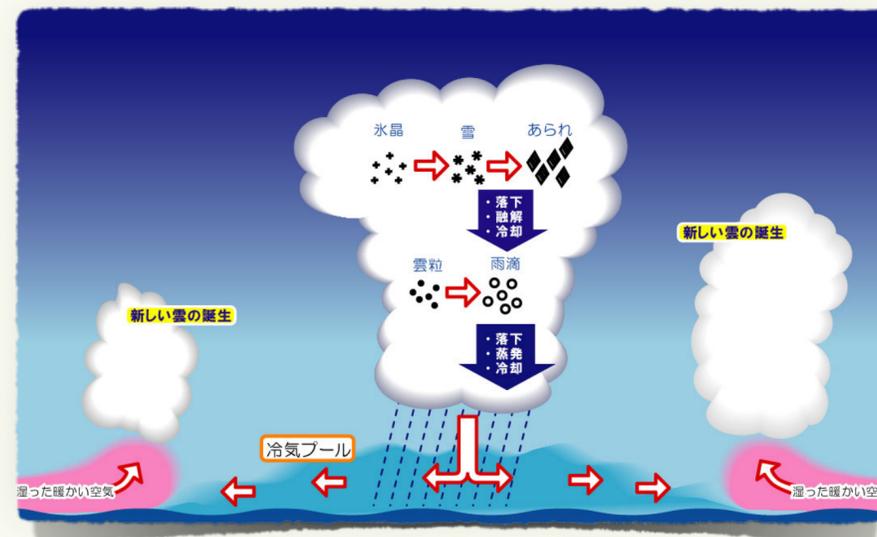
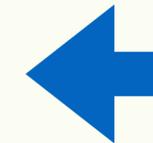
- 雲域の形成と積雲による対流を巨視的にモデル化
- 集中豪雨の空間スケールに対して大きすぎる
: 強い降水量を再現できない
- 計算量が少なく軽量



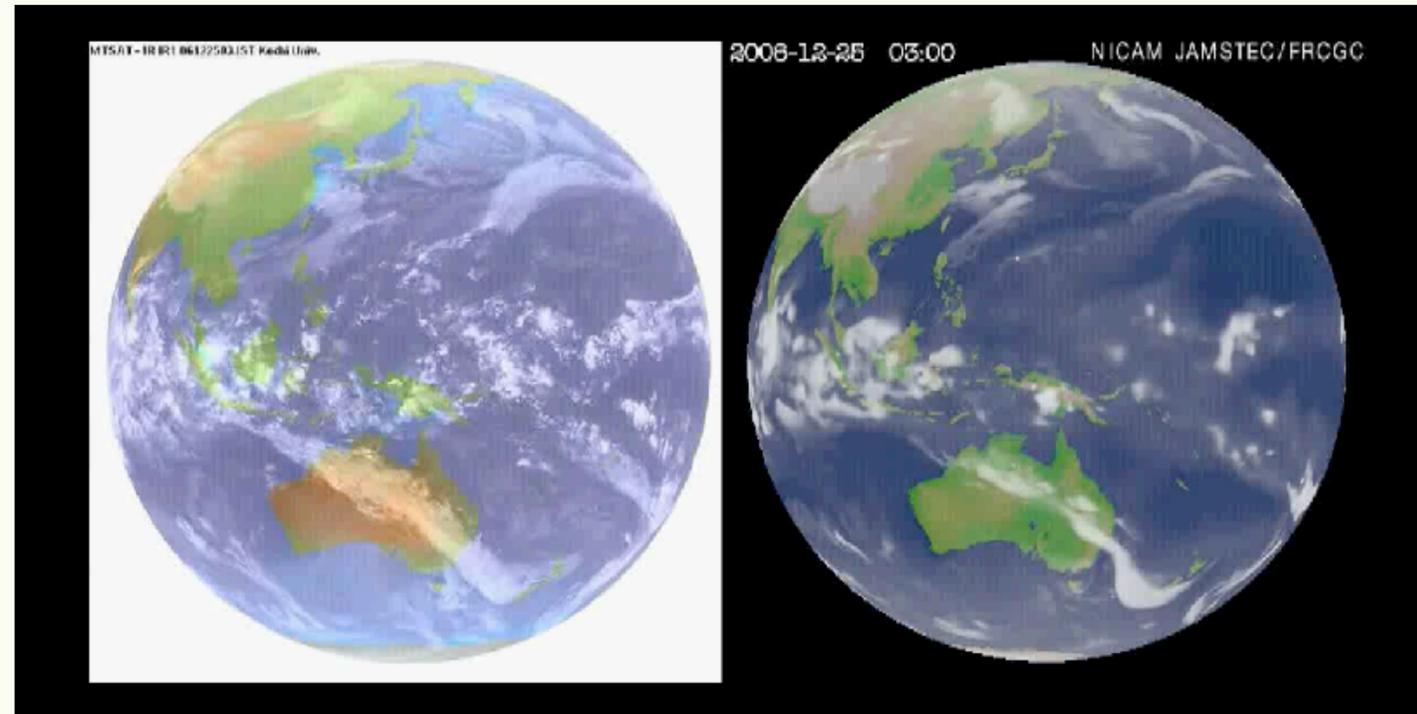
高解像度大気モデル



- 水蒸気・雲水・雨・雲氷・雪・あられ・ひょう等のカテゴリ遷移を陽に表現、鉛直対流も陽に扱う
- 局所的な気象現象をよりよく再現
- 計算量が多い（解像度2倍で計算量8倍）



全球高解像度シミュレーションによる飛躍的な再現性の向上（1）



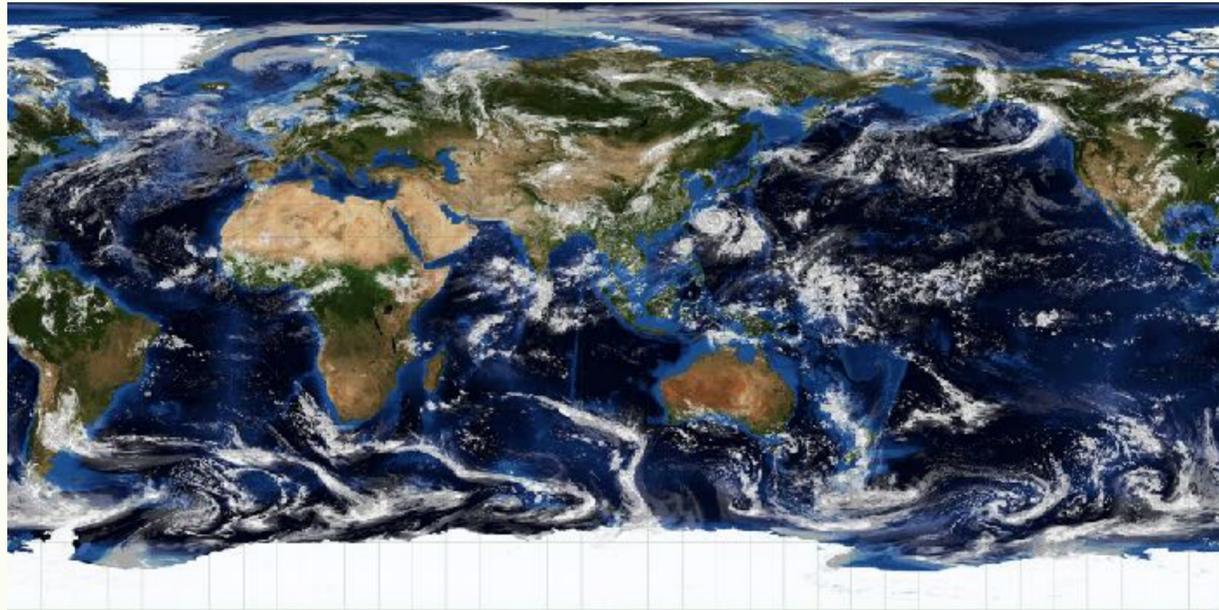
Miura et al. (2007) Science

- 世界初の3.5km全球実験
 - 地球シミュレータ（初代）320ノード
 - 水平4300万格子点、3次元18億格子点
 - 1週間積分（4万ステップ）

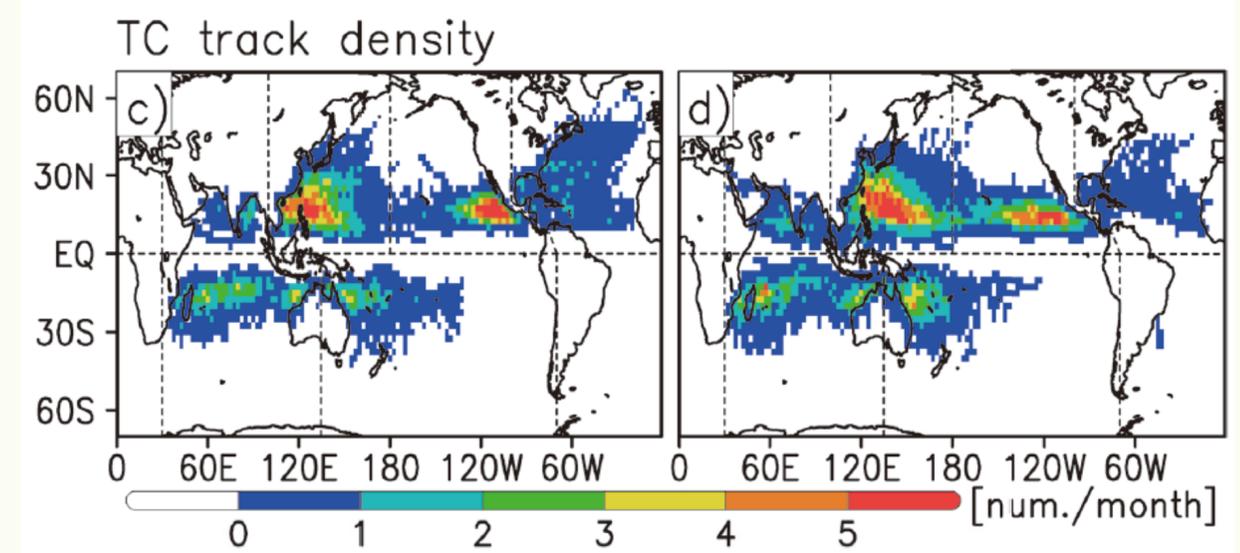


**Earth Simulator
(2002-2009)**

全球高解像度シミュレーションによる飛躍的な再現性の向上 (2)



Miyamoto et al. (2013) GRL



Kodama et al. (2015) JMSJ

- 世界初の870m全球実験
 - 京コンピュータ20480ノード
 - 水平7億格子点、3次元660億格子点
 - 2日積分 (8.6万ステップ)



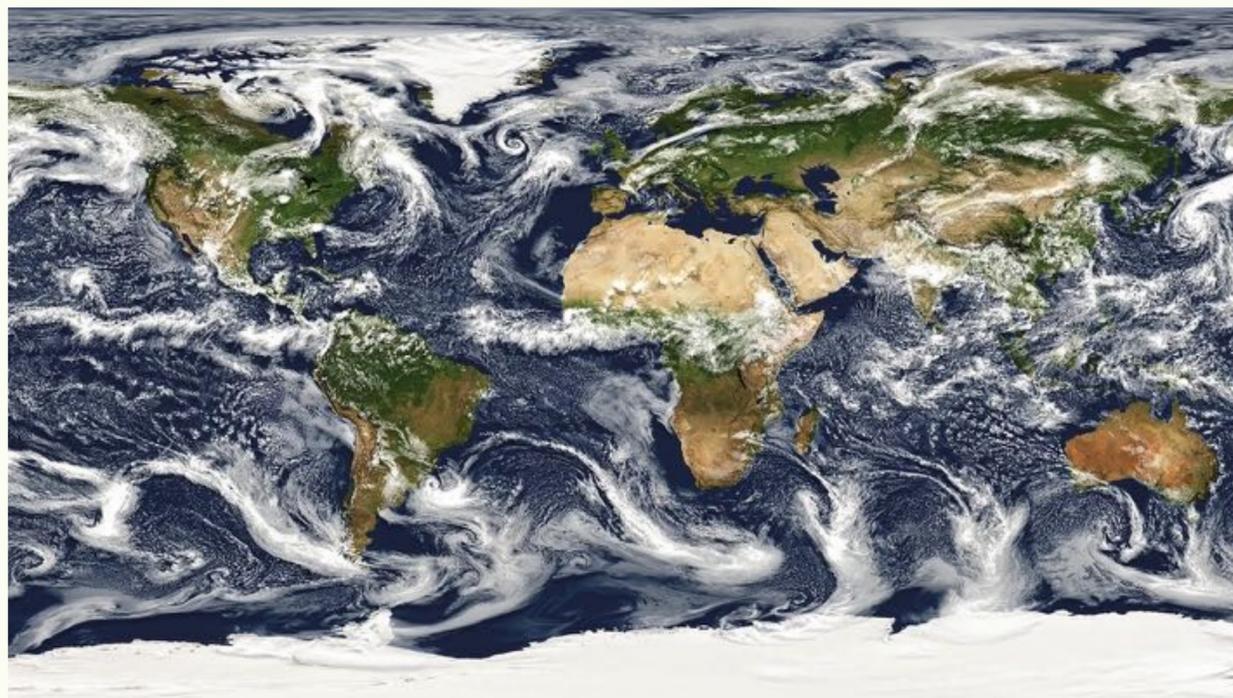
The K computer
(2011-2019)



©RIKEN

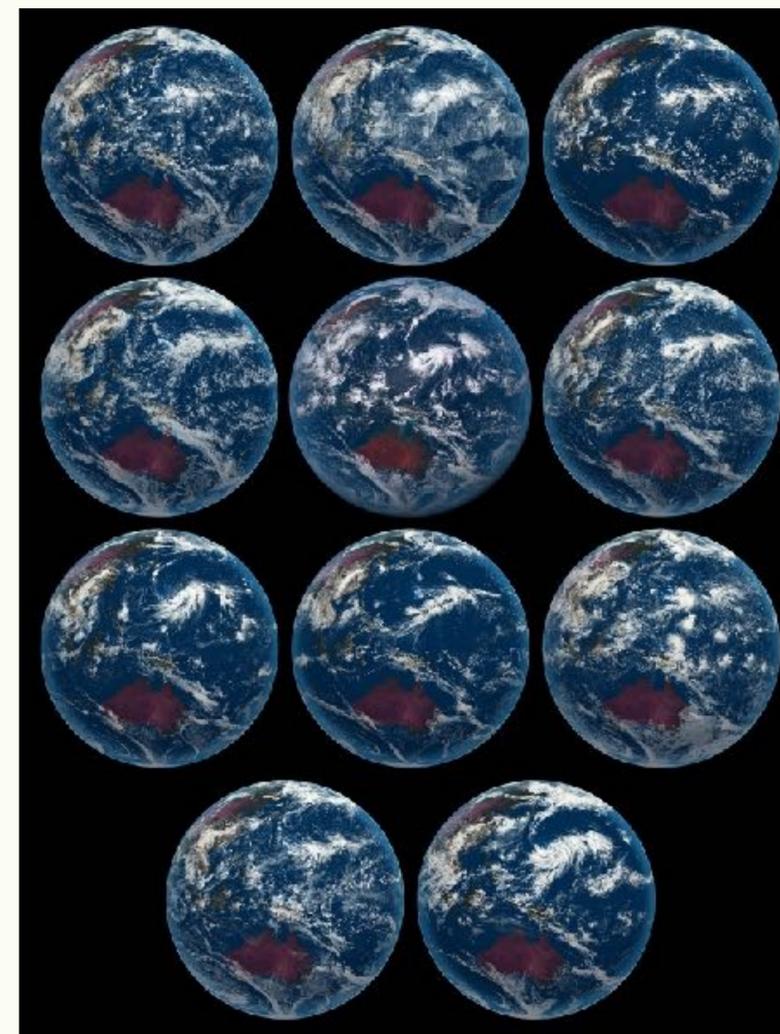
- 世界初の14km全球気候実験
 - 京コンピュータ640ノード
 - 水平280万格子点、3次元1億格子点
 - 30年間積分 (1600万ステップ)

世界的な動向



Slingo et al. (2022) Nature Climate Change

- km-スケールの気候シミュレーションが、気候予測精度向上にとって大きな飛躍となる点について強調
- km-スケールが見えてきたのは計算機性能の向上によるものが大きいですが、この先は計算機科学分野とより強い連携をとらなければ実現は難しい



Stevens et al. (2019) PEPS

DYAMOND initiative:

全球嵐解像モデル (GSRM; $dx < 5\text{km}$)

による国際比較実験

全球1km気候実験への道：NICAMの場合

- 富岳での計測結果
 - 870mメッシュ, 94層, 1年積分 (混合精度)
 - 総浮動小数点演算量：23ZFLOP
 - 総メモリ転送量：51EByte
 - Application B/F：2.2
- 1/4系を用いた場合 (SYPD: Simulation-Year Per wall-clock Day)
 - 京 (20480node): 0.0006SYPD
 - 富岳 (32768node): 0.024SYPD
- 2030年には1SYPDを達成したい
 - 富岳の40倍
 - 大気だけではなく海洋やエアロゾルモデルも含めた気候シミュレーション
 - 京→富岳ではノードあたりメモリ性能16倍 x ノード数1.6倍 x 混合精度で1.6倍 = 40倍

全球1km気候実験への道：立ちはだかる壁

- 次の40倍をどう達成する？
 - メモリ性能のトレンドは1年で1.16倍程度：10年で4倍
 - ノード数を増加させる？
 - ストロングスケール問題になり、性能劣化の影響が大きい
 - 総電力消費量をより増やす
 - 演算密度の高いアルゴリズムに切り替える？
 - 考慮する変数が多い中、頑張っても時間があまり減らない
 - ステンシル計算は全体の一部でしかない
 - より低精度の演算を活用する？
 - 継続的な効果が見込める一方、労力が大きい

気象・気候シミュレーションモデルの抱える問題（1）

- データ律速のアルゴリズム

- 演算密度の低い、多数のカーネル：Flat Profile
- 考慮する変数が多い
- 時系列でファイル出力したい情報もたくさん
- 解像度の向上＝格子数の増加が表現力に直結する：さらに増えるデータ数

- データの移動＝電力消費

- 今のHPCは省エネルギーこそ至上命題
- 気候計算は演算加速器からファイルシステムまで、巨大なデータを行ったり来たり
：エネルギーを喰らう重量級の計算

これから10年の気象・気候シミュレーション高速化の方向性

「京」以前の時代

- 計算時間を定式化と離散化のアイデアで削減

「京」の時代

- 計算時間のムダの削減：Flat Profileの認識とループレベルのアルゴリズムクリーンアップ

「富岳」の時代

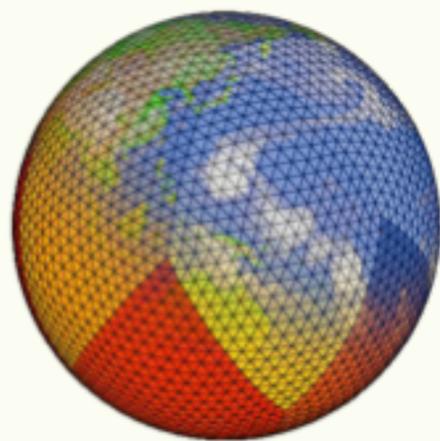
- データ転送時間の削減：データ中心のソフトウェア設計、低精度演算

「富岳NEXT」の時代

- もう削れるものがない！：複雑で時間のかかる物理モデルの利用そのものを削減する？

「富岳」でのグラウンドチャレンジ

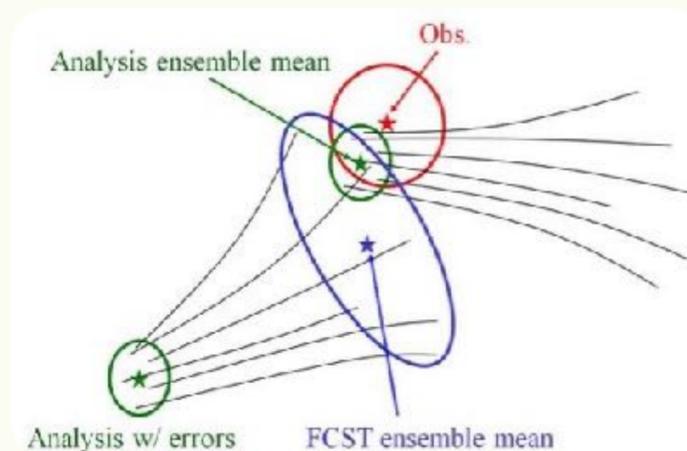
HPC



NICAM

(Global Storm Resolving Model)

+



LETKF

(Ensemble Data Assimilation System)

HPDA

- 気象予報で実運用に耐えるレベルでのベンチマークとして、データ同化サイクル実験を選択
 - カーネルのみの高速化では全く科学的成果に繋がらないという「京」での教訓
 - 今のスパコンを使えば、ここまで出来るという実現例を気象業界に示す必要があった
- アンサンブルデータ同化は単一のシミュレーションよりも、さらにデータ移動の多い実験設定
 - データ中心デザインの体現：混合精度化、分散・並列I/O、データ移動距離の最小化、等

→2020 Gordon bell Finalist paper (Yashiro et al.,2020)

Big News!

Specific Types of Contributions

ACM Gordon Bell Prize for Climate Modelling

Innovations in applying high-performance computing to climate modelling applications

Nominations

Committee Members

- 2023年より「気候モデリングのためのACMゴードン・ベル賞」
 - 気候科学と計算機科学との連携を促進する起爆剤となるだろう
 - 我々が頑張らないで誰がやる

気象・気候シミュレーションモデルの抱える問題（2）

- 性能ポータビリティの問題は根深い

- 計算機の多様化についていくのが大変
- 数十万行のソースコードを新しいアーキテクチャにどう対応させていくか
 - もう一つの重量級問題
 - 海外では人海戦術でポーティングを完了させる例も：TaihuLightのケース等
 - 日本のモデル開発コミュニティはコード全面書き換えに耐えうるだけの人材を確保しきれていない

- 開発体制の問題

- 日本の気象・気候業界はスパコンに恵まれているが、ソフトウェアエンジニアリングに関する努力に乏しい
- アルゴリズム開発者がアプリケーションの実装コードを書き、シミュレーションの結果で分野科学の論文を書く「フルスタック研究者」はそうそう居ないし、生きづらい

→絶滅危惧種

海外の動向

• CLiMA (米)

- Caltech, MIT, NASA/JPLによるアライアンス、Schumidt財団等の巨大投資会社の後押し
- 地上・衛星による地球観測情報から学習した地球システムモデルの構築
 - 機械学習やデータ同化を用いて、モデル化やパラメータ推定を進めている
- プログラムコードはJulia言語を用いて構築



• Vulkan Inc. (米)

- ポール・アレン氏が設立した投資会社+NOAA/GFDLによるタッグ
- NOAAの新気象予報システムの中核となった大気モデルFV3を、C++テンプレートフレームワーク群 (GridTools>4Py) で置き換える
 - 主要な物理過程はほぼ置き換え済み



海外の動向

• E3SM (米)

- DOEによる地球システムモデル構築のための大プロジェクト
- 主要な過程はC++ライブラリ(kokkos)を用いて書き換えを完了
- 陸面や海洋はOpenACCで



• ESCAPE1/2 (欧)

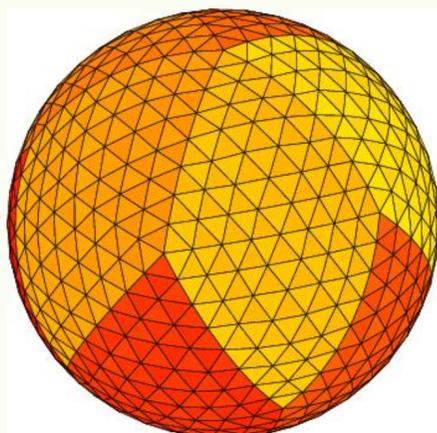
- ヨーロッパECMWF主導によるプロジェクト (2015-2018, 2018-2021)
- 大気モデルのための基本コンポーネントをDwarfと呼び、将来のマシンに最適なアルゴリズム、コーディングを模索→ベンチマークセットHPCWへ
- DSLによる記述：ATLAS, CLAW, GridTools, PSyclone
- 低精度でのシミュレーション・データ同化の積極的な推進



データ駆動型モデルの利用による制約の打破

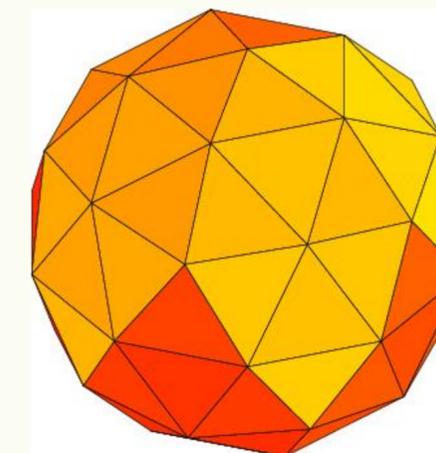
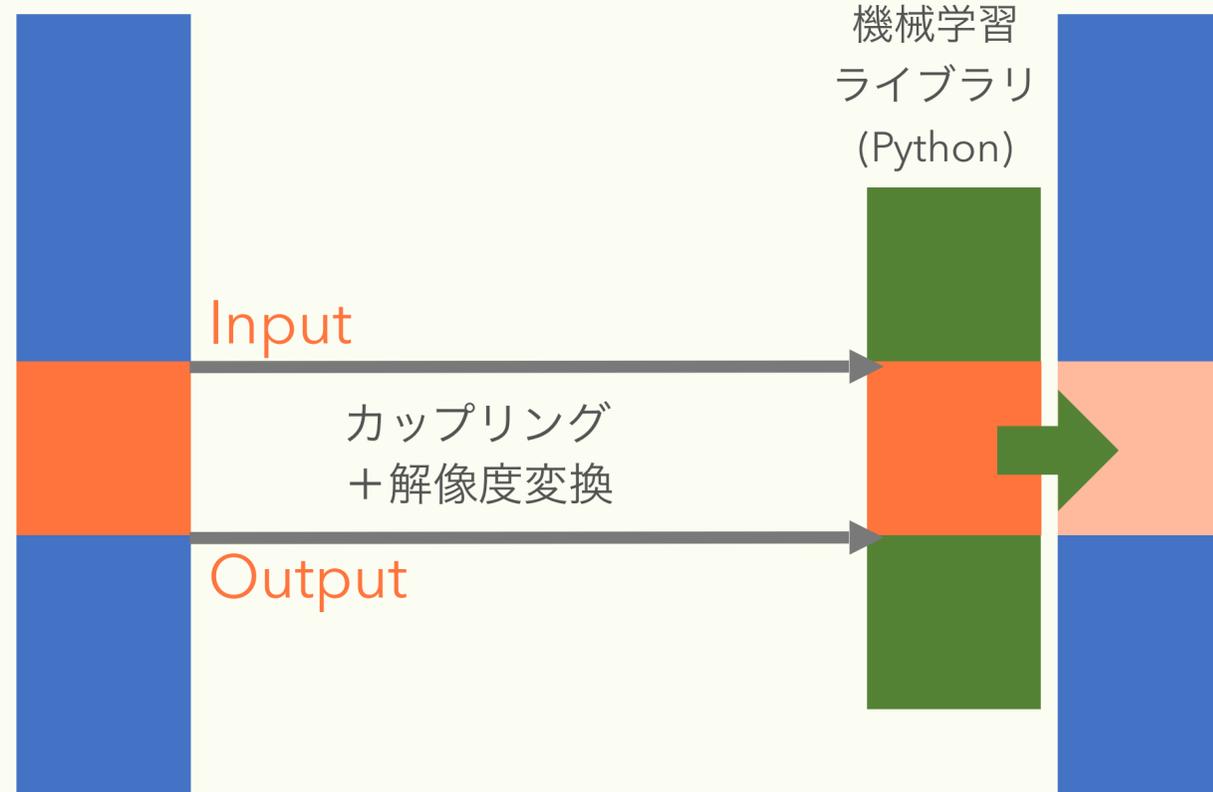
- 実現したいことの本質は何か
 - 物理法則に則った気象と気候のモデリング
 - 今よりも高速なシミュレーションの実施
 - メモリの制約、電力の制約を受けにくいアルゴリズム
 - 計算機アーキテクチャに左右されにくいアルゴリズム
 - フルスタック研究者に頼らない開発
- 「精緻で重い」計算と、「近似で軽い」計算を使い分ける
 - これまでの物理モデルはさらに精緻に、さらに重く、精度を追い求めることでよい
 - 長期シミュレーションでは、物理モデルの近似モデルをデータ駆動型アプローチで作成し利用する
 - 機械学習技術を用いた代理モデルの作成
 - はじめからシミュレーションモデルすべてを代替しようとするしない：Step-by-stepの変更
 - 少数・短期の重い計算によって、多数・長期の軽い計算を実現する

代替モデル構築のためのプラグイン環境構築



より高解像度のシミュレーション

コード内の好きな区間



低解像度のシミュレーション
(GCM解像度とは限らない)



h3-open-UTIL/MP
弱連成ライブラリ

Fortran APP
(NICAM)

h3open modules
Jcup modules
jcup_mpi_lib.f90

Python APP
(PyTorch)

h3opp.py
h3open_py.f90
h3open modules
Jcup modules
jcup_mpi_lib.f90

- 科研費S (PI: 東大中島研吾教授) での開発
富岳NEXT-FS(理研)でワークフローベンチマークとして登録
- 機械学習のためのデータを「その場」供給
- 任意のコンポーネントから入出力データを取り出す

気象学会としての取り組み

- 計算科学研究連絡会 (<https://metsoc-hpc.github.io/>) の発足 (2022年)
 - 第1回連絡会会合 (2022.5)
 - 気象業界と関連の深い、世界のHPC開発計画・シミュレーションモデル開発計画の紹介
 - NGACI White Paperの紹介、計算科学ロードマップ (気象・気候分野) の執筆状況
 - 第2回連絡会会合 (2022.8) 「地球デジタルツインの動向に関するワークショップ」
 - デジタルツイン、データプラットフォーム、社会とのインタフェース、シミュレーションモデル開発
- 次回以降予定
 - 初心者から見たGPUプログラミング、GPUハッカソン報告
 - 令和時代の解析・可視化ツールランキング：愛用ユーザによる布教付き
 - 参加資格は日本気象学会の会員、非会員を問いません：ご興味のある方は是非

地球デジタルツイン (Digital Earth)

"Digital Twin Engine"



観測

データ同化



シミュレーション

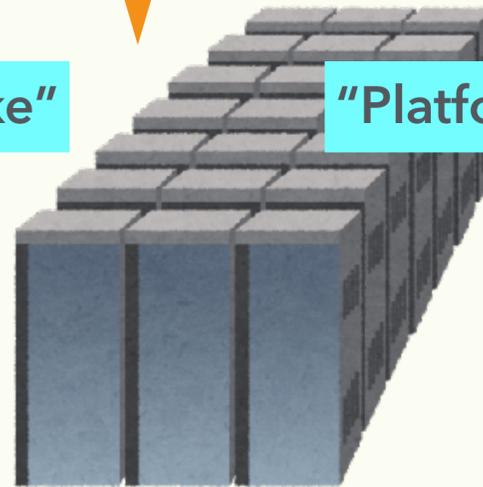
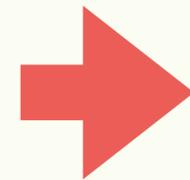
巨大化するデータの
その場解析・キュレーション

"Data Lake"

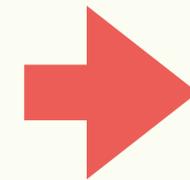
"Platform Service"

数十TB~数PB

~数MB



データ共有



迅速なデータ配信

オンデマンドな簡易シミュレーション?



社会への貢献

革新的な精度の観測・シミュレーション技術
→まるで双子

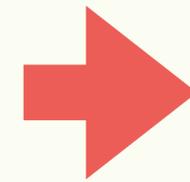
階層的なデータ生成・保存・圧縮・変換・流通



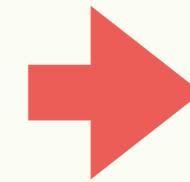
シミュレーションから出力されるデータ
(TB~PB級)



高速エミュレーション



解析・統計・付加価値追加
された応用データ
(MB~GB級)



ユーザが必要とするデータ
(KB~MB級)

地球デジタルツインシステムの構築に向けて

- 古くて新しい議論

- データの収集、シミュレーションの実施、結果の公開は気象・気候分野が今までもやってきたこと
- リアルタイム性や自動化、データのキュレーションが重要な要素

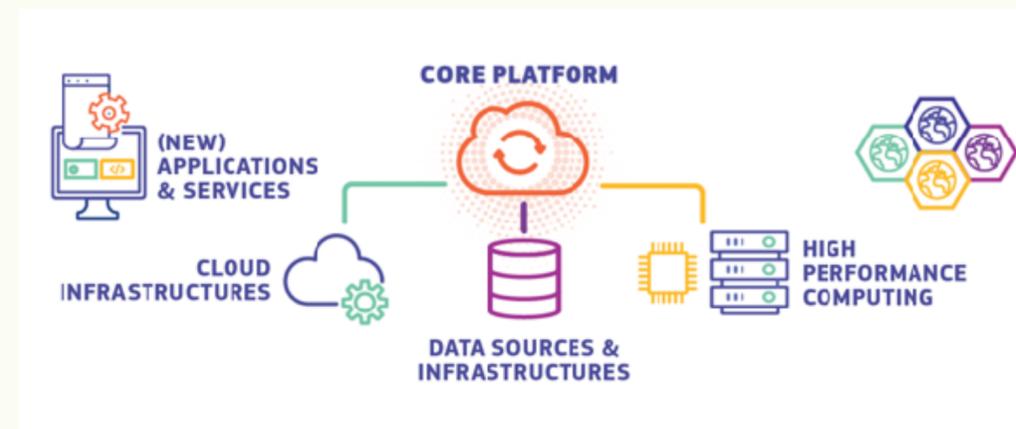
- 気象・気候分野と他分野の研究者が連携する場を形成する

- 地球デジタルツインをハブとして、共同研究の起点を作る
- 今ないコラボレーションを推進するには、呼び水が大事：着実に事例を増やす
- DIAS、mdxのようなプラットフォームをさらに活用したい
- 気象庁も新たな気象データ共同利用に関する枠組みを準備中

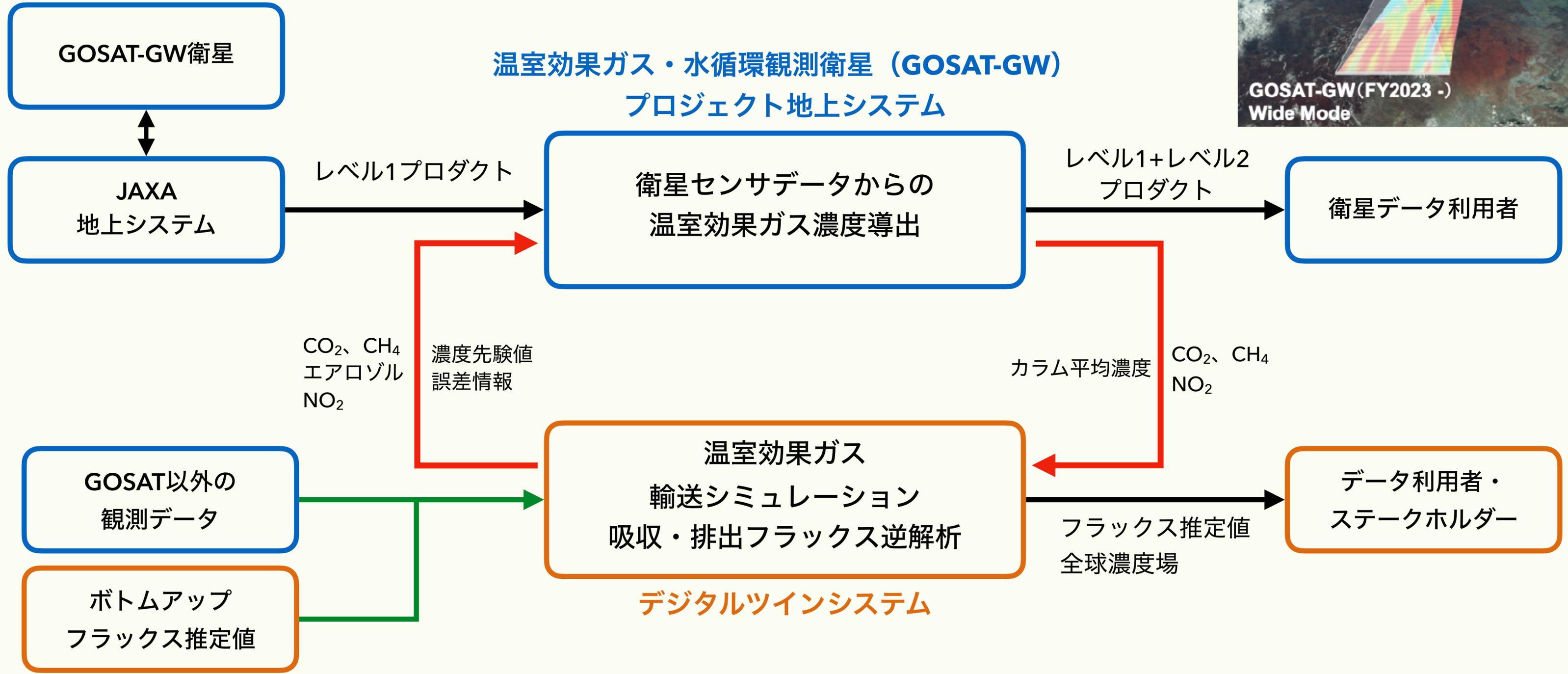
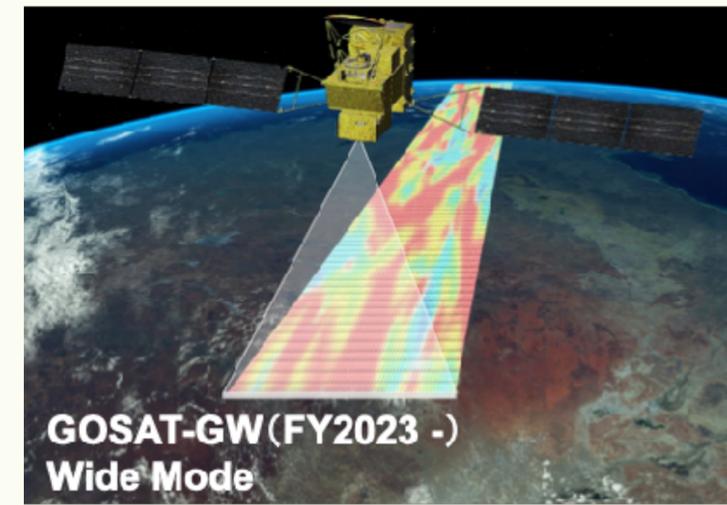
海外の動向

- **Destination Earth (欧)**

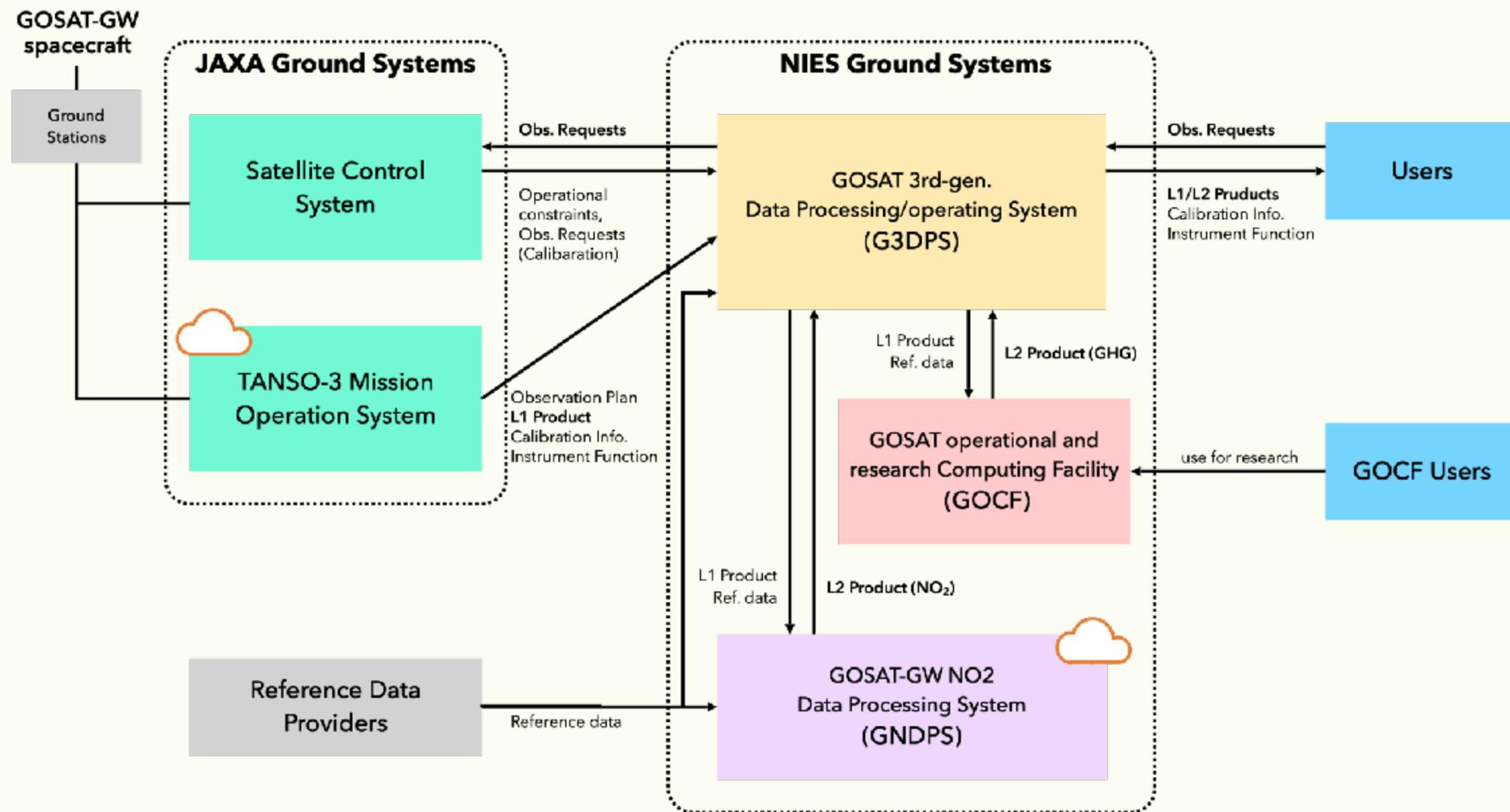
- ECのデジタル戦略に基づくプラットフォーム構築 (2022-2024, 2030年目標)
 - ESA : プラットフォームの開発運営。
 - EUMESAT : データレイクの運営。Copernicus社が協力サポート
 - ECMWF : デジタルツイン (気象・気候) の開発。



国立環境研における取り組み： 温室効果ガス観測衛星プロジェクト



GOSAT-GWプロジェクト地上システムの開発



- 国環研2システムは筑波大CCS
計算機室に設置予定

- G3DPS（基幹システム）：
HCIとオブジェクトストレージで構成
- GOCF（スパコン）：
衛星センサデータ処理を担当

衛星データからの温室効果ガス濃度導出の高速化

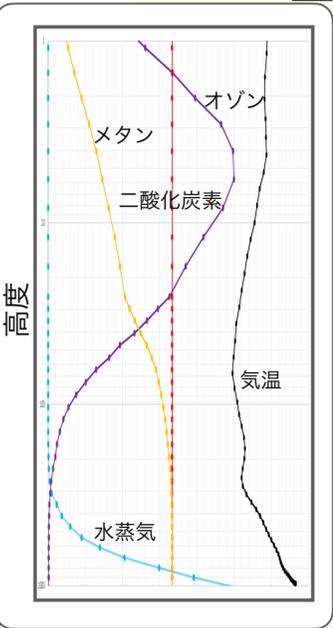
A prioriを修正して
収束するまで反復計算

幾何パラメータ

- 太陽天頂角・方位角 [scl]
- 衛星天頂角・方位角 [scl]

大気の状態

- 地表面気圧 [scl]
- 地上風速とその分散 [scl]
- 気圧 [vec]
- 気温 [vec]
- H₂O, O₃, CO₂, CH₄濃度 [vec]
- エアロゾル(19種)の濃度 [vec]



入力情報：
100-200次元

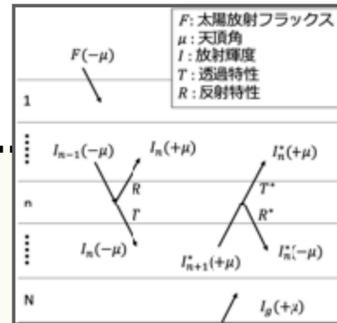
放射パラメータ

- 各気体の波長毎の吸収係数 (圧に非線形に依存)
- エアロゾルの波長毎の散乱係数 (粒径(∝濃度)や湿度に依存)

放射伝達モデル [物理モデル]

Nakajima and Tanaka (1986), Ota et al. (2010)

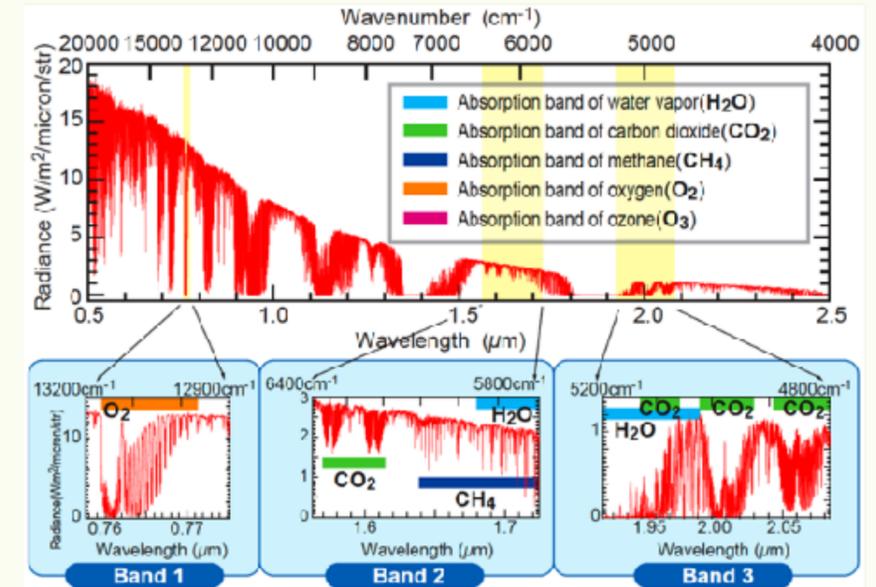
- 鉛直一次元の大気柱を光路とし、離散座標系(DOM)で定式化
- 細かく分割した波長帯毎に、大気各層での吸収・散乱の強さを計算
- 光路に沿って、順に放射の吸収・散乱を積算
- 多重散乱も考慮



最大事後確率(MAP)推定

- 放射スペクトルから大気状態を逆推定

シミュレートされた放射スペクトル：数万次元



観測される放射スペクトル

Approach 1: 代理モデルの構築

- カーネル法を用いた関数回帰によって、放射伝達モデルを模擬するエミュレータを作成
- 物理モデルによる多数のシミュレーションを学習データに利用

Approach 2: 物理モデルを介さない温室効果ガス濃度の逆推定

- カーネル法と近似ベイズ計算(ABC)を用いた推定
- まずは次元数の少ないSynthetic Dataで検証→複雑度を上げていく
- 10年以上にわたる実データの蓄積を元に、衛星測器の持つ観測誤差が影響する場合の推定精度についても検証を進めていく

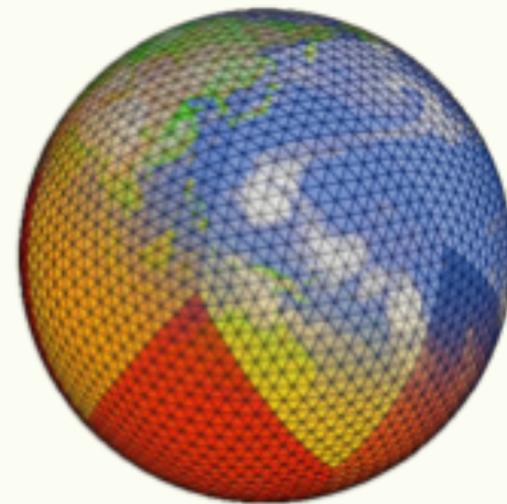
- NEC、仏EURECOMとの共同研究を開始

アーキテクチャに期待すること

- **メモリ性能の継続的な向上**
 - ムーアの法則はまだ続くことを期待
 - 3次元実装
- **ストリーム演算コア？**
 - 富岳エコモードの経験より、メモリ律速のアプリは複数の演算器を使いきれていない
 - メモリやキャッシュに書き戻す回数が減る演算器？
- **階層化されたストレージ・メモリを楽に活用できるシステムソフトウェア**
 - 不揮発性メモリまでを利用して、総メモリサイズを拡張したい

まとめ

- 気象・気候シミュレーションモデルはデータ律速であり、コードベースが巨大でモデル最適化のためにはアプリケーション全体を書き換える必要がある
→HPC業界の重量級実社会アプリケーション
- 富岳ではデータ中心のデザインにより高速化を達成
- 富岳NEXTでは必要な演算・データ移動すら削減する方向での最適化が必要
- DSLによるソースコード書き換えは、国内ではほとんど進んでいない
：機械学習技術を用いたコンポーネントの代替では、あまり必要でなくなる？
- デジタルツインでは、データのキュレーションも重要な要素に



ご清聴ありがとうございました