

jh241009

全球 km スケールモデルの共通解析基盤の確立

佐藤正樹（東京大学大気海洋研究所）

概要

地球全体を km スケールで解像する全球モデルの開発と国際比較が進められており、極端気象の理解や地球デジタルツインへの貢献が期待されている。本課題では、mdx を活用し、国内外 10 機関程度が参加する国際比較実験データのアーカイブと配信サイトの構築を行う。本年度は、日本の全球 km スケールモデル NICAM を含む複数のデータを対象として 100TB の領域を確保し、FTP サーバ機能を備えた仮想マシンを構築、データのアップロード・ダウンロードの運用体制を整備した。今後は、計算機能を備えた解析サーバとしての機能拡張を進め、2025 年 5 月に開催予定の Global km-scale Hackathon 東京ノードでの活用を目指す。

1. 共同研究に関する情報

(1) 共同利用・共同研究を実施している拠点名

mdx

(2) 課題分野

データ科学・データ利活用課題分野

(3) 参加研究者一覧と役割分担

佐藤正樹：課題代表

大野知紀：副代表

Woosub Roh：課題実施者・アドバイス

宮川知己：課題実施者・アドバイス

松岸修平：課題実施者・アドバイス

高須賀大輔：課題実施者・アドバイス

八代尚：課題実施者・アドバイス

小玉知央：課題実施者・アドバイス

2. 研究の目的と意義

地球全体を km スケールのメッシュで覆う超高解像度の大気海洋モデル・地球システムモデルによる気象気候研究の新たな展開が始まっている。進行する地球温暖化に伴って激甚化している台風や豪雨等の極端気象現象のリアルタイムモニタリングや将来

変化予測において、地球デジタルツインへの貢献等、全球 km スケールモデルに期待される役割は大きい。全球 km スケールモデルによるシミュレーションの性能評価のためには、世界各国の全球 km スケールモデルの比較実験を推進し、そのデータをアーカイブし、解析利用に資する必要がある。本課題では、この目的のために JHPCN の mdx を利用申請した。

3. 当拠点の公募型共同研究として実施した意義

世界各国の 10 機関以上で「全球 km スケールモデル」の開発を進めている。本申請は、世界各国の全球 km スケールモデルの出力結果を JHPCN mdx に集積し、世界中の関係する研究コミュニティへの利用を図るものである。

日本における「全球 km スケールモデル」は東京大学大気海洋研究所、東北大学、海洋研究開発機構、国立環境研究所、理化学研究所等の研究者の連携によって、大気海洋モデル NICAM および NICOCO を共同開発してきた。これらの数値モデルは、構成拠点のスーパーコンピュータ「富岳」や東京大学情報基盤センターWisteria等を用いて実施している。これらの構成拠点において実施した数値シミュレーション結果を、mdx を利用して公開することが必要とされ

る。その上に、世界各国で実施される「全球 km スケールモデル」の数値シミュレーション結果を mdx 上にアーカイブし、世界各国の研究者の利便性に資することが必要である。

全球 km スケールモデルの出力は膨大であり、アーカイブされたサーバからダウンロードすることなく、サーバ上でデータの解析する方法の確立が必要である。諸外国では、ドイツ DKRZ、イギリス JASMINE 等、先行する取り組みがある。JHPCN mdx では、これら諸外国の研究拠点との連携を図り、世界的な全球 km スケールモデルのシームレスな利便性の向上を図る必要がある。

4. 前年度までに得られた研究成果の概要

該当なし

5. 今年度の研究成果の詳細

本年度は、国際比較実験「全球 km スケールモデル」のデータアーカイブ、配信サイトの構築を行う。10 機関による全球 km スケールの計算結果を格納する予定であり、総容量はおよそ 200TB となる見込みである。今年度は、日本におけるモデルデータ NICAM のみのデータアーカイブを行い、mdx の大規模ストレージから 100TB を課題用に確保した。

基本機能としての FTP サーバの構築を当初目標とした。mdx 上に FTP サーバとして仮想マシンを構築した。さらに、サーバサイド解析やデータ提供フォーマットの検討を行い、機能強化を行った。当初の仮想マシンの用途はファイルサーバでありストレージが重視されるが、今後は解析サーバとしての機能を充実させることを考えている。

構築した仮想マシンに大規模ストレージをマウントし、FTP サーバの使用領域に設定した。またファイルのアップデートは指定されたユーザが、ダウンロードは不特定のユーザが実行できるように FTP サーバソフトウェアの設定および mdx のネットワーク設定を行った。これにより、本研究の第一ステップであるデータサーバ構築に関する最低限の要求を満たす。

mdx サーバに ssh によりアクセスし、全球 km スケール実験が可能な数値モデル NICAM によるシミュレ

ーションの稼働が可能な計算環境を構築した。netCDF 等の全球 km スケールデータの解析に必要なツール・ライブラリを整備した。

複数の研究機関が個別にアーカイブしている総容量 200TB におよぶデータ群に対して、利用者が目的のデータに迅速にアクセスするためには、適切なタグ付けと、それに基づくユーザナビゲーションの仕組みが不可欠である。そのため、Zarr 形式、HEALPix フォーマット、STAC 等を活用したデータ整備を進める。今後は、JupyterHub や VSCode を利用したユーザ環境の整備を図るとともに、データアーカイブの構造設計およびシステム構築についても検討を進めていく。

6. 進捗状況の自己評価と今後の展望

従来のデータ解析の手順はアーカイブされたデータをダウンロードし、ローカルなマシンで計算を行うというものである。全球 km スケールのデータは容量が大きく、アーカイブされたデータを逐一ダウンロードすることはネットワークトラフィックやディスクスペースを著しく消費することになる。そこで、必要なデータの切り出しや簡単な統計計算をサーバサイドで実行するための仕組みを mdx 上に構築する。また異なる格子系で計算された結果を統一的に取り扱うためのデータ構造やソフトウェアについても設計/構築を行う。

本年度は mdx サーバ構築を開始した初年度であったため開発目標を最低限の FTP サーバ機能の構築とした。一方で、2025 年 5 月に国際的な複数拠点（日本、ドイツ、イギリス、アメリカ、オーストラリア、ブラジル、中国等）において、協調的な国際同時全球 km-スケールモデル解析イベント The World Climate Research Programme Global KM-scale Hackathon の開催が決定し、本 mdx は日本拠点として利用することになった。このため、当初の開発目標を超えて、解析機能の強化を前倒しして進める必要が生じた。本年度末の時点では、課題の継続を念頭に、5 月の本イベントに間に合わせるべく、mdx サーバの構築を進めている。