

jh240079

AI 気象モデルを用いた気候シミュレーションの高速化に関する研究

八代 尚（国立環境研究所）

従来の物理法則に基づく気象・気候シミュレーションモデルとは異なる手法で構築された、気象予測のためのデータ駆動型モデルについて、新たな教師データや前処理された気象変数を用いることによる予測精度の改善について解析を行なった。まず、並列での学習実行環境を整備し、地表面近くの地形に伴う値の欠損処理について改善を行ない、新たな教師データである全球高解像度大気モデル NICAM の 30 年気候実験結果に対しても同様の改善を適用して安定した学習を行うことに成功した。また、データの高解像度化、大アンサンブル化に伴う計算のスケーラビリティやファイルシステム性能についても評価を行った。気象変数の前処理改善では、大気中に偏在する降水量等の学習結果を向上すべく、長期データに基づく時間方向統計量を用いた変数変換を適用し学習・予測を行った結果、予測結果が改善する変数とそうでない変数の存在が明らかになった。

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

東京大学 情報基盤センター

名古屋大学 情報基盤センター

(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

国立環境研究所 八代尚（代表）：課題統括

名古屋大情報基盤センター 片桐孝洋（副代表）

：自動チューニング、AI 高性能化

東京大学大気海洋研究所 宮川知己（副代表）

：AI 気象モデル評価

東京大学大気海洋研究所 上条藍悠

：AI 気象モデル開発、評価

名古屋大学大学院情報学研究科 水島慎吾

：自動チューニング、AI 高性能化

名古屋大学大学院情報学研究科 任軒正博

：自動チューニング、AI 高性能化

東京大学情報基盤センター 山崎一哉

：AI 気象モデル評価、AI 高性能化

東京大学情報基盤センター 中島研吾

：AI 高性能化

2. 研究の目的と意義

気候・気象シミュレーションは人類の生命と財産を守る上で非常に重要なツールである。更なる精度向上のために、より高解像度・大アンサンブルのシミュレーションの実施が求められるが、これまでの物理法則・経験則に基づくアルゴリズムと汎用 CPU を用いた計算では、ゆっくりとしたメモリ転送性能の向上に計算速度が律速され、計算の大規模化が妨げられることにより、大幅な精度向上が今後見込めなくなるであろうことが将来予測として示されている。一方で、最新のスーパーコンピュータは GPU 等の演算加速機の発展により、機械学習で用いるような低精度の密行列計算では年々着実に性能を伸ばしており、将来の計算機開発においても重視されることが予想される。上記の背景をふまえると、これからの気候予測精度の向上に向けて、従来の物理モデルのアルゴリズムをデータ駆動型モデルで用いるアルゴリズムに模倣させ、より現代

の計算機を活用できる形に変換 (Algorithm Transformation) し、演算の高密度化、高速化を図ることが必要となる。本研究では、近年活発な開発が進められているデータ駆動型の気象・気候予測モデル (AI 気象モデル) を用い、その計算速度とシミュレーション・解析結果再現性について評価を行うことを目的とする。

3. 当拠点の公募型共同研究として実施した意義

本研究は従来の気候科学研究者による気候予測精度の向上のための研究という形だけでは不十分であり、計算科学・計算機科学を専門とする研究者との協働があって初めて進展するものである。また JHPCN 拠点が有する GPU 計算資源は、本研究内で実施するデータ駆動型モデルの機械学習計算を実施するために不可欠である。本研究では国立環境研・東大大気海洋研から気候学・大気科学の専門家、東大・名大の JHPCN センターから高性能計算の専門家が参画し、密な連携をとることで研究を加速することが可能になった。

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

5.1 AI 気象モデルの動作検証と初期的性能確認

AI 気象モデルとして FourCastNet (Pathak et al., 2022) を用い、提供済みの学習済データをそのまま使わず、ヨーロッパ中期予報センター (ECMWF) が配布する気象再解析データセット ERA5 (緯度経度格子、水平 720x1440 格子点、6 時間間隔) を教師データとして機械学習を実施し、想定通りの予測性能を示すかについて確認した。使用した変数は気圧面変数として 1000hPa・925hPa・850hPa・700hPa・500hPa・200hPa 基準面での東西風速・南北風速・鉛直風速・気温・比湿、単一平面変数として地表面気圧・海面更正気圧・2m 気温・表面気温・10m 東西及び南北風速・海水・海面水温・鉛直積算水量・鉛直積算水蒸気量・雲被覆率・降水量・潜熱及び顕熱フラックス・大

気上端及び地表面下向き短波放射フラックス・大気上端及び地表面上向き短波放射フラックス・地表面下向き長波放射フラックス・大気上端及び地表面上向き長波放射フラックスの計 51 変数である。1979 年から 2004 年までの 26 年分のデータを学習に用い、2008 年のデータを推論性能評価に使用した。計算には東京大学 Wisteria/BDEC-01 Aquarius を用いた。FourCastNet のモデルアーキテクチャについては、当初公開されていたフーリエ変換手法に基づく AFNO (Adaptive Fourier Neural Operators) から球面調和関数展開に基づく SFNO (Spherical Fourier Neural Operators) に切り替え、両極点での数値不安定に起因する長期予測の不安定を解消した。学習後の予測結果を解析した結果、各気象変数 (1000hPa 高度：およそ地表付近、850hPa 高度：海上 1.5km 付近、500hPa 高度：海上 5km 付近での風速、気温、相対湿度等) の多くは最初の 48 時間で精度が急激に悪化し、中でも地表に近い変数の予測精度の時間変化が滑らかではないことを確認した (図 1)。これは地形によって値が欠損している変数の取り扱いに起因するものであった。そこで等圧面の入力変数を地形に沿った σ -p 座標面の値に変換し、欠損値が適切に補完された学習データを用いることで、ほとんどの変数の長期予測精度を向上させ、変数によっては 10 日程度にわたり高い精度を保つことに成功した (図 2)。

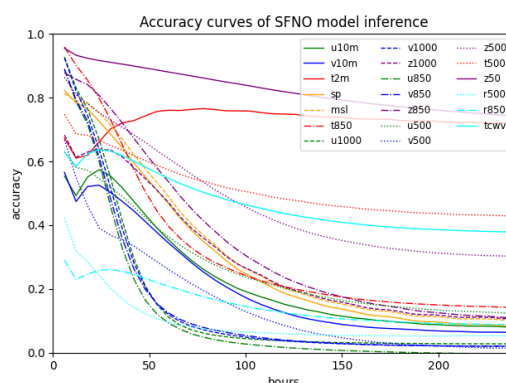


図 1 : ERA5 大気再解析データを教師データとして学習した FourCastNet の、10 日間予報の精度推移。

予測期間の後半における精度は、気温・ジオポテンシャル高度、風速、相対湿度の順に高く保たれており、空間的に偏在する水物質とその相変化を予測することが相対的に難しいことを確認した。

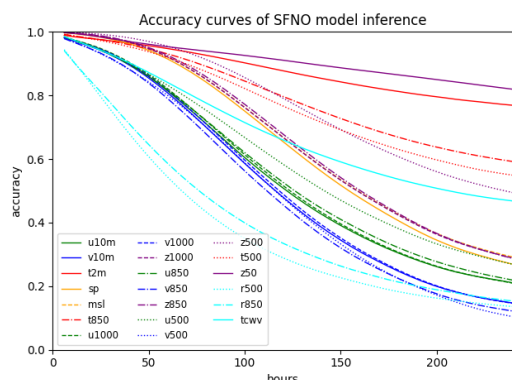


図2：図1と同じ、ただし、地表面近くの欠損データの扱いを改良した結果。

5.2 NICAM 高解像度データを用いた学習

次に、より高解像度で、水平格子点数が ERA5 の 4 倍程度ある全球高解像度大気モデル NICAM の 30 年気候シミュレーション結果を新たな教師データとして FourCastNet の学習を試みた。学習時において、ERA5 を用いた時に比べて計算時間がかかなり増大することが判明した。これはモデルに含まれる SFNO 部分に、水平解像度に対して $O(N^3)$ で計算量が増大するルジャンドル変換処理を含むことが要因であった。このような計算量の増大は従来のスペクトル法を用いた気候モデルが直面してきた問題と同一である。球面上で定義された関数を球面調和関数で展開するスペクトル法では FFT とルジャンドル変換を使用しており、高速化のために高並列化を進めていった場合もノードを跨ぐ大域通信がボトルネックとなる。一方でスペクトル法は 1 時刻ステップをかかなり長くとれるという利点があり、目的とするシミュレーション時間を少ない総ステップ数＝総計算量で計算出来るため高速である。大気モデルの計算では、スペクトル法以外の方法として隣接する格子点の情報をを用いる格子法がある。スペクトル法と格子法にはいずれもメリット・デメリットがあり、格子法を用いた場合は

大域通信が必要なく高並列性を保てるが、CFL 条件のために 1 時刻ステップを長く取れない。総合的な計算速度では、両者は今でも拮抗している。しかし AI 気象モデルの場合は、格子法に近い手法をマルチスケールで適用することにより、CFL 条件に囚われずに 1 時刻ステップを予測することが可能となるため、格子法のデメリットが解消され、高解像度化を行った場合にスペクトル法よりも優位であることが明らかになった。

5.3 FourCastNet 推論時のファイルシステム性能評価

名大スパコン「不老」を用いて、FourCastNet の推論部分の性能評価を行った。AI 気象モデルは物理モデルと比較して計算時間が圧倒的に早いいため、入力データが少しずつ異なるデータアンサンブルでの推論を多数行い、統計的により強固な予測結果を得るような利用方法がたいへん有効である。結果として、推論時にモデル重み情報等を格納したファイルに同時アクセスが集中することが考えられる。このようなファイルアクセスの状況を仮定し、ファイルシステムの性能評価を行った。スパコン「不老」のファイルシステム性能評価として、次の 4 つの異なるファイル性能領域に数 GB のファイルデータを置き同時に FourCastNet の推論を行う場合の性能評価を行った。すなわち、(1)/home 領域、(2)/data 領域、(3)ローカル SSD 領域、(4)分散 SSD ファイルシステムを形成する NVMesh 領域、である。結果より、複数同時実行が性能低下の要因となることを確認し、また(3)SSD 領域のファイルアクセス性能速度について有意な向上を確認した。これにより、並列実行性能と並列ファイルアクセス性能を考慮した実行形態を選択することが良いことを示し、成果は名古屋大学大学院情報学研究科の修士論文としてまとめられた。

5.4 非ガウス分布変数の変換操作が推論性能に与える影響の評価

現状の AI 気象モデルは大規模な風の場合や気圧・

気温分布を再現することが相対的に得意であるが、5.1 節で得られた結果からもわかるように、降水量や雲分布等の空間的に滑らかに分布せず、出現頻度としてゼロを多く含むような気象要素については予測性能に改善の余地がある。本研究ではそのような非ガウス分布傾向が強い変数について、気象データ同化で近年用いられるようになったガウス分布変換手法（例えば Lien et al., 2013; Kotsuki et al., 2017）を適用することで、予測性能が向上するかについての実験と評価を行った。対照実験として、1) ERA5 を用いて通常の学習を行ったモデル、2) ERA5 の学習変数のうち、降水量のみにガウス分布変換を適用して学習を行ったモデル、3) ERA5 の 51 個の学習変数すべてにガウス分布変換を適用して学習を行ったモデル、の 3 種類の推論結果を比較した。

ガウス分布変換の事前準備として、各変数の 26 年分のデータに対し、その最大・最小値をもとに 1024 階級の度数分布表を作成した。ガウス分布変換では変換前の値が度数分布表内で上位何%の値になるかを計算し（度数分布表の階級値間は線形補間）、ガウス分布関数上でその階級に対応する値を返す。一方でその逆変換は値をガウス分布関数上の値であるとみなして、その値がガウス分布関数上で上位何%に属するかを計算し、対応する度数分布表内の値を返す。学習においては通常の学習データの代わりにガウス分布変換された値を AI モデルに入力して学習を実施した。推論時においても入力データに変換を適用し、出力値には逆変換を適用して元の変数の形での推論結果を得た。

推論結果の性能比較には変数それぞれの「正答率」という指標を用いた。この正答率は推論データと正答データとの間の相関係数で表されるが、事前に両データそれぞれを学習データ期間全体での標準偏差と平均で正規化し、全格子点を面積重み付け平均した値を用いる。評価期間に含まれるすべての初期値を用いて、それぞれ 2 週間の推論を実施し正答率を求め、全てのケース

の平均値を計算した。図 3 は初期値からの正答率の時間変化を各変数に対して計算した結果を示す。縦軸は降水量のみを変数変換したモデルの正答率を、変数変換を行わない従来モデルの正答率で除した値を示す。結果より、予測精度が向上することを最も期待した降水量（紫実線）が予測期間を通して常に悪化する結果となった。このほか、上層の南北風についても悪化したが、一方で鉛直風速や水蒸気量については、若干の改善がみられた。

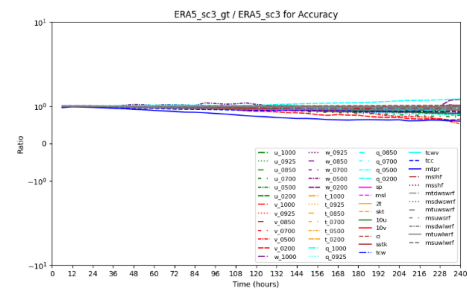


図 3：初期値からの正答率の時間変化を各変数に対して計算した結果を示す。縦軸は降水量のみを変数変換したモデルの正答率を、変数変換を行わない従来モデルの正答率で除した値を示す。

降水データだけでなく、全データにガウス分布変換をかけた場合の性能比較結果を図 4 に示す。結果より、地表気圧（ピンク実線）の精度が予測開始直後から急激に悪化し、その他の変数についても結果がより改悪される傾向にあった。これは地表気圧の予測精度悪化の影響がより上層の大気の高度位置を歪め、精度改悪に繋がった可能性がある。地表気圧の予測精度悪化は、度数分布の階級幅が地表気圧のダイナミックレンジ（1000hPa から 300hPa 程度まで）に対して粗す

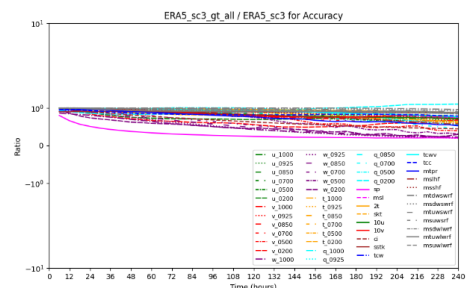


図 4：図 3 に同じ。ただし、縦軸は全ての変数を変数変換したモデルの正答率を、変数変換を行わない従来モデルの正答率で除した値を示す。

ぎたことが考えられる。今回の実験の場合、度数分布表の階級数は GPU のメモリ容量の制約によって決められており、今後より大きなメモリを持つ GPU の利用や、メモリを節約するための手法を適用することで、改善が可能であるか検証を進める。

5.5 AI 気象モデル実行環境構築について

本研究で用いた FourCastNet の実行のための環境構築には、手動ビルドが必要なライブラリも多く、C 言語及び MPI ライブラリ、Python、HDF5 ライブラリ等のビルドを試みたが、MPI ライブラリのビルドで解決しがたい問題が発生した。そこで NVIDIA が配布している学習環境を整備したコンテナの利用を検討し、実際に動作確認をしながら利用するバージョン調整を行い、最終的に `nver.io/nvidia/pytorch:24.03-py3` を用いた。また研究ではオリジナルの FourCastNet ではなく、FourCastNet の研究開発のために作成された makani ライブラリ (<https://github.com/NVIDIA/makani>) をベースに、米国ローレンスバークレイ国立研究所が大アンサンブルでの実行を可能にするために変更を加えたコード (<https://github.com/ankurmahesh/modulus-makani-fork>) を用いた。本研究での実行環境構築手順について、以下に詳細に示す。

- ① singularity コマンドを用い、sandbox フラグを有効にしてコンテナファイルを build することで、編集可能なコンテナを作成する。
- ② makani リポジトリ、HDF5-1.14.5 のリポジトリをローカルにクローンする
- ③ 学習用データが入ったディレクトリ、作業用ディレクトリ、およびコンテナの HOME ディレクトリを用意する。
- ④ ②及び③のディレクトリを `--bind` オプションを付けてコンテナ内の適当な場所にマウントしつつ、`--writable` オプションを付けて書き込み可能な状態で、①で作成した singularity コンテナの shell を起動する。

- ⑤ コンテナ内にて HOME ディレクトリの環境変数を指定し、次のライブラリをインストールする。

- git
- wget (ここまで apt で)
- `nvidia-modulus==0.9.0`
- `torch-harmonics==0.6.5`
- `moviepy==1.0.3`
- `jsbeautifier==1.15.1`
- `tensorly==0.8.1`
- `tensorly-torch==0.4.0`
- `netCDF4==1.6.5` (ここまで pip で)

- ⑥ コンテナ内にて HDF5-1.14.5 のディレクトリに移動し、手動でのビルドを行う。この時、`configure` コマンドの引数は次の通り。

- `export CC=/usr/local/mpl/bin/mpicc`
- `./configure --disable-fortran --prefix=/usr/local/ --enable-shared --enable-parallel`

- ⑦ pip にて hdf5 の Python バインドをソースパッケージからインストールする。

- `export LD_LIBRARY_PATH=$LD_LIBRARY_PATH:/usr/local/lib`
- `export CC=mpicc`
- `export HDF5_MPI=ON`
- `pip3 install --no-binary=h5py h5py==3.12.1`

- ⑧ コンテナ内にて makani のディレクトリに移動し、編集可能な形式でインストールする。

- `pip install -e .`

さらに、次のコード改善を実施した。

- 学習及び推論時、プロセスグループそれぞれが持つデータバッチに対し処理が行われ、最終的に統計情報等を交換するというアルゴリズム仕様の結果、ロードインバランスが大きくなった際に同期のタイミングでタイムアウトが頻発した。そこで一定ステップ毎の同期回数を増やして待ち時間をこまめにとった。

- 検証時には検証用データ全てに対し一定ステップ分の予測を行い、それらの統計情報を保存する仕様であったが、推論時に各プロセスが読み込んだ初期値との対応が取りにくかったため、ファイル内のどの時刻のデータを読み込んだか分かりやすくなるようデータローダに変更を加えた。

ERA5(水平方向解像度約 28km)の 51 変数、時間間隔 6 時間の 30 年分のデータに対し、打ち切り次数 240、埋め込み次元数を 128 に設定して 90 エポック学習させる際、必要な計算資源はおおよそ Wisteria/BDEC-01 Aquarius の 16GPU を用いて 810 時間=10,800GPU 時間であった。一方、異なる 1400 個の初期値に対し 10 日間の予報を行うのに必要な計算資源は同じく 16GPU を用いておおよそ 7 時間=112GPU 時間であった。ガウス分布変換に伴う計算量の増加は全変数を変換する場合においても 10%程度であった。

パターンの学習を実施し比較を行うことが出来た。結果では想定した予測性能の改善がみられず、その原因について特定するに至っていないが、予定した研究をほぼ達成したといえる。

今後の展望としては、本研究によって明らかになった AI 気象モデルのスケラビリティの問題を克服するため、グラフニューラルネットワーク (GNN) を用いたモデルの評価を実施したい。また、予測期間として「季節内から季節 (S2S: Sub-seasonal to Seasonal)」スケールでの予測性能の解析を進め、数日よりさらに長い期間における予測の性能や、AI 気象モデルが得意・不得意とする現象について評価を進めたい。

6. 進捗状況の自己評価と今後の展望

今年度計画として設定した 2 つの研究課題のうち、一つ目の「新たなデータセットを用いたモデルの高解像度化とその評価」については、利用した AI 気象モデルを並列 GPU 計算機で安定して学習させるための環境整備に予定以上の準備期間を要したが、最終的に 4 倍程度の格子点数を持つデータセットの学習を行うことが可能になった。また、従来のデータセットを用いた予測においても、気候科学の知見に根ざした解析の結果、改善すべき点が明らかになったため、それらの開発を優先して実施した。そのため新たなデータセットに対して、特に数ヶ月スケールの現象の再現性能について、最新の気候学研究に用いられる解析手法を適用することで性能評価を行いそれらをまとめるには至らなかった。全体として 70%程度の達成であったと評価する。二つ目の研究課題である「予測する変数の特性に合わせた前処理手法の検討」については、必要な開発を完了し検証のための計算についても複数