

AI気象モデルの利用可能性調査

令和7年2月
February 2025

気象庁 数値予報開発センター

AI気象モデルの利用可能性調査

目次

1	はじめに	1
2	AI気象モデルを取り巻く状況と先行研究	1
2.1	AI気象モデルの発展と先行研究	1
2.2	海外数値予報センターの動向	3
2.3	その他のAI気象モデル	3
3	AI気象モデルを用いた調査	4
3.1	今回検証したAI気象モデルとその仕様	4
3.2	代表的要素の予測精度	5
3.3	台風予測の基礎検証	7
3.4	予測特性と課題	9
3.5	AI気象モデルを用いたアンサンブル予測	11
3.6	AI気象モデルを境界とするMSM予測	12
3.7	ガイダンス	14
3.8	長期積分	18
3.9	AI気象モデルの利点と課題	19
4	AI気象モデル開発試行	20
4.1	学習データの検討	20
4.2	ファインチューニング	21
4.3	全球AI気象モデルの試作	23
5	まとめ	24

1 はじめに

気象庁における AI(Artificial Intelligence) 利用の歴史は古く、1977 年から機械学習を用いたガイダンスの運用を行い、1996 年からは現在の AI の基盤技術であるニューラルネットワークの利用を開始している(石川 2018)。2006 年に Hinton and Salakhutdinov (2006) による多層のニューラルネットワーク(深層学習)の手法が発表され、2012 年の画像認識大会において深層学習を用いた AlexNet(Krizhevsky et al. 2012) が圧勝したことをきっかけに大きな注目を集めてから、深層学習の改良が急速に進んでいる。また、汎用計算への GPU(Graphics Processing Unit) 利用が進展し、計算手法自体の改良も進んだことにより、深層学習の計算を高速に行うことが可能となり、様々な分野において深層学習が実用化されてきている。このような状況を受け、気象庁は「2030 年に向けた数値予報技術開発重点計画」において、最新 AI 技術を活用した効率的なモデル開発や、更に高精度かつ利用者にとって利便性の高い数値予報プロダクトの作成を掲げている。この計画に従い、気象庁では物理過程を模倣する AI(サロゲートモデル)による計算の高速化、モデルのバイアス補正、観測データの品質管理、最新の AI 技術を用いたガイダンスなどの開発を進めてきた。

一方、近年深層学習による気象予測、特に過去の解析値・観測値の時系列を学習したデータ駆動型気象予測モデル(本報告ではこれを AI 気象モデルと呼ぶ)の研究開発が急速に発展している。このような背景を踏まえ、数値予報課ではその利用可能性について調査を進めてきた。この調査においては、AI 気象モデルに関する先行研究の情報を収集・分析するとともに、公開されている AI 気象モデルの FourCastNetv2(Bonev et al. 2023)¹、Pangu-Weather(Bi et al. 2023)、GraphCast(Lam et al. 2023) を現業利用可能な設定で気象庁のスーパーコンピュータを用いて実行し、数値予報モデル開発者の視点で予測性能や特徴の検証を行い、数値予報への利用可能性を調査して整理した。また、AI 気象モデル開発に関する知見を得ることを目的に、GraphCast のファインチューニングや AI 気象モデルの作成の試行も実施した。

本報告では、AI 気象モデルを取り巻く状況と先行研究についてまとめるとともに、AI 気象モデルの予測精度やその特徴、他のモデルやガイダンスに対する影響等について紹介する。また、ファインチューニングや AI 気象モデルの試行から得られた知見についても報告する。

2 AI 気象モデルを取り巻く状況と先行研究

2.1 AI 気象モデルの発展と先行研究

深層学習の発展を受け、2010 年代後半から深層学習による予測(データ駆動型気象予測)の研究は活発に行われてきた(例えば Dueben and Bauer 2018; Scher 2018; Scher and Messori 2019; Weyn et al. 2019 など)。しかし、この段階では現業予測において実用的と言えるほどの予測精度を持たず、また水平格子間隔・鉛直層などの仕様も現業数値予報モデルに比べて粗いものに留まっていた。

2020 年代に入り、AI 気象モデルの研究が進み、深層学習による実用的な気象予測が大きく発展した。その理由として、計算機の性能向上と自然言語処理や画像処理分野での深層学習技術の発展が挙げられるが、それに加えて WeatherBench(Rasp et al. 2020) の登場は特筆すべき点である(de Burgh-Day and Leeuwenburg 2023)。WeatherBench は全球中期予測を対象としたベンチマークセットで、欧州中期予報センター(ECMWF)の第5次の長期大気再解析(ERA5, Hersbach et al. 2020)を基に AI 気象モデル開発用に整備したデータセットも提供している。また、WeatherBench は欧米の大学や ECMWF などの気象関係者を中心に作成されたものであるが、気象分野の知識を持たない機械学習研究者にも簡単に利用できるように工夫されているのが大きな特徴である。WeatherBench は登場後すぐに研究で利用され、例えば、Weyn et al. (2020) では結果の相互比較を容易にするために WeatherBench と一致する検証指標とデータセットが用いられている。

2022 年には、既存の各国の数値予報センターが運用する物理シミュレーションベースの数値予報の予測精度を上回る革新的な AI 気象モデルが相次いで登場した(Rasp et al. 2024)。これらモデルの多くは気象関係の研究開発機関ではなく、NVIDIA, Google, Huawei といった大手 IT 企業を中心に研究が進められたものである。Keisler (2022) は、ERA5 を学習して水平格子間隔 1° 鉛直層 13 層の AI 気象モデルを開発し、一部の現業全球モデルに匹敵する予測精度を持つことを示した。FourCastNet(Pathak et al. 2022) は 6 時間間隔かつ世界で初めての水平 0.25° 格子の高解像度予測を可能にし、Pangu-Weather や GraphCast は中期予測において多くの要素で全ての現業全球モデルを上回る予測精度を示した。FourCastNet, Pangu-Weather, GraphCast はどれも 1979 年から 2017 年(FourCastNet は 2015 年)の 40 年程度の ERA5 を学習データに用いて、主に全球中期予測を対象として作成されている。

¹ ここでは FourCastNet のうち、Pathak et al. (2022) と Bonev et al. (2023) を区別する場合にそれぞれ FourCastNetv1, FourCastNetv2 と呼ぶこととする。FourCastNetv1 も公開はされているものの、利用は非推奨となっていることから、今回の調査対象からは外している。

表 2.1 全球を対象とする代表的な AI 気象モデルの仕様

名称	FourCastNetv1	FourCastNetv2	Pangu-Weather	GraphCast
開発元	NVIDIA	NVIDIA	Huawei	Google DeepMind
登場時期	2022 年 2 月	2023 年 6 月	2022 年 11 月	2023 年 11 月
学習データ	ERA5	ERA5	ERA5	ERA5, IFS Analysis
AI 手法	AFNO ² (Transformer)	SFNO ³ (Transformer)	3DEST ⁴ (Transformer)	Graph Neural Network
学習に用いた計算機	64 × NVIDIA A100 ⁵	8 × NVIDIA DGX ⁶	192 × NVIDIA V100 ⁵	32 × Google TPU ⁷ v4
学習時間	16 時間	4 時間	16 日	4 週間
論文	Pathak et al. 2022	Bonev et al. 2023	Bi et al. 2023	Lam et al. 2023
名称	FuXi	GenCast	AIFS	Aurora
開発元	Fudan Univ.	Google DeepMind	ECMWF	Microsoft
登場時期	2023 年 10 月	2023 年 12 月	2023 年 10 月	2024 年 5 月
学習データ	ERA5	ERA5	ERA5, IFS Analysis	ERA5, CMCC ⁸ , IFS-HRES Forecast ⁹ , GFS ¹⁰
AI 手法	U-Transformer	Graph Transformer	Graph Neural Network	Multi-scale 3D Swin Transformer + U-Net
学習に用いた計算機	8 × NVIDIA A100	32 × Google TPUv5	64 × NVIDIA A100	32 × NVIDIA A100
学習時間	30 時間	5 日	1 週間	2.5 週間
論文	Chen et al. 2023c	Price et al. 2023	Lang et al. 2024	Bodnar et al. 2024

2023 年においても AI 気象モデルの発展のペースは衰えず、様々な AI 気象モデルが発表され予測精度の改善が続いた (Bonev et al. 2023; Chen et al. 2023a,b,c; Price et al. 2023 など)。これら AI 気象モデルの発展を受け、各国の数値予報センターでも AI 気象モデルの研究開発プロジェクトが 2023 年以降本格化しており、ECMWF による独自の AI 気象モデルである AIFS(Lang et al. 2024) が登場している。引き続き 2024 年においても、Microsoft が複数の解析値やモデル予測の結果を学習して通常の天気予測に加えて大気汚染予測にも適用可能な Aurora(Bodnar et al. 2024) を開発するなど発展が続いている。Aurora は他にも、波浪や熱帯低気圧進路の予測などの複数タスクに適用可能であり、基盤モデルを名乗っている。

表 2.1 に全球を対象とする代表的な AI 気象モデルの仕様を示す。2024 年 11 月の時点では、MetNet-

3(Andrychowicz et al. 2023) など特定の領域を対象とする AI 気象モデルも存在するが、多くの AI 気象モデルは全球予測を対象としている。表に示したように、学習データとして ERA5 を用いたものが多いが、これは ERA5 データの取得のしやすさに加えて、ERA5 を基にしたデータセットを提供する WeatherBench が機械学習の研究者向けに問題を定義したことの影響があると考えられる。

表 2.1 に示した AI 気象モデルでは、Deep Neural Network (DNN) のアーキテクチャに Transformer(Vaswani et al. 2017) および、その画像適用版である Vision Transformer(ViT, Dosovitskiy et al. 2020) を拡張した手法を用いるものが多い。Transformer はもともと自然言語処理分野で革新を起こしたと言われる DNN 技術で、既存の畳み込み/再帰型ネットワークに比べて軽量かつ並列計算がしやすく学習時間を大幅に短縮可能であり、またデータの依存関係について、例えば自然言語処理における長い文章の文頭と文末にある情報のように、距離のある情報の関連性をより正確に捉えることができる。ViT はそれを画像処理において効率的に適用可能にしたものである。これら革新的な深層学習技術が気象予測にも転用された結果、AI 気象モデルの急速な発展がもたらされたことが見て取れる。他

² Adaptive Fourier Neural Operators

³ Spherical Fourier Neural Operators

⁴ three-dimensional Earth-Specific Transformer

⁵ NVIDIA 社の Tensor コア GPU

⁶ NVIDIA の AI 向け GPU プラットホーム

⁷ Tensor Processing Unit

⁸ CMIP6 の気候シミュレーションデータ

⁹ ECMWF の高解像度全球予報値

¹⁰ Global Forecast System の予報値

にも、GraphCast や AIFS のようにグラフニューラルネットワーク (GNN: Graph Neural Network, Scarselli et al. 2009; Pfaff et al. 2020) を採用したモデルも存在する。GraphCast は全球を水平格子間隔を変えつつ階層的に覆うマルチメッシュ構造を採用して効率的な局所および長距離の情報伝播を可能にしている。

表 2.1 に示したすべての AI 気象モデルは学習において GPU またはニューラルネットワークの行列計算に特化した TPU (Tensor Processing Unit) を用いており、例えば比較的登場時期の遅い Aurora では、モデルを生成するための事前学習において NVIDIA A100 を 32 枚利用して 2.5 週間の大規模な計算を行っている。なお、学習時間は DNN アーキテクチャだけでなく、学習手法、学習データ量にも依存するので一概には言えないが、Transformer 系も GNN も長いものでは 1 週間から 1 か月程度の学習時間となっている。

近年登場した実用的な AI 気象モデルは、いずれも基本的には Iterative prediction と呼ばれる手法を採用している。これは、例えば 6 時間など特定の時間間隔を決めて時間発展演算子として DNN を構成し、それを初期値から繰り返し適用することで予測を進める手法であり、物理モデルが時間積分間隔に対する時間発展演算子として構築されるのと似た手法である。一般的には、DNN を用いる場合には必ずしもこの繰り返し手法である必要はなく、例えば 5 日先などの予測結果を直接予測する仕組みも可能であるが、これは対象とする予測時間ごとに DNN を構築し学習することを意味するため、学習コストを考えるとあまり現実的ではない。一方で、繰り返し予測には手法上誤差の蓄積が存在するため、予測精度は直接予測の方が高くなる。そのため、例えば Pangu-Weather では予測時間間隔を 1, 3, 6, 24 時間とする DNN を 4 つ作成し、それらを間隔が長い方から階層的に組み合わせることで、長時間の予測については繰り返し回数を減らす工夫をしている (Bi et al. 2023)。

2.2 海外数値予報センターの動向

海外数値予報センターにおいても AI 気象モデルの研究開発が始まっている。ECMWF は 2023 年 10 月に独自の AI 気象モデル AIFS の最初のバージョンを完成させ (Lang et al. 2024)、アンサンブル予報システム用の AIFS も合わせ、実験的に現業 IFS(Integrated Forecasting System) 解析値を初期として日々実行して予測結果を公開している¹¹。また、Aurora, FourCast-Net, FuXi, GraphCast, Pangu-Weather の各 AI 気象モデルの研究者と連携して、それらモデルについても IFS 初期値から実験的に日々実行して結果の比較を公

開し¹²ている。さらに、世界中の開発者がそれらのモデルを ECMWF のオープンデータを初期値として簡単に実行できる環境を公開する¹³など、AI 気象モデルのオープンな発展を推進している。同時に、今後も AI 気象モデルの開発を推進し、AIFS の高解像度化や気象-気候に関する様々なタスクに対応可能な基盤モデルを開発することを公表している。

英国気象局は 2023 年にアランチュエリング研究所と共同で AI4NWP(Artificial Intelligence for Numerical Weather Prediction) というプロジェクトを開始し、2024 年には独自の AI 気象モデル (FastNet) を開発したと発表した。FastNet の詳細は 2024 年 11 月時点では公開されていないが、発表によると全球を対象とする予測では現業の物理モデルと同等の予測性能が得られているとのことである。また、FastNet を基にして、全球に加えて英国領域を高解像度に扱う領域 AI 気象モデルや、アンサンブル予報モデルの開発を進めていることを報告している。

米国海洋大気庁は 2023 年 11 月に各国の数値予報センターの研究開発者や、巨大 IT 企業の研究者を集めて数値予報のための AI ワークショップ (Artificial Intelligence for Numerical Weather Prediction Workshop; AI4NWP¹⁴) を開催した。2024 年 3 月には AI4NWP の報告書として AI 気象モデルの研究開発戦略をまとめ (Frolov et al. 2024)、研究開発のグランドチャレンジとして AI 気象モデルによる全球アンサンブル予報システム (GEFS-AI) や km スケール AI 気象モデルの開発を目指すことを発表した。

他にも、カナダ気象局や中国気象局、台湾中央気象局などの幅広い国の機関において AI 予測の研究開発・活用を進めていることが報告されている。

2.3 その他の AI 気象モデル

既存の物理モデルにおける力学過程においては、計算安定性や偽の解を抑制するため格子配置に強い制限が掛かるが、DNN を用いた予測ではそのような制限はなく、既存の物理シミュレーションと比較して格子点配置の自由度が高い。特に、画像処理を前提としない GNN 型のネットワークは自由な格子配置が可能であり、一部の領域だけ解像度が高いストレッチ格子など可変格子も採用可能である。ただし、学習データよりも細かい水平格子間隔を扱っても意味がなく、また初期値が必要になるため、ストレッチ格子で高解像度化する対象は高解像度な学習データが存在する領域に限られる。全球モデルの領域の一部を詳細に扱うストレッチ格子を利用した AI 気象モデルの開発が ECMWF と欧州各国の気象局の連携の下で進められており、その

¹¹ https://charts.ecmwf.int/catalogue/packages/ai_models/

¹² 同上、また https://charts.ecmwf.int/products/cyclone?product=tc_tracks_ml_forecast

¹³ <https://github.com/ecmwf-lab/ai-models>

¹⁴ 英国気象局のプロジェクトと名前は同じであるものの、別物である。

枠組を利用したノルウェー気象局による北欧を対象とした全球・領域一体型予測を行う AI 気象モデルが報告されている (Nipen et al. 2024)。このモデルでは事前学習のデータとして水平格子間隔 31km の 43 年間の ERA5 データと水平格子間隔 2.5km の 3.3 年分の領域解析データを学習している。学習には約 6300GPU-hour (AMD Instinct MI250X) の計算機資源を利用しており、表 2.1 で紹介した AI 気象モデルの中では中程度の資源量である。予測精度は地上要素を中心に観測値を真値として検証されており、要素によって良い悪いは異なるが、既存の現業領域モデルに十分に匹敵する予測精度を持つことが報告されている。AI 気象モデルは予測計算のコストが高解像度物理モデルに比べてかなり小さく、この手法は現在別々に計算されている全球予測と領域予測を一体的に計算可能で、単一のモデルをさまざまな用途で使用できるようになる可能性がある。

GenCast(Price et al. 2023) は、水平 0.25° 格子で 15 日先までの確率的な全球アンサンブル予測を行う AI 気象モデルである。GenCast は、生成 AI の一種である条件付き拡散モデル (Karras et al. 2022; Song et al. 2021; Sohl-Dickstein et al. 2015) として実装されており、大気の過去状態と現在状態を入力に用いて、乱数の分布を繰り返し修正することで次の時刻の予測を生成する。この修正過程は Transformer の一種を用いる DNN で、気象場にノイズを加える過程の逆過程となっており、ノイズから気象場を生成することが出来る。乱数を用いるため、予測を複数回実行すると自然とアンサンブル予測 (モデルアンサンブル) を構成することが可能である。GenCast のように、アンサンブル予測を対象としてモデル摂動を表現できる AI 気象モデルの研究開発も進められている。

全球中期予測を対象とする AI 気象モデル以外にも、特定の領域を対象とするものや季節予測を対象とするもの、数値予測モデルの物理過程全体を DNN で置き換えるものなどが存在する。MetNet-3 は Google DeepMind が開発する AI 気象モデルで、米国もしくはヨーロッパの一部を対象に 24 時間先までの領域予測を行う。MetNet-3 は観測データ (レーダーや衛星データ含む) を数値予測の解析値とともに学習している。地上の降水、気温、風、露点温度の確率分布予測を出力するのが特徴的である。

Ocean-linked-atmosphere (Ola) モデルは大気と海洋を扱う AI モデルを結合し大気海洋結合予測を行う AI 気象モデルである (Wang et al. 2024)。学習データは大気が ERA5、海洋は米国大気海洋庁が作成した海洋再解析¹⁵(UFS-replay: Unified Forecast System replay, Orbe et al. 2017) であり、それぞれを個別に学習して AI モデルを作成している。海洋の初期値と内

部情報を扱うため大気の予測可能性を超えて延長予測を行える可能性がある。ただし、まだ研究段階の手法であり予測精度の検証は十分にはされていない。

NeuralGCM は既存の物理モデルと DNN のハイブリッドモデルである (Kochkov et al. 2024)。既存の物理モデルのうち力学過程を用い、それ以外の物理過程とバイアス補正に DNN を用いる。物理過程の一部で DNN を用いる手法 (物理過程のサロゲートモデル) とは異なり、力学過程と DNN の両方を合わせた結果を一体的に学習するため、学習手法の面で AI 気象モデルに近い。力学部分は計算内容が明確であり、既存の知見をもとに解釈や改良が可能である一方で、DNN が持っている計算安定性や小資源で実行可能な点などの優位性は低下する。

他にも、気象予測だけでなく、ダウンスケーリングタスクにも適用可能な基盤モデルを名乗る Prithvi WxC(Schmude et al. 2024) など、様々なコンセプトのモデルが登場している。

3 AI 気象モデルを用いた調査

3.1 今回検証した AI 気象モデルとその仕様

ここでは、第 2 章で報告した先行研究による AI 気象モデルで公開されているもののうち、精度が高いと報告されている FourCastNetv2, Pangu-Weather, GraphCast を対象に利用可能性に関する調査を行った。なお、これらのモデルは、ECMWF によって開発された ai-models という Python パッケージの形で学習済みモデルが公開されている。

今回用いた AI 気象モデルの論文では初期値として学習データと同じ ERA5 を用いている。ERA5 では、なるべく多くの観測データを解析で利用するために、観測データの入電を十分に待ってから解析を行う (遅延解析)。一方で日々の現業の速報解析では、ある時刻で観測の入電を打ち切って解析を行う必要があり、利用できる観測データ (特に衛星データ) が少なくなり、遅延解析に対しては精度面で劣るという特徴がある。遅延解析の方が多くの観測データを同化利用できるため解析精度は高いもの、解析値を利用可能になるまでに時間がかかるため (ERA5 の場合は 5 日以上)、現業では利用することができない。今回の調査では現業で実現可能な設定での評価を行うため、AI 気象モデルの初期値には全球モデル GSM の現業の解析値 (速報解析) を用いた。特に明記がない場合には GSM 解析値を初期値とした結果であり、AI 気象モデルの初期値に ERA5 を用いた結果を示す場合には都度補足する。

各モデルの解像度及び実行に要した計算機資源を表 3.1 に示す。物理モデルでは大気の支配方程式を離散化して計算を行うため、離散化誤差を小さくするために鉛直層数を多くとる必要があるが、AI 気象モデルでは出力データとして必要な鉛直層だけを計算すればよい。いずれの AI 気象モデルも作成時の学習にあたっては

¹⁵ https://psl.noaa.gov/data/ufs_replay/

表 3.1 モデルの解像度及び全球 11 日予測に要した計算機資源

モデル	AI 気象モデル	GSM	
水平解像度	0.25° × 0.25°	約 27km	約 13km
鉛直総数	13 層	128 層	
使用計算機資源	1CPU(Intel Xeon Gold 6338) と 1GPU(NVIDIA A100)	10 CPU (Intel Xeon Max 9480)	98 CPU
実行時間 (I/O 含む)	約 120–290 秒	約 2000 秒	約 1150 秒

表 3.2 GSM と調査に用いた AI 気象モデルの出力データ。GSM では記載された全ての面・要素を出力しているのに対し、AI 気象モデルは下線の面・要素のみ出力可能。* のついている要素のうち、雨と鉛直速度は GraphCast、可降水量は FourCastNetv2 のみ出力可能。

面	要素
地上面	海面更正気圧、 <u>地上気圧</u> 、 <u>気温</u> 、 <u>湿度 (比湿)</u> 、 <u>風</u> 、 <u>雨*</u> 、 <u>雲量</u> 、 <u>可降水量*</u>
気圧面 (hPa) 1000, 975, 950, <u>925</u> , 900, <u>850</u> , 800, <u>700</u> , <u>600</u> , 500, <u>400</u> , 300, <u>250</u> , 200, <u>150</u> , <u>100</u> , 70, <u>50</u> , 30, 20, 10, 7, 5, 3, 2, 1, … (0.01hPa まで)	<u>高度</u> 、 <u>気温</u> 、 <u>湿度 (比湿)</u> 、 <u>風</u> 、 <u>鉛直速度*</u> 、 <u>雲水量</u> 、 <u>雲量</u> 、 <u>渦度</u> 、 <u>発散</u> 、 <u>流線関数</u> 、 <u>速度ポテンシャル</u> 、 <u>ジオポテンシャル</u>

AI 計算に特化した大規模な計算資源を要する (表 2.1) のに対し、実行においては 1CPU と 1GPU で 11 日予測の計算が 5 分以内に終了した。物理モデル (GSM) に比べて、実行に必要な計算資源が非常に少ないことがわかる。このように、AI 気象モデルの作成には大規模な AI 計算機資源を必要とするものの、一度作ってしまえば少ない資源で高速に実行可能という特筆すべき点を確認できる。

次に、調査に用いた AI 気象モデルの出力データを表 3.2 に示す。出力データの水平解像度はいずれの AI 気象モデルも ERA5 プロダクトと同じ 0.25° × 0.25° (約 28km)、出力層は WeatherBench が提案する気圧面 13 層、出力時間間隔は 6 時間 (Pangu-Weather は 1 時間間隔も可能) である。出力要素は AI 気象モデル間で大きな違いはないものの、GSM に比べて出力要素・面が限られていることがわかる。現状で AI 気象モデルが出力していない要素や出力時間間隔が必要になった場合には、何らかの方法で診断して求めるか、AI 気象モデルを再構築する必要がある。

3.2 代表的要素の予測精度

500hPa 高度場などの代表的な要素の予測精度を確認するため、FourCastNetv2, Pangu-Weather, GraphCast に対して、数値予報モデル開発でも標準的に用いられている検証手法や WMO 標準検証¹⁶ による統計的な検証を実施した。

まずは、対 GSM 解析値検証の結果について述べる。図 3.1 に北半球領域 (20°N-90°N) における 500hPa 高度場の対解析値アノマリー相関係数を示す。Pangu-Weather は 5 日先予測以降で GSM を大きく上回って

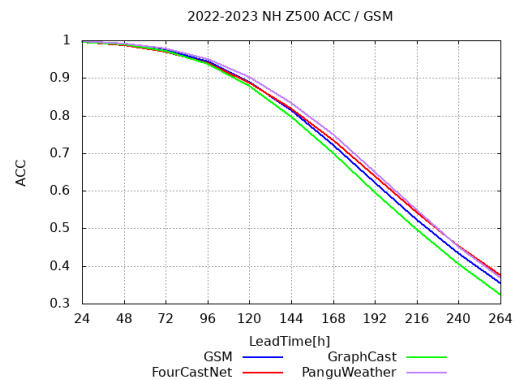


図 3.1 北半球領域 (20°N-90°N) における 500hPa 高度場についての対解析値アノマリー相関係数 (2022 年-2023 年平均値) の予測時間依存性。初期時刻は 00, 12UTC。

おり、高度場のアノマリーを精度良く予測していることが分かる。他の要素についても、特に対流圏中上層の総観規模の予測精度で GSM を上回る要素が複数見られ、AI 気象モデルの予測精度の高さが伺えた。一方で、対流圏下層においては急峻な山岳地帯 (例えばロッキー山脈やチベット高原) を中心に誤差の大きな領域が存在する (図略)。この領域では下層になるほど GSM 解析値と ERA5 解析値の差が広がっており、AI 気象モデルでは ERA5 解析値に近い予測となっていることを確認している (図略)。このことから、検証の真値としている GSM 解析値との差異が下層ほど顕著になったものと考えられる。日本周辺などの山岳地帯が少ない地域に限れば、Pangu-Weather と FourCastNetv2 は GSM と同程度の性能を示した (図略)。GraphCast については、下層から上層に至るまで多くの要素で GSM よりも予測精度が低かった。

¹⁶ 検証の概要については金浜貴史 (2013) を参照。

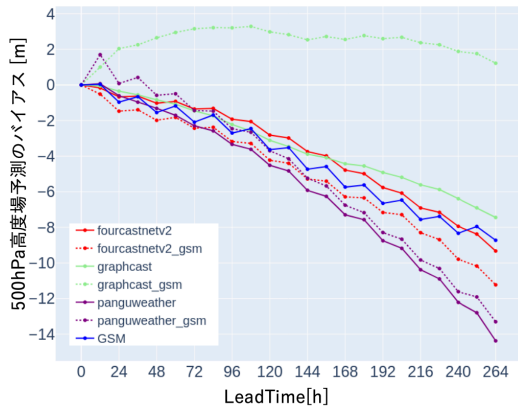


図 3.2 2021 年冬季 (2021 年 12 月~2022 年 2 月) の北半球領域 (20°N-90°N) における 500hPa 高度場予測のバイアス。初期時刻は 12UTC。横軸は予測時間 (時間)。各線はそれぞれ、青線が GSM、紫線が Pangu-Weather、赤線が FourCastNetv2、緑線が GraphCast の結果であり、AI 気象モデルについては実線 (点線) が ERA5 解析値 (GSM 解析値) を初期値とする場合の ERA5 解析値 (GSM 解析値) に対するバイアスを表す。

FourCastNetv2 と GraphCast では予測時間が進むほど予報活動度の著しい低下が見られたが、Pangu-Weather では予報活動度の低下が比較的に緩やかであった (図略)。図 3.2 に 500hPa 高度予測のバイアスを示す。Pangu-Weather と FourCastNetv2 では、予測時間の経過とともにバイアスの拡大が見られた。GraphCast は初期値によってバイアスの予測時間依存性に違いが現れ、GSM 解析値を初期値として用いた場合のバイアス (緑点線) は、11 日先予測まで一貫して ERA5 を用いた場合のバイアス (緑実線) と符号が逆で、その変化傾向も異なっていた。AI 気象モデルでは物理過程に対する特定のスキームやパラメータセット等に仮定を置かないことから、物理過程由来のバイアスは持たない。しかしながら、Pangu-Weather (紫線) や FourCastNetv2 (赤線) では冬季の 500hPa 高度場等において予測時間が進むほどバイアスが拡大する傾向があり、GSM (青線) と同程度かより大きなバイアスを示した。

次に、図 3.3 対ラジオゾンデ検証における北半球 500hPa 高度場の二乗平均平方根誤差 (RMSE: Root Mean Square Error) の改善率を月別で示す。AI 気象モデルは北半球の夏季を中心に GSM より精度が良く、冬にやや悪いことが見て取れる。図 3.4 のスコアカードで対流圏での主要要素の精度を見ると、対 GSM 解析値検証の結果と異なり、北半球領域において下層の要素の精度は特に悪くないことが分かる。スコアの検証対象地点の違いや GSM 解析値との特性の違いの影響が大きいと考えられる。図 3.5 の改善率でみた場合には下層の要素の方が改善幅が大きい、日本付近に着目すると要素ごとに良し悪しがある結果となった。

最後に地上実況気象通報 (SYNOP: surface synoptic observations) に対する地表面要素の検証結果を示す。

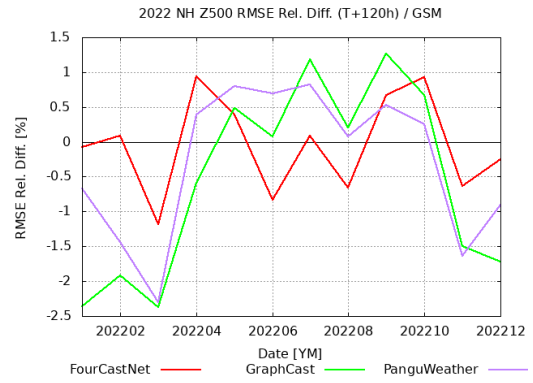


図 3.3 北半球 500hPa 高度場の 5 日先予測における対 GSM での月別 RMSE 改善率 (2022 年: 対ラジオゾンデ検証)。初期時刻は 00, 06, 12, 18UTC。縦軸の値が 0 よりも大きければ (小さければ) AI 気象モデルで改善 (改悪) を意味する。

		Score-Differences Confidence [PanguWeatherG] scores compared to [GSM] period: 202201 / Daily Snapshot Scores from D+1 to D+11				
		NH(90N-20N)	TR(20N-20S)	SH(20S-90S)	JP(110E-150E-20-50)	NWP(100-180E-0-60)
Z500	RMSE	▲▲▲▲▲	▲▲▲▲▲	▲▲▲▲▲	▲▲▲▲▲	▲▲▲▲▲
	CC	▲▲▲▲▲	▲▲▲▲▲	▲▲▲▲▲	▲▲▲▲▲	▲▲▲▲▲
	ME	▲▲▲▲▲	▲▲▲▲▲	▲▲▲▲▲	▲▲▲▲▲	▲▲▲▲▲
T700	RMSE	▲▲▲▲▲	▲▲▲▲▲	▲▲▲▲▲	▲▲▲▲▲	▲▲▲▲▲
	CC	▲▲▲▲▲	▲▲▲▲▲	▲▲▲▲▲	▲▲▲▲▲	▲▲▲▲▲
	ME	▲▲▲▲▲	▲▲▲▲▲	▲▲▲▲▲	▲▲▲▲▲	▲▲▲▲▲
T850	RMSE	▲▲▲▲▲	▲▲▲▲▲	▲▲▲▲▲	▲▲▲▲▲	▲▲▲▲▲
	CC	▲▲▲▲▲	▲▲▲▲▲	▲▲▲▲▲	▲▲▲▲▲	▲▲▲▲▲
	ME	▲▲▲▲▲	▲▲▲▲▲	▲▲▲▲▲	▲▲▲▲▲	▲▲▲▲▲
Ws250	RMSE	▲▲▲▲▲	▲▲▲▲▲	▲▲▲▲▲	▲▲▲▲▲	▲▲▲▲▲
	CC	▲▲▲▲▲	▲▲▲▲▲	▲▲▲▲▲	▲▲▲▲▲	▲▲▲▲▲
	ME	▲▲▲▲▲	▲▲▲▲▲	▲▲▲▲▲	▲▲▲▲▲	▲▲▲▲▲
Ws850	RMSE	▲▲▲▲▲	▲▲▲▲▲	▲▲▲▲▲	▲▲▲▲▲	▲▲▲▲▲
	CC	▲▲▲▲▲	▲▲▲▲▲	▲▲▲▲▲	▲▲▲▲▲	▲▲▲▲▲
	ME	▲▲▲▲▲	▲▲▲▲▲	▲▲▲▲▲	▲▲▲▲▲	▲▲▲▲▲
RH700	RMSE	▲▲▲▲▲	▲▲▲▲▲	▲▲▲▲▲	▲▲▲▲▲	▲▲▲▲▲
	CC	▲▲▲▲▲	▲▲▲▲▲	▲▲▲▲▲	▲▲▲▲▲	▲▲▲▲▲
	ME	▲▲▲▲▲	▲▲▲▲▲	▲▲▲▲▲	▲▲▲▲▲	▲▲▲▲▲

図 3.4 Pangu-Weather における対流圏主要要素のスコアカード (2022 年 1 月: 対ラジオゾンデ検証)。初期時刻は 12UTC。スコアは上から RMSE、相関係数 (CC: Correlation Coefficient)、平均誤差 (ME: Mean Error) を表す。欄内の黄色・灰色はそれぞれ対象の予測時間において統計的に有意に改善・改悪するかを表しており、着色面積が大きいほど統計的有意性が高いことを示す。検証対象の領域はそれぞれ、NH が北半球 (20°N 以北)、TR が熱帯 (20°S-20°N)、SH が南半球 (20°S 以南)、JP が日本周辺 (110°E-150°E, 20°N-50°N)、NWP が北西太平洋領域 (100°E-180°E, 0°-60°N) を意味する。

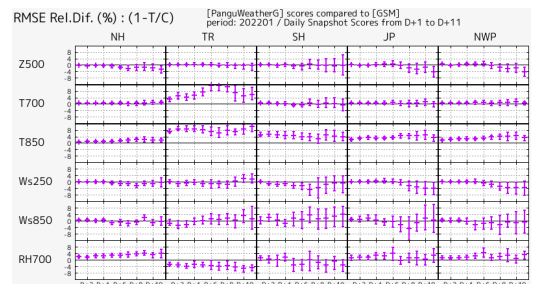


図 3.5 Pangu-Weather における対流圏における主要要素の、予測時間ごとの対 GSM の RMSE 改善率 (2022 年 1 月: 対ラジオゾンデ検証)。初期時刻は 12UTC。改善率が 0 より大きい (小さい) ほど Pangu-Weather の方が改善 (改悪) していることを意味する。検証領域は図 3.4 と同様。

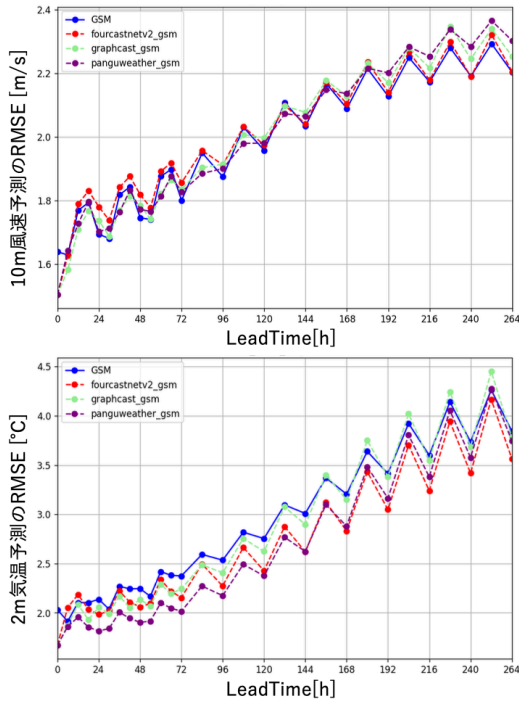


図 3.6 2022 年夏季 (6 月～8 月) の北半球における 2m 気温と 10m 風速予測の RMSE。初期時刻は 00UTC。横軸は予測時間 (時間)。モデル予測値は観測点の最近接格子点の値を使用。ただし、2m 気温予測値についてはモデルと観測点の標高差に応じて気温減率 0.0065 K/m を用いて鉛直外挿した値を使用。

図 3.6 に 2022 年夏季の北半球における 2m 気温と 10m 風速の RMSE を示す。2m 気温では GraphCast 以外の AI 気象モデルがどの予測時間でも GSM を上回る精度を示しているが、10m 風速については、予測時間初期の 1 日目までを除き AI 気象モデルと GSM の精度は同程度となっている。AI 気象モデルの学習元データである ERA5 解析値では SYNOP の 2m 気温が同化利用されているため、AI 気象モデルが 2m 気温予測で高い精度を示した一方で、陸上の 10m 風速観測値を同化利用していないことから¹⁷(Hersbach et al. 2020)、10m 風速については予測精度はそれほど高くなかった要因の一つと考えられる。

3.3 台風予測の基礎検証

数値予報システムの評価において台風予測に関する精度は重要な指標の一つとなっている。AI 気象モデルの先行研究においても、台風予測に関する指標が精度評価の指標の一つとして用いられており、例えば Bi et al. (2023) では、2018 年台風第 25 号や第 26 号の事例を取り上げ、ECMWF の現業モデルである IFS の高解像度予測 (high-resolution forecast; HRES) の進路予測結果よりも Pangu-Weather の予測結果の方が優れていたことを報告している。その他、Lam et al. (2023)

¹⁷ 観測船や係留・漂流ブイによる海上観測値は同化利用している。

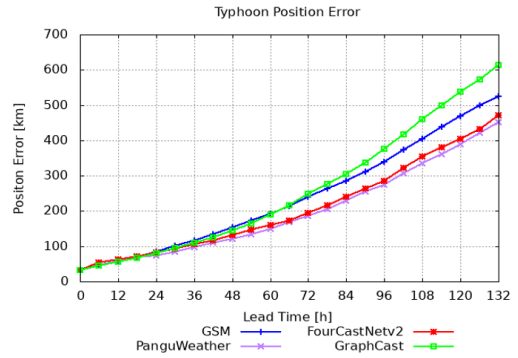


図 3.7 各 AI 気象モデル及び GSM の予報時間ごとの台風進路予測の RMSE。

でも顕著現象の予測精度検証の 1 つとして、熱帯低気圧の進路予測誤差について言及し、GraphCast の進路予測精度は HRES より良好であったことを主張している。ただし、これらの論文では HRES の初期値は IFS の速報解析を用いているのに対し、AI 気象モデルの初期値は ERA5、すなわち遅延解析を用いているため、公平でない検証になっていることに留意が必要である。

ここでは、上記留意点を踏まえつつ、GSM 解析値を初期値とした AI 気象モデルの台風予測について基礎的な検証を行ったので、その結果の概要を報告する。なお、検証期間は 2021 年 7 月 20 日から 2023 年 12 月 31 日とし、その期間において台風が存在した 1339 初期値分¹⁸の予測結果を検証対象とした。

最初に進路予測検証結果について示す。図 3.7 は気象庁ベストトラックを用いた進路予測誤差の統計検証結果を表している。3 日予測の結果に着目すると、Pangu-Weather や FourCastNetv2 の結果は GSM の予測精度を 20 % 程度上回った一方で、GraphCast は同様の精度向上を示さなかった。Pangu-Weather に関しては Bi et al. (2023) で報告されていた熱帯低気圧の進路予測精度の高さが、GSM 解析値からの実行でも確認できた一方で、GraphCast については Lam et al. (2023) での報告と異なる結果となった。後者の要因としては、GraphCast は学習したデータと初期値に利用したデータの特性が異なることによる不整合に鋭敏に反応し、そのことが進路予測精度に影響したことが考えられるが、正確な原因は不明である。また、統計的な予測精度の良かった Pangu-Weather においても全ての事例で進路予測誤差が GSM よりも小さかったわけではなく、例えば、2023 年台風第 2 号の事例では台風の転向を上手く予測できなかったり、2023 年台風第 11 号の事例では複雑な迷走傾向を示したりと、進路予測誤差が大きくなった事例が少なからずあった (図略)。このことは、AI 気象モデルにおいても物理モデルと同じく予測誤差の事例依存性があることを示している。

次に進路予測検証の中でも、系統的な誤差に関する

¹⁸ 00,06,12,18 UTC の初期値の該当分。

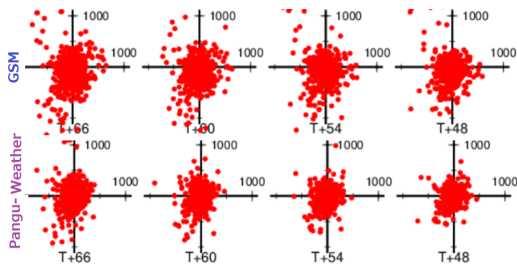


図 3.8 GSM 及び Pangu-Weather の台風進路予測における位置誤差 [km] の散布図。縦軸がベストトラックの台風の進行方向に沿う方向 (AT) の誤差、横軸が同じくベストトラックの台風の進行方向に直交する方向 (CT) の誤差を表す。上段が GSM、下段が Pangu-Weather の結果を示し、右側からそれぞれ FT=48,54,60,66 の結果を示す。

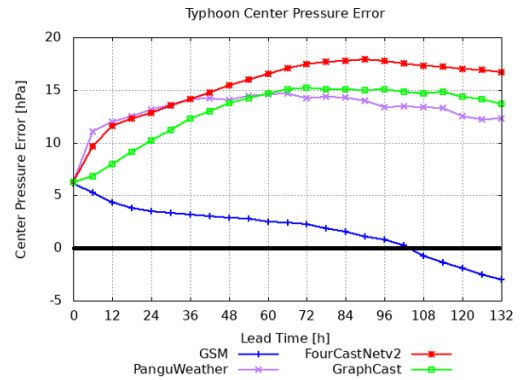


図 3.10 各 AI 気象モデル及び GSM の予報時間ごとの台風中心気圧の予測値の ME。

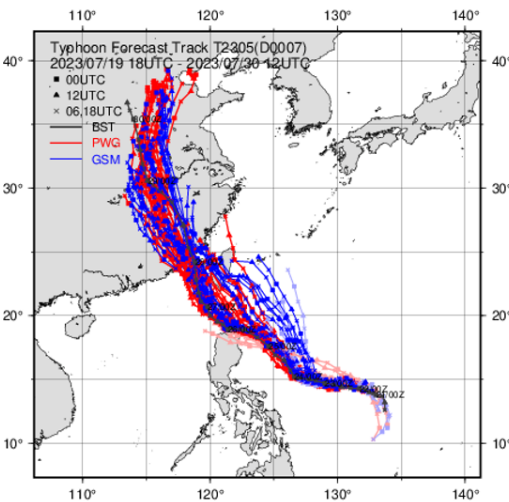


図 3.9 2023 年台風第 5 号の台風経路図。黒線がベストトラック、薄赤線・赤線が 2023 年 7 月 19 日 18UTC から同年 7 月 30 日 12UTC 初期値の Pangu-Weather による予測経路、薄青線・青線が同じ初期値における GSM の予測経路を示す。

調査結果を示す。図 3.8 は進路予測誤差を台風の進行方向に沿った方向の成分 (AT: Along-Track) と進行方向に直交する方向の成分 (CT: Cross-Track) に分けて検証した結果であり、GSM 及び AI 気象モデルの中でも特に進路予測の精度が良かった Pangu-Weather の結果を示している。まず各散布図の CT 誤差 (横軸方向の誤差) に着目すると、どちらのモデルもある程度ばらつきが見られるが、右側の象限に注目すると Pangu-Weather は GSM に比べ比較的 CT 誤差が小さい傾向が確認できる。このことは、経田・越智 (2014) などで GSM の進路予測の課題として指摘されていた台風西進時での北上バイアスが Pangu-Weather では見られないことを示している。ここで一例として図 3.9 に、2023 年台風第 5 号における GSM 及び Pangu-Weather の進路予測の経路図を示す。それぞれの予測を比較すると、フィリピンの東において、台風第 5 号が西進から北西進に変わるタイミングの前後で GSM の進路予測が概ねベストトラックに比べ北側を進む予測となっていたの

に対し、Pangu-Weather は概ね北西進する予測となっており、北上誤差が小さかったことがわかる。一方で、図 3.8 の AT 誤差 (縦軸方向の誤差) の特に 66 時間予測での下側の象限に注目すると、CT 誤差に比べあまりばらつきに差がないことがわかる。これは北上バイアス同様 GSM の課題として知られている転向後のスローバイアスが AI 気象モデルでも少なからず見られることを示していると考えられる。

最後に強度予測検証の結果について述べる。図 3.10 に GSM 解析値を初期値としたときの中心気圧予測の統計検証結果¹⁹を示す。まず 6 時間予測の結果を見ると、どの AI 気象モデルも正のバイアスが大きくなっていることがわかり、特に Pangu-Weather と FourCastNetv2 で顕著である。また、どの AI 気象モデルもおおよそ 3 日予測あたりからバイアスが +15hPa 付近で落ち着くようになり、一貫として中心気圧をベストトラックよりも浅めに予測する傾向にあることがわかる。言い換えると、今回調査した AI 気象モデルはどのモデルにおいてもベストトラックほど発達することを予測できない傾向にある。この傾向の要因として第一に考えられるのが、学習元データセットの特性である。学習元データセットである ERA5 は、1979 年以降のデータにおいてベストトラック (IBTrACS) の中心気圧の同化を行っておらず、Bell et al. (2021) において ERA5 で再現された 1979 年以降の熱帯低気圧の中心気圧はベストトラックよりもかなり浅く表現されていることが報告されている。今回調査した AI 気象モデルはいずれも 1979 年以降のデータを学習していることから、この特性を学習している可能性は十分考えられ、ベストトラックほど発達することを予測できないという傾向の大きな要因であると推察できる。その他、AI 気象モデルが小さいスケールの大気現象を表現できていないことや初期値として与えている解析値の情報不足等も原因と考えられる。

¹⁹ 進路予測検証と同じく気象庁ベストトラックを参照値としている。

3.4 予測特性と課題

AI 気象モデルの各初期値の予測を GSM の予測と比較すると、GSM とは異なった予測特性が複数見られた。用いた初期値は同じであるため、これらの違いは DNN と物理シミュレーションの予測手法の違いに基づくが、特に学習データである ERA5 解析値の水平格子間隔・鉛直層数と表現能力が原因となっていると考えられる。

図 3.11 に東北地方の大雨事例における GraphCast と現業 GSM の降水予測を比較したものを示す。GraphCast の初期値には GSM 解析値を用いており、両者の初期値は同一である。GraphCast も大まかには 6 時間 20 mm を超える降水を予測しているが、6 時間 50 mm を超える強い雨は予測しておらず、ピーク表現が GSM よりも弱く解析雨量との対応が良くない。また、降水域がより広がっている。このようなピークが弱く降水域が広がる GraphCast の傾向は他の事例でも概ね同じであり、ERA5 を入力とした場合でも同様の傾向が報告されている (Liu et al. 2024)。

また、冬型の事例を見ると地上風の収束帯を十分に表現せず、弱い雨も GSM よりも広く予測するなど、面的なメリハリが弱かった (図略)。他にも細かいスケールの擾乱・降水、前線の強化 (frontogenesis) を比較すると、GSM に比べて現象そのものを表現しない事例が目立つ、海上風のピークが弱いなど、小スケールの現象や極端現象の表現は GSM に比べて弱い特徴が見られた。

これらの特徴は、学習データである ERA5 の降水量や地上風速は水平格子間隔が約 30km の全球モデルによる予測結果であり、GraphCast はそれを学習していることや、二乗平均誤差 (MSE: Mean Square Error) 型の損失関数を最小化する DNN ではメリハリを出さず鈍った予測を行う方が有利になる点を考えると自然な結果であると考えられる²⁰(Bonavita 2024)。

図 3.12 に台風について GraphCast とメソモデル MSM の解析値による海面更正気圧と地上風速を比較したものを示す。この事例では、GraphCast は水平格子が約 28 km 相当の解像度であるにもかかわらず、5 km の MSM と比べても台風の等値線に大きながたつきが見られており不自然である。また、GraphCast やこの事例に限らず、どの AI 気象モデルも等値線を細かく見ていくと物理モデルには見られないがたつきが見られた (図略)。この細かいスケールのノイズは海面更正気圧以外の要素でも同様に見られ、空間的な物理量の整合性が小さいスケールまでは十分には取れていないことを示している。なお、ここでの GraphCast の予測では初期値に ERA5 を用いており、DNN を学習した

²⁰ 特に、MSE 型の損失関数を採用する限り、解析時系列における解析誤差や時間発展に不足している情報量の不確実性に対応して平均化された予測になると推測される。学習データの解析解像度程度が表現可能な限界と考えられる。

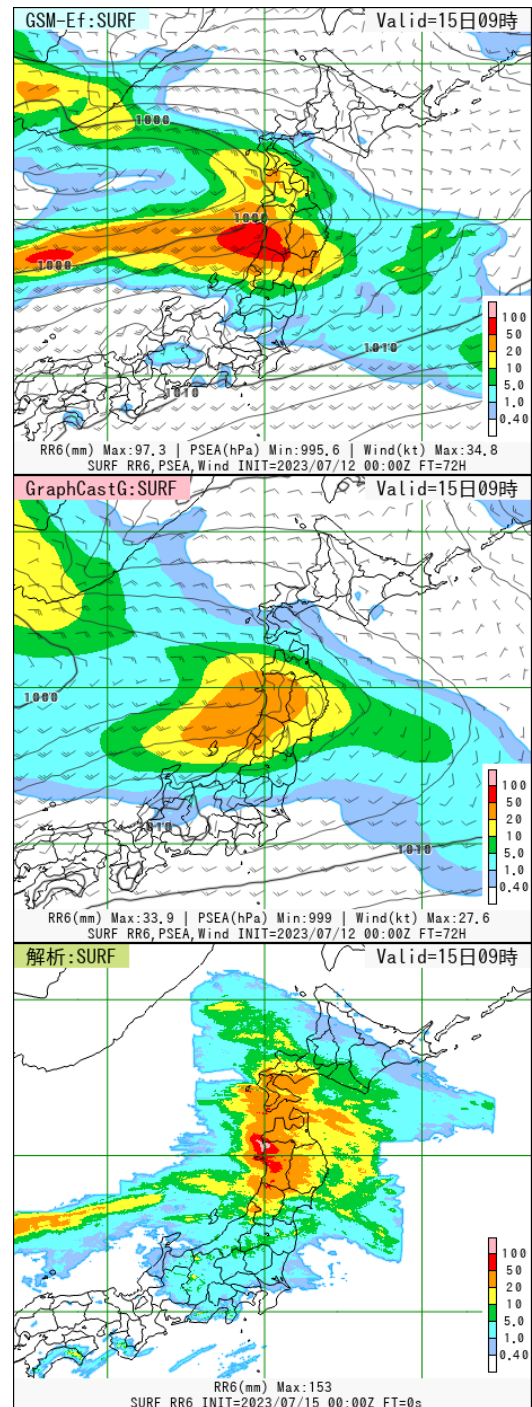


図 3.11 2023 年 7 月 15 日 00UTC を対象時刻とする前 6 時間降水量 [mm] の比較 (塗り潰し)。上段が GSM の予測で中段が GraphCast の予測、下段が解析雨量。GSM と GraphCast は 2023 年 7 月 12 日 00UTC を初期時刻とする 72 時間予測の結果で、コンターは海面更正気圧 [hPa]、矢羽は地上風 [knot] である。

データと異なる特性の初期値を用いていることは原因ではない。このような学習データには見られない特性が表現される理由は不明であり、AI 気象モデルの予測が学習データよりも鈍る傾向とは反しており理由は不明である。

図 3.13 に FourCastNetv2, Pangu-Weather, Graph-

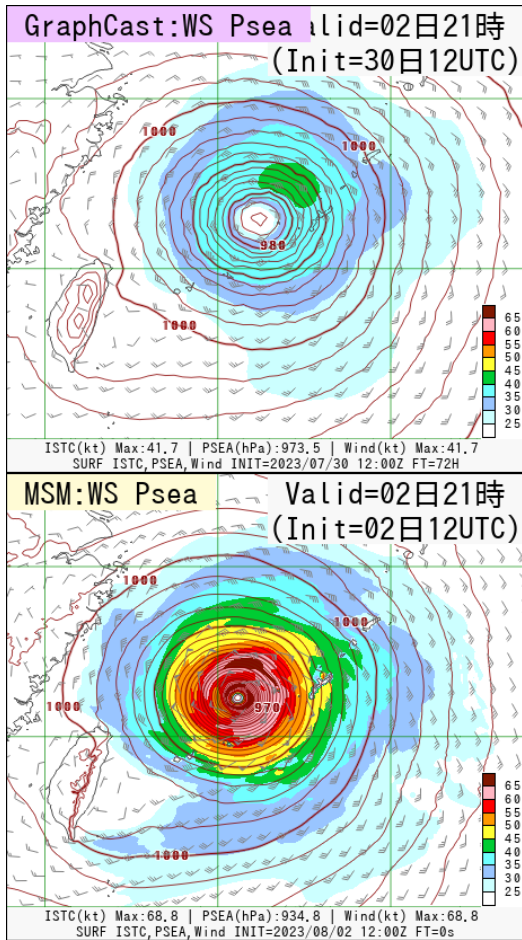


図 3.12 2023 年台風第 6 号についての GraphCast 予測と MSM 解析の比較。コンターが海面更正気圧 [hPa] で塗り潰しと矢羽が地上風速 [knot] (矢羽) の比較。上段が GraphCast の 2023 年 7 月 30 日 12UTC を初期値とする 72 時間予測で、下段が対象時刻の MSM 解析値。GraphCast の初期値には ERA5 を用いている。

Cast それぞれの 500hPa 高度での全球領域の運動エネルギーのスペクトルを GSM と比較したものを示す。どの AI 気象モデルのスペクトルも波数 100 より低波数側で GSM と比べると大きな差はないが、それよりも高波数側では減衰が大きく、波数 400 より高波数側では振動が目立つ。また、特に FourCastNetv2 の高波数側での減衰が大きく、減衰特性はそれぞれの AI 気象モデルの違いに依存している。高波数側での減衰や振動は、活動度の低さ (予測が鈍る) 特性や等値線がたつく特性を波数空間で見ているものだと考えられる。物理モデルの予測と比較してスペクトルの分解能が低い点は Bonavita (2024) で指摘されているものと同様の結果である。

今回の調査では検証対象からは外しているが、FourCastNetv1 を実行した際には予測結果に格子状のノイズが発生する事例がみられた (図 3.14)。このようなノイズが予測に生じると物理モデルでは計算不安定になって異常終了するが、異常終了することなく計算が進むというのも AI 気象モデルの特徴の一つである。

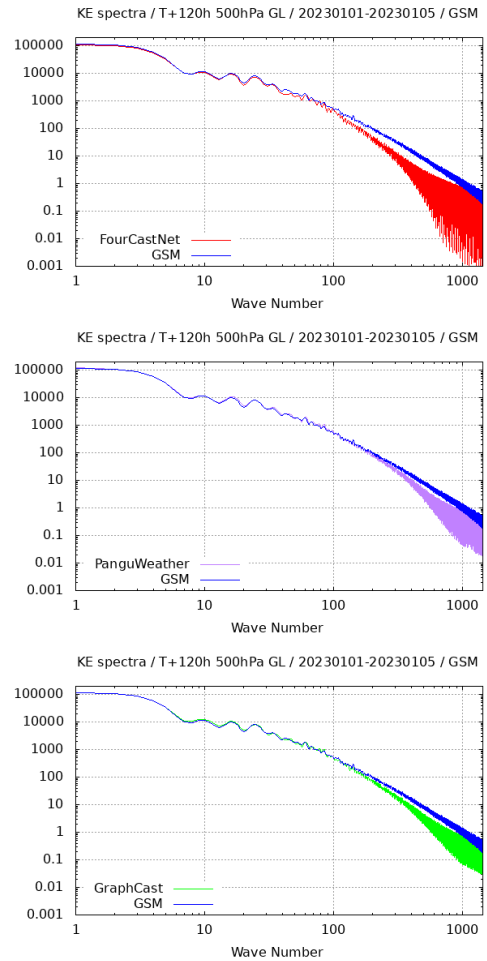


図 3.13 500hPa 高度の単位密度あたりの運動エネルギー (KE) のスペクトルを全球領域で平均したもの。縦軸がスペクトルで横軸が波数番号。KE を東西方向 (格子点数 1440) でフーリエ変換して 2 乗して波数ごとに平均している。上段は FourCastNetv2 (赤線)、中段は Pangu-Weather (紫線)、下段は GraphCast (緑線) の結果。それぞれの図で GSM の結果は共通して青線で示している。2023 年 1 月 1 日 00UTC 初期値の予測について、5 日先までの 6 時間ごとの予測を平均している。3 つの AI 気象モデルの初期値は GSM と同一であり初期値が原因の差異は含まれない。

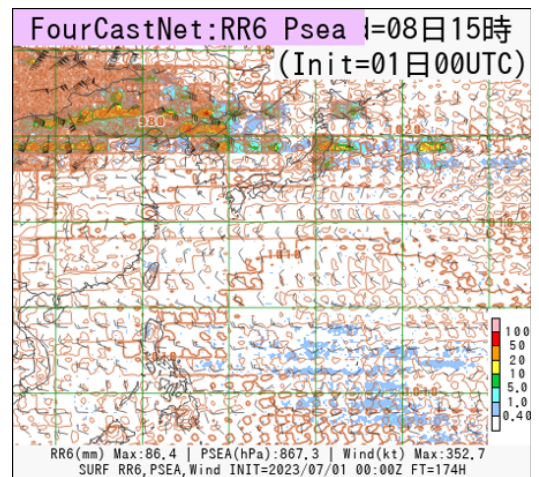


図 3.14 FourCastNetv1 で予測結果にノイズが生じた例。実線は海面更正気圧 [hPa]、塗り潰しは前 6 時間降水量 [mm]。

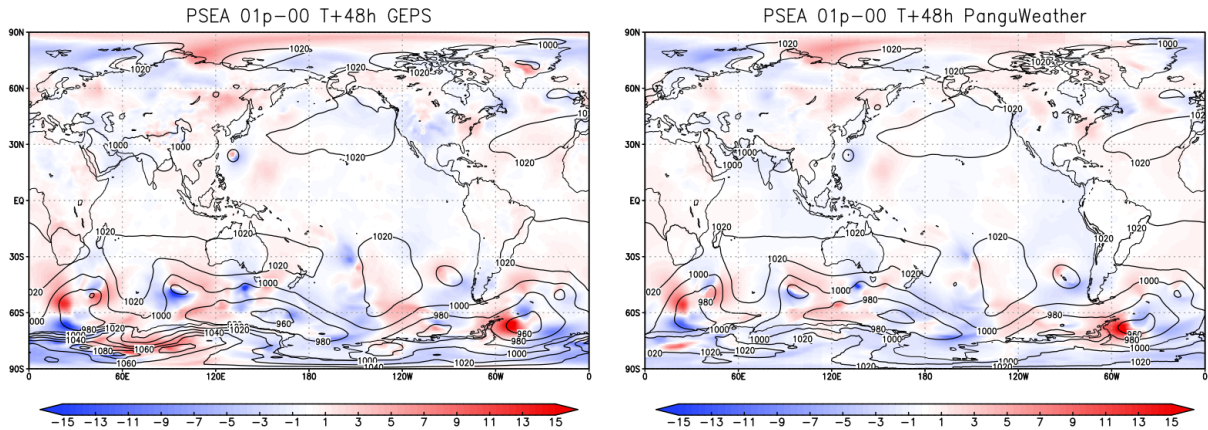


図 3.15 GEPS と Pangu-Weather それぞれでの海面更正気圧 [hPa] の 2023 年 7 月 29 日 00UTC 初期値からの 48 時間予測。左図が GEPS のもので右図が Pangu-Weather のもの。黒実線は摂動なし予測の海面更正気圧 [hPa] を、塗り潰しは摂動の時間発展（摂動あり予測から摂動なし予測を引いたもの）を表す。摂動は GEPS のこの初期値での 01p メンバーのものをを用いている。

3.5 AI 気象モデルを用いたアンサンブル予測

はじめに、アンサンブル予報システム (EPS) で AI 気象モデルを利用する場合の一般論について整理する。アンサンブル予報は多数の予測 (メンバー) を実行するため、計算機資源の使用量が仕様上のボトルネックになる。そのため、現業運用する EPS では、予報モデルの水平格子間隔や予報時間などの仕様を落とし、またメンバーの数を制限することにより計算機資源の量を現業的に実行可能な範囲に収めている。AI 気象モデルの予測計算は物理モデルに比べて使用する資源量が非常に少ないことから (第 3.1 節)、高解像度かつ長時間を予測し、また大メンバー数の構成とすることが可能となる。このため、アンサンブル予報における AI 気象モデルの利用価値は高いと考えられる。ただし、EPS では予測の不確実性を扱うための摂動を用いるが、AI 気象モデルを用いた EPS において物理モデルを対象とした既存の摂動手法が必ずしも利用できるわけではなく、また利用すべきかには議論の余地がある。数値予報モデルの予測誤差の原因は、(1) 初期値の不確実性、(2) 予報モデルの不完全性、(3) 境界値の不確実性に分類できる。物理モデルを用いた EPS では、それぞれを (1) 解析値誤差からの誤差成長を捉える初期値摂動、(2) 物理過程の不完全さを考慮して加えるモデル摂動、(3) 海面や陸面などモデルの境界値の不確実性を表現した境界値摂動により表現する。初期値摂動については、AI 気象モデルが物理モデルに似た摂動応答をするのであれば利用可能であると考えられる。ただし、大メンバーのアンサンブルを構成したい場合、メンバー数分の初期摂動をどのようにして作成するかは検討しなければいけない。モデル摂動は、そもそも予測を行う仕組みが全く異なるため既存の手法は使えない。境界値摂動は、時系列データを学習する AI 気象モデルにおいては予測値とは別に境界値を与える必要

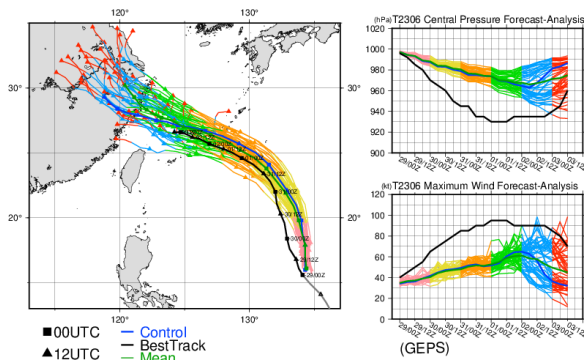
性が弱いため、摂動手法として考慮すべきにそもそも検討の余地がある。

次に、AI 気象モデルがアンサンブル予測にどの程度利用できるかの初期的な確認として、気象庁全球アンサンブル予報システム (GEPS: Global Ensemble Prediction System) の初期摂動を用いてアンサンブル予報実験を行った結果を示す。使用する AI 気象モデルには GSM 初期値からの予測精度が良かった Pangu-Weather を用いた。GEPS の初期摂動は GSM の成長モードや解析誤差を表現したものであり、一般には他のモデルに適用可能とは限らない。特に、初期摂動を加えた初期値は解析値が表現しないような、例えば擾乱が分裂した場など、教師データに含まれない現実には出現しない気象場を表現することがあるため、AI 気象モデルの信頼性は下がると考えられる。

図 3.15 に Pangu-Weather の摂動成長の様子を確認するため、GEPS と Pangu-Weather の海面更正気圧の摂動予測の例を摂動なし予測からの差で示す。図を見ると、AI 気象モデルは摂動のスケールの大きな構造について GEPS と同様のパターンで成長させており、実用的に誤差を補足できる可能性を示唆する結果である。一方で、台風周辺など相対的にスケールの小さい擾乱に伴う摂動は GEPS とは異なり摂動が成長しにくくなっている。これは、擾乱が維持されにくい特性が原因の一つと考えられる。この、総観規模の大きなスケールの誤差は成長するが、小さいスケールの誤差成長は表現しない特徴は先行研究と整合的である (Selz and Craig 2023)。

図 3.16 に 2023 年の台風第 6 号に対する台風進路予測について Pangu-Weather によるアンサンブル予測と GEPS を比較したものを示す。図を見ると、Pangu-Weather も初期摂動に応じて台風進路の不確実性を表現しているが、GEPS に比べて進路の広がりやや小

T2306(D0008) Ensemble Typhoon Forecast
2023/07/29 00UTC init T+0-T+132h/DT6h



T2306(D0008) Ensemble Typhoon Forecast
2023/07/29 00UTC init T+0-T+132h/DT6h

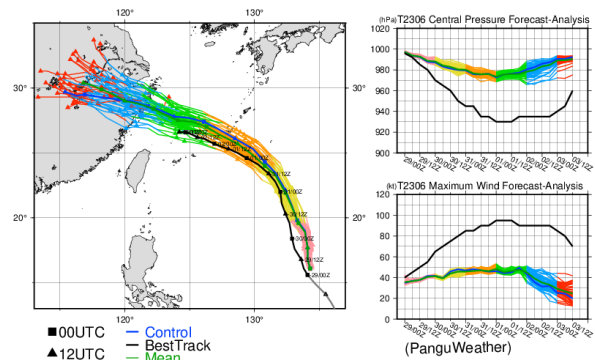


図 3.16 2023 年 7 月 29 日 00UTC 初期値の GEPS と Pangu-Weather それぞれでの 2023 年台風第 6 号に対するアンサンブル予報の比較。左図が GEPS、右図が Pangu-Weather の結果。両者とも GEPS の初期値を用いて 51 メンバーで 132 時間先までの予測を行っている。黒線はベストトラック、青線はコントロールラン (摂動なし予測)、緑線はアンサンブル平均を表す。桃、黄、橙、緑、水、赤線はそれぞれ各アンサンブルメンバーの 24, 48, 72, 96, 120, 132 時間先までの予測を表す。

さい。特に、Pangu-Weather はこの事例において途中で転向するメンバーはほとんど表現していない。また、台風中心気圧の予測はどのメンバーも浅く、海上ではあまりばらつかない。

3.6 AI 気象モデルを境界とする MSM 予測

これまで見てきたように、今回調査した AI 気象モデルにはぼやけた予測となり、スケールの小さい現象や強雨・暴風のような極端現象の予測が苦手であるという課題がある一方で、総観規模の現象の予測精度が高いという利点がある。このため、AI 気象モデルの予測結果を境界値として領域モデルを実行することで、AI 気象モデルによる精度の高い総観場を境界から得つつ、小スケール現象や極端現象を領域モデル内で表現できる可能性がある。ここでは、AI 気象モデルを境界値として MSM を実行した結果について報告する。

ここでの調査では、AI 気象モデルを境界値に用いた場合の影響を簡潔に調べるため、メソ予報のみの評価を行った。現業 MSM では GSM のモデル面データを 1 時間毎に参照しているのに対し、AI 気象モデルのデータを MSM の境界値とする場合は、出力データの制約 (第 3.1 節) からいずれも P 面 13 層、6 時間間隔で参照することになるため、境界値の鉛直解像度及び時間間隔が粗くなる。また、現業 MSM はモデルトップを約 35km としているのに対し、AI 気象モデルは上空 50hPa までしか出力していないため、AI 気象モデルを境界とする実験では MSM のモデルトップを 50hPa より低い約 19.8km に下げ、これに伴い上部境界におけるレイリーダンピング (河野・荒波 2014) をかける層も下げる変更を行った。以後、GSM モデル面データを 1 時間ごとに境界値として参照する MSM (すなわち現業 MSM と同じ構成) を GSM 境界 MSM、AI 気象モデルのデータを境界とする MSM を AI 気象モデル境界 MSM と呼ぶこととする。実験期間は、夏は 2023

年 6 月 1 日 00UTC から 2023 年 8 月 15 日 18UTC、冬は 2022 年 12 月 15 日 00UTC から 2023 年 2 月 28 日 18UTC である²¹。

まず、AI 気象モデル境界 MSM の実行結果について示す。AI 気象モデルの予測では物理的整合性が担保されているとは限らない (第 3.4 節) ことから、その予測値を MSM の境界値とした場合に計算安定性に問題が生じる可能性²²があったが、いずれの実験でも実験期間を通して問題なく予測計算が終了した。MSM の境界値作成処理では境界値データの気圧を直接利用せず、境界値の密度や気温等のプロファイルから静力学平衡を元に気圧の再診断を行っている。このようにして物理的整合性を担保するようになっていたことが、安定に計算できた要因として大きいと考えている。

MSM では予測後半ほどスケールの大きな場の予測が境界値とする GSM と似た予測となることが知られている (越智・石井 2013; 原 2016) が、AI 気象モデル境界 MSM でも、予測が進むにつれて次第に境界値と似た予測となる傾向が見られた。図 3.17 に 2023 年台風第 7 号に対する GSM 及び AI 気象モデルの予測と、それらを境界値に用いた MSM の予測結果を示す。境界値の予測 (図 3.17 左図) と、MSM (図 3.17 右図) で、似た台風進路予測となっていることがわかる。一方で、原 (2016) で指摘されていることと同様に、偏西風の影響が弱い南海上に台風があり、MSM における台風周辺の非断熱加熱の表現が境界値と異なるような場合には、MSM と境界値で異なる進路予測となる事例が見られた (図略)。次に、図 3.17 と同じ初期時刻にお

²¹ AI 気象モデルの出力間隔が 6 時間毎であることと、現状の MSM の境界値処理の都合から、00, 06, 12, 18UTC 初期値の MSM のみを実行している。

²² 例えば、境界値の鉛直プロファイルが静力学平衡から大きく外れた場合、MSM の境界付近で大きな鉛直流が生じうる。

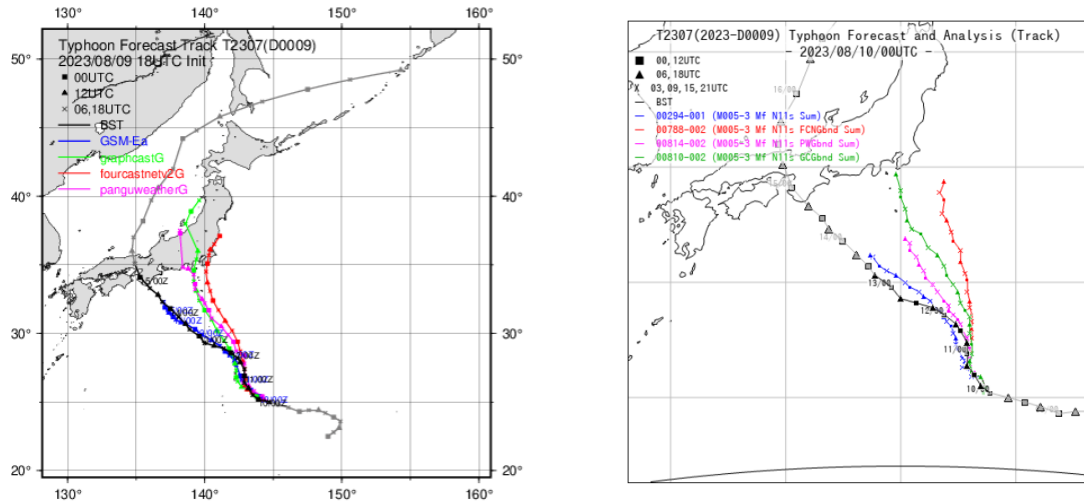


図 3.17 2023 年台風第 7 号進路に対する GSM と AI 気象モデル (左図) と、それらを境界とする MSM の予測 (右図)。黒線・灰線が気象庁ベストトラックを、青線・赤線・桃線・緑線がそれぞれ GSM, FourCastNetv2, Pangu-Weather, GraphCast の予測 (左図)、もしくはそれらを境界とする MSM の予測 (右図) を表す。モデルの初期時刻は左図は 2023 年 8 月 9 日 18UTC、右図は 2023 年 8 月 10 日 00UTC。左図は 132 時間、右図は 78 時間先までの予測。

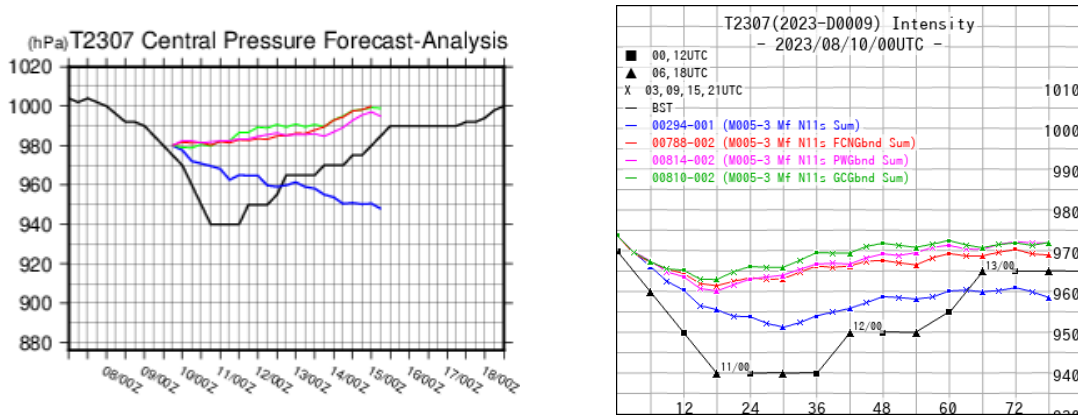


図 3.18 2023 年台風第 7 号に対する GSM と AI 気象モデルの台風の中心気圧の予測 (左図) と、それらを境界とする MSM の予測 (右図)。線の色とモデル初期時刻・予測時間は図 3.17 と同じ。

る台風の中心気圧の予測を図 3.18 に示す。台風の強度予測については、第 3.3 節で述べたように AI 気象モデルの予測では台風の中心気圧が浅くなる特徴があるが、AI 気象モデル境界 MSM では境界値の影響を受けて浅くなる傾向は見られるものの、AI 気象モデルよりも深い台風を予測できていることがわかる。

次に、統計検証の結果を見ていく。図 3.19 に夏・冬それぞれについて、代表的な要素の予測結果に対するスコアカードを示す。FourCastNetv2 と GraphCast を境界にした実験では、ほぼ全ての要素が夏・冬の両方で改悪する結果となっている。一方で、Pangu-Weather 境界実験では予測後半に対ゾンデ観測で改善が見られ、総観場の予測精度が向上する傾向が見られるものの、降水予測は中立から改悪傾向となっている。AI 気象モデルは GSM とは異なるバイアス特性があり、それらを境界とする MSM では予報が進むにつれて徐々に気温や水蒸気場のバイアス特性が変わっていた (図略)。

モデル内の水の相変化を伴う物理過程の応答は特に水蒸気場に対して敏感であり、水蒸気量のバイアスが変化すると降水予測に大きく影響する²³。総観場の予測の精度が良くなった一方で、水蒸気量のバイアスが変化したことで物理過程の応答が変わり、降水予測が悪化したと考えられる。すなわち、現在の MSM は GSM 境界に対して最適化されており、バイアス特性の異なる境界値を用いると予報精度が悪化する可能性があることを示唆している。

台風進路予測では、GraphCast と FourCastNetv2 境界では悪化する傾向が見られるのに対し、Pangu-Weather 境界 MSM では中立の結果となった (図略)。台風中心気圧 (図 3.20) と最大風速 (図略) の予測では、GSM 境界 MSM には予測が進むにつれて台風の中心気

²³ 例えば、下層乾燥バイアスがあるモデルに対して下層水蒸気量の流入を増加させて乾燥バイアスが改善するような変更を行うと、降水が過大になるといった問題が生じうる。

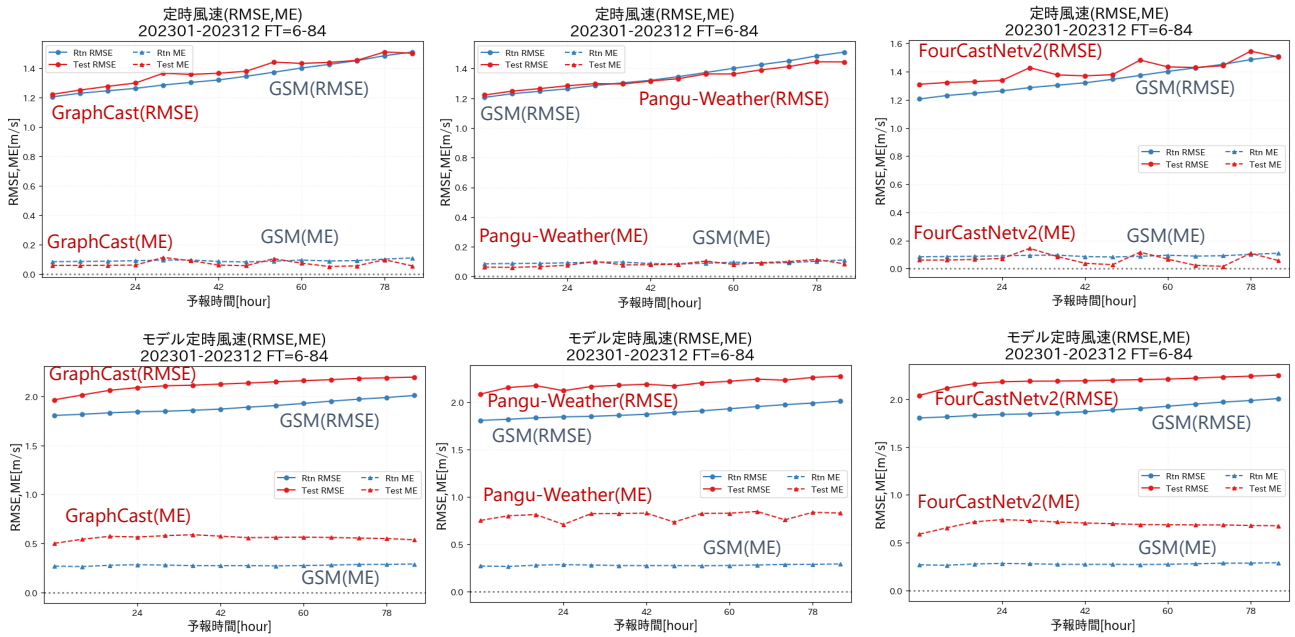


図 3.21 アメダス地点で検証した、予報時間別の定時風速の RMSE (実線) と ME (破線) で、上段はガイダンス、下段は AI 気象モデルと GSM の結果。左から GraphCast, Pangu-Weather, FourCastNetv2 の結果 (赤線) と比較のためそれぞれに GSM の検証結果 (青線) を示す。

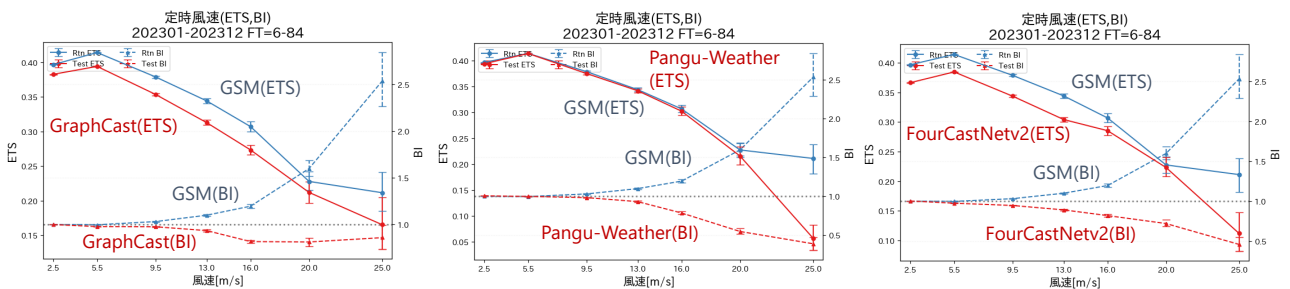


図 3.22 風速別の定時風速の ETS (実線) と BI (破線) で、左から GraphCast, Pangu-Weather, FourCastNetv2 の風ガイダンスの結果 (赤線) と、比較のためそれぞれに GSM 風ガイダンスの検証結果 (青線) を示す。エラーバーは 95 % 信頼区間を表す。

スト、前 3 時間内の最大風速とその風向を予測する最大風速ガイダンスを作成した。AI 気象モデルを入力とした風ガイダンスは 2021 年 7 月 20 日から作成を開始した。初期係数は前日 19 日の GSM 風ガイダンスの係数を利用し、逐次学習により係数を更新している。なお、GSM 風ガイダンスの予報時間間隔は 1 時間ごとだが、AI 気象モデルの予報時間間隔が 6 時間ごとであるため、AI 気象モデルを入力とした風ガイダンスも予報時間間隔は 6 時間ごととなっている。

最初に AI 気象モデルを入力とした風ガイダンスの 2023 年の検証結果を示す。定時風ガイダンスの結果のみを示すが、最大風速ガイダンスについてもおおむね同様の検証結果となることを確認している。

図 3.21 にアメダス地点で検証した、定時風速の予報時間別の RMSE と ME を示す。上段はそれぞれのガイダンス、下段は AI 気象モデルと GSM の検証結果

で、AI 気象モデルと GSM はアメダス地点に最も近い格子の地上風で検証した。RMSE では、GraphCast と FourCastNetv2 の風ガイダンスは GSM ガイダンスに比べて同程度かやや精度が低くなっているのに対し、Pangu-Weather の風ガイダンスについては同程度かやや精度が高くなっている。ME についてはおおむねどの AI 気象モデルの風ガイダンスも GSM 風ガイダンスと同程度となっている。GraphCast と FourCastNetv2 の風ガイダンスについて、FT=30, 54, 78 のスコアでは他の予報時間と比べてギャップが見られるが、これは風ガイダンスが予報対象時刻で層別化しており、FT=6, 12, 18, 24 で学習した係数を、FT=30 以降は同じ対象時刻の係数を繰り返し利用しているためである。入力となる GraphCast と FourCastNetv2 は FT=6 と FT=30 で ME の特性がやや異なっており、この影響でそれぞれのガイダンスでギャップができたと考えられる。Pangu-

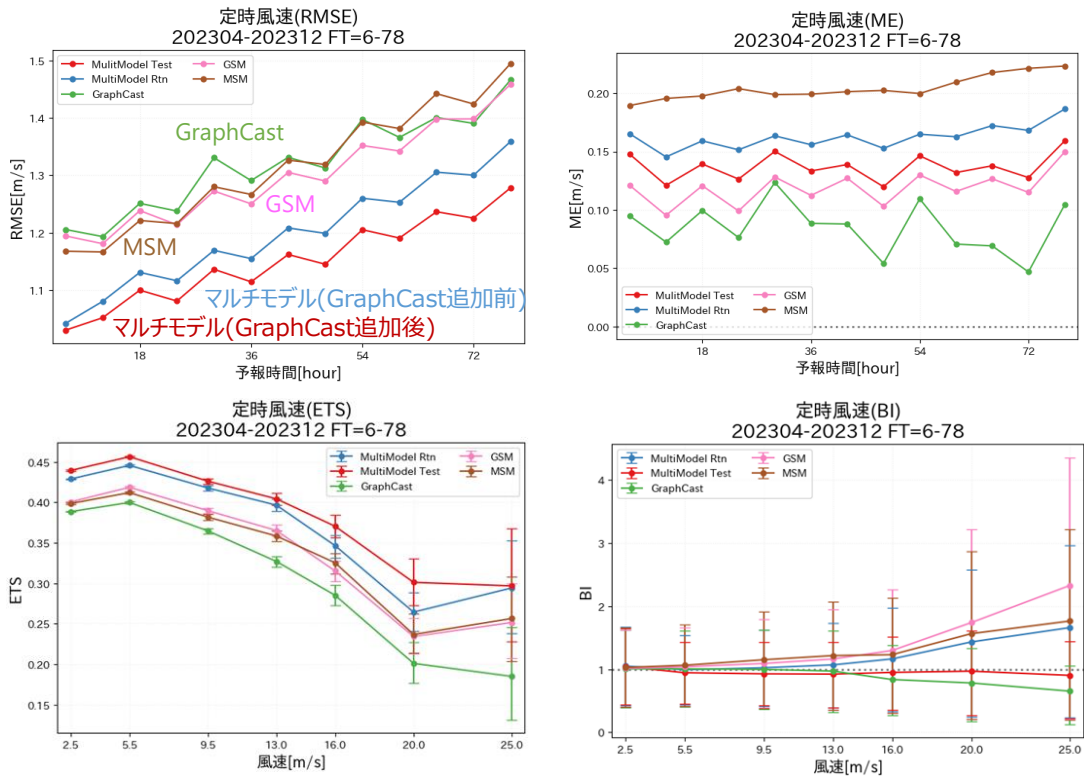


図 3.23 マルチモデル風ガイダンスと GSM, MSM, GraphCast の風ガイダンスの検証結果。マルチモデル風ガイダンスは GraphCast 追加前の結果 (青線) と、GraphCast 追加後の結果 (赤線) を示す。予報時間別の RMSE (左上) と ME (右上)、風速別の ETS (左下) と BI (右下)。エラーバーは 95 %信頼区間を表す。

Weather については予報時間の異なるモデルの組み合わせで長時間の予測をする手法 (第 2.1 節) のため、FT=24, 48, 72 でギャップが見られるが、ガイダンスで同じ係数を利用する予報時間 (例えば FT=18 と 42 や、FT=24 と 48 など) で比較すると同じような予測特性となっているため、風ガイダンスではギャップは見られなくなる。

図 3.22 に、定時風速の風速別のエクイタブルスコア (ETS) とバイアススコア (BI) を示す。GraphCast と FourCastNetv2 の風ガイダンスは、GSM 風ガイダンスと比べて ETS が低くなっている。Pangu-Weather の風ガイダンスについては、風速が 13m/s 以下では GSM 風ガイダンスと同程度となっているが、それより風速が強くなると ETS が低くなる。AI 気象モデルの風ガイダンスは全般に強風の BI が低くなっており、これは AI 気象モデルが台風を浅めに予測する傾向があり (第 3.3 節)、強風の予測が不十分となることや、ガイダンスで強風事例の学習が不足して強風の補正が十分に出来ていないことが影響しているためと考えられる。

図 3.23 に、マルチモデル風ガイダンスと GSM, MSM, GraphCast の風ガイダンスの検証結果を示す。マルチモデル風ガイダンスは GSM, MSM, LFM 風ガイダンスを算術平均したガイダンスである (気象庁 2021)。今回はマルチモデル風ガイダンスに GraphCast の風ガ

イダンスを追加して検証を行った。左上図は予報時間別の RMSE で、GraphCast を追加したマルチモデル風ガイダンスの RMSE が最も小さく、GSM, MSM, GraphCast の風ガイダンスや GraphCast 追加前のマルチモデル風ガイダンスよりも精度が高いことが分かる。右上図は予報時間別の ME で、GraphCast を追加することでマルチモデル風ガイダンスの ME は 0 に近くなる。左下図は風速別の ETS で、こちらも GraphCast を追加したマルチモデル風ガイダンスの ETS が最も高くなっている。右下図は風速別の BI で、GraphCast を追加したマルチモデル風ガイダンスはおおむね 1 に近くっており、強風時の GSM, MSM 風ガイダンスの過大なバイアスを、GraphCast を追加することで緩和することができている。

次に事例を確認する。図 3.24 に、2023 年 8 月 15 日 00UTC の地上天気図と、2023 年 8 月 13 日 00UTC 初期値の FT=48 の GSM, GraphCast, Pangu-Weather, FourCastNetv2 の海面更正気圧と地上風の予測、各モデルの風ガイダンスと GraphCast を追加したマルチモデル風ガイダンスの関西空港と神戸空港の予測と実況を示す。地上天気図では台風第 7 号が淡路島の南にあって、中心気圧は 980hPa で解析されている。FourCastNetv2 では台風が三重県の東に予測されているが、その他のモデルはおおむね実況と同じ位置に予測できている。中心気圧は GSM が 968hPa、GraphCast が

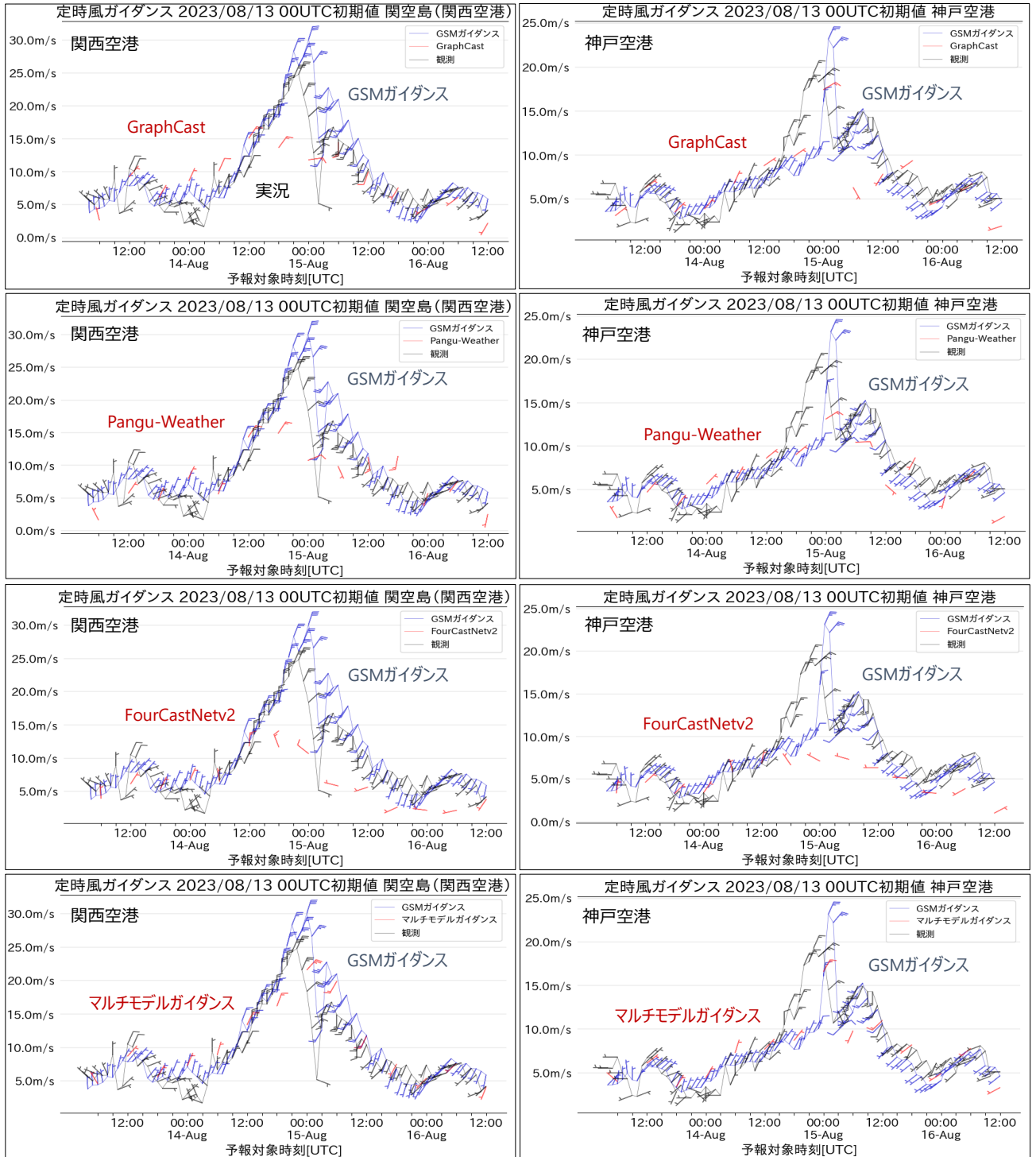
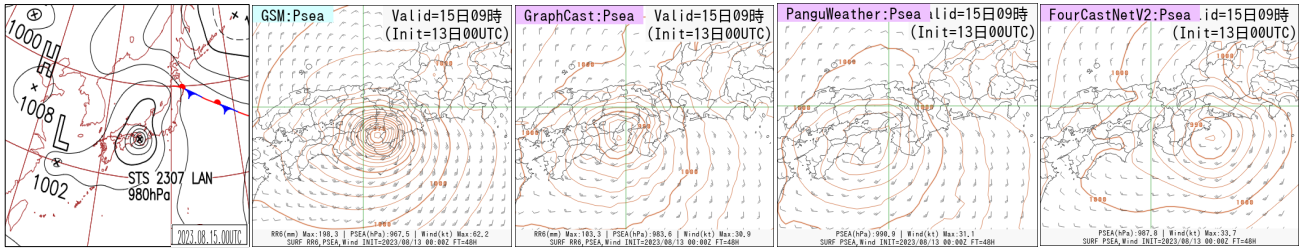


図 3.24 上段は 2023 年 8 月 15 日 00UTC の地上天気図と、2023 年 8 月 13 日 00UTC 初期値の FT=48 の GSM, GraphCast, Pangu-Weather, FourCastNetv2 の海面更正気圧と地上風の予測を、2 段目以降は各モデルの風ガイダンスと GraphCast を追加したマルチモデル風ガイダンスの関西空港 (左列) と神戸空港 (右列) の予測と実況を示す。GSM 風ガイダンス (青矢羽) と実況 (黒矢羽) は 1 時間ごと、GraphCast, Pangu-Weather, FourCastNetv2 の風ガイダンスとマルチモデル風ガイダンス (赤矢羽) は 6 時間ごとに描画している。

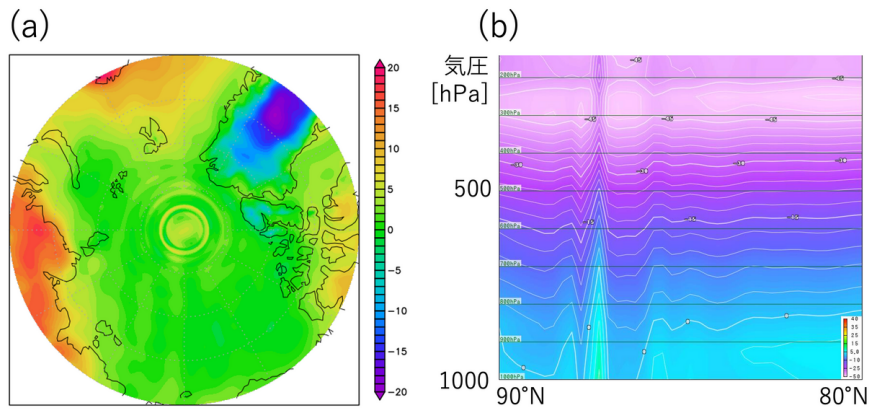


図 3.25 FourCastNetv2 の予測場に見られた不自然な分布の例。2021 年 7 月 20 日 12UTC からの 15 日（360 時間）予測における北極付近の気温分布。(a) 850 hPa 気温 [°C]、(b) 本初子午線に沿った緯度-鉛直断面 [°C]。

984hPa、Pangu-Weather が 991hPa、FourCastNetv2 が 988hPa となっており、GSM は実況と比べ台風を深めすぎて、AI 気象モデルはやや浅めの予測となっている。関西空港と神戸空港ではピーク時にそれぞれ 24.8m/s、19.3m/s の強風を観測したが、各モデルの風ガイダンスの予測を見ると、GSM 風ガイダンスは実況と比べやや強め、AI 気象モデルの風ガイダンスはいずれも弱めの予測となった。この事例に限らず台風の予測では、AI 気象モデルの風ガイダンスは強風が予測されにくい傾向があった。マルチモデル風ガイダンスについては、おおむね実況に近い風速を予測できている。風向については、FourCastNetv2 は台風中心位置の予測がずれているため、北東風系から南西風系に変わるタイミングが早い、GraphCast、Pangu-Weather、マルチモデル風ガイダンスについてはおおむね実況に近い風向を予測できている。

以上の結果から、AI 気象モデルの風ガイダンスはおおむね GSM 風ガイダンスと同程度かやや精度が低く、特に強風の予測頻度が低いことが分かった。学習が進めばある程度改善する可能性はあるが、台風事例については強風が予測されにくい傾向が残ると思われる。マルチモデル風ガイダンスについては AI 気象モデルを入力とした風ガイダンスを追加することで、マルチモデル風ガイダンスの精度が上がることを確認した。

3.8 長期積分

FourCastNetv2, Pangu-Weather, GraphCast の 3 つの AI 気象モデルを用いた一か月程度の長期積分の特性を検証した。

これらの AI 気象モデルでは、海洋内部や陸面過程等の初期値情報を持たないため、それらの効果（第二種の予測可能性）が支配的になる長いリードタイムでの予測には原理的な限界がある。よって、現在大気海洋結合モデルが担っているような季節予報は調査対象外とし、全球アンサンブル予報システム（全球 EPS）が担っている 34 日目までの季節内予報を対象とした。

なお、これらの AI 気象モデルは決定論的な中期予報を想定して開発されたモデルであり、11 日目より先の予報は文献でも詳細に調べられていない想定外の利用である。

長期積分実験として、3 つのモデルに対して 2021 年 7 月 20 日から 2023 年 12 月 31 日の 12UTC を初期時刻とする 895 事例を実行した。積分は全ての事例で正常に完了し、いくつかの要素の予測場を目視で確認した限りでは、多くが自然な予測となっていた。その一方で、FourCastNetv2 の予測場を細かく見ると、予測が進むにつれて両極で同心円状の歪みが発達する様子が多くの事例で確認された（図 3.25）。

統計的な評価として、同期間のうち初期時刻が火曜または水曜であるものを全球 EPS ルーチンのコントロールランと比較した。なお、ここで検証の対象とした 2021 年以降の事例は各モデルの学習期間と重複しないため、長期積分の検証で通常行われるような過去数十年間に遡る再予報型の実験よりも全球 EPS との比較検証に適していると考えられる。検証には GSM 解析値の等気圧面データを 1.25° の等緯度経度格子に変換したものをを用いて対解析誤差を評価した。

図 3.26 に、夏を初期値とする 3-4 週目の 500 hPa ジオポテンシャル高度と 850 hPa 気温の、平均誤差 (ME: Mean Error) を例として示す。気温に着目すると、チベット高原やグリーンランド、南極大陸と言った標高の高い地域で AI 気象モデルの ME が大きくなっていることがまず目に付くが、これは学習データである ERA5 と検証データである GSM 解析値の差に起因することが分かっている（第 3.2 節）。これを差し引いて見比べると、夏を初期値とする 850 hPa 気温の ME は、FourCastNetv2 が全球 EPS と同程度、Pangu-Weather と GraphCast は明らかに全球 EPS より大きくなっている。空間分布を見ると、南太平洋（ペルー沖）の高温など一部の ME は AI 気象モデルに共通しており ERA5 と GSM の差に起因するとみられる。一

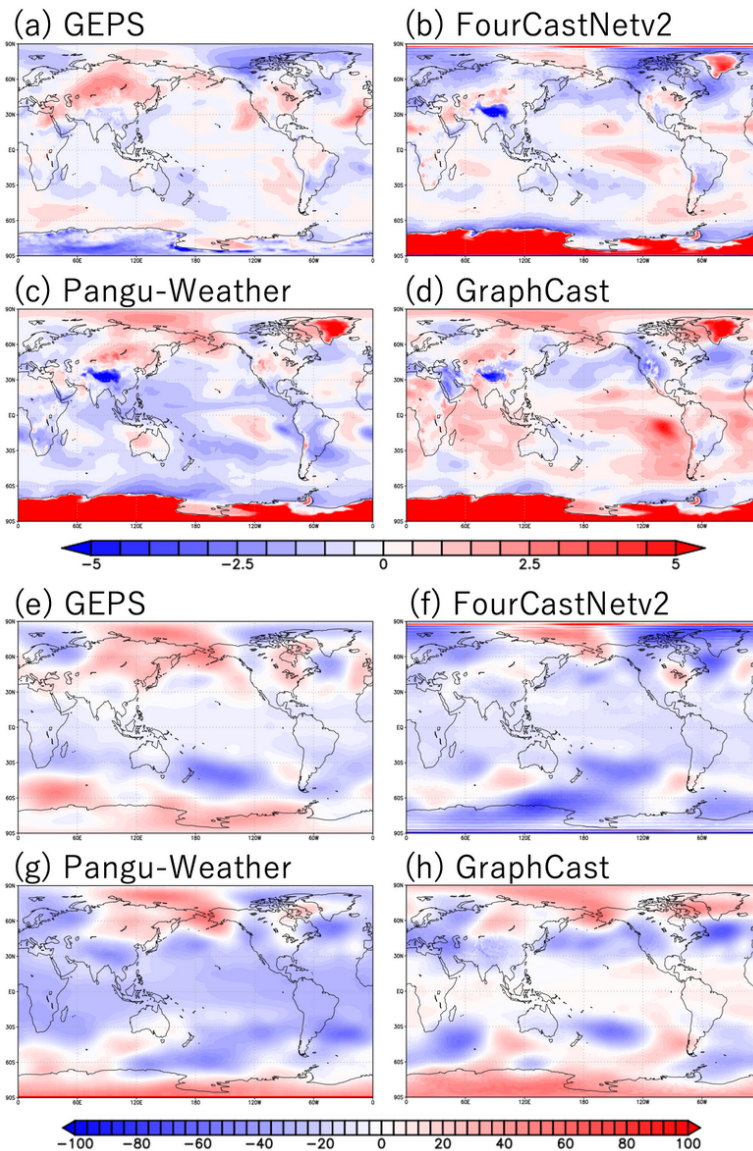


図 3.26 夏(6-8月)を初期値とする全球 EPS コントロールラン及び AI 気象モデルの 3-4 週目(17-30 日目) 予測の GSM 解析値に対する ME。(a-d): 850 hPa 気温 [°C]、(e-h): 500 hPa ジオポテンシャル高度 [m]。モデル名は各パネルに記した通り。

方で、同じ ERA5 データから学習したはずのモデル間にも大きな ME の差がある点は興味深い。予報時間に伴う全球的な ME の発展を見ると、Pangu-Weather は予報時間に伴い低温方向にドリフトする傾向が卓越し、GraphCast は予報初期に一度高温化した後次第に低温化する様子が見られていた。夏初期値の 500 hPa ジオポテンシャル高度の ME は、GraphCast が全球 EPS と同程度で、FourCastNetv2 と Pangu-Weather は全球 EPS よりも大きな傾向だった。各モデルの ME の傾向は要素・季節に依存するが、一般的には AI 気象モデルの ME が全球 EPS コントロールランより速く拡大するものが多かった。RMSE の大小に関しては、ME の大小と概ね整合的で、全球 EPS コントロールランに及ばない要素・季節が多かった(図略)。

長期積分の結果から、中期予報用に開発された AI 気象モデルの予報時間延長には限度があることが確かめ

られた。長期予報を精度良く実行するには、それに合わせた設計(繰り返し予報の時間間隔、評価関数、学習対象の予報時間など)が必要と考えられる。

3.9 AI 気象モデルの利点と課題

AI 気象モデルは既存の物理モデルに比較して少ない資源で高速に動作するのが大きな利点の一つである。今回調査した FourCastNetv2, Pangu-Weather, GraphCast はどれも 1GPU と 1CPU を用いて 5 分以内に全球の 11 日予測が可能であり、これは同程度の水平格子間隔を持つ GSM の低解像度版が 10CPU を用いて 35 分程度であるのに比べてかなり高速である(表 3.1)。なお、この比較はファイル I/O を含むものであり、予測演算のみで比較すればその差はさらに広がる。一方で、使用メモリに関しては、実行に大きなメモリの搭載されたノードを必要とする点は GSM と同じであった。

今回調査した3つのAI気象モデルはどれもGSM解析値を初期値として安定に動作し、大規模な時空間スケールの流れの表現等において既存の数値予報モデルのそれを良く模倣できていた。1日4回約3年間分の初期値に対して予測を実行したが、それらの結果において大規模なスケールに限れば不自然な予測は見られず、利用可能性は高いと考える。特に、台風進路予測の精度は高く、Pangu-WeatherとFourCastNetv2の予測精度がGSMを20%程度上回り、また多くの物理モデルが持つ北進バイアスがほぼ見られなかった。

一方で、3つのAI気象モデルとも発達した台風が予測されず、暴風が予測されない、極端現象が十分に表現されない、JPCZや前線の降雨による強化など小スケールの現象が出ないなど、学習元データであるERA5解析値の解像度・表現能力や、学習手法に由来すると考えられる課題が見られた。また、AI気象モデルには、等値線が波打つ、台風が移動した後小渦が残るなど非物理的な挙動が稀に見られるものがあった。学習データの信頼できるスケールに対してスケールが小さい現象の予測には問題が多いと考えられるため、より高解像度で精度の高い解析値を学習するのが良いと考えられる。また、小スケールの現象を良く表現するための学習手法の改良も重要になると考える。

AI気象モデルはGEPSの初期摂動を加えた初期値から動作し、大きなスケールでの初期摂動の成長を表現した。一方で小さなスケールの現象では摂動が十分に成長せず、初期摂動以外の摂動を考慮できていないことも合わせ、スプレッドの広がりは十分ではなかった。GenCast(Price et al. 2023)やFuXi(Chen et al. 2023c)などの先行研究ではそれらを補うAI気象モデルの研究が進められている。

AI気象モデルの予測をMSMの境界値として用いた場合、いずれのAI気象モデル境界値でも、GSM境界値と比べて、台風進路予測と降水予測はともに中立から悪化であった。これは現在のMSMがGSM境界に対し最適化されていることを示唆しており、AI気象モデルを領域予測で活用するには、第2章第2.3節で紹介した全球・領域一体型の予測手法を含めて検討することが良いと考える。

AI気象モデルの予測値をガイダンスの入力として利用する場合、計算されている要素の数や予測時間間隔が不十分であることが課題である。既存のガイダンスのうち、3つのAI気象モデルの予測結果に対して実行可能なものは風ガイダンスのみであった。今後、雲量や雲水量をはじめとする予測要素の追加が課題になると考えられる。各AI気象モデルの風ガイダンスの精度は、GSM風ガイダンスと比べ同等かやや低かったが、マルチモデル風ガイダンスの入力として追加するとその精度を改善した。

3つのAI気象モデルともGSM解析値からの34日積分を正常に完了したが、モデルによっては予報時間

が進むにつれて局所的に不自然な分布となるなどの問題も見られた。また、3から4週目の予測にかけて多くの要素でMEやRMSEがGEPSコントロールメンバよりも早く悪化した。これらから、中期予報用に開発されたAI気象モデルの予報時間延長には限度があると考えられる。また、物理法則の制約を考えると、更に必要な予測精度を確保しつつ予報時間を延長するためには、大気にとっての境界値である海洋や陸面等との相互作用を考慮する必要がある。

4 AI気象モデル開発試行

4.1 学習データの検討

AI気象モデルは基本的にはIterative prediction手法を採用しており時間発展演算子としてDNNを構成する。そのDNNは大気の時系列データを学習しており、時系列データを生成する時間発展演算子を近似しているものになる。大気を一様にカバーする時系列データとしては、データ同化による解析値の時系列と、物理シミュレーションによる予測値の時系列が存在する。物理シミュレーションによる時系列データの生成における時間発展演算子は物理モデルであり、データ同化による時系列の時間発展演算子は第一推定値を観測情報で修正する手法から物理モデルと解析を合成したものになる。両者とも物理モデルを用いているが、データ同化では物理モデルの時間発展を観測情報で修正している分、時間発展演算子としての精度が高い(現実大気系に近い)。本稿で紹介したAI気象モデルは、解析値の時系列を学習することで物理シミュレーション(既存の物理モデル)による予測の精度を超えることを可能にしており、その精度の改善量は解析値の精度、つまり物理モデルが表現する気象場の時間発展を観測情報により修正した精度向上量に制限される。また同時に、水平格子間隔・鉛直層数、予測時間間隔等の仕様も解析値の仕様に縛られる。その解析値の精度や仕様は、解析値を作成する解析システムに依存し、物理シミュレーションにはない制限に拘束される。

表4.1でデータ同化による解析値の時系列と、物理シミュレーションによる予測値の時系列について、学習に関連する特徴を比較する。解析値は観測データの情報を含む精度が相対的に高い点が大きな利点である。一方、物理シミュレーションの結果を観測情報を用いて解析で修正する仕組み上、精度を修正できる水平スケールが観測密度に実質的に制限される点や、計算コストの高さから高解像度の計算が難しいこと、観測データが存在しない期間は作成できないこと、特に衛星データが豊富にある期間が限られるため精度の良い解析値を作成できる期間が限られる点が不利な点である。このため、AI気象モデルで既存の物理モデルの精度を改善できる可能性がある対象には、解析システムの解像度程度の大きなスケールまで、対象とする現象が観測データの豊富に存在する40年程度の期間で十

表 4.1 学習データの比較

作成手法	データ同化	物理シミュレーション
作成手法の詳細	物理シミュレーションの結果を観測データにより修正	流体の支配方程式(力学過程)と物理過程を数値的に計算
データの精度	高い (観測データの量・質とデータ同化システムの性能による)	低い(予報モデルの性能による)
水平解像度	低い(計算コストの高さと観測密度が束縛)	高い
期間の長さ	短い(観測データの存在期間による)	長い(計算資源の許す範囲で可能)

分に学習可能な程度に高頻度であるなどの制限が掛かると考えられる。

近年登場した多くの AI 気象モデルは ERA5 を学習データとして利用しているが(表 2.1)、一般的に再解析とは現業の客観解析(現業解析)に対して過去の期間について客観解析を再び行うことおよびその生成物を指し、主に長期再解析と機械学習(ガイダンス)用データセットの 2 種類が存在する。長期再解析は、数十年以上の長い期間を対象とした均質なデータセットであり、気候変動や予測可能性などの研究用途に加え、GPV 気候値(平年値)の作成や気候予報モデルの初期値としても利用される。通常、長期再解析の開発を開始した時点での現業解析システムをベースに、利用可能な計算機資源を考慮してその低解像度版を用いて作成される。機械学習用データセットは、数値予報モデル更新に際して主に各種ガイダンスを再学習するためのデータセットである。再解析を初期値とする予報(再予報)を行い、ガイダンスでは主にその再予報を利用するが、再解析部分の作成コストが大きい。通常、数値予報システムが改良され、現業導入が決まった段階でその次世代の数値予報システムを用いて作成される。AI 気象モデルの学習のためには、解析時系列の期間は長い方が良く、一方でより新しく解像度などが高い方が良いため、長期再解析と機械学習用データセットの中間の性質を必要とする。

現業解析値も学習データの候補である。再解析と比べて、現業解析システムは運用中に水平格子間隔などの仕様が変わることがある点や、解像度等のスペックが高いため保存コストが高く保存期間が短くなりやすいなどの欠点もあるが、現時点で最も解像度が高くまた精度が高い利点がある。また、AI 気象モデルを実際に運用する場合には、その解析値を初期値として AI 気象モデルを実行するため学習する価値が高い。再予報も現業の解析システムがベースとなるため、学習データの精度や水平格子間隔等の仕様を向上させるためには、いずれにせよ現業解析システムの高度化が必要になる。AI 気象モデルの学習データを作るニーズの面から、解析システムの高精度化や高解像度化など仕様向上に対する期待が高いと考えられる。

AI 気象モデルの最新の動向として、複数の学習デー

タを組み合わせる学習する発展を見せており、大規模言語モデルなど他の AI と同様に、今後学習データの大規模化・多様化(基盤モデル化)を志向する可能性がある(Bodnar et al. 2024; Schmude et al. 2024; Nipen et al. 2024)。学習データの組み合わせとしては、再解析と現業解析、全球解析と領域解析、解析システムの異なる解析値、解析値と予測シミュレーション、解析と観測リトリバーなど多様である。この方向で発展が続くならば、AI 気象モデルを作成するにあたっては、既存の様々な解析値を入手して目的に合わせて適切に組み合わせることが基本になる。AI 気象モデルは、水平格子間隔や出力時間間隔などの仕様が学習データの仕様に、予測精度が学習データの精度にそれぞれ制限されるため、よりスペックと精度が高い学習データをより多く集めて作成することが重要である。特に、日本周辺の気象を高解像度かつ高精度に予測する場合、解像度と精度がより高い領域解析値を利用することが必須である。

4.2 ファインチューニング

ファインチューニングとは、モデルを生成する際の学習(事前学習)に用いたデータセットとは異なるデータセットを用いて追加学習を行い、異なるデータセットの入力に対して適応できるようにパラメータの調整を行う手法のことである。AI 気象モデルにおいても GraphCast の他、AIFS についても予測精度向上を目的に事前学習は ERA5 で行うものの、IFS 速報解析値を用いてファインチューニングが行われている(Lam et al. 2023; Lang et al. 2024)。今回、学習用コードや学習済みモデルが公開されている GraphCast を用いて、AI 気象モデルの学習に関するノウハウの蓄積や、ファインチューニングによる予報結果に対するインパクトを調査することを目的に、GSM 解析値を学習データとしたファインチューニングを行った。

今回試行したファインチューニングの仕様について表 4.2 に示す。AI モデルの予測と学習データとの誤差の大きさを表す損失関数(Loss function)には MSE を用いることや、学習要素別の損失関数への寄与率を表

²⁴ 前 6 時間降水量は、該当時刻の 6 時間前の解析値からの 6 時間予測、すなわち解析-予報サイクルの予報値を利用。

表 4.2 ファインチューニングの仕様

学習データ	GSM 解析値
解像度 (格子、時間間隔)	0.25°、6 時間間隔
学習要素 (地表面)	海面更正気圧、2m 気温、10m 風 (U,V)、前 6 時間降水量 ²⁴ 、ジオポテンシャル、海陸比
学習要素 (気圧面)	気温、比湿、水平風 (U,V)、鉛直速度、ジオポテンシャル
気圧面	1000, 925, 850, 700, 600, 500, 400, 300, 250, 200, 150, 100, 50 [hPa]
学習データ期間	2018 年 4 月 1 日 00UTC～2020 年 3 月 31 日 18UTC
検証データ期間	2020 年 4 月 1 日 00UTC～2021 年 3 月 31 日 18UTC
学習回数	1000 回

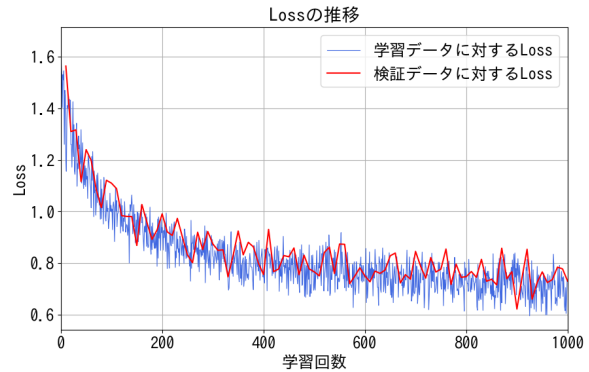


図 4.1 学習データに対する Loss(青線) と検証データに対する Loss(赤線)

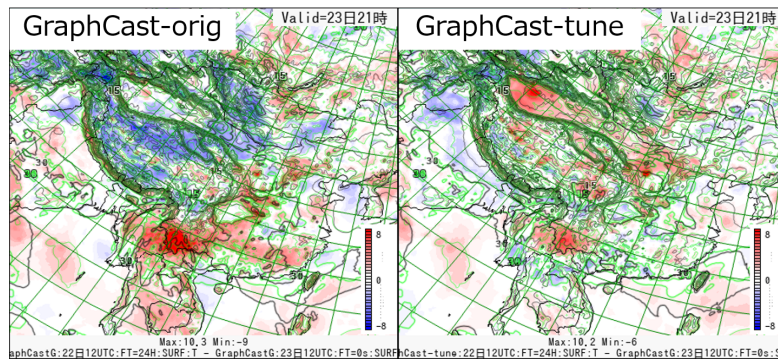


図 4.2 2023 年 7 月 23 日 12UTC を対象時刻とする 2m 気温 [°C] の予測値 (黒実線)、全球速報解析値 (緑実線)、予測値と全球速報解析値との差分 (塗り潰し)。左図が GraphCast-orig の予測で、右図が GraphCast-tune の予測。それぞれ 2023 年 7 月 22 日 12UTC を初期時刻とする 24 時間予測の結果。

す係数の重みについてはオリジナルの GraphCast から変更していない。

ファインチューニングを実施するにあたっては未知のデータに対する予測性能を確認するため、学習に用いるデータとは別に検証データを用意し、学習を 10 回実行する毎に 1 回検証データにより損失関数の値 (Loss) の計算を行った。図 4.1 に学習回数に対する Loss の推移を示す。学習データと検証データに対する Loss は学習回数を進めるごとに減少しており、意図通りの学習が進んでいると判断できる。

ファインチューニングを実施したモデルを用いていくつかの事例について単発実験を行い、予報へのインパクトを確認した。図 4.2 にチベット高原付近の 2m 気温の GSM 解析値に対する差分を示す。調査を行った 3 つの AI 気象モデルに共通して、学習元データとなった ERA5 と GSM との地形の差や地表面気温の診断手法の違いから、標高の高い地域において大きな気温のバイアスが存在することが分かっている (第 3.2 節)。GSM 解析値を入

力としたオリジナルの GraphCast(以後、GraphCast-orig) では予報開始後速やかに ERA5 の地表面気温の特徴に変化していくため、対 GSM 解析値との差分で見るとヒマラヤ山脈付近で大きな低温バイアスが見られるが、ファインチューニングした GraphCast(以後、GraphCast-tune) では GSM 解析値に対するバイアスが大きく減少し、GSM の予測特性に近づいていることが確認できた。

図 4.3 に台風事例の一例を示す。GraphCast に限らず今回調査を行った AI 気象モデルでは台風の中心気圧を浅く予測する傾向があるが、GraphCast-tune ではより深い中心気圧を予測し、GSM 解析値や実況に近づいている様子が見られた。一方で、線状降水帯といった水平スケールの小さい顕著現象での予測性能については、GraphCast-tune は GraphCast-orig と比べるとやや強めの降水を予測できるようになるものの、実況と比べると大幅に少ない降水量しか予想できておらず、学習元データとなる GSM の降水表現性能に制約され

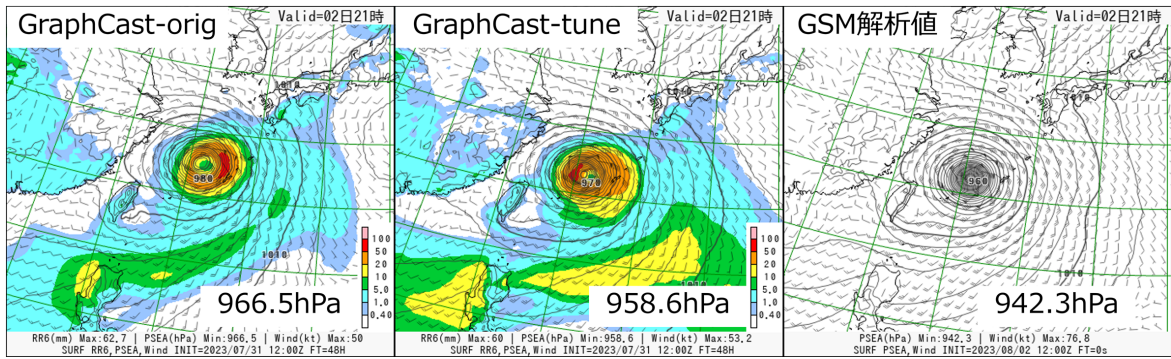


図 4.3 2023 年 8 月 2 日 12UTC を対象時刻とする前 6 時間降水量(塗り潰し)と海面更正気圧(コンター)、地上風(矢羽)。左図が GraphCast-orig の予測で、中図が GraphCast-tune の予測、右図が GSM 解析値。それぞれ 2023 年 7 月 31 日 12UTC を初期時刻とする 48 時間予測の結果で、各図の右下に台風を中心気圧を示している。

ていた(図略)。

今回実施したファインチューニングの 1000 回という学習規模は一般的な AI 気象モデルのファインチューニングと比べて非常に小規模である一方、GPU(NVIDIA A100)を 1 枚利用して約 35 時間の実行時間を要した。学習回数をより大規模とするファインチューニングを行うにあたっては、複数 GPU を利用した分散学習²⁵の導入や I/O 時間を隠蔽する等の実行時間削減の工夫が必要になることに加え、学習データの期間や損失関数の選び方、適切な係数の重みについても検討する必要があると考えられる。

4.3 全球 AI 気象モデルの試作

GNN を用いた小規模な全球 AI 気象モデルの試作について報告する。設計の大半は Keisler (2022) と Lam et al. (2023) に基づいている。

一般的に画像のように空間的な広がりを持つデータをニューラルネットワークで取り扱う際には、畳み込みニューラルネットワーク (CNN) が用いられることが多い。しかしながら、CNN は格子点の一様性を仮定しているため、球面上のデータに対して適用するためには球面をいくつかの領域に分割した上でそれぞれの領域を平面で近似する必要がある (Weyn et al. 2021; Karlbauer et al. 2024)。一方、グラフというデータ構造は任意の点集合と、それらの間の任意の 2 点間のつながりで相互作用する系を表現したものである。特に、球面上の準一様格子点上に分布する特徴量を考え、その近傍との相互作用を学習する GNN は、CNN の曲面への自然な拡張と捉えることができる。

図 4.4 に、試作した GNN モデルのネットワーク構造を模式的に示す。内部格子系には各頂点の担当領域が等面積である等の特徴を持つ HEALPix (Gorski et al. 2005) を用い、グラフ上の畳み込み (近傍頂点の特徴

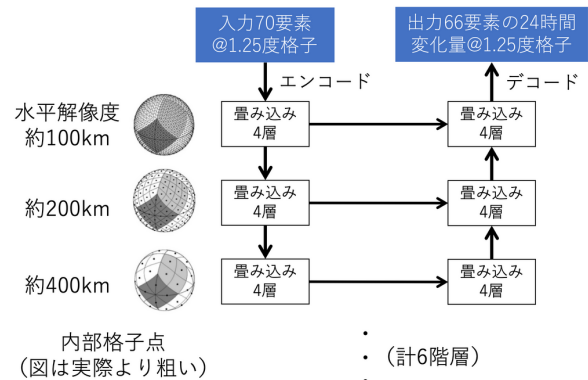


図 4.4 試作した GNN モデルのネットワーク構造。

量や頂点の相対位置等の学習可能な関数で頂点の値を更新する)を基本処理としている。解像度変換に相当するエンコード/デコードも、始点と終点が別の点集合に属する(二部グラフと呼ばれる)GNN で実装している。ここでは GraphCast と同様に、長距離相互作用を表現しつつ計算量を抑えることを狙って 6 階層の内部格子系を併用しているが、全ての階層が相互に接続されている GraphCast と違い、解像度変換を順次行う構造を採用した。これによりデコードに用いる粗い格子が切り替わる場所で学習初期に不連続な場となる問題が解消することを確認している。同様の階層構造を採用した Oskarsson et al. (2023) でも偽の空間パターン解消が報告されている。

JRA-3Q (Kosaka et al. 2024) の 1.25° 気圧面データのうち、1979-2014 年を学習用データ、2015-2016 年を検証用データとして用いた。ここでは簡単のために問題を 00UTC からの 24 時間予測に限定し、繰り返し予測も行わない。入出力する変数には、海面更正気圧に加え、13 の気圧面における気温、比湿、ジオポテンシャル高度、東西風、南北風を用いる。同じ変数であっても、異なる気圧面上にあるものはそれぞれ独立な要素として扱う(計 66 要素)。入力には、上記に加えて地形、海陸マスク、大気上端下向き短波放射、1979 年 1

²⁵ ネットワークで接続された複数の計算機(ノード)で並列に行う学習。並列化の手法は各プロセスに別のデータを参照させるデータ並列と、1つのモデルを複数の GPU またはノードに分割して動作させるモデル並列に大別される。

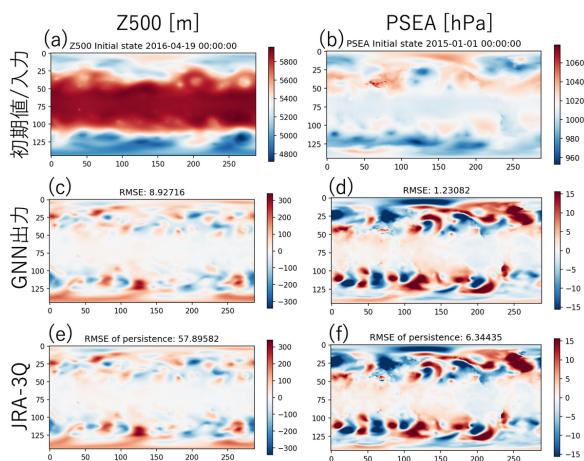


図 4.5 GNN による予測の例。(a, c, e) 2016 年 4 月 19 日 00UTC と 24 時間後の 500 hPa ジオポテンシャル高度。(b, d, f) 2015 年 1 月 1 日 00UTC と 24 時間後の海面更正気圧。(a, b) 入力に用いた初期値。(c, d) GNN から出力された 24 時間変化量。(e, f) JRA-3Q 解析値における 24 時間変化量。

月 1 日からの経過日数の 4 要素を用いた。損失関数には各要素の気候学的な 24 時間変化量（標準偏差）の逆数と緯度に応じた格子面積を乗じた二乗平均誤差を用いる。4GPU（NVIDIA A100）によるバッチ並列で約 26 時間（約 24 万反復）学習させたところで検証用データの Loss が下げ止まったと判断し学習を打ち切った。

図 4.5 に検証用データに対する GNN モデルの予測例を示す。ここに示す両要素とも、JRA-3Q で解析されている時間変化を概ね捉え、持続予報に対して RMSE を数分の 1 に減らしていることから、意味のある予測が出来ていると言える。一方で、空間的に細かな時間変化が解析値に比べて鈍っている等の問題も見られる。なお、検証データをランダムに 100 事例用いた簡易的な統計検証では 500 hPa ジオポテンシャル高度の 24 時間予測の RMSE は 8.8 m となり、最新の AI 気象モデルで報告されているような精度には及ばない。更なる高精度化には学習データと初期値の増強、内部格子点の高解像度化や注意機構を用いるといったモデル（ネットワーク構造）の高度化、学習や推論手法の高度化などを含む総合的なアプローチと、それらに見合った学習量が必要であると考えられる。

試作中に得られた知見のうち業務上重要な事柄を述べる。まず一点目に、ネットワーク設計等を検討する場合には、試行錯誤は不可欠である。特に、学習初期の Loss の下がり方から学習完了後の精度の良し悪しを推測するのは難しいため、試行ごとに学習をある程度完了させる必要がある。二点目として、学習のための誤差逆伝播には計算過程での変数保存が不可欠であり大量のメモリを必要とすると同時に、gradient checkpointing²⁶や

²⁶ 前方計算の過程で全ての変数をメモリ上に保存せずの一部の変数を逆伝播時に再計算する方法。計算量が増える代わりに

分散学習といった大規模モデル特有の技術習得が必要になる。三点目に、比較的小さい GPU 時間の学習でも学習用データと検証用データの Loss の違いからデータ量が不足している兆候が見られており、学習データの時空間解像度の増強と同時に、複数のデータを収集し学習に活用することも精度向上の助けになると考えられる (Bodnar et al. 2024)。

5 まとめ

近年急速に発展した、機械学習による気象予測、特に過去の解析値・観測値の時系列を学習したデータ駆動型気象予測モデルである AI 気象モデルについて、先行研究のレビューや公開されている AI 気象モデルの実行と検証、学習や作成の試行などを通じて、その利点と課題の整理を進めてきた。

AI 気象モデルの研究は 2020 年代から加速し始め、2022 年には実用的な気象予測を可能とするものが複数登場し、それ以降も様々な AI 気象モデルが登場し続けている。この分野の発展は速く、調査を進めている間にも新しい成果が毎週のように報告されている状況であった。今後も先行研究のレビューや海外数値予報センターの動向を注視し、情報収集を行うことが重要と考える。

今回、公開されている AI 気象モデルを用いて可能な調査を実施し、先行研究を再現することを確認し、また精度評価で不十分な点や細かい特徴の把握を進めた。調査対象とした FourCastNetv2, Pangu-Weather, GraphCast はどれも GSM 解析値を初期値として安定に動作し、大規模なスケールの流れの表現等において既存の数値予報モデルのそれを良く模倣できていた。台風進路予測など、大規模なスケールの特定要素の予測に限れば精度が高くまた不自然な予測は見られず、利用可能性は高いと考える。同時に、小スケールの現象については GSM 予測と比較して複数の異なった予測特性を示した。これらの違いは DNN と物理シミュレーションの手法の違いに基づくが、特に学習データである ERA5 解析値の水平格子間隔・鉛直層数と表現能力が原因となっていると考えられる。また、今回の調査では AI 気象モデルの開発元の論文では紹介されていない違いも見られたため、今後も自ら評価を行うことが重要であると考えられる。

AI 気象モデルは既存の物理モデルと比較して少ない資源で高速に動作するのが大きな利点の一つであることをあたらめて確認した。今回調査した FourCastNetv2, Pangu-Weather, GraphCast は、どれも GPU1 枚を用いて、同程度の水平格子間隔を持つ GSM と比較してかなり高速に実行可能であった。

一方で、AI 気象モデルの予測値を MSM 境界値として用いた場合に MSM の予測精度を必ずしも改善しなかった点、ガイダンスの入力として利用するには計算に必要なメモリが減る。

している要素が不足している点、11日を超える長期積分ではバイアスの拡大がGSMよりも大きい点など、作成時の目的外利用では様々な課題が見られた。

また、AI気象モデルの学習についても検討した。AI気象モデルの仕様や精度は学習データに拘束されるため、学習データの仕様・精度は非常に重要である。最新の動向として、複数の学習データを組み合わせて学習し、学習データの大規模化・多様化（基盤モデル化）を志向する可能性があるため、学習データの収集・整備が重要になる可能性が高い。今回、GraphCastをGSM解析値を用いてファインチューニングする調査を行ったが、試行の結果は良好で、この技術の可能性を示すものであった。蓄積された現業解析値を十分に活かしていくことも重要であると考えられる。また、GNNを用いた小規模な全球AI気象モデルを試作した。業務的な利用はできない簡素なモデルの試作であるが、ネットワーク設計等を検討する場合には試行錯誤は不可欠であることや、実用的なAI気象モデルの開発には大規模な計算機資源や大規模モデル特有の技術習得が必要になること、十分な量の質の高い学習データを確保することが重要であるなどの知見が得られた。

今回の調査を通じて、DNNによるデータ駆動型の気象予測が非常に有望な技術であることを理解すると共に、既存の数値予報システムと併用して活用する価値が高いことが確認できた。この分野の発展は速く、今後も様々なアイデアに基づくDNNの利用方法が提案され発展が続くと考えられ、現在課題となっている点や達成できていない項目についても今後新しい成果が登場する可能性が高い。引き続きAI気象モデルの活用や開発の観点での検討に務めたい。

参考文献

- Andrychowicz, Marcin, Lasse Espeholt, Di Li, Samier Merchant, Alexander Merose, Fred Zyda, Shreya Agrawal, and Nal Kalchbrenner, 2023: Deep Learning for Day Forecasts from Sparse Observations. *arXiv*, <https://arxiv.org/abs/2306.06079>.
- Bell, B., H. Hersbach, A. Simmons, P. Berrisford, P. Dahlgren, A. Horányi, J. Muñoz-Sabater, J. Nicolas, R. Radu, D. Schepers, C. Soci, S. Villaume, J. Bidlot, L. Haimberger, J. Woollen, C. Buontempo, and J. Thépaut, 2021: The ERA5 global reanalysis: Preliminary extension to 1950. *Quart. J. Roy. Meteor. Soc.*, **147**, 4186–4227.
- Bi, Kaifeng, Lingxi Xie, Hengheng Zhang, Xin Chen, Xiaotao Gu, and Qi Tian, 2023: Accurate medium-range global weather forecasting with 3D neural networks. *Nature*, **619**, 533–538.
- Bodnar, Cristian, Wessel P. Bruinsma, Ana Lucic, Megan Stanley, Johannes Brandstetter, Patrick Garvan, Maik Riechert, Jonathan Weyn, Haiyu Dong, Anna Vaughan, Jayesh K. Gupta, Kit Tamberatnam, Alex Archibald, Elizabeth Heider, Max Welling, Richard E. Turner, and Paris Perdikaris, 2024: Aurora: A Foundation Model of the Atmosphere. *arXiv*, <https://arxiv.org/abs/2405.13063>.
- Bonavita, Massimo, 2024: On Some Limitations of Current Machine Learning Weather Prediction Models. *Geophysical Research Letters*, **51**, e2023GL107377.
- Bonev, Boris, Thorsten Kurth, Christian Hundt, Jaideep Pathak, Maximilian Baust, Karthik Kashinath, and Anima Anandkumar, 2023: Spherical Fourier Neural Operators: Learning Stable Dynamics on the Sphere. *arXiv*, <https://arxiv.org/abs/2306.03838>.
- Chen, Kang, Tao Han, Junchao Gong, Lei Bai, Fenghua Ling, Jing-Jia Luo, Xi Chen, Leiming Ma, Tianning Zhang, Rui Su, Yuanzheng Ci, Bin Li, Xiaokang Yang, and Wanli Ouyang, 2023a: FengWu: Pushing the Skillful Global Medium-range Weather Forecast beyond 10 Days Lead. *arXiv*, <https://arxiv.org/abs/2304.02948>.
- Chen, Lei, Fei Du, Yuan Hu, Fan Wang, and Zhibin Wang, 2023b: SwinRDM: Integrate SwinRNN with Diffusion Model Towards High-Resolution and High-Quality Weather Forecasting. *arXiv*, <https://arxiv.org/abs/2306.03110>.
- Chen, Lei, Xiaohui Zhong, Feng Zhang, Yuan Cheng, Yinghui Xu, Yuan Qi, and Hao Li, 2023c: FuXi: a cascade machine learning forecasting system for 15-day global weather forecast. *npj Climate and Atmospheric Science*, **6**, 190.
- de Burgh-Day, C. O. and T. Leeuwenburg, 2023: Machine learning for numerical weather and climate modelling: a review. *Geoscientific Model Development*, **16**, 6433–6477.
- Dosovitskiy, Alexey, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby, 2020: An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv*, <https://arxiv.org/abs/2010.11929>.
- Dueben, P. D. and P. Bauer, 2018: Challenges and design choices for global weather and climate models based on machine learning. *Geoscientific Model Development*, **11**, 3999–4009.
- Frolov, Sergey, Kevin Garrett, Isidora Jankov, Daryl Kleist, Jebb Q. Stewart, and John Ten Hoeve, 2024:

- Integration of emerging data-driven models into the NOAA research to operation pipeline for numerical weather prediction. *Bull. Amer. Meteor. Soc.*
- Gorski, K. M., E. Hivon, A. J. Banday, B. D. Wandelt, F. K. Hansen, M. Reinecke, and M. Bartelmann, 2005: HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere. *The Astrophysical Journal*, **622**, 759–771, doi:10.1086/427976.
- Hersbach, Hans, Bill Bell, Paul Berrisford, Shoji Hirahara, András Horányi, Joaquín Muñoz-Sabater, Julien Nicolas, Carole Peubey, Raluca Radu, Dinand Schepers, Adrian Simmons, Cornel Soci, Saleh Abdalla, Xavier Abellan, Gianpaolo Balsamo, Peter Bechtold, Gionata Biavati, Jean Bidlot, Massimo Bonavita, Giovanna De Chiara, Per Dahlgren, Dick Dee, Michail Diamantakis, Rossana Dragani, Johannes Flemming, Richard Forbes, Manuel Fuentes, Alan Geer, Leo Haimberger, Sean Healy, Robin J. Hogan, Elías Hólm, Marta Janisková, Sarah Keeley, Patrick Laloyaux, Philippe Lopez, Cristina Lupu, Gabor Radnoti, de Patricia Rosnay, Iryna Rozum, Freja Vamborg, Sebastien Villaume, and Jean-Noël Thépaut, 2020: The ERA5 global reanalysis. *Quart. J. Roy. Meteor. Soc.*, **146**, 1999–2049.
- Hinton, G. E. and R. R. Salakhutdinov, 2006: Reducing the Dimensionality of Data with Neural Networks. *Science*, **313**, 504–507.
- 金浜貴史, 2013: 全球決定論予報の WMO 標準検証. 数値予報課報告・別冊第 59 号, 気象庁予報部, 122–126.
- Karlbauer, Matthias, Nathaniel Cresswell-Clay, Dale R. Durran, Raul A. Moreno, Thorsten Kurth, Boris Bonev, Noah Brenowitz, and Martin V. Butz, 2024: Advancing Parsimonious Deep Learning Weather Prediction Using the HEALPix Mesh. *Journal of Advances in Modeling Earth Systems*, **16**, e2023MS004021.
- Karras, T., M. Aittala, T. Aila, and S. Laine, 2022: Elucidating the design space of diffusion-based generative models. *Advances in Neural Information Processing Systems*, **35**, 26 565–26 577.
- Keisler, R., 2022: Forecasting Global Weather with Graph Neural Networks. *arXiv*, <https://doi.org/10.48550/arXiv.2202.07575>.
- 越智健太, 石井憲介, 2013: 予報時間を 39 時間に延長した MSM の初期時刻別統計検証. 平成 25 年度数値予報研修テキスト, 気象庁予報部, 14–17.
- 気象庁, 2021: 統合型ガイドンス. 数値予報開発センター年報 (令和 2 年), 気象庁 数値予報開発センター, 107–110.
- Kochkov, Dmitrii, Janni Yuval, Ian Langmore, Peter Norgaard, Jamie Smith, Griffin Mooers, Milan Klöwer, James Lottes, Stephan Rasp, Peter Düben, Sam Hatfield, Peter Battaglia, Alvaro Sanchez-Gonzalez, Matthew Willson, and Stephan Michael P. Brenner Hoyer, 2024: Neural general circulation models for weather and climate. *Nature*, **632**, 1060–1066.
- 河野耕平, 荒波恒平, 2014: 側面・上部境界条件. 数値予報課報告・別冊第 60 号, 気象庁予報部, 57–61.
- Kosaka, Yuki, Shinya Kobayashi, Yayoi Harada, Chiaki Kobayashi, Hiroaki Naoe, Koichi Yoshimoto, Masashi Harada, Naohika Goto, Jotaro Chiba, Kengo Miyaoka, Ryohei Sekiguchi, Makoto Deushi, Hirotaka Kamahori, Tosiya Nakaegawa, Taichu Y. Tanaka, Takayuki Tokuyoshi, Yoshiaki Sato, Yasuhiro Matsushita, and Kazutoshi Onogi, 2024: The JRA-3Q Reanalysis. 気象集誌. 第 2 輯, **102**, 49–109.
- Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E Hinton, 2012: ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems*, **25**.
- 経田正幸, 越智健太, 2014: 台風アンサンブル予報システム. 平成 26 年度数値予報研修テキスト, 気象庁予報部, 62–71.
- Lam, Remi, Alvaro Sanchez-Gonzalez, Matthew Willson, Peter Wirnsberger, Meire Fortunato, Ferran Alet, Suman Ravuri, Timo Ewalds, Zach Eaton-Rosen, Weihua Hu, Alexander Merose, Stephan Hoyer, George Holland, Oriol Vinyals, Jacklynn Stott, Alexander Pritzel, Shakir Mohamed, and Peter Battaglia, 2023: Learning skillful medium-range global weather forecasting. *Science*, **382**, 1416–1421.
- Lang, Simon, Mihai Alexe, Matthew Chantry, Jesper Dramsch, Florian Pinault, Baudouin Raoult, Mariana C. A. Clare, Christian Lessig, Michael Maier-Gerber, Linus Magnusson, Zied Ben Bouallègue, Ana Prieto Nemesio, Peter D. Dueben, Andrew Brown, Florian Pappenberger, and Florence Rabier, 2024: AIFS – ECMWF’s data-driven forecasting system. *arXiv*, <https://arxiv.org/abs/2406.01465>.
- Liu, Cheng-Chin, Kathryn Hsu, Melinda Peng, Der-Song Chen, Pao-Liang Chang, Ling-Feng Hsiao, Chin-Tzu Fong, Jing-Shan Hong, Chia-Ping Cheng, Kuo-Chen Lu, Chia-Rong Chen, and Hung-Chi Kuo, 2024: Evaluation of five global AI models

- for predicting weather in Eastern Asia and Western Pacific. *npj Climate and Atmospheric Science*, **7**.
- 松本智文, 2018: 風ガイドンス. 数値予報課報告・別冊第 64 号, 気象庁予報部, 144–153.
- Nipen, Thomas Nils, Håvard Homleid Haugen, Magnus Sikora Ingstad, Even Marius Nordhagen, Aram Farhad Shafiq Salihi, Paulina Tedesco, Ivar Ambjørn Seierstad, Jørn Kristiansen, Simon Lang, Mihai Alexe, Jesper Dramsch, Baudouin Raoult, Gert Mertes, and Matthew Chantry, 2024: Regional data-driven weather modeling with a global stretched-grid. *arXiv*, <https://arxiv.org/abs/2409.02891>.
- 石川宜広, 2018: ガイドンス. 平成 30 年度数値予報研修テキスト, 気象庁予報部, 106–116.
- Orbe, Clara, Luke D. Oman, Susan E. Strahan, Darryn W. Waugh, Steven Pawson, Lawrence L. Takacs, and Andrea M. Molod, 2017: Large-Scale Atmospheric Transport in GEOS Replay Simulations. *Journal of Advances in Modeling Earth Systems*, **9**, 2545–2560.
- Oskarsson, Joel, Tomas Landelius, and Fredrik Lindsten, 2023: Graph-based Neural Weather Prediction for Limited Area Modeling. *arXiv*, <https://arxiv.org/abs/2309.17370>.
- Pathak, Jaideep, Shashank Subramanian, Peter Harrington, Sanjeev Raja, Ashesh Chattopadhyay, Morteza Mardani, Thorsten Kurth, David Hall, Zongyi Li, Kamyar Azizzadenesheli, Pedram Hassanzadeh, Karthik Kashinath, and Animashree Anandkumar, 2022: FourCastNet: A Global Data-driven High-resolution Weather Model using Adaptive Fourier Neural Operators. *arXiv*, <https://arxiv.org/abs/2202.11214>.
- Pfaff, Tobias, Meire Fortunato, Alvaro Sanchez-Gonzalez, and Peter W. Battaglia, 2020: Learning Mesh-Based Simulation with Graph Networks. *arXiv*, <https://arxiv.org/abs/2010.03409>.
- Price, Ilan, Alvaro Sanchez-Gonzalez, Ferran Alet, Tom R. Andersson, Andrew El-Kadi, Dominic Masters, Timo Ewalds, Jacklynn Stott, Shakir Mohamed, Peter Battaglia, Remi Lam, and Matthew Willson, 2023: GenCast: Diffusion-based ensemble forecasting for medium-range weather. *arXiv*, <https://arxiv.org/abs/2312.15796>.
- Rasp, Stephan, Peter D. Dueben, Sebastian Scher, Jonathan A. Weyn, Soukayna Mouatadid, and Nils Thuerey, 2020: WeatherBench: A Benchmark Data Set for Data-Driven Weather Forecasting. *Journal of Advances in Modeling Earth Systems*, **12**, e2020MS002203.
- Rasp, Stephan, Stephan Hoyer, Alexander Merose, Ian Langmore, Peter Battaglia, Tyler Russell, Alvaro Sanchez-Gonzalez, Vivian Yang, Rob Carver, Shreya Agrawal, Matthew Chantry, Zied Ben Bouallegue, Peter Dueben, Carla Bromberg, Jared Sisk, Luke Barrington, Aaron Bell, and Fei Sha, 2024: WeatherBench 2: A Benchmark for the Next Generation of Data-Driven Global Weather Models. *Journal of Advances in Modeling Earth Systems*, **16**, e2023MS004019.
- Scarselli, Franco, Marco Gori, Ah Chung Tsoi, Markus Hagenbuchner, and Gabriele Monfardini, 2009: The Graph Neural Network Model. *IEEE Transactions on Neural Networks*, **20**, 61–80.
- Scher, S., 2018: Toward Data-Driven Weather and Climate Forecasting: Approximating a Simple General Circulation Model With Deep Learning. *Geophysical Research Letters*, **45**, 12,616–12,622.
- Scher, S. and G. Messori, 2019: Weather and climate forecasting with neural networks: using general circulation models (GCMs) with different complexity as a study ground. *Geoscientific Model Development*, **12**, 2797–2809.
- Schmude, Johannes, Sujit Roy, Will Trojak, Johannes Jakubik, Daniel Salles Civitarese, Shradha Singh, Julian Kuehnert, Kumar Ankur, Aman Gupta, Christopher E Phillips, Romeo Kienzler, Daniela Szwarcman, Vishal Gaur, Rajat Shinde, Rohit Lal, Arlindo Da Silva, Jorge Luis Guevara Diaz, Anne Jones, Simon Pfreundschuh, Amy Lin, Aditi Sheshadri, Udaysankar Nair, Valentine Anantharaj, Hendrik Hamann, Campbell Watson, Manil Maskey, Tsengdar J Lee, Juan Bernabe Moreno, and Rahul Ramachandran, 2024: Prithvi WxC: Foundation Model for Weather and Climate. *arXiv*, <https://arxiv.org/abs/2409.13598>.
- Selz, T. and G. C. Craig, 2023: Can Artificial Intelligence-Based Weather Prediction Models Simulate the Butterfly Effect? *Geophysical Research Letters*, **50**, e2023GL105747.
- Sohl-Dickstein, J., E. Weiss, N. Maheswaranathan, and S. Ganguli, 2015: Deep unsupervised learning using nonequilibrium thermodynamics. *In International conference on machine learning*, PMLR, **37**, 2256–2265.
- Song, Y., J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole, 2021: Score-based generative modeling through stochastic differential equations. *In International Conference on*

- Learning Representations, 2021.*, URL <https://openreview.net/forum?id=PxTIG12RRHS>.
- 原旅人, 2016: 渦位の追跡によって見る MSM における境界値の影響. 平成 28 年度数値予報研修テキスト, 気象庁予報部, 89–99.
- Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin, 2017: Attention is All you Need. *Advances in Neural Information Processing Systems*, **30**.
- Wang, Chenggong, Michael S. Pritchard, Noah Brenowitz, Yair Cohen, Boris Bonev, Thorsten Kurth, Dale Durran, and Jaideep Pathak, 2024: Coupled Ocean-Atmosphere Dynamics in a Machine Learning Earth System Model. *arXiv*, <https://arxiv.org/abs/2406.08632>.
- Weyn, Jonathan A., Dale R. Durran, and Rich Caruana, 2019: Can Machines Learn to Predict Weather? Using Deep Learning to Predict Gridded 500-hPa Geopotential Height From Historical Weather Data. *Journal of Advances in Modeling Earth Systems*, **11**, 2680–2693.
- Weyn, Jonathan A., Dale R. Durran, and Rich Caruana, 2020: Improving Data-Driven Global Weather Prediction Using Deep Convolutional Neural Networks on a Cubed Sphere. *Journal of Advances in Modeling Earth Systems*, **12**, e2020MS002109.
- Weyn, Jonathan A., Dale R. Durran, Rich Caruana, and Nathaniel Cresswell-Clay, 2021: Sub-Seasonal Forecasting With a Large Ensemble of Deep-Learning Weather Prediction Models. *Journal of Advances in Modeling Earth Systems*, **13**, e2021MS002502.