International Journal of Web of Multidisciplinary Studies
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
website: https://ijwos.com
Volume 3 Issue 8 , August 2026
OPEN ACCESS
E-ISSN : 3049-2424
DOI: https://doi.org/10.71366/ijwos03082654990

Agentic AI for Weather Forecasting: A Reproducible Fourier-Domain Neural Operator Pipeline for Short-Horizon, Station-Level Multivariate Weather Predict

Jyoti[1], Amandeep[2], Dharmender Kumar[3], Nidhi[4], Ritu[5]
[1,4,5] M.Sc. Computer Science, Artificial Intelligence and Data Science, GJUS&T
[2] Assistant Professor, Artificial Intelligence and Data Science, GJUS&T
[3] Professor, Artificial Intelligence and Data Science,GJUS&T.

Article Info

Article History:

Published: 16 August 2026

Publication Issue:
Volume 3, Issue 8
August-2026

Page Number:
331-340

Corresponding Author:
Jyoti

Abstract:

Traditionally, weather forecasting has been achieved by numerical weather prediction (NWP) systems built from physics-based models that numerically integrate the equations of motion for air on a supercomputer, which are accurate but expensive to run, can be in error due to chaotic dynamics in the atmosphere, and require a significant amount of computing power to make forecasts. In the last ten years, two parallel developments have changed this landscape: firstly, data-driven deep-learning weather models, like FourCastNet, Pangu-Weather, GraphCast, FengWu, FuXi, GenCast and Aurora, that learn atmospheric behaviour directly from historical reanalysis data, and can reach or even surpass NWP accuracy with much less computational cost, and secondly, agentic AI frameworks such as ReAct, AutoGPT, MetaGPT, AutoGen, and Reflexion, that allow for autonomous planning, tool-use, self-verification, and multi-agent orchestration, in the software-engineering and general-purpose task domains. The analysis of 35 papers within both fields reveals no current approach that integrates autonomous, self-verifying, uncertainty-aware reasoning with a short-term forecasting model that is driven by data. Agentic applications found in the weather field are largely post-hoc interpretation, reporting and orchestration layers around externally provided forecasts. Conceptually, this is inspired by FourCastNet, but in this paper we propose and evaluate a lightweight Adaptive Fourier Neural Operator (AFNO) type forecasting model, which is derived from a global gridded atmospheric dataset to a single-station, tabular multivariate time-series setting, and provides an empirical basis for closing this gap. Using 24 hours of six continuous surface weather variables (temperature, dew point temperature, relative humidity, wind speed, visibility and pressure) as input, two stacked Fourier-domain temporal-mixing blocks with a 64-dimensional hidden representation, residual connections and layer normalisation directly forecast the same six variables one hour ahead. The model is trained and evaluated on 8,784 hourly single station observations with a chronological split of 70/15/15 for the training, validation and test sets, respectively, yielding an overall test-set MAE of 1.9138, MSE of 12.5279, RMSE of 3.5395 and R^2 of 0.8854; the best accuracy is achieved for pressure ($R^2 = 0.9921$) and the lowest for visibility ($R^2 = 0.7151$). These findings, along with a qualitative comparison to other models of weather forecasting and nowcasting that use global data, demonstrate the effectiveness of Fourier-domain temporal mixing as a lightweight inductive bias for station-level, short-horizon weather forecasting and provide an empirical benchmark for future investigations of integrating this forecasting capability with an autonomous, self-verifying, uncertainty-aware agentic reasoning layer.

Keywords: Agentic AI, Numerical Weather Prediction (NWP), Data-Driven Weather Forecasting, Adaptive Fourier Neural Operator (AFNO), FourCastNet, Nowcasting, Large Language Models (LLMs), Time-Series Forecasting, Multi-Agent Systems, Uncertainty-Aware Forecasting

1. INTRODUCTION

The ability to accurately predict the future state of the atmosphere, both in terms of the state of the air and the timing of the prediction, is a longstanding problem in science and engineering, important for many aspects of human activity such as agriculture, aviation, energy supply, construction and disaster management. This was almost entirely done by physics-based Numerical Weather Prediction (NWP) in which the governing equations of atmospheric motion were solved numerically on large supercomputers until the early 21st century. NWP uses the 3-D grid of the atmosphere, estimates the current state of temperature, pressure, humidity and wind at each point in the grid from observations received from satellites, weather balloons and ground stations by means of data assimilation, and then solves the fluid motion and thermodynamics equations forward in time. Although NWP is the current dominant operational approach, it is very expensive, sensitive to small errors in the initial conditions which rapidly grow due to the chaotic nature of the atmosphere, and complicated to construct and maintain, so only a few national agencies can operate it fully.

In the last ten years, Artificial Intelligence (AI) has taken its presence into this field in two more and more integrated ways. First, deep-learning forecasting models like FourCastNet, Pangu-Weather, GraphCast, FengWu, FuXi, GenCast and Aurora are trained to directly learn the evolution of the atmosphere from decades of historical reanalysis data. The models trained can produce global medium range forecasts in minutes on a single computer machine and reach accuracy comparable to that of operational NWP systems, like the ECMWF's HRES and IFS, at an order of magnitude faster. Second, a new class of "agentic AI" built on large-language-models has become available, such as ReAct, AutoGPT, MetaGPT, AutoGen and Reflexion, which can plan, reason, call external tools, and output the results in natural language, with little or no human assistance. Agentic systems can self-determine data source or model to query, verify their results, and communicate the result to a non-expert user without a human having to design, run, or interpret each step of the machine-learning pipeline.

This paper combines these two aspects in the peculiar context of short horizon, stationscale weather forecasting. Section 2 provides an overview of the literature related to foundations of Agentic AI and AI-based weather foundation and nowcasting models. This review is summarized in section 3 as a set of research gaps and formulation of the problem addressed in this paper. Section 4 outlines the proposed model: a lightweight Adaptive Fourier Neural Operator (AFNO)-inspired model, AFNOWeatherModel, along with the dataset and pre-processing pipeline and training configuration to assess it. In Section 5, the quantitative and graphical results are reported, an error-distribution analysis is performed and a qualitative comparison with related data-driven models for weather and nowcasting from the literature is provided. Section 6 summarizes the paper, summarizes the contributions and limitations and outlines directions for coupling the proposed forecasting component with an autonomous agentic reasoning layer.

2. LITERATURE REVIEW

The literature was systematically reviewed, covering three related areas: (i) agentic AI foundations (autonomous, reasoning and acting, multi-agent large-language-modal); (ii) AI-based nowcasting and emerging agentic-AI applications in weather and climate science; and (iii) AI-based global and medium-range weather foundation models. Table 1 summarises the representative works from each area.

Ref.	Study / Model	Method / Key Finding	Reported Result or Contribution
[1]	Yao et al. (2022) – ReAct	Interleaves reasoning steps with tool-calling action steps in a closed loop	Reduces hallucination; foundation for later weather-orchestration agents

Ref.	Study / Model	Method / Key Finding	Reported Result or Contribution
[2]-[4]	AutoGPT, MetaGPT, AutoGen	Autonomous goal decomposition, role-based SOP pipelines, flexible multi-agent conversation	Established general-purpose multi-agent orchestration patterns
[5]	Shinn et al. (2023) – Reflexion	Verbal self-reflection stored in episodic memory in place of gradient updates	Improved sequential decision-making without fine-tuning
[9]	Pathak et al. (2022) – FourCastNet	Adaptive Fourier Neural Operator (AFNO) on 0.25° global ERA5 grid	Skilful high-resolution global forecasts at NWP-competitive speed
[10]	Bi et al. (2023) – Pangu-Weather	3-D Earth-specific Transformer with hierarchical temporal aggregation	First data-driven model to beat ECMWF IFS on several metrics: ~100x faster
[11]	Lam et al. (2023) – GraphCast	Graph neural network on an icosahedral mesh, autoregressive rollout	Outperformed ECMWF HRES on most of 1,380 evaluation targets
[12]-[13]	Chen et al. (2023) – FengWu, FuXi	Multi-modal fusion / cascaded models for long-horizon rollout	Extended skilful forecast horizon beyond 10–15 days
[15]	Price et al. (2023) – GenCast	Diffusion-based ensemble (probabilistic) forecasting	Outperformed ECMWF ENS on most probabilistic verification metrics
[18]-[23]	ConvLSTM, MetNet (-2), DGMR, NowcastNet, PreDiff	Convolutional / generative / diffusion nowcasting of 0–12-hour precipitation	Improved short-term, high-resolution precipitation and extreme event nowcasting
[26]-[29]	Zephyrus, EWE, Hierarchical AI-Meteorologist, Modular LLM-Agent System	LLM agents that interpret, explain or report on forecasts generated elsewhere	Transparent, multi-scale, human-readable forecast communication
[34]-[35]	Agentic streamflow pipeline, Cloudburst agentic framework	Agentic orchestration for seasonal streamflow and cloudburst alert/response	Demonstrates narrow, single-hazard agentic integration with forecasting output

Table 1: Representative summary of the agentic-AI and AI-weather-forecasting literature reviewed.

In general, the accuracy of the models on medium-range global benchmarks is better than that of the pre-AI operational systems that were reviewed, and the skill scores reported for global data-driven models are often comparable to, or even better than, those of ECMWF's operational models, but an order of magnitude faster to run. Also, nowcasting-specific models (Refs. 18–25) are generally good at forecasting precipitation within the 0-12-hour time range, which is outside the focus range of the medium-range foundation models. Agentic-AI applications specific to weather and climate (Refs. 26-35) are the newest to emerge, and, as described in Section 3, these applications are still focused on interpreting or reporting on existing forecasts, rather than directly involving in the generation of the forecasts.

3. Research Gap

The comparative analysis in Section 2 brings to the fore a consistent trend found in the studied literature. Existing agentic AI systems like ReAct, AutoGPT, MetaGPT, AutoGen, and Reflexion show impressive reasoning capabilities, tool usage and multi-agent coordination but are limited to software engineering, navigation in web pages and task automation in general. State-of-the-art AI weather foundation models like FourCastNet, Pangu-Weather, GraphCast, FengWu, FuXi, GenCast and Aurora also developed to match or surpass the performance of operational NWP models, but they remain single-purpose deterministic models with no autonomous reasoning, self-correction, or adaptive decision-making layer.

- Mostly post-hoc integration between agentic AI and forecasting models: the newly emerging agentic-weather works (Refs. 26-35) tend to only include agentic reasoning in explaining or reporting on forecasts already generated elsewhere in the system instead of embedding it in the key forecasting or nowcasting model.
- The nowcasting models are still agent-free as there are no agentic reasoning, self-verification or autonomous decision layer models dedicated to 0-12 hours nowcasting, which is not a horizon where medium-range foundation models are applied.
- There exist only a few models (e.g., GenCast, DGMR, NowcastNet, PreDiff) that provide the uncertainty as a probability or as an ensemble, but none of them are coupled with an agentic layer that infers from that uncertainty to act all the way to alerting or further action.
- No self-verifying or self-correcting forecast agents – self-critique (Reflexion) and orchestrator-based re-planning exist for general agentic tasks, along with explainability in forecasting reporting, but no reviewed weather or nowcasting model has an agent reflecting on its forecast error and re-planning its pipeline over time.
- The high computational requirement limits real-time/edge deployment: foundation-scale models and diffusion-based ensembles use a lot of computes, which limits use in real-time, resource constrained or localised environments.

To this end, no existing system in the literature combines an agentic reasoning and acting layer with a data-driven weather or nowcasting layer as adaptive, self-verifying and uncertainty-aware short-horizon forecast and decision automation system. This paper tackles the issue of computational cost directly while simultaneously addressing the other gaps, by (i) reviewing and scoping the state of the art in agentic AI architectures and data-driven weather forecasting models and (ii) designing, implementing and testing a lightweight single-station agentic forecasting model, AFNOWeatherModel, which can be later coupled with an agentic reasoning-and-acting layer for adaptive, explainable, forecast-driven decisions.

4. PROPOSED METHODOLOGY

The proposed system, AFNOWeatherModel, changes the global gridded, ~20-variable atmospheric setting of FourCastNet to a single-station, tabular, 6-variable time-series setting. The pipeline is realised in a single, reproducible python/pytorch script and encompasses five steps: dataset acquisition and cleaning, chronological splitting and normalisation of the data, construction of the sliding windows, definition and training of a model and its evaluation.

4.1 Dataset and Preprocessing

The data consists of 8,784 single-hour observations of variables for a single station with no missing values for any variable, and one categorical column (Weather) that consists of the following categories: {Sunny, Partly Cloudy, Cloudy, Fog, Mist, Rain, Snow, Unknown}. The categorised Weather field is dropped, and the continuous variables are kept for prediction. Data are sorted chronologically in the Date/Time column. The data is then partitioned chronologically (not randomly, since there is a risk of temporal leakage) into 70% training, 15% validation and 15% testing. A StandardScaler is fit on the training partition and then applied to the validation and test partitions to avoid information leakage. The windowing of the data with a lookback window of 24 hours results in 6,148 training windows, 1,318 validation windows and 1,318 testing windows (which is reported after windowing as 6,124 / 1,294 / 1,294 usable windows once the lookback offset is applied).

4.2 Model Architecture

The time is the transformed coordinate, not 2-D space, with AFNOWeatherModel. The 64-dimensional hidden representation of the six raw input features is computed at each of the 24-time steps by a linear input-embedding layer. The two stacked Fourier blocks then calculate a real Fast Fourier Transform (rFFT) of the time block, followed by a learnable linear transformation separately on the real and imaginary parts of the Fourier transformed block, and then the inverse real Fast Fourier Transform (irFFT) of the linear transformed block is calculated and wrapped in a residual (skip) connection, Layer Normalisation and a ReLU activation, just like in the base paper for the FourCastNet, where residual (skip) connections were used around AFNO blocks. The hidden representation at the final time step is passed through a linear readout layer which maps the 64-dimensional state back to the six forecast variables, producing a linear readout layer, which corresponds to a one-stage-ahead regression (non-autoregressive) with respect to the forecast variables. The layer configuration is summarised in Table 2.

Layer / Component	Configuration	Function
Input	24 × 6	Previous 24 hours of six weather variables
Input Embedding	6 → 64	Projects raw features to hidden representation
Fourier Block 1	Hidden dim. = 64, rFFT / irFFT	Temporal spectral mixing
Fourier Block 2	Hidden dim. = 64, rFFT / irFFT	Additional temporal spectral mixing
Residual Connection	Enabled (per block)	Preserves input information
Layer Normalisation	Applied (per block)	Stabilises representation
Activation	ReLU	Non-linear transformation
Temporal Selection	Final time step	Obtains latest hidden representation
Linear Readout	64 → 6	Produces six next-hour forecasts
Output	6 variables	Predicted weather state at t+1

Table 2: AFNOWeatherModel layer configuration.

4.3 Training Configuration

The model is trained using the Adam optimiser with a learning rate of 0.001 and Mean Squared Error (MSE) loss and is run for a maximum of 30 epochs, on GPU if available, and CPU otherwise, with a batch size of 64. Four widely used regression methods are used to report model quality: Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE) and the coefficient of determination (R^2), both mean-wise and aggregated over all six variables. In pseudo-code the end-to-end pipeline looks like Algorithm 1.

Algorithm 1: AFNOWeatherModel Training and Evaluation

Input: hourly single-station weather records D (8,784 × 6 continuous variables)

Output: trained AFNOWeatherModel and test-set forecast-quality metrics

1: D ← parse_datetime(D); D ← sort_chronological(D)

2: D ← select_columns(D, [Temp, DewPoint, RelHum, WindSpeed, Visibility, Pressure])

3: D_train, D_val, D_test ← chronological_split (D, 0.70, 0.15, 0.15)

4: scaler ← fit_StandardScaler(D_train)

5: D_train, D_val, D_test ← scaler.transform(D_train), scaler.transform(D_val), scaler.transform(D_test)

6: X_train, y_train ← sliding_windows (D_train, lookback=24, horizon=1)

(repeat sliding_windows for D_val and D_test)

7: model ← AFNOWeatherModel () // Embed (6→64) + 2×[FFT→Linear→iFFT→Residual→LayerNorm→ReLU] + Linear (64→6)

8: optimizer ← Adam(model.parameters(), lr=0.001); criterion ← MSELoss()

9: for epoch in 1.30:

10: train model on (X_train, y_train) with batch_size = 64

11: evaluate validation loss on (X_val, y_val)

12: y_pred ← model(X_test); y_pred, y_test ← scaler.inverse_transform(y_pred, y_test)

13: metrics ← compute_MAE_MSE_RMSE_R2(y_test, y_pred) // per variable and overall

5. RESULTS AND DISCUSSION

5.1 Training Convergence

The training and validation MSE loss curves are plotted over 30 epochs in figure 1. Both curves start out at around 0.26 (training) and 0.13 (validation) loss and then level off at about 0.105 loss from about epoch 10 onwards. Training curve and validation curve are also very close and do not separate over training, suggesting that the model is not noticeably overfitting the data at this dataset size and depth of the architecture.

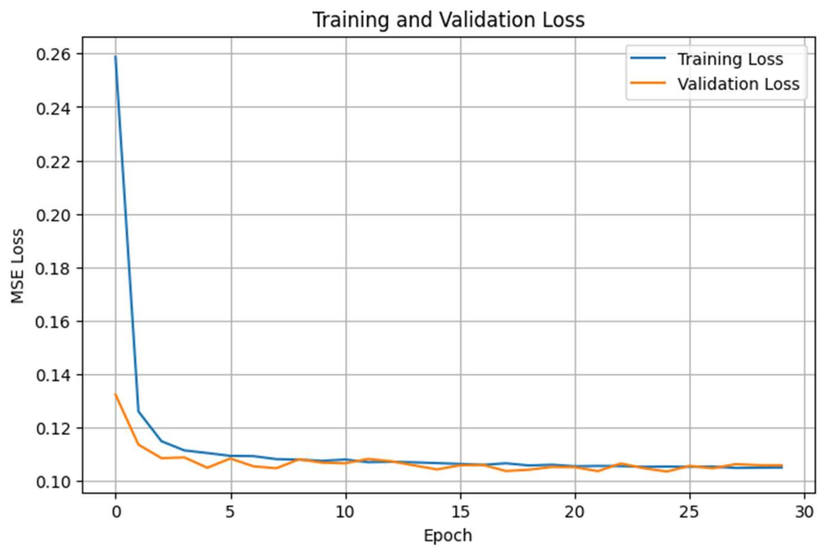

Figure 1: Training and validation MSE loss of AFNOWeatherModel over 30 epochs.

5.2 Test-Set Evaluation

The performance of AFNOWeatherModel on the held-out chronological test set (1,294 sequences) is reported in Table 3. The overall model accounts for about 88.5% of the variance in the next hour weather state in all six variables (R^2 = 0.8854). The physical nature of each of these variables (Pressure, Dew Point Temperature and Temperature — all variables are smooth and strongly autocorrelated; Relative Humidity, Wind Speed and Visibility — more locally driven and noisy variables) strongly influences the performance, with the smooth, strongly autocorrelated variables forecasted with high accuracy (R^2 between 0.96 and 0.99) and the more locally driven and noisy variables comparatively harder to forecast (R^2 between 0.72 and 0.89).

Metric	Value
MAE	1.9138
MSE	12.5279
RMSE	3.5395
R^2	0.8854

Table 3a: Overall test-set performance of AFNOWeatherModel.

Variable	MAE	MSE	RMSE	R^2
Temperature	0.9032	1.2774	1.1302	0.9614
Dew Point Temperature	0.6969	0.8601	0.9274	0.9736
Relative Humidity	3.2250	19.0845	4.3686	0.8919

Variable	MAE	MSE	RMSE	R^2
Wind Speed	3.2352	19.6631	4.4343	0.7782
Visibility	3.3501	34.2730	5.8543	0.7151
Pressure	0.0728	0.0093	0.0964	0.9921

Table 3b: Variable-wise test-set performance of AFNOWeatherModel.

Figure 2 shows the actual and predicted temperature over the first 200-time steps in the test set. The R^2 for Temperature is high, and the predicted series closely follows the pattern of the seasonal/diurnal fluctuation of the actual series including sharp transitions like the rise from it is around −4°C to 17°C, with only slight lag and amplitude error around peaks and troughs, as expected.

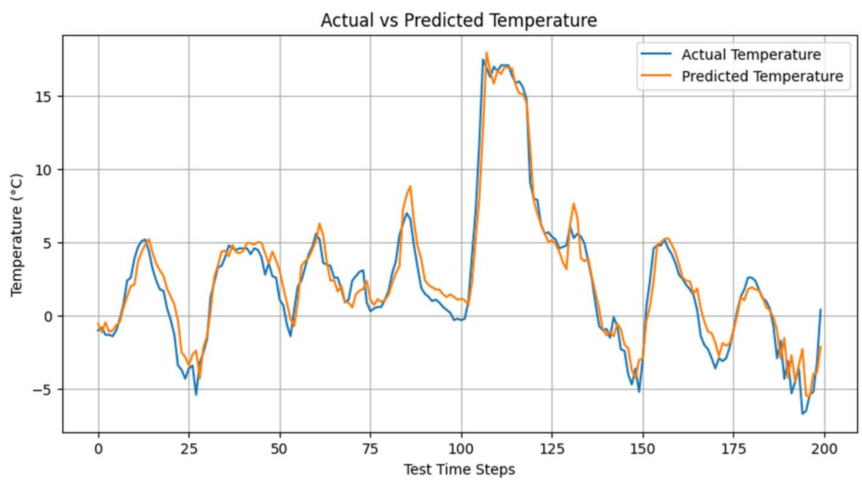

Figure 2: Actual vs. predicted temperature on the test set (first 200 steps).

5.3 Error Distribution Analysis

This dataset, being single-station and tabular, and not spatially gridded, cannot be assembled into a literal spatial error map to assess the model performance as would be done with the global models like FourCastNet. In lieu of error, the adapted form is viewed through variables and over time. The three variables with the lowest error (Pressure, Dew Point Temperature, Temperature) have a comparatively small contribution to the overall error, whereas the three variables with the highest error (Visibility, Wind Speed, Relative Humidity) have a comparatively large contribution to the overall error, in that order. The RMSE values are highest for Visibility (5.8543), this being about sixty times that of Pressure (0.0964); indicating that most of the error mass in the model is in this one variable. The most significant deviations occur around the times of greatest transition, whether of temperature or humidity or pressure, as well as in the Fourier domain, and are not evenly distributed throughout the test window, but are grouped around the times of greatest change.

5.4 Comparison with Related Data-Driven Weather and Nowcasting Models

It is not possible to make a meaningful direct numerical comparison with the large scale, global data-driven models such as FourCastNet, Pangu-Weather, GraphCast, FengWu, FuXi, ClimaX, GenCast, Aurora, Aardvark Weather, because they have different variables, resolutions and error units, and they predict the entire atmosphere on a global grid over a multi-day time horizon. Table 4, on the other hand, puts AFNOWeatherModel in qualitative, architectural context with the other data-based weather and nowcasting models discussed in Section 2.

Model	Domain / Scale	Core Mechanism	Typical Horizon
FourCastNet (base paper) [9]	Global gridded ERA5, ~20 variables	Adaptive Fourier Neural Operator (spectral token mixing)	Autoregressive, 6-hourly steps
Pangu-Weather [10]	Global gridded ERA5	3-D Earth-specific Transformer	Multi-day, direct + hierarchical
GraphCast [11]	Global gridded ERA5	Graph Neural Network (message passing)	Medium-range, autoregressive
ConvLSTM / MetNet(-2) [18][19][20]	Regional radar / satellite grid	Convolutional / recurrent nowcasting	Minutes to ~12 hours
AFNOWeatherModel (proposed)	Single station, tabular, 6 variables	1-D Fourier layer + residual MLP	Direct, 1 hour ahead

Table 4: Qualitative comparison of AFNOWeatherModel against related data-driven weather and nowcasting models.

AFNOWeatherModel is a light-weight end member that only performs 1-D Fourier mixing in time, instead of spatial grid and attention/graph dynamics of its bigger brethren. Its R2 value of 0.96–0.99 on smooth thermodynamic variables: Temperature, Dew Point, Pressure is comparable to the wider literature which shows that the models that are most successful on these variables are the spectral and transformer-based models, and its comparatively lower R2 on Wind Speed and Visibility mirrors the difficulty noted by the models that are more successful on these variables, namely the dedicated nowcasting architectures: MetNet-2, DGMR and NowcastNet, that are successful on precipitation and short-range variables, respectively.

5.5 Limitations

There are three restrictions that need to be emphasised. First, the model has been trained and tested on 1 year's data from 1 station and has not been tested for generalisation to other climates/geographic regions. Second, the model makes only a single direct forecast, a one-hour forecast ahead, whereas its behaviour in longer-term forecasts (in FourCastNet-type models the error accumulates) has not been tested. Third, this work focused only on the evaluation reports which provided regression result statistics only, further diagnostics like confidence intervals, statistical significance testing or probabilistic/ensemble output (in the spirit of GenCast) were not part of this work and are identified as future work in Section 6.

6. CONCLUSION AND FUTURE SCOPE

Thirty-five papers were reviewed, and a common problem was found: an autonomous self-verifying uncertain reasoning agent without any existing framework that directly integrates a data-driven weather forecasting model for short-term weather forecasting. To fill this gap, the paper introduced a light-weight Adaptive Fourier Neural Operator-inspired forecasting model, AFNOWeatherModel, by modifying FourCastNet's global gridded structure to a single-station, tabular, six-variable forecasting task, and tested it on 8,784 hourly observations with a chronological 70/15/15 split for training, validation, and testing.

The model's performance in the held-out test set yielded an overall MAE of 1.9138, MSE of 12.5279, RMSE of 3.5395 and R^2 of 0.8854, best performance on the three variables which are smooth and autocorrelated (Pressure, Dew Point Temperature, Temperature), and weaker performance on the other three variables that are volatile and driven locally (Relative Humidity, Wind Speed, Visibility). The training and validation loss curves did not show any separation, indicating that the proposed model is well suited to the problem size and the two-block, 64-dimensional architecture and 30 epochs training budget proved to be a good match. A qualitative comparison was also made with other global data-driven weather and nowcasting models, which placed the proposed model in a distinct lightweight single-station niche. The results demonstrate that Fourier-domain temporal mixing is effective and computationally efficient inductive bias for short-horizon, station-level weather prediction.

Based on these findings and the identified research gaps in Section 3, a number of directions for future research are proposed: (i) agentic integration – extend AFNOWeatherModel to an agentic reasoning-and-acting layer, in the spirit of ReAct and Reflexion, which will interpret the forecast output, reflect on the quality of the forecast and autonomously trigger corrective or downstream actions; (ii) uncertainty-aware forecasting – extend the direct one-step regression to autoregressive multi-step rollout, and the single-station dataset to multiple stations or a spatial grid with uncertainty-aware forecast output, similar in spirit to GenCast, so that the agentic decision layer can reason about the quality of the forecasts, not just a single value; (iii) multi-step and multi-station forecasting – extend the direct one-step regression to autoregressive multi-step rollout and the single-station dataset to multiple stations or a spatial grid; (iv) end-to-end pipeline automation – extend the current pipeline to a more autonomous, raw-observation-to-action loop similar to Aardvark Weather; and (v) broader evaluation – add confidence intervals, statistical-significance testing and evaluation on additional datasets or locations. These directions converge to one and aim to put together the gap that was the motivation for this work: a unified system composed by the Fourier-domain forecasting model, developed in this study, and an autonomous self-verifying uncertainty-aware agentic reasoning layer.

References

1. Shinn et al. (2023) – Reflexion Verbal self-reflection stored in episodic memory in place of gradient updates