Zeus: A Decentralized Mixture-of-Experts Approach for State-of-the-Art Weather Forecasting

July 17, 2025

Ørpheus AI

Authors

Wouter Haringhuizen Egill de Visser

Abstract— **Continuous improvement in weather forecast accuracy is a great challenge with profound economic and societal implications. Machine learning has emerged as a powerful paradigm for weather forecasting. This paper introduces a novel decentralized mixture-of-experts approach for machine learning-based weather prediction (MLWP). We evaluate the performance of the Zeus subnet, a distributed network on the Bittensor blockchain that incentivizes weather forecasting. Using a robust Best Recent Performer (BRP) metric to identify consistently superior models, we benchmark the network against a dynamic state-of-the-art (SOTA) baseline. Our results, based on 77,208 forecast events, show that the Zeus BRP reduces the root mean square error (RMSE) by a statistically significant 39.8% ($p < 0.001$) compared to the SOTA baseline. This demonstrates that a decentralized, incentive-driven network can produce forecasts that are significantly more accurate than current SOTA systems, establishing a new frontier in predictive meteorology.**

1 Introduction

Timely and accurate weather forecasts are indispensable for modern society and influence decision making in sectors ranging from agriculture and energy to transportation and disaster management. Historically, the field has been dominated by physics-based numerical weather prediction (NWP) models. Although these models have steadily improved, they are computationally intensive and operate within large, centralized institutions. Recently, machine learning-based weather prediction (MLWP) has emerged as a transformative paradigm [4], [5]. By training deep learning models on vast amounts of historical data, MLWP systems can produce forecasts of comparable or superior accuracy to NWP models at a fraction of the computational cost. This has opened the door to new methods of model development and deployment.

This paper explores a novel approach to MLWP by leveraging a decentralized mixture-of-experts approach. We present a comprehensive performance analysis of the **Zeus subnet**, a distributed network on the Bittensor blockchain that incentivizes global operators (“miners”) to develop, host and continuously improve competing weather forecasting models. The central hypothesis is that this decentralized, incentive-driven competition can produce a collective intelligence that is more accurate and adaptive than any single, centralized forecasting system.

Our analysis evaluates the Zeus subnet’s performance in predicting 2m temperature for a 6-day future horizon, benchmarking its accuracy against a strong, dynamic state-of-the-art (SOTA) baseline.

2 Methodology

Our evaluation framework is designed to rigorously assess the Zeus subnet’s 2m temperature forecasting capabilities against a SOTA baseline using robust metrics. The analysis covers all 2m temperature forecast data from the Zeus subnet between **July 07, 2025** to **July 15, 2025**, comprising **77,208** distinct forecast events.

2.1 Data Sources and Baseline Definition

- **Ground Truth:** The final, quality-controlled **ERA5 reanalysis dataset** is used as the single, definitive ground truth for evaluating all forecasts [1]. This data is provided by the Copernicus Climate Change Service (C3S) [2].
- **SOTA Baseline:** We utilize **Open-Meteo** as a SOTA benchmark [3]. This service provides a powerful aggregated forecast by intelligently selecting from a suite of leading global models (e.g., ECMWF-IFS, DWD-ICON, GFS) based on the specific context of the query.

Disclaimer: While Open-Meteo represents a robust, globally-optimized weather service that aggregates forecasts from SOTA operational models, we acknowledge that specialized, proprietary models may outperform it in specific, narrow contexts. However, our goal is to benchmark against a transparent and widely recognized global standard.

2.2 Performance Metrics and the BRP

Our evaluation is primarily based on the root mean square error (RMSE), a metric that quantifies the average magnitude of forecast errors. In addition, we introduce the **Best Recent Performer (BRP)** metric to robustly measure the optimal capability of the subnet. In contrast to a naive ”best-of-the-moment” score, which is susceptible to noise, the BRP is defined as the single miner who achieved the *lowest average RMSE over the preceding 24-hour rolling window* for a given forecast horizon. This metric rewards sustained excellence and identifies consistently reliable models within the network. We compare the SOTA baseline against both the mean performance of all Zeus miners and the more rigorous BRP.

3 Results and analysis

The Zeus subnet demonstrates a decisive and statistically significant performance advantage over the SOTA baseline in 2m temperature forecasting.

3.1 Quantitative Performance Analysis

On average, the Zeus BRP achieved an RMSE of **1.05 Kelvin (K)**. This constitutes a remarkable **39.8%** improvement over the SOTA baseline's average RMSE of **1.74 K**. The collective mean of all subnet miners also outperformed the baseline with an RMSE of 1.16 K, which corresponds to a 33.7% improvement over the baseline. To validate this finding, a Wilcoxon signed-rank test was performed on the paired error values of the BRP and the SOTA baseline, confirming that the BRP's outperformance is highly statistically significant ($p < 0.001$).

This superior accuracy is consistent across all forecast horizons, as shown in Figure 1. The performance gap is substantial at every time step from one to six days out. For instance, for forecasts 1 to 2 days into the future, a critical period for many operational decisions, the BRP improves upon the baseline by an exceptional **42.8%**, as detailed in Figure 2.

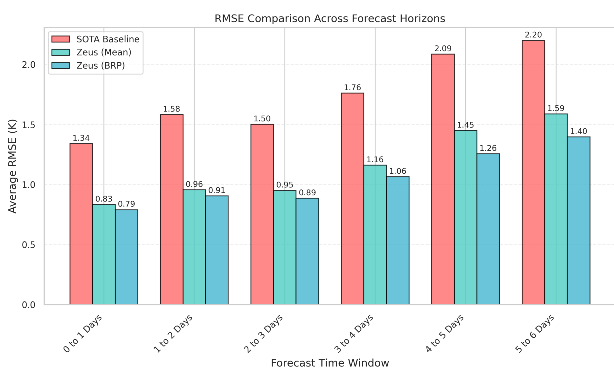

Fig. 1. Average RMSE for the forecasting task. Lower values indicate higher accuracy. The Zeus BRP consistently delivers more accurate forecasts than the SOTA baseline across all forecast horizons.

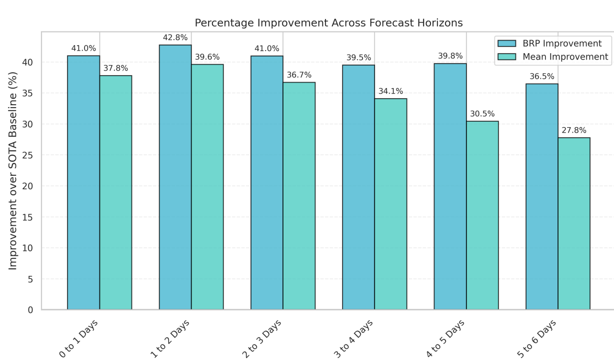

Fig. 2. Percentage improvement over the SOTA baseline for forecasts. Higher values indicate better performance, highlighting the subnet's superior and statistically significant predictive capability.

3.2 Performance Across Forecast Horizons

A key strength of the decentralized approach lies in its ability to identify models specializing in different forecast horizons and collectively harness their complementary capabilities. Our analysis reveals that the BRP maintains a strong advantage across all evaluated time horizons. As shown in Table I, the BRP's RMSE is significantly lower than the SOTA baseline's at every forecast step.

TABLE I
PERFORMANCE OF THE BRP VS. SOTA ACROSS FORECAST HORIZONS

Forecast Window	**SOTA Avg. RMSE (K)**	**BRP Avg. RMSE (K)**
0 to 1 Day	1.34	0.79
1 to 2 Days	1.58	0.91
2 to 3 Days	1.50	0.89
3 to 4 Days	1.76	1.06
4 to 5 Days	2.09	1.26
5 to 6 Days	2.20	1.40

4 Discussion

The empirical results provide powerful validation of the decentralized AI paradigm for predictive tasks. The ability of the Zeus subnet, particularly the BRP, to consistently and significantly outperform a strong SOTA baseline confirms that the network's models possess genuine, state-of-the-art predictive power.

The insights provided by the BRP metric highlight the effectiveness of the subnet's incentive mechanism. It demonstrates that the network is not just a collection of models, but a dynamic system capable of identifying and elevating its most consistently reliable contributors in near-real-time. This ”wisdom of the crowd” mechanism, where the network as a whole can perform better than its average participant by leveraging its top talent, is a core feature of the Bittensor protocol.

Furthermore, the detailed performance breakdown (Table I) shows that the competitive environment produces a model that is robust across a variety of forecast lengths, rather than one that only excels at short-term or long-term forecasts. This leads to a more reliable and multifaceted network.

5 Conclusion

This paper has evaluated the Zeus subnet, demonstrating its exceptional performance in weather forecasting. The network delivers forecasts that are significantly more accurate than a dynamic, state-of-the-art baseline. Notably, the BRP achieves an average RMSE reduction of 39.8%, while the collective mean of all subnet miners improves upon the baseline by 33.7%. This outperformance is statistically significant ($p < 0.001$) and is driven by the network's ability to dynamically leverage diverse, high-performing models. This capability, driven by the powerful mechanisms of decentralized competition, collective intelligence, and emergent specialization, validates the Bittensor approach and showcases the profound potential of decentralized AI to solve complex, real-world scientific problems.

References

[1] Hersbach, H., et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730), 1999-2049.

[2] Copernicus Climate Change Service (C3S). (2017). *ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate*. [Online]. Available: https://climate.copernicus.eu/. [Accessed: July 15, 2025].

[3] Open-Meteo. (2025). *Weather Models Documentation*. [Online]. Available: https://open-meteo.com/en/docs/model-updates. [Accessed: July 15, 2025].

[4] Bi, K., et al. (2022). Pangu-Weather: a 3D High-Resolution model for fast and accurate global weather forecast. *arXiv:2211.02556*.

[5] Lam, R., et al. (2023). Learning skillful medium-range global weather forecasting. *Science*, 382(6677), 1416–1421.