Assessing AI weather models' tropical cyclone intensity forecasts using autonomous surface vehicle data

Toshiyuki Bandai (toshiyuki.bandai@ntt.com), Naoko Kosaka, Naoto Endo, Tsuneko Kura, Ryusuke Yamamoto, Yusuke Umemiya, Gota Yamasaki, Hiroshi Matsubara (NTT)

Background and Objective

Recent AI weather models have shown strong skill in forecasting tropical cyclones (TCs) on standard benchmarks, which often use operational analysis or best-track data as verification targets. However, these references do not necessarily represent the highly variable surface conditions observed near the TC core. This study evaluates whether model skill and ranking change when forecasts are verified against direct observations from autonomous surface vehicles (Wave Gliders; Liquid Robotics, USA) rather than ECMWF analysis or best-track data.

Conclusions

AIFS produced the lowest track errors for the named TCs in the North Pacific in 2024, while Aurora and IFS were competitive for intensity forecasts. However, model ranking changed depending on the verification reference. Verification against Wave Glider observations revealed that near-core surface variability is not fully captured by model outputs and the best-track data. Wave Glider observations therefore provide a complementary and more discriminating benchmark for evaluating AI weather models, particularly for TC intensity forecasts.

References and Acknowledgments

[1] Blaylock, B. K. (2025). [2] Ullrich, P. A., Zarzycki, C. M., McClenny, E. E., Pinheiro, M. C., Stansfield, A. M., Reed, K. A. (2021). We acknowledge the Okinawa Institute of Science and Technology (OIST) for their cooperation when operating the Wave Gliders.

Methods

Models: We evaluated AIFS, Aurora, FengWu, GraphCast, and ECMWF IFS. The AI weather models were fine-tuned on ECMWF operational analysis. The temporal and spatial resolutions are 6 hours and 0.25° , respectively, for all models.

Weather Prediction: We initialized Aurora, FengWu, and GraphCast with IFS 0-hour forecast (HRES_T0) to predict all named TCs in the North Pacific in 2024 (26 TCs in total). For each TC, forecasts were initialized every 12 hours (00 and 12 UTC), starting from 5 days before the beginning of the best-track data provided by the Japan Meteorological Agency (JMA) and ending when the minimum central pressure was recorded. For each initialization time, forecasts were produced for up to 10 days or until the end of the best-track record, whichever came first. For the same initialization times, we obtained forecasts from AIFS and IFS via Herbie[1].

Evaluation: TCs were identified and tracked with TempestExtremes[2]. We evaluated TC track and intensity forecasts against JMA best-track data and ECMWF operational analysis regridded to 0.25° . We also evaluated model outputs against sea-surface observations collected by a MaxiMet compact weather station GMX560GPS (Gill Instruments Limited, UK) mounted on Wave Gliders (SV2 and SV3, also informally nicknamed OISTER and Seiuhi-san, respectively) operated near Okinawa, Japan. We used observations collected between 2024-07-17 and 2024-09-26, which included five TCs that passed near Okinawa (202403, 202409, 202410, 202413, and 202414). For Wave Glider verification, model outputs were interpolated in space and time to the observation locations and times, and wind speed and pressure were adjusted to 1.3 m above sea level to match the sensor height.

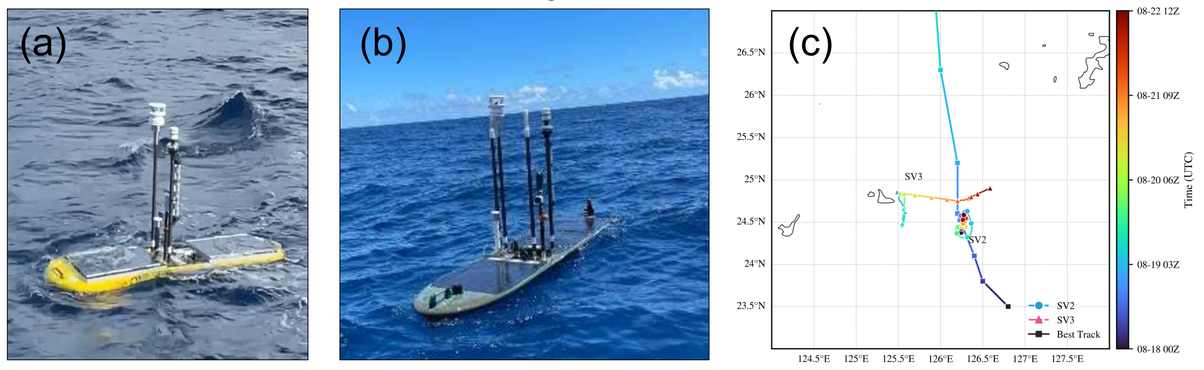

Figure 1. Wave Gliders operated near Okinawa, Japan: SV2 (a) and SV3 (b). Panel (c) shows the best track of Typhoon 202409 and the tracks of Wave Gliders SV2 and SV3.

Results

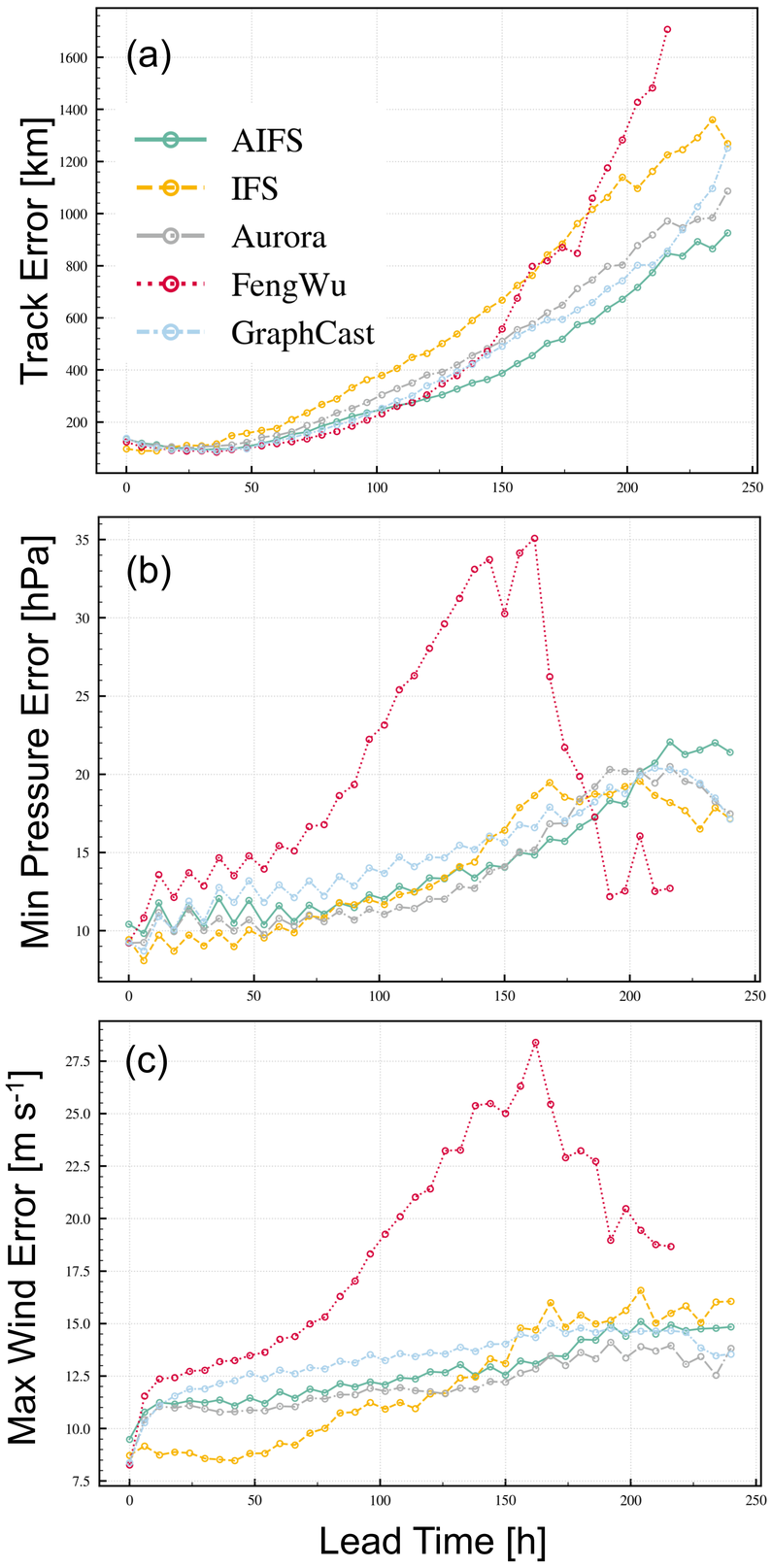

Figure 2. Mean absolute TC forecast errors for all named TCs in the North Pacific in 2024 as a function of lead time, evaluated for track position (a), minimum pressure (b), and maximum wind speed (c) against JMA best-track data.

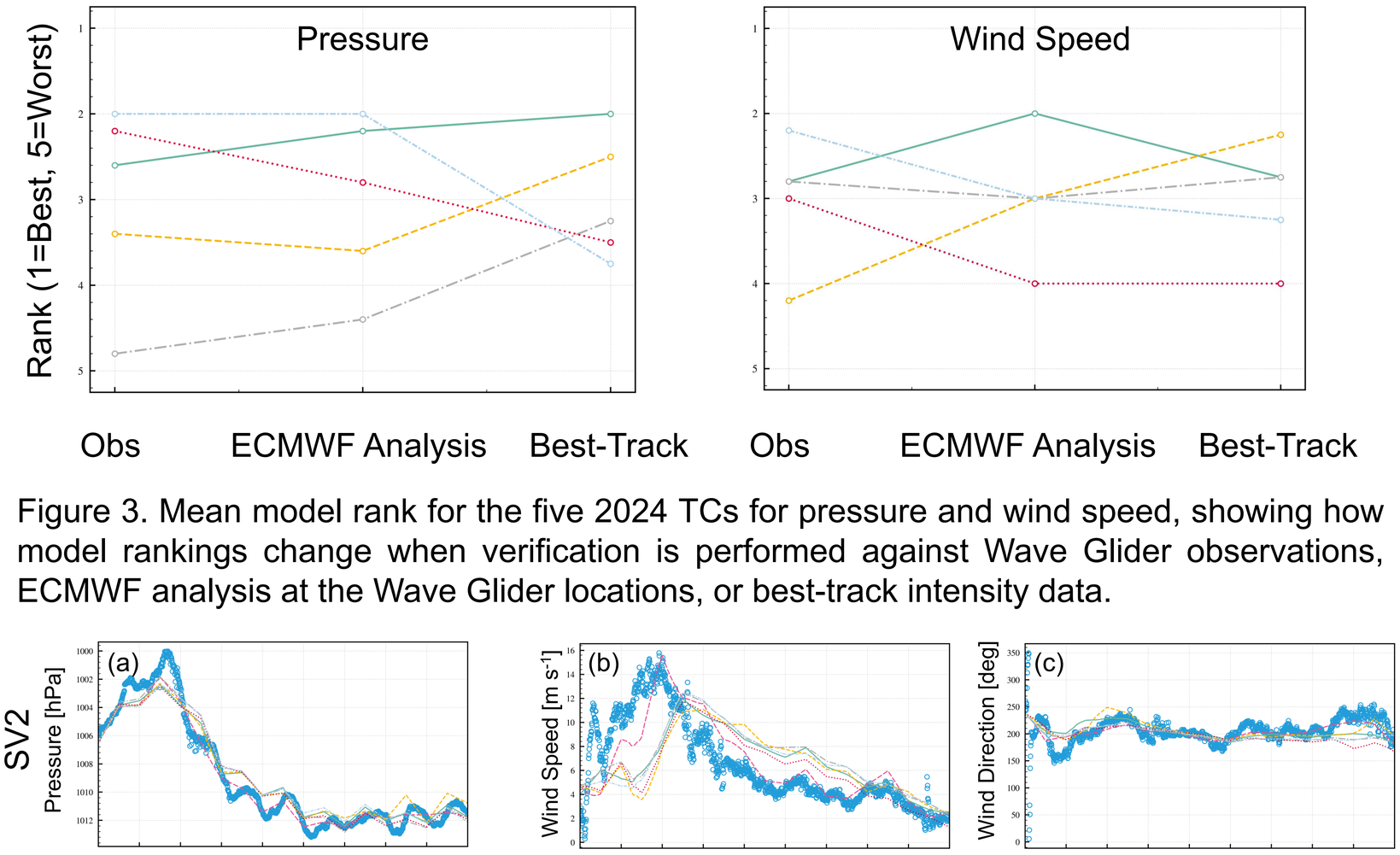

Figure 3. Mean model rank for the five 2024 TCs for pressure and wind speed, showing how model rankings change when verification is performed against Wave Glider observations, ECMWF analysis at the Wave Glider locations, or best-track intensity data.

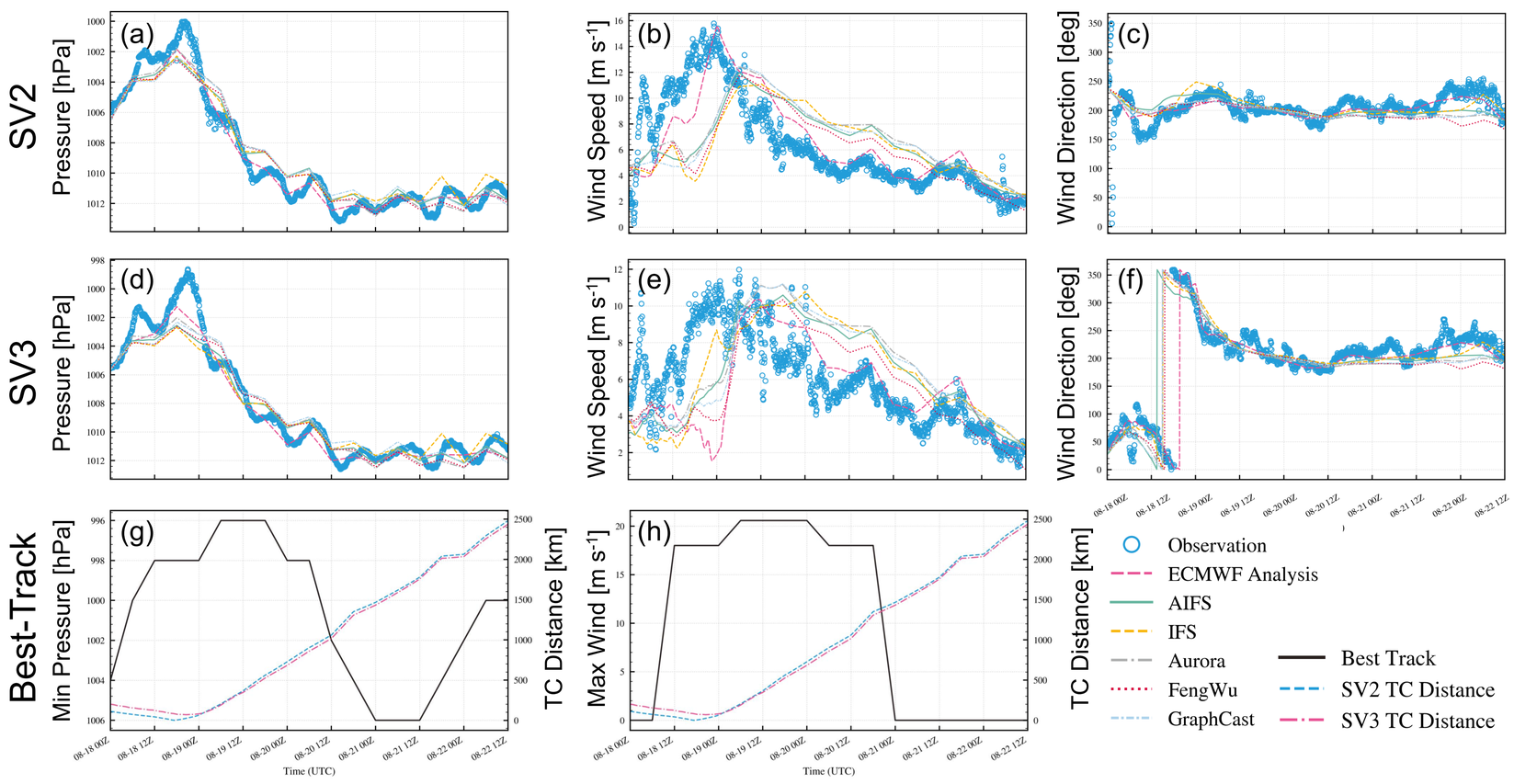

Figure 4. Verification of model output for TC 202409 against Wave Glider observations. The models were initialized at the beginning of the best-track record. Panels (a)-(c) show SV2 observations for pressure, wind speed, and wind direction, respectively. Panels (d)-(f) show SV3 observations for pressure, wind speed, and wind direction, respectively. Panels (g) and (h) show best-track minimum pressure and maximum wind speed, respectively, together with the Wave Glider distance from the TC center.

© NTT, Inc. 2026