

EGU24-13853, updated on 23 Jul 2024

<https://doi.org/10.5194/egusphere-egu24-13853>

EGU General Assembly 2024

© Author(s) 2024. This work is distributed under the Creative Commons Attribution 4.0 License.



A Research Agenda for the Evaluation of AI-Based Weather Forecasting Models

Imme Ebert-Uphoff^{1,2}, Jebb Q. Stewart³, Jacob T. Radford^{1,3}, and the CIRA-NOAA team^{*}

¹Cooperative Institute for Research in the Atmosphere, Colorado State University, Fort Collins, CO, United States of America

²Electrical and Computer Engineering, Colorado State University, Fort Collins, CO, United States of America

³Global Systems Laboratory, NOAA, Boulder, CO, United States of America

^{*}A full list of authors appears at the end of the abstract

Over the past few years purely AI-driven global weather forecasting models have emerged that show increasingly impressive skill, raising the question whether AI models might soon compete with NWP models for selected forecasting tasks. At this point these AI-based models are still in the proof-of-concept stage and not ready to be used for operational forecasting, but entirely new AI-models emerge every 2-3 months, with rapidly increasing abilities. Furthermore, many of these models are orders of magnitude faster than NWP models and can run on modest computational resources enabling repeatable on-demand forecasts competitive with NWP. The low computational cost enables the creation of very large ensembles, which better represent the tails of the forecast distribution, which, if an ensemble is well calibrated, allows for better forecasting of rare and extreme events.

However, these AI-based weather forecasting models have not yet been rigorously tested by the meteorological community, and their utility to operational forecasters is unknown. In this presentation we propose several studies to address the above issues, grouped into two central foci:

(1) Nature of AI models: AI-based models have very different characteristics from NWP models. Thus, in addition to applying evaluation procedures developed for NWP models, we need to develop procedures that test for AI-specific weaknesses. For example, NWP models and their physics backbone guarantee certain properties - such as dynamic coupling between fields - that AI-based models are not required to uphold. Developing suitable tests is based on a fundamental understanding of the AI-based models.

(2) Forecaster Perspective: Evaluation of weather forecasting models should be performed with respect to particular applications of weather forecasts, and it is critical to have research meteorologists and operational forecasters involved in the evaluation process. Our initial evaluation of AI-based models in CIRA weather briefings revealed that these models have characteristics that make interpretation of their forecasts fundamentally different from the physics-based NWP model predictions meteorologists are familiar with. For example, the increasing “blurriness” of AI-based predictions with longer lead times is not a reflection of weaker

atmospheric circulations, but rather a reflection of uncertainty. Evaluations aimed at specific meteorological phenomena and atmospheric processes will allow the community to make informed decisions in the future regarding in what environments and for which applications AI-based weather forecasting models may be safe and beneficial to use.

In summary, AI-based weather forecasts have different characteristics from familiar dynamically-based forecasts, and it is thus important to have a robust research plan to evaluate many different characteristics of the models in order to provide guidelines to operational forecasters and feedback to model developers. In this abstract we propose a number of characteristics to evaluate, present results we already obtained, and suggest a research plan for future work.

CIRA-NOAA team: Kyle A. Hilburn, Robert T. DeMaria, Randy Chase, Ryan A. Lagerquist, Charles White, Yoonjin Lee, Jason Apke, Kate D. Musgrave, Lander Ver Hoef, Christina E. Kumler, Matthew S. Wandishin, Jeffrey D. Duda, Isidora Jankov, David D. Turner, Marie McGraw, Mark DeMaria, Galina Chirokova, Sergey Frolov, Laura Slivinski, Tim Smith, Niraj Agarwal