NOAA SCIENCE ADVISORY BOARD REPORT ON NOAA INVESTMENT IN DEEP LEARNING NUMERICAL WEATHER PREDICTION

PRESENTED TO THE NOAA SCIENCE ADVISORY BOARD
BY THE ENVIRONMENT INFORMATION SERVICES WORKING GROUP

JUNE 18, 2024

NOAA Science Advisory Board (SAB)
Environmental Information Services Working Group (EISWG)

Statement on NOAA Investment in Deep Learning Numerical Weather Prediction

18 June 2024

Problem statement

NOAA's mission is to understand and predict changes in climate, weather, oceans, and coasts, and to share that knowledge and information with others. Numerical weather prediction (NWP) is the underpinning technology behind accurate predictions of weather changes. Products and services that NOAA provides improve when NOAA upgrades its NWP guidance; improving NWP is the rising tide that lifts many NOAA service boats. While humans provide interpretations of numerical guidance, our advancing skill in predicting hurricanes, tornadoes, floods, and snowstorms are largely driven by NWP advances. These improvements have accumulated at a rate of about one day a decade, i.e., a four-day forecast now is as accurate as a three-day forecast produced a decade ago. The slow accumulation of skill, which translates into improved products and services, represents a "quiet revolution" in weather prediction.

This quiet revolution began in the 1950's with very simple computer models that started by applying simplified approximations of Newton's Laws of motion applied to atmospheric circulation. With the steady increase in computer power, scientists developed ever more sophisticated NWP models, and in the past few decades, these now encompass ensembles of predictions of ocean, land, and atmosphere that facilitate an estimation of the variety of forecast scenarios that may unfold.

The quiet revolution has come through a massive public investment, whether measured in the number of staff involved in system development, in the computational and observational data resources needed, or the number of lines of code in the prediction system. Even with the computational horsepower now available, it is impossible to accurately model all of the myriad processes that may affect a forecast; the wind blowing through every tree, or the life cycle of individual water droplets. Approximations are necessary in these NWP approaches – the average dissipation of winds from trees in an area several km on a side, or the average heating in a volume caused from the condensation of water vapor across a collection of clouds in a grid cell.

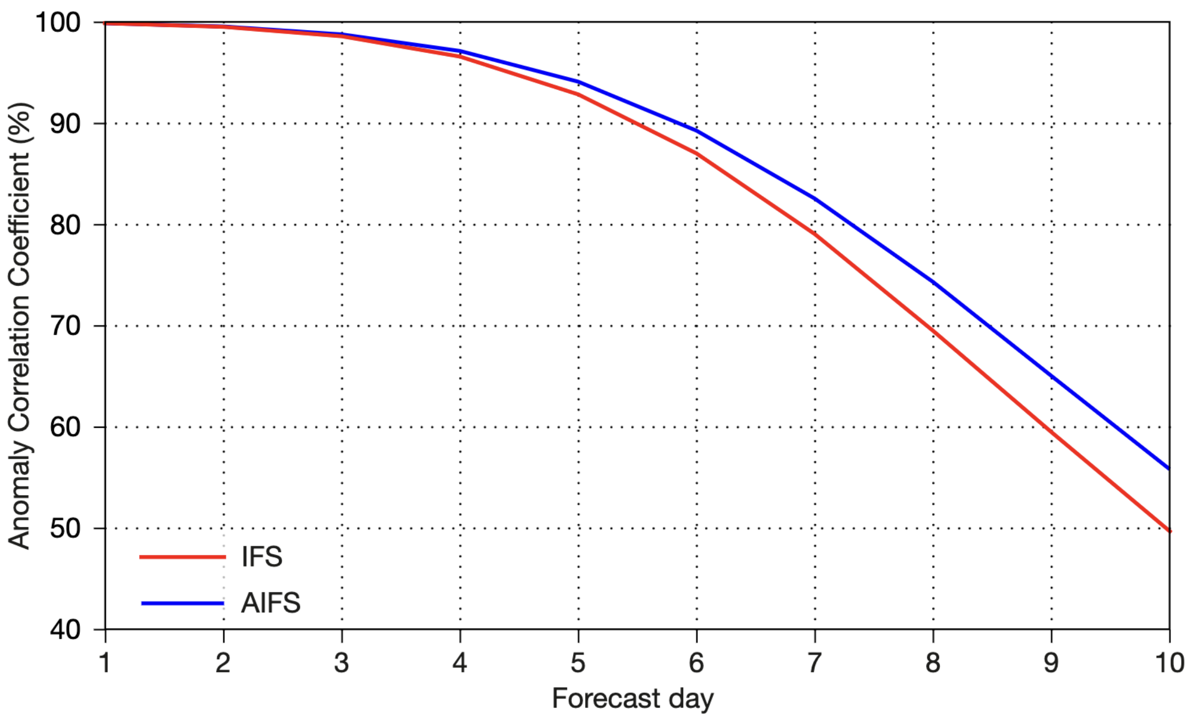
Despite enormous investment by NOAA and peer institutions in developing complicated approximations (or "parameterizations") as faithful to known physics as possible, they remain a simplification, a source of error in the NWP system, and a maintenance burden. With many interacting approximations in the system, isolating and correcting sources of error is an ongoing and increasing challenge in the coupled prediction systems. This requires lots of experts in individual processes, it requires lots of thorough and computationally expensive testing and iteration. This is, in part, why system upgrades in NWP systems happen only at a cadence of once every several years, not every few weeks or months.

Within the last few years and with the advance of artificial intelligence, a radically different approach to NWP has been developed. The new models are *data-driven*; they do not represent a complex human codification of the physical laws of motion and parameterized processes. Instead, comparatively simple neural-network models are trained. In a common method of coding these models, this provides a sophisticated mapping from the current atmospheric state to the state a few hours or days hence; the weights used in the neural network are chosen to minimize error, often root-mean square error. These mappings are typically chained together to provide a prediction; from the current state, a forecast is made to three hours in the future; from the forecast at three hours, a forecast is made to six hours hence, and so forth. Henceforth we will refer to these data-driven models as deep-learning NWP models, or DLNWP. The complexity of these models is great but is hidden within the neural network; the actual number of lines of bespoke code written is very small compared to conventional NWP, perhaps by an order of 100 or more.

The first low-resolution, proof-of-concept DLNWP models were developed only in the late 2010's and were significantly less accurate than conventional NWP forecasts. Informed by these proofs of concept, more computational horsepower was made available to train more sophisticated DLNWP models, and these have increased in skill at a dramatic rate. In selected ways of measuring weather forecast skill, several of these are now competitive with or more skillful than forecasts from the world-leading European Centre for Medium-Range Weather Forecasts (ECMWF; discussed here). Advanced DLNWP development was mostly led by initiatives in private industry including, for example, GraphCast, Met-Net-3, NeuralGCM, and GenCast at Google/DeepMind and FourCastNet and CorrDiff from Nvidia. DLNWP represents a radical change for weather prediction – simplified code that bypasses the many complex parameterizations of physical processes and yet may produce more accurate forecasts.

Understanding of the tradeoffs for conventional versus DLNWP is necessarily tentative given the novelty of the latter. The DLNWP approach is surprisingly accurate, given its comparatively short history of development. The forecasts have less systematic error and hence may not need the substantial statistical postprocessing of conventional NWP guidance. After training, DLNWP predictions are much, much more computationally efficient than conventional NWP, perhaps by a factor of 1,000 or more. Suggested limitations to early DLNWP forecasts have been quickly addressed by industry developers. Many of the complexities with conventional NWP parameterizations are elided in the data-driven approach, leading to a leaner, easier-to-maintain software stack, one that leverages efficient open-source libraries. Testing new models is more straightforward without explicitly parameterizing processes and their interactions, as in conventional models. New product generation paradigms may be possible for NOAA customers. For example, with the reduced computational expense, the local computation of forecasts may be possible, thereby avoiding the cost of massive storage and internet transfer of data from centralized production facilities; only model initial conditions need be transferred. This may facilitate local computation of probabilistic forecasts that greatly aid improved customer decisions based on the forecast guidance.

There are disadvantages to DLNWP as we understand it in 2024. It front-loads computations into a model training phase which can be computationally very expensive. It requires different computational hardware, GPUs versus NOAA's current investment in CPUs. Many of the first-generation DLNWP models do not provide guidance with the small-scale weather variability forecasters prefer to see, details such as supercell thunderstorms, hurricane eye walls, valley fog, or the information on wind gusts. These models are garbage-in, garbage-out; the training data for DLNWP must be of high quality, comprehensive, and spanning a wide range of weather scenarios, otherwise the trained model may not be trustworthy. Despite these problems, the skill of these systems (Fig. 1) and the pace of industry development suggests that overcoming many of these limitations will be possible in the next few years.



AIFS forecast skill. We show the northern hemisphere Anomaly Correlation Coefficient (ACC) for geopotential height at 500 hPa of IFS forecasts (red, dashed) and AIFS forecasts (blue) for 2022. Higher values indicate better skill.

Figure 1: Reproduced from Lang et al., *ECMWF Newletter,* **178**, Winter 2023-2024. The IFS is the conventional NWP system of ECMWF, the AIFS is the deep-learning, artificial-intelligence version of the system. Higher skill is better.

Because DLNWP is so novel, too early to know for sure whether it will play a limited role in weather prediction or whether it will represent a core capacity years hence. However, the rapid advancement of these data-driven methods argues for a substantial NOAA investment. This emergence of DLNWP has come so rapidly that NOAA has yet to develop a comprehensive agency DLNWP plan, with its existing plans still largely reflecting conventional NWP development priorities, or with the use of AI in NWP subspecialties rather than in holistic DLNWP. For example, in late 2021, the NOAA Science Advisory Board prepared a comprehensive report for NOAA, Priorities for Weather Research (PWR). While AI/deep learning was mentioned in this report, this preceded many of the developments in private industry noted above, and DLNWP was not highlighted as a key priority for future NWP. NOAA has its Center for AI, but again, DLNWP developments are so novel that existing AI funding within NOAA is not concentrated on DLNWP but instead facilitates other smaller projects. The NOAA Weather Program Office mentions AI in its plan, but in limited capacities, such as in model statistical post-processing. DLNWP activity is yet to be mentioned as a priority in the NWS Strategic Plan or in the OAR Strategic Plan.

Recognizing the radical potential of the DLNWP approach, peer institutions such as the European Center for Medium-Range Weather Forecasts (ECMWF), the UK Met Office, and Environment Canada have recently made major investments in DLNWP, exploring the efficacy of DL-based global probabilistic prediction, data assimilation, and coupled seasonal and climate prediction.

Motivated by these developments, NOAA has conducted several recent meetings (here, here) to discuss potential investment and focus areas, but NOAA has not yet made a core investment directly in DLNWP. This delay represents a substantial risk to NOAA, which is expected to be the authoritative provider of weather forecasts for the US.

In the statement to follow, we provide a concise set of basic recommendations for near-term NOAA actions specifically related to DLNWP. These will facilitate NOAA in its ability to evaluate DLNWP and put it on a sound foundation should this approach be as revolutionary as early results indicate. These recommendations are preliminary steps – they are not a comprehensive list of all activities NOAA may require to effectively incorporate DLNWP into regular operational production and product generation. The recommendations presume that once DLNWP is better understood within NOAA and a base capacity is established, NOAA will be well positioned to address how these new forecast capacities are turned into improved products and services.

Recommendations

1. A substantial investment in new staff with DLNWP expertise. NOAA will need scientists with different skills in deep learning, whether achieved by recruiting from other parts of the enterprise or through retraining. We encourage NOAA to identify how many they believe are needed to establish a base DLNWP capacity; what skill set is necessary from those FTE's; and we recommend that NOAA briskly hire or retrain this core staff. Some of the initial work will be managerial in character: Writing a roadmap, facilitating collaborations, elucidating what is possible with the new technology to NOAA staff and stakeholders. Hence, the team should include scientist leaders as well as scientists familiar with deep learning.

An appropriate level of initial staffing to develop and evaluate a base capacity is not easily determined and is dependent upon several factors. Based on the very promising DLNWP results, ECMWF recently added 15 FTE's working at the intersection of atmospheric predictions and deep learning, this for an organization much smaller than NOAA, with a narrower mission. After the DLNWP technology is thoroughly evaluated and advanced in readiness level, more staff may need to be reoriented to the transition to operational DLNWP-based products and services, but decisions on these can await a more comprehensive understanding and development of a core DLNWP capacity.

2. Roadmap. NOAA should recognize the potential revolutionary character of DLNWP and incorporate this into relevant strategic plans as well as more directed documents. This would include preparing a living roadmap document for DLNWP evaluation and development, with the recognition that this topic area is rapidly developing and that adjustments to it are to be expected. A DLNWP roadmap should identify key research questions, necessary data sets, DLNWP focus areas, and a plan of action. The roadmap should outline how NOAA will train a fewer number of prediction systems while serving the diversity of NOAA operational NWP needs, given that DLNWP model training can be computationally expensive. It should also include an early assessment of the investments necessary to transform today's operational system to a future system that is potentially supported by DLNWP. Since DLNWP systems may have very different software architectures and pipelines, DLNWP scientists should not be unduly constrained by conventional NWP approaches.

3. Data collation and reanalysis. DLNWP methodological development can benefit from large, curated data sets including reanalyses and long time series (at least 20 years, ideally 40 years) from major observing systems such as satellite radiances, radars, and conventional data. With ECMWF/Copernicus providing high-quality, lower-resolution multi-decadal global reanalyses, there is a particular need for very high resolution, high-quality, unbiased reanalyses and data collections covering areas of US interest, including CONUS, hurricane basins, Alaska, and Hawaii, designed to serve multiple use cases and made readily available to all developers across the enterprise. These will facilitate the generation of products such as Warn on Forecast and high-resolution precipitation guidance for NOAA's hydrologic predictions. The rapid production and public dissemination of such data should be an initial priority. We note that DLNWP and reanalyses are not necessarily independent activities. For example, as the data assimilation underpinning reanalyses requires a prior forecast and an estimate of its uncertainty, the advance of DLNWP may facilitate the more computationally efficient reanalysis production.

4. Changes in computational resourcing. NOAA has largely assumed that model development and operational production costs would be comparable and would be conducted primarily using many CPUs, and it has purchased supercomputer resources with this in mind. DLNWP breaks this paradigm. Surges of GPU-based computations will be required for model training, but radically less computational resources may be needed for prediction. Re-thinking NOAA high-performance computational resourcing will be needed alongside the development of a DLNWP capacity and its evaluation. NOAA should evaluate how much computational resources will be

needed for initial DLNWP training and data set development to get it to a point of being able to more comprehensively evaluate DLNWP and estimate longer-term computational resourcing. It should provide those resources to the DLNWP developers.

5. Partnerships. Collaboration is in the mutual interests of partners across the enterprise, given the cross-disciplinary and rapidly evolving nature of DLNWP and the natural synergies. Industry partners want ready access to NOAA data for their own development, and NOAA wants access to early industry experience with DLNWP. NOAA should identify barriers to industry, academic, and pan-government collaborations and work to address them. NOAA should make all its DLNWP data readily available to others through partnerships with cloud providers. Pan-government collaborations are greatly encouraged where they advance NOAA interests; however, it is important for NOAA to proceed briskly with DLNWP. Other federal agencies may be slower to pivot.

6. Management of DLNWP. As NOAA DLNWP will be built from the ground up, there is an opportunity to improve upon past practices. We recommend that NOAA build and manage its DLNWP as a coherent team, working collaboratively in pursuit of the common goal, with the enthusiastic support of all affected line offices. Any ancillary seed projects (within or across line offices) need to advance NOAA's overall DLNWP position and be aligned with the main agency DLNWP plan and resource pool.

Summary

To better serve its customers with state-of-the-science forecast products, NOAA has an opportunity and a responsibility to explore deep learning numerical weather prediction, resourcing it commensurate with its game-changing potential. The NOAA Science Advisory Board's Environmental Information Services Working Group has outlined what we believe are constructive steps to this end.