BAMS
Meeting Summary

Integration of Emerging Data-Driven Models into the NOAA Research-to-Operations Pipeline for Numerical Weather Prediction

Sergey Frolov,[a] Kevin Garrett,[b] Isidora Jankov,[c] Daryl Kleist,[d]
Jebb Q. Stewart,[c] and John Ten Hoeve[e]

KEYWORDS:
Numerical weather prediction/ forecasting;
Coupled models;
Machine learning

Meeting Title: AI4NWP: Integrating Emerging Machine Learning Tools into NOAA's Research-to-Operations Pipeline for Numerical Weather Prediction

What: Identify a research and development roadmap and priorities for integrating emerging deep learning tools for numerical weather prediction into the NOAA production pipeline.

When: 28–30 November 2023

Where: Boulder, Colorado, and online

DOI: 10.1175/BAMS-D-24-0062.1
Corresponding author: Sergey Frolov, sergey.frolov@noaa.gov
In final form 29 February 2024

AFFILIATIONS: [a] Physical Sciences Laboratory, National Oceanic and Atmospheric Administration, Boulder, Colorado; [b] National Weather Service, National Oceanic and Atmospheric Administration, Silver Spring, Maryland; [c] Global Systems Laboratory, National Oceanic and Atmospheric Administration, Boulder, Colorado; [d] Environmental Modeling Center, College Park, Maryland; [e] Weather Program Office, National Oceanic and Atmospheric Administration, Silver Spring, Maryland

1. Imperative to act

Over the last 18 months, a tidal wave of data-driven models revolutionized the way we think about the future of the weather prediction enterprise (Bauer et al. 2023). Groundbreaking data-driven models (Pathak et al. 2022; Bi et al. 2023; Lam et al. 2023) showed that machine learning (ML) emulators trained on the ERA5 reanalysis are capable of achieving the skill of the flagship deterministic weather forecast from the European Centre for Medium-Range Weather Forecasts (ECMWF) at ~1/1000 cost of the traditional model. Early results also show enhanced skill of the new data-driven models in predicting extreme life-threatening events like hurricanes, winter storms, and heatwaves.

The National Oceanic and Atmospheric Administration (NOAA) is dedicated to understand how the climate, weather, and oceans are changing; issue effective forecasts of these changes from minutes to decades into the future; preserve living marine resources and their habitats; disseminate this information to the Nation; enhance the Nation's economy; and save lives and property. For NOAA to continue to be effective, it is imperative to embrace the emerging data-driven modeling technology to deliver environmental forecasts and datasets to the Nation more rapidly, efficiently, and with more skill. The velocity with which the field of data-driven models is developing is unprecedented and transformative and requires a reimagining of how innovations can be incorporated into operational practice and service delivery.

By embracing this opportunity and establishing collaborations across government agencies and with the private and academic sectors, NOAA can play a key role in what could be a once-in-a-generation innovation in numerical weather prediction. Seizing this opportunity will ensure that the nation's people and commerce have access to accurate and timely weather forecasts and are prepared for emerging threats from changing climate and extreme weather conditions.

2. Workshop objectives and format

To respond to the emergence of the data-driven models for numerical weather prediction, NOAA Research, NOAA's National Weather Service, and NOAA Cooperative Institutes convened a 2-day hybrid workshop in Boulder, Colorado. Participants discussed the following:

- Opportunities and barriers for incorporating emerging data-driven models into the NOAA research-to-operations pipeline.
- Pathways for connecting existing NOAA investments in basic research, Earth system observations, and traditional models with new opportunities in data-driven modeling.
- Partnerships needed within NOAA, private industry, and academia.
- Scope and focus of investment areas required to develop a viable research initiative in data-driven models at NOAA.

The first day of the workshop started with a clear message from the Assistant Secretary of Commerce for Environmental Observation and Prediction on the importance of incorporating the data-driven models into NOAA research and operations and relevance of this emerging technology to our mission and our Nation. On the first day of the workshop, keynote speakers from NOAA, ECMWF, and the private industry described the state of science in data-driven modeling. NOAA speakers described the existing research-to-operations pipeline at NOAA, NOAA weather testbeds, and NOAA-native datasets for training of the data-driven models. The breakout discussions during day 1 focused on identifying opportunities and barriers for integrating data-driven models into NOAA's research-to-operations pipeline.

Day 2 of the workshop focused on identifying opportunities and research efforts in data-driven modeling that can lead to improved skill of the NOAA forecasts. Keynote speakers from NOAA reviewed the relevance of data-driven models to NOAA's priorities such as the NOAA weather testbeds; data assimilation; short-range, medium-range, and extended-range prediction; reanalysis and training dataset generation; and hurricane prediction. The breakout sessions provided an opportunity for participants to further ideate research efforts that can advance the priority areas above.

The workshop was closed with a panel discussion that included workshop organizers from NOAA, academia, and private industry. The panel summarized high-level action items for the workshop audience and to the Assistant Administrators from the Office of Atmospheric Research and the National Weather Service.

3. Workshop findings

As a result of these 2 days of detailed presentations, in-depth discussions, and ideation sessions, workshop participants outlined several opportunities, barriers, future strategies, and grand challenges. These items below are intended to present NOAA with potential pathways forward on how best to integrate data-driven models into numerical weather prediction, inform existing operations, and help guide future conversations around this innovative initiative. This report provides recommendations to NOAA but does not necessarily reflect the views of NOAA or the Department of Commerce.

a. Opportunities. Workshop participants identified the following opportunities for data-driven models to enhance NOAA's ability to deliver on its mission at lower cost and with more fidelity:

- Ensemble forecasts represent the majority of computational expenses in NOAA's research and operational enterprises. Replacing ensemble forecasts with data-driven models presents a great opportunity to reduce computational costs across the enterprise. This can potentially result in redirecting of freed-up computational resources to currently underresourced research and operational priorities.
- Skill improvements can be achieved by using significantly larger ensemble forecasts performed with data-driven models (~1000 ensemble members). These larger forecast ensembles will improve probabilistic forecasts of extreme events.
- Skill improvements can be achieved by improving initial conditions for weather forecasts by using data-driven models to generate large ensembles for the currently operational ensemble-based data assimilation or for enabling a future implementation based on more advanced particle filters. Data-driven models also show promise for automatic differentiation of the forecast model, which can facilitate the development of 4DVAR-based algorithms for NOAA applications (Xiao et al. 2023).
- Further improvements in the forecast skill of data-driven models can be achieved by integrating feedback from NOAA forecasters and NOAA weather testbeds. This feedback

can be further formalized in a form of the tuned ML cost function that can hyperfocus data-driven forecast skill on the priorities identified by NOAA forecasters.

- The current pace of major model upgrades at NOAA is measured in years and is, in large part, driven by the high computational and wall-clock costs of testing innovation as they propagate from low to high readiness levels. Lower costs and rapid testing with data-driven models provide a unique opportunity for NOAA to accelerate innovation and transition of research to operations. Furthermore, the rapid pace and low cost of innovation in industry, while leveraging decades of government investment in research, shows that it is possible to advance innovations to operations at a significantly faster pace by embracing the industry-standard frameworks for ML operations, continuous integration (CI), and continuous deployment (CD).
- Development of NOAA-native benchmark datasets for training artificial intelligence (AI) models (Dueben et al. 2022) can facilitate further research in ML and numerical weather prediction across government, academia, and private industry. NOAA is uniquely positioned to develop these benchmark datasets given its extensive archive of observational and modeling data.

b. Barriers. Workshop participants identified the following barriers to integrating data-driven models into the existing NOAA research-to-operations pipeline:

- The existing concepts of readiness levels and transition gates were designed based on the assumption that testing of innovations is slow and computationally expensive. Innovation in the data-driven modeling field showed that data-driven models are evolving significantly faster than any of the existing roadmaps for future operational systems at NOAA. For NOAA to successfully embrace the data-driven models, alternative testing and transition to operations procedures will need to be established.
- The current generation of data-driven models is focused on deterministic forecasts for the global atmosphere. To incorporate data-driven models into the NOAA operational suite will require development of data-driven models for ensemble forecasting, coupled Earth system prediction, and kilometer-scale applications (e.g., hurricane and regional storm scale).
- For NOAA to continue benefiting from its investment in the basic science of physical model development, NOAA will need to develop and produce native benchmark training datasets that are generated based on NOAA native modeling and data assimilation technologies.
- NOAA does not have assured access to sufficient computing on platforms with modern graphics processing unit (GPU) and tensor processing unit (TPU) processors required for training of the ML models. The existing NOAA-owned GPU capacity is insufficient. Additional agreements are needed to assure sufficient access to GPU, TPU, and ML operations (MLOps) resources utilizing cloud providers or partnerships with other government agencies.
- NOAA has very limited human capital and know-how required to develop and innovate in the field of data-driven models. It is essential that NOAA invests in recruitment, training, and retention of new and existing talent to develop native skill sets in data-driven modeling. It is also essential that NOAA develops native data-driven modeling capabilities based on the best practices available in the open literature.
- There is currently limited cultural conditioning of the NOAA's internal and external user base to data-driven models. While workshop attendees indicated a palpable excitement about the potential of data-driven models to help NOAA to more effectively deliver on its mission, a certain cultural resistance is expected from the workforce, decision-makers, and users to new technology.

c. Strategies for moving forward. Workshop participants identified several strategies to overcome barriers and incorporate data-driven models into the NOAA research-to-operations pipeline:

- Establish a working group that cuts across line offices and divisions. One way that organizations respond to the introduction of innovations (Christensen 2016) is by establishing cross-division work groups that can cut across traditional (often siloed) organizational structures. It is recommended that NOAA leadership charters a cross-line office team that can help coordinate the development and integration of data-driven models into the NOAA research-to-operations pipeline.
- Focus on the development of AI-ready datasets and metrics for training of the data-driven models. Achieving this outcome will require sufficient computational and human resources focused on the production of NOAA-native training datasets. A similar effort in Europe focused on reanalysis development is resourced on the order of 15 full-time employees and ~1 billion CPU hours annually. It is essential that the new dataset include both global coupled and kilometer-scale regional datasets produced with the best versions of the United Forecast System (UFS) modeling and Joint Effort for Data assimilation Integration (JEDI) systems.
- Develop a set of metrics for training of kilometer-scale models. The current generation of data-driven models for the global atmosphere benefited greatly from the availability of standard score cards used by centers like the ECMWF to evaluate improvements in the global modeling system. A similar quantitative metric needs to be developed for kilometer-scale models.
- Focus resources across NOAA on two Grand Challenges (described in detail in the next section) that will exercise the new data-driven models in the entire NOAA pipeline, including research, transition to operations, computational resources, and weather testbeds.
- Introduce data-driven capabilities to the NOAA stakeholders through a combination of the parallel forecast suite and through existing NOAA testbeds.
- Mutualize AI expertise across laboratories and line offices. Unlike the traditional modeling expertise that can be very application specific (e.g., global vs hurricane modeling), ML expertise tends to be more cross-cutting. In addition, recruiting and retaining ML specialists will be challenging due to high competition with the private sector. To address the shortage of the human capital, it is suggested that NOAA establish a cross-line office development group that can benefit from the ML workforce distributed across line offices and research laboratories. This working group can further leverage resources available through the NOAA Center for Artificial Intelligence and through the National AI Research Resource.
- Build partnerships with the private sector, across government, and with academia. This includes actions to
- Establish and adjust existing cooperative research and development agreements with industry players that are at the forefront of the data-driven model development, such as Google, NVIDIA, and Microsoft.
- Develop partnerships across government agencies that are also invested in kilometer-scale data-driven models and training dataset development, such as the National Science Foundation, Department of Energy, Department of Defense, and the National Aeronautics and Space Administration. These cross-government partnerships can be focused on mutual science interests or on assuring access to mutual computational resources, training datasets, and MLOps frameworks.
- Secure compute and tooling support for the ML-based research-to-operations pipeline from large-scale cloud compute providers. This includes assured access to both sufficient ML training hardware and MLOps frameworks such as AzureML, VertexAI, or SageMaker.

- Accelerate transition of existing small-scale (individual Principle Investigator) projects that add ML to existing NOAA products (e.g., emulation of physics, observational operators, or radiation modeling). NOAA Center for Artificial Intelligence can serve as an effective incubator and accelerator of these small-scale AI projects through the NOAA research-to-operations pipeline by helping teams update NOAA processes and guidelines for transition of data-driven models; interface with NOAA's Office of Research Transition and Application; coordinating trustworthy AI training; and curating a library of contributed open science outcomes including reusable software and Jupyter notebooks.

4. Suggested grand challenges

a. GEFS-AI. Ensemble forecasting represents the bulk of the computational cost in the NOAA research and operations. Ensemble forecasts are also at the heart of the future transitions at NOAA and peer weather centers such as ECMWF, the Met Office (UKMO), and Deutscher Wetterdienst. Prior research also indicates that our ability to deliver better forecast guidance is limited by our inability to afford ensemble forecasts with significantly larger number of ensemble members (>1000) (Miyoshi et al. 2014; Craig et al. 2022).

In recognition of these opportunities, taking into account the state of science, and recognizing the feedback from the workshop participants, it is recommended that NOAA deploys a Global Ensemble Forecast System (GEFS)-AI near-real-time system that will parallel the operational GEFS. This GEFS-AI system consists of a complete end-to-end system that can gradually replace/augment all components of the existing GEFS system with data-driven counterparts. The parallel development of the GEFS-AI system will provide NOAA with the experience required to evaluate the feasibility of transitioning data-driven components into the operational GEFS systems.

The initial implementation of the GEFS-AI system might look like existing data-driven models that were trained on the ERA5 data—like the GraphCast (Lam et al. 2023), the NeuralGCM (Kochkov et al. 2023), or the GenCast (Price et al. 2023)—as an on-demand service deployed on commercial cloud providers. It is recommended that this first system will use NOAA initial conditions for the ensemble forecast and provides a rigorous automated evaluation of the forecasts. It is also recommended that this system uses MLOps and CI/CD frameworks early on allowing for seamless upgrade to newer versions of the data-driven models as they become available from NOAA researchers or from the external community. This early prototype is similar to the Artificial Intelligence/Integrated Forecasting System (AIFS) system already developed at ECMWF. This initial prototype will also allow NOAA to evaluate critical aspects of the data-driven model deployment, including ensuring access to computing resources, data transfer, software stack deployment, automated evaluation pipeline, and cultural conditioning of the NOAA workforce to the new product.

An interim prototype of the GEFS-AI system should mimic the configuration of the future GEFSv13 system that will include a fully coupled UFS model with atmosphere, ocean, ice, land, wave, and aerosol components. This prototype will force NOAA to develop a native data-driven model development capability and the workforce capable of extending existing data-driven models to the new components of the Earth system. This prototype will also ensure that NOAA develops the capacity to produce native training datasets and secure access to compute and software frameworks for training of the data-driven models.

An advanced GEFS-AI system will integrate data-driven models within the GEFS data assimilation framework, resulting in improved quality of ensemble initial conditions. This will require further extending existing data-driven models to make them compatible with the current and future constellation of satellite- and ground-based remote sensing platforms. This will most likely include extending the number of vertical levels and model variables within the data-driven framework.

Another possible extension could be the integration of data-driven models within the data assimilation framework, for example, direct simulation of linear algebra and observational operators used by the JEDI software. This advanced prototype will require development of the internal NOAA capacity and partnerships that can focus on the problems unique to NOAA that might not get sufficient attention from the private industry or academia.

b. Kilometer-scale models. Kilometer-scale models form the backbone of the forecast guidance for high-impact events that NWS provides to end-users. This includes continental-scale forecasts from Rapid Refresh Forecast System (RRFS), High-Resolution Rapid Refresh (HRRR), and High-Resolution Ensemble Forecast (HREF) systems, storm-scale systems like the Warn-on-Forecast (WoFS), and the hurricane prediction systems like the Hurricane Analysis and Forecast System (HAFS). The depth of expertise in these traditional kilometer-scale prediction systems places NOAA in a favorable position to lead the development of data-driven systems for kilometer-scale prediction. At the time of the writing, most of the data-driven models are either developed for global forecasting at the resolution of the ERA5 reanalysis dataset (25 km) or are variants of the nowcast systems like the MetNet3 (Andrychowicz et al. 2023) that excel at propagating radar-like images based on the observations from the dense radar network and model guidance from NOAA systems like HRRR. Preliminary research from NVIDIA suggests that development of the data-driven ensemble forecast system at the kilometer scale is possible (Mardani et al. 2023), however for only a very small model domain trained over Taiwan, for a very limited set of key variables, and at a high computational cost.

Embracing kilometer-scale data-driven modeling will require NOAA to leverage existing technology for generating benchmark ML training datasets at the appropriate resolution. These datasets should include the landmass of the United States as well as the ocean basins where impactful events such as landfalling hurricanes and atmospheric rivers develop. Generation of these datasets will have prodigious computational and data storage requirements that will likely require NOAA to establish partnerships with external organizations that share similar challenges [like the Department of Energy in the United States and the Destination Earth (DestinE) program in Europe].

Development of the kilometer-scale models will also require substantial advancements in the science of data-driven modeling. NVIDIA's research results indicate (Mardani et al. 2023) that substantial research, engineering, and computing resources are needed to develop data-driven models at the kilometer scale to achieve NOAA's mission. Development of these capabilities will require a healthy mix of internal resources and external partnerships with the leading ML research organizations and ML training compute providers.

Acknowledgments. NOAA Global Systems Laboratory, NOAA Physical Science Laboratory, The Cooperative Institute for Earth Systems Research in Environmental Sciences and Data Science, and NOAA Office of Science and Technology all provided financial support for the workshop. The authors also acknowledge feedback from internal and external reviewers. The scientific results and conclusions, as well as any views or opinions expressed herein, are those of the author(s) and do not necessarily reflect the views of NOAA or the Department of Commerce.

References

Andrychowicz, M., L. Espeholt, D. Li, S. Merchant, A. Merose, F. Zyda, S. Agrawal, and N. Kalchbrenner, 2023: Deep learning for day forecasts from sparse observations. arXiv, 2306.06079v3, https://doi.org/10.48550/arXiv.2306.06079.

Bauer, P., P. Dueben, M. Chantry, F. Doblas-Reyes, T. Hoefler, A. McGovern, and B. Stevens, 2023: Deep learning and a changing economy in weather and climate prediction. *Nat. Rev. Earth Environ.*, **4**, 507–509, https://doi.org/10.1038/s43017-023-00468-z.

Bi, K., L. Xie, H. Zhang, X. Chen, X. Gu, and Q. Tian, 2023: Accurate medium-range global weather forecasting with 3D neural networks. *Nature*, **619**, 533–538, https://doi.org/10.1038/s41586-023-06185-3.

Christensen, C. M., 2016: *The Innovator's Dilemma.* Harvard Business Review Press, 288 pp.

Craig, G. C., M. Puh, C. Keil, K. Tempest, T. Necker, J. Ruiz, M. Weissmann, and T. Miyoshi, 2022: Distributions and convergence of forecast variables in a 1,000-member convection-permitting ensemble. *Quart. J. Roy. Meteor. Soc.*, **148**, 2325–2343, https://doi.org/10.1002/qj.4305.

Dueben, P. D., M. G. Schultz, M. Chantry, D. J. Gagne II, D. M. Hall, and A. McGovern, 2022: Challenges and benchmark datasets for machine learning in the atmospheric sciences: Definition, status, and outlook. *Artif. Intell. Earth Syst.*, **1**, e210002, https://doi.org/10.1175/AIES-D-21-0002.1.

Kochkov, D., and Coauthors, 2023: Neural general circulation models for weather and climate. arXiv, 2311.07222v3, https://doi.org/10.48550/arXiv.2311.07222.

Lam, R., and Coauthors, 2023: Learning skillful medium–range global weather forecasting. *Science*, **382**, 1416–1421, https://doi.org/10.1126/science.adi2336.

Mardani, M., and Coauthors, 2023: Residual diffusion modeling for Km-scale atmospheric downscaling. arXiv, 2309.15214v3, https://doi.org/10.48550/arXiv.2309.15214.

Miyoshi, T., K. Kondo, and T. Imamura, 2014: The 10,240-member ensemble Kalman filtering with an intermediate AGCM. *Geophys. Res. Lett.*, **41**, 5264–5271, https://doi.org/10.1002/2014GL060863.

Pathak, J., and Coauthors, 2022: FourCastNet: A global data-driven high-resolution weather model using adaptive Fourier neural operators. arXiv, 2202.11214v1, https://doi.org/10.48550/arXiv.2202.11214.

Price, I., and Coauthors, 2023: GenCast: Diffusion-based ensemble forecasting for medium-range weather. arXiv, 2312.15796v1, https://doi.org/10.48550/arXiv.2312.15796.

Xiao, Y., L. Bai, W. Xie, K. Chen, T. Han, and W. Ouyang, 2023: FengWu-4DVar: Coupling the data-driven weather forecasting model with 4D variational assimilation. arXiv, 2312.12455v1, https://arxiv.org/abs/2312.12455.