NOAA SCIENCE ADVISORY BOARD
RAPID RESPONSE REPORT

Transforming NOAA Data Assimilation with Machine Learning

BY THE ENVIRONMENTAL INFORMATION SERVICES WORKING GROUP (EISWG)
APPROVED BY THE NOAA SCIENCE ADVISORY BOARD

MARCH 10, 2026

Transforming NOAA Data Assimilation with Machine Learning

EXECUTIVE SUMMARY

This report describes options for NOAA to accelerate the integration of Machine Learning (ML) into its data assimilation (DA) systems to improve the nation's weather prediction capabilities.

Key Findings

While efforts have been made and promise has been shown for utilizing ML to accelerate the development of and improve DA, the following issues were found during the study of this report:

(i) Progress in utilizing ML to more effectively ingest observations in the DA system is still limited;

(ii) Key gaps exist in (a) funding to support fundamental research to advance the use of ML to transform DA, (b) the development of a transdisciplinary workforce, and (c) an efficient research-to-operations (R2O) transition pipeline;

(iii) High-quality training datasets, including easily accessible and consistently formatted long-duration observations and high-resolution reanalysis, as well as increased GPU computing power, are needed;

(iv) Challenges also exist that hinder effective collaboration between NOAA, academia, and industry.

Strategic Recommendations

The report contains four potential components of a strategy for the incorporation of ML to improve DA:

(i) Accelerate the use of ML in observation pre-processing and quality control, as well as in the emulation of computationally expensive and approximate forward models;

(ii) Establish significant new support for fundamental research in ML-based DA, for transdisciplinary workforce development, and for an efficient research-to-operation (R2O) partnership;

(iii) Expand online observation repositories and high-resolution reanalyses to allow broad community participation in improving ML tools for DA, and increase availability of dedicated GPU clusters for NOAA and community DA development; and

(iv) Launch a community-endorsed partnership scorecard to align incentives and track progress in incorporating ML in DA to foster partnerships with the private sector, academia, other agencies, and international partners.

We expand on each of these overarching recommendations in the rest of the report.

Recommendation 1:

Accelerate the use of ML in observation pre-processing and quality control, as well as in the emulation of computationally expensive and approximate forward models.

Findings:

(i) ML methods are especially effective at detecting when data is out of sample. As such, they are potentially highly useful in quality control of observations. While current techniques are already designed to look for extreme values, AI is capable of examining the entire spatial field as an image, making it possible to flag entire regions that are suspect, as well as putting extreme values in context. Treatment of the entire spatial field may also make it possible to more effectively represent spatial correlations in observation error, reducing the need for observation thinning or "super-obbing". In addition, current observation pre-processing codes are extensive and complicated. For example, in the Navy DA system, more code is devoted to observation pre-processing than is in the actual DA system itself. There is an opportunity to simultaneously simplify the observation pre-processing activity and to make it more effective. Such activities could be added to existing activities, including: 1) use of anchor observations, including radiosondes and GNSS-RO; 2) use of variational bias control within the DA system; 3) and use of independent data from field campaign data.

(ii) In addition to observation quality control and screening, ML can be used to correct for observation biases. In the same manner in which ML methodologies excel at pattern recognition for outliers, they can also be trained to identify, and subsequently mitigate, biases in observations. Non-ML methods are already in use to detect and correct biases, but ML methods have the potential to make this process faster and more effective, as well as to use sources of information that are related in complex ways.

(iii) ML has been used to develop fast and accurate emulators of entire numerical weather prediction models, and NOAA has already produced AI versions of its global model and two of its regional models. Similarly, it is possible to design and implement fast AI-based emulators of the forward models used to convert model to observations and back. For example, it is necessary to apply complex, and often computationally expensive, radiative transfer models that convert the model temperature, pressure, and water vapor, along with surface properties, into top of the atmosphere upwelling radiances. These so-called forward models must by necessity be simplified so that they run quickly, and this in turn leads to additional sources of error in the DA procedure. Emulation of radiative transfer has the potential to greatly accelerate forward models and to make them more accurate. In addition, many AI methods yield not only a forward model, but also its Jacobian or adjoint; crucial in converting observation information into changes to the model state variables. One complication to the use of AI for forward modeling is the fact that it is not just the computation of simulated observations that is important, but also the sensitivity of the observations to changes in the state. In practice, AI models often produce highly noisy sensitivities, making them challenging to apply in practice.

(iv) If an AI version of the forecast model is available, then it is possible to conduct a much larger number of model integrations within one DA update cycle. This means that, potentially, the ensemble spread may more fully represent the variability in the true system. Making the distribution of simulated states in the ensemble more complete makes it possible to more effectively use observations because 1) the covariance structures become better resolved and less noisy, and 2) the influence of any particular observation is able to be more effectively utilized.

(v) Observing system simulation experiments (OSSEs) represent a powerful tool for evaluating candidate, not yet operational, observations in the context of all current observations. OSSEs have the advantage of a known reference state (referred to as the nature run) and can measure the impact of a new observation dataset in the context of all existing observations. However, OSSEs are often hampered by computational constraints and also by inaccurate or over-simplified treatments of uncertainty. The availability of an AI version of the forecast and forward models (for current future observations) could allow forecast OSSEs to be faster and more representative.

Critical Actions:

(i) A high priority should be placed on using ML techniques in observation pre-processing in DA systems, including in the areas of quality control, bias detection and correction, and observation thinning/super-obbing.

(ii) It would be beneficial for ML experts in other agencies and academia to work directly with NOAA to find optimal ways to emulate computationally expensive radiative transfer models such that they produce non-noisy Jacobians/adjoints. Partnerships could be enabled by leveraging resources within the Earth Prediction and Innovation Center (EPIC).

(iii) NOAA has already made significant progress in the development of AI-based numerical weather prediction. Research into the use of AI-based ensembles of model predictions in the DA process could result in enhanced use of observation information and potentially more effective representation of the potential forecast outcomes.

Recommendation 2:

Establish significant new support for fundamental research in ML-based DA, for transdisciplinary workforce development, and for an efficient research-to-operation (R2O) partnership

Findings:

(i) Although efforts by NOAA and the broad community have shown promise, the field of cross-fertilizing ML and DA, including the use of ML to improve DA and vice versa, is still in its early stages (e.g. Arcucci et al. 2025). A significant amount of fundamental research is required to advance this field. There is a recognized funding gap for this fundamental research.

(ii) ML and DA share a common theoretical foundation and use similar computational tools (Geer 2021; Bonativa et al. 2021). Close collaboration between the ML and DA communities is essential and mutually beneficial. For example, DA can benefit from ML to improve its cost-efficiency and accuracy; ML can benefit from adapting techniques and ideas from DA, e.g., in terms of uncertainty quantification and incorporating proper prior information.

(iii) The rapid pace of ML development makes traditional, slow, and sequential funding and development models for R2O (e.g. traditional NOAA Readiness Level (RL) model) ineffective. In that model, efforts for low RL, high risk, and high return research are often disconnected from those for high RL work that is ready to be transitioned to operations. A new paradigm is needed that emphasizes co-development and co-ownership among research and operational entities. For example, NOAA operational centers should collaborate with academia to identify high-risk but high-return research areas that address NOAA priorities. Academia should get involved with NOAA's operational planning and implementation. Additionally, the R2O2R pipeline should have a greater risk tolerance for early-stage research.

(iv) Advancement in utilizing ML to transform DA requires scientists with highly transdisciplinary expertise. DA itself is already transdisciplinary, requiring an understanding of mathematical algorithms, physical processes, observations, modeling, and high-performance computing (NOAA Science Advisory Board, 2021). Transforming DA with ML requires additional skills in data science, software, and cloud computing engineering, etc (Garett and Frolov 2025). There is a serious lack of expertise and talent in this highly transdisciplinary area in the workforce. Cultivating the new generation of scientists and retooling the skills of mid-career and senior scientists are both needed.

Critical Actions:

(i) Establish a new, sustained, and collaborative NOAA funding program to support the fundamental research on the novel use of ML in DA, and its rapid R2O process. This new program should

- Secure dedicated and long-term funding
- Develop high-level strategy to determine low-hanging fruit research vs. longer term transformative research areas
- Replace the traditional sequential RL approach and allow co-development and co-ownership through aligning goals among NOAA, academia and the private sector
- Increase tolerance of risks from early-stage research to bring in broader swathes of potential investigators
- Leverage NOAA EPIC for building community infrastructure

(ii) Expand the University DA consortium that is currently partnered between universities and NOAA to include more ML experts and industry partners to foster fundamental research, rapid R2O and workforce growth. Goals of the expanded University DA consortium should include

- Collaborating with NOAA to prioritize critical research areas in transforming DA with ML
- Tackling critical issues on novel use of ML in DA through innovative research
- Facilitate efficient R2O through implementing and testing novel approaches in JEDI and UFS models
- Increasing significantly the number of graduate students and postdocs that are trained with this highly transdisciplinary expertise
- Enhancing this transdisciplinary workforce pipeline through public training and outreach (e.g. undergraduates)
- Assisting NOAA in training its current employees to acquire ML-related expertise
- Identifying new ways to bridge the timeline gaps between academia, private sector and NOAA operations
- Enabling a co-development, co-ownership and co-student advisement infrastructure among NOAA, universities and the private sectors

Recommendation 3:

Expand online observation repositories and high-resolution reanalyses to allow broad community participation in improving ML tools for DA, and increase the availability of dedicated GPU clusters for NOAA and community DA development.

Findings:

(i) Machine learning data assimilation techniques depend on the existence of high-quality training data sets. To assure continued rapid progress in ML-based DA tools, comprehensive, public, easily accessible and consistently formatted long-duration observation data sets are needed. While deep-learning forecast models have typically been trained on at least 20 years of reanalysis data, the degree to which the skill of these models depends on the resolution and length of available training data has not been clearly established in the literature. It is clear that some saturation of the skill of trained models can be expected as the number, resolution and volume of training data is increased. For example, Karlbauer et al. (2024) showed that a model trained on only 7 variables at ~1 degree horizontal resolution showed nearly comparable forecast skill to leading ML models trained on far more data at lead times of about a week. It is also worth noting that the training for extreme weather or high-resolution processes may require greater datasets.

(ii) High-resolution reanalysis data sets against which ML-based DA processes can be compared and trained are lacking. The HRRR initial state is a good resource, but it suffers from a lack of consistency as the model evolves over the years and covers only a limited area. A high-resolution reanalysis covering the period from the year 2012 or earlier to the present would be of great use in both ML DA and forecast training.

(iii) NOAA PSL is presently developing extensive archives of observations for Earth system reanalysis (the NOAA NASA Joint Archive, NNJA), accessible to the broader community via the AWS S3 service, and NOAA has a Cooperative Research and Development Agreement with Brightband to develop a repackaged version of this data that is easily accessible using Python code (nnja-ai.readthedocs.io). These efforts are important steps in the direction of allowing NOAA, university, and private sector scientists to develop efficient tools for ML-based DA. Additional steps should include extension of the geostationary satellite data archive back to at least 2012, incorporation of radar reflectivities and derived quantities that are assimilated into the HRRR model, GPS occultation data, as well as privately maintained observations (such as those of tomorrow.io and PlanetIQ.com). Such a comprehensive collection of data would allow a thorough assessment of the additional accuracy and precision of the analysis attainable through these data sources.

(iv) Although NOAA has devoted substantial resources to GPU facilities for AI research (c.f., https://www.noaa.gov/organization/information-technology/hpcc-locations-and-systems/ursa), limited access to sufficient GPU-based computational power is currently limiting progress on ML DA research, especially in University settings. All of our subject matter experts emphasized the need for additional large GPU facilities for use by NOAA, university, and private-sector scientists in their collaborative efforts to maximize the potential of ML in DA. As of January 2025, Frovlov et al. (2025) reported that "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."

Critical Actions:

(i) High-resolution reanalysis (either using the HRRRDAS or a cloud-permitting global system) should be carried forward to create a multi-year (back to 2000) consistent reanalysis dataset against which ML-based DA techniques can be trained and compared.

(ii) The NNJA and NNJA-AI projects should be supported and expanded to include radar and additional geostationary satellite data.

(iii) Federal support and facilitation of the development of AI data centers should include support for competitive grant-funded access to sufficient GPU power to allow training of ML DA.

(iv) NOAA's own GPU facilities should be enhanced to allow multiple candidate ML-based DA models to be trained on high-resolution datasets.

Recommendation 4:

Launch a community-endorsed partnership scorecard to align incentives and track progress in incorporating ML in DA to foster partnerships with the private sector, academia, other agencies, and international partners.

Findings:

NOAA's potential collaborators each bring unique strengths and needs:

(i) Universities provide novel algorithms, state-of-the-art research expertise, and a pipeline of skilled students, and they require timely access to NOAA datasets and sustained funding to perform research and assist NOAA in transitioning research into operational experiments.

(ii) Industry partners bring scalable compute infrastructure, engineering expertise for rapid prototyping, and productization capabilities, and they require clear performance benchmarks, transparent intellectual property terms, and predictable engagement pathways.

(iii) Government laboratories contribute access to operational datasets, long-term verification records, and experience with rigorous evaluation frameworks, and they require sustained funding and programmatic support to integrate ML advances into operational systems.

(iv) Operational forecasters supply domain knowledge, real-time feedback, and practical requirements for usability and interpretability, and they require ML outputs with granular features, predictable latency, and interfaces that support rapid human-in-the-loop decision making.

(v) International partners offer complementary tools, diverse case studies, and opportunities for cross-border validation, and they require streamlined collaboration processes, harmonized metrics and standards, and clear data-sharing agreements to enable joint implementation and evaluation.

Critical Actions:

(i) Common test sets and funding: Define common test sets that reflect operational needs and issue joint funding calls tied to public benchmarks and open-science outputs.

(ii) International exchanges: Expand and enhance exchange programs with leading international forecast organizations such as ECMWF, UKMET, DWD, etc.

(iii) Balanced metrics and low friction: Adopt balanced metrics that cover accuracy, cost, and robustness, and reduce friction with predefined IP arrangements and minimal NDAs while mandating open-source code outputs.

(iv) Annual ML-DA bake-off: Host an annual ML-DA bake-off with rotating tests, public leaderboards, prizes to incentivize participation, blind evaluations, and published transparent results.

(v) Partner requirements and governance: Codify value propositions and requirements for each partner type and outline agreements for transparent governance.

APPENDIX A: Processes for developing recommendations

The following approach and procedures were adopted by this MLDA rapid response team (RRT):

- The RRT defined the planning approach, scope, and timeline, focusing on how to utilize ML to improve NOAA DA and considering the ecosystem from agencies to universities and the private sector.
- RRT interviewed with 8 Subject Matter Experts (SMEs) from US agencies, universities, the private sector, and ECMWF, including Steve Greybush (The Pennsylvania State University, Dan Hodyss (Naval Research Lab), Daryl Kleist (NOAA NWS), Dan Holdaway (NOAA NWS), Peter Jan Van Leeuwen (Colorado State University), Thomas Vandal (Zeus AI), Sergey Frolov (NOAA OAR), Massimo Bonavita (ECMWF)
- RRT studied reports and publications
- RRT summarized findings and created the recommendations and critical actions
- RRT's draft report was reviewed by EISWG members and SAB

References

Arcucci, et al., 2025: The Convergence of Machine Learning and Data Assimilation in Earth System Science. Nature Communications, to be submitted.

Bonavita, et al, 2021: Data assimilation or machine learning? https://www.ecmwf.int/sites/default/files/elibrary/2021/19997-data-assimilation-or-machine-learning.pdf

Frolov, S., et al., 2025: Integration of Emerging Data-Driven Models into the NOAA Research-to-Operations Pipeline for Numerical Weather Prediction, BAMS, published.

Garrett, K. and S. Frolov, 2025: NOAA 10-year strategy for data assimilation. DOI : https://doi.org/10.25923/yz0v-my29

Geer A.J., 2021 Learning earth-system models from observations: machine learning or data assimilation? Phil. Trans. R.Soc. A 379: 20200089.https://doi.org/10.1098/rsta.2020.0089

Karlbauer et al., 2024: Advancing Parsimonious Deep Learning Weather Prediction Using the HEALPix Mesh, JAMES, published

Kleist, et al. 2024: Data Assimilation Strategy and Development Plan for NCEP's Environmental Modeling Center, https://doi.org/10.25923/kqra-6z66

NOAA Science Advisory Board, 2021: A Report on Priorities for Weather Research. 119 pages. https://sab.noaa.gov/wp-content/uploads/2021/12/PWR-Report_Final_12-9-21.pdf