NOAA Technical Memorandum
NWS MDL 86

A REVIEW OF ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING ACTIVITY ACROSS THE UNITED STATES NATIONAL WEATHER SERVICE

Paul J. Roebber

Meteorological Development Laboratory
Silver Spring, MD
October 2022

U.S. DEPARTMENT OF COMMERCE / National Oceanic and Atmospheric Administration / National Weather Service

NOAA Technical Memorandum
NWS MDL 86

A REVIEW OF ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING ACTIVITY ACROSS THE UNITED STATES NATIONAL WEATHER SERVICE

Paul J. Roebber

Meteorological Development Laboratory
Silver Spring, MD
October 2022

UNITED STATES
Department of Commerce
Gina Raimondo
Secretary

UNITED STATES
Department of Commerce
Richard W. Spinrad
Under Secretary for Oceans and Atmospheres and NOAA Administrator

National Weather Service
Ken Graham
Assistant Administrator

Table of Contents

A REVIEW OF ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING ACTIVITY ACROSS THE UNITED STATES NATIONAL WEATHER SERVICE

Paul J. Roebber

ABSTRACT

This report was commissioned to summarize current artificial intelligence and machine learning activity within the U.S. National Weather Service (NWS) with a view towards identifying existing obstacles and recommending future directions. Artificial intelligence and machine learning activity is growing rapidly within the NWS, but is fragmented and lacks the needed infrastructure for improved coordination of effort. Current obstacles to future progress include: lack of workforce training in artificial intelligence and machine learning, lack of curated datasets and software that can be used for development and evaluation of artificial intelligence/machine learning approaches, absence of a centralized clearing house available to weather service personnel for technical expertise/consultation, limited operational compute resources, and lack of a clear end-to-end project pathway that encompasses exploration, development, testbed/proving ground and operational implementation.

Each of these limitations is addressable. Partnering with the NOAA Center for Artificial Intelligence to develop National Weather Service specific training materials, using "learning journey" style materials, is of interest to that group and would help address the current knowledge gap within the weather service. The development of reference software and datasets and the establishment of a consulting team to work on specific projects with operational units will reduce siloed efforts and enhance productivity. By establishing funding vehicles for theme-based projects, and for which there is a sustainable pathway from initial exploration all the way through operational implementation, will help bridge the "valley of death" between research and operations. Agent-based modeling capability with the weather service is currently limited. Given NWS emphasis on Impact-based Decision Support Services (IDSS), agent-based modeling capability should be developed, since this approach can directly link natural and human systems, and can reveal non-intuitive, emergent properties of complex systems like decision support. Collaboration with academic experts in this area, through the above-mentioned sustainable funding pathways, can help to build this expertise.

1. INTRODUCTION

The charge for this report is to review activity in the areas of artificial intelligence (AI), machine learning (ML), and most broadly, post-processing, within the groups reporting to the National Oceanographic and Atmospheric Administration (NOAA) Office of Science and Technology Integration (OSTI). This syncs with one of the recommendations of the Priorities for Weather Research report (NOAA Science Advisory Board 2021, hereafter PWR-2021), which states a need to “target the understanding and prediction of high-impact weather to match the urgent need imposed by climate trends, population and infrastructure increases, and disproportionate impacts on vulnerable communities; including exploring new innovations with AI and machine learning applications.” Likewise, this effort is consistent with the 2020 NOAA Artificial Intelligence Strategy (NAIS), which states a vision that the “expansion of Artificial Intelligence (be) accelerated across the entire agency to make transformative improvements in NOAA mission performance and cost effectiveness.”

In order to perform such a review, and to point towards viable future directions, it is worth taking a step back and considering some recent history in the AI/ML space. A first step in this regard is to define ML and its distinction from AI. A number of definitions and understandings exist, for example, from the NAIS: “Artificial Intelligence refers to computational systems able to perform tasks that normally require human intelligence, but with increased efficiency, precision, and objectivity. A subset of AI called machine learning refers to mathematical models able to perform a specific task without using explicit instructions, instead relying on patterns and inference.” A similar, common definition of these two terms posits that AI is “any technique that enables computers to mimic human intelligence, using logic, if-then rules, decision trees, and ML” while ML is “a subset of AI that includes abstruse statistical techniques that enable machines to improve at tasks with experience.”

To some degree, each of these terms and their accompanying definitions are overstatements of what is actually being accomplished with present technologies. Indeed, in the early days of AI, one precise name that was proposed for the field (and which obviously did not win out) was “complex information processing.” To be sure, evolutionary history has ingrained the importance of pattern recognition as an element of human intelligence, including the tendency towards overforecast bias – it has famously been observed that when the grass is waving at the edge of the savannah, it is evolutionarily more adaptive to assume that pattern is the result of a stalking lion rather than a gust of wind (e.g. Foster and Kokko 2008; Shermer 2009) - but true intelligence cannot be divorced from context. The waving of sea grasses in a shallow lagoon is not likely to spark a similar alarm in our brains, or at least not alarms associated with a large, *land-based* predator. It is the context that allows us to make this distinction. Presently, AI/ML is not capable of this form of generalization. Ecologically speaking, the mind is gauging value, which is context dependent (Barrett 2021). For example, whether or not it is worth expending energy to obtain food depends on the current and, critically, future state (hunger) and how much energy expenditure is needed. Our brains are essentially prediction engines, trying to anticipate what sensory inputs it will receive next.

In contrast, AI/ML approaches today are based upon “learning” patterns in a narrow and rigorously defined framework, such as skill games like chess or Go. Strictly speaking, an ML algorithm does not truly learn these games, but instead performs some kind of optimization based on a defined measure of success while following the specified rules of that game. Change a rule slightly and the algorithm will no longer function or if it does function, likely will not perform as well.

An excellent demonstration of both the strength and weakness of AI/ML is, ironically, provided by the game of chess. For two decades, chess grandmasters have employed ML to gain insight into particular lines of play, finding the optimal moves in the most probable situations. The most powerful such chess engines now use the particular form of ML known as reinforcement learning, which essentially instructs the computer to play hundreds of millions of games against itself, building its expertise through these trials. Reigning five-time World Chess Champion Magnus Carlsen effectively weaponizes this practice against his opponents by finding lines that are useful but have not necessarily been favored in computer evaluations. This is reminiscent of the experienced forecaster who uses computer guidance to inform but not supplant personal judgement. This syncs with the need to understand why an AI/ML tool is suggesting certain outcomes, since trustworthiness will depend on that understanding. Another definition, provided by the NSF AI Institute for Research on Trustworthy AI in Weather, Climate, and Coastal Oceanography (AI2ES), perhaps best captures the current state of ML by defining it as "a field of study involving computer algorithms that can improve and adapt automatically through continual experience with data."

In the atmospheric sciences, data analysis and post-processing have been rooted in traditional statistical methods. The most noteworthy and longstanding example of this is Model Output Statistics (MOS; Glahn and Lowry 1972), which uses the well-established method of multiple linear regression (MLR) to produce forecast variables in an operational context. There are two primary differences between MOS and AI/ML techniques. A key element of the AI2ES definition – automatic adaptation and improvement with continual experience with the data – is not satisfied by the NOAA/NWS implementation of MOS, although experiments with an updateable form have been trialed in Canada (Wilson and Vallée 2002 and 2003). The second, and perhaps most relevant difference, is that AI/ML is intrinsically nonlinear. Indeed, a standard feedforward neural network has MLR as one solution, which can be directly produced by *removing the hidden layers* in the network. To the extent that a forecast problem is inherently linear, this is not a limitation of MOS, and in fact, explains the broad success of this approach. One main interest in AI/ML work, however, is to go beyond the linear restriction. While it is possible to overcome this restriction in MLR, to do so requires the *a priori* transformation of variables to nonlinear forms, while AI/ML automatically learns the nonlinear mapping of inputs to outputs.

More generally, AI/ML differs most importantly from traditional statistical methods in its focus on making predictions by finding generalizable patterns rather than by drawing inferences from a sample. While this focus seems particularly well-suited to a field organized around forecasting, the physics basis of atmospheric science argues for a need to understand as well as predict, and the confidence that ensues from understanding *why* a prediction is being made.

In recent years, most particularly in social science applications, it has become recognized that the so-called objective nature of AI/ML does not imply a lack of bias, since these techniques fundamentally depend on the choices of the developers in terms of what data to use, what metrics define success, and the fundamental limitations of the data itself (quantity, quality, collection patterns, etc.). This issue will become increasingly important as social science applications become more common in the decision support context of weather forecasting. For example, as noted in PWR-202, the risks of extreme weather fall disproportionately on historically underserved and socially vulnerable communities, so an understanding of how best to engage these communities is needed.

Some degree of institutional resistance to AI/ML may be a residue of history. In the 1980s, AI work was largely devoted to so-called expert systems, which began with considerable promise and high expectations, but in the end did not deliver. However, particularly with the advent of the backpropagation method, in which artificial neural network (ANN) errors are used to define algorithm weights, those earlier failures were largely overcome, and today there are many real-world applications of effective pattern recognition (e.g., speech recognition, facial recognition, map routing, etc.).

Further, there are now a variety of machine learning approaches available, with research continuing on new forms. As previously described, these approaches perform a mapping of inputs [for example, numerical weather prediction (NWP) model output] to outputs (some desired forecast which may or may not be explicitly included in the model data). As noted in the introduction. one form of this post-processing approach is quite familiar to meteorologists – MOS, which uses the long-established method of multiple linear regression to perform this mapping. It is extremely risky, and ill-advised, however, to suggest that there is a direct step between performing that mapping and making decisions with little-to-no human intervention. Rather, ML algorithms should be considered another useful tool for informing the decision-making process, but in making those decisions, a deep understanding of both the strengths and weaknesses (such as inherent biases) of these algorithms is necessary.

This speaks to the need for a trained workforce – not necessarily algorithm developers, but coordinated efforts between those developers and domain experts who are tasked with using these tools in the decision support process. This perspective will inform the contents of this report, which was assembled with contributions from a large cohort of National Weather Service (NWS) practitioners and external (academic) collaborators, whose collective work involves research, implementation, and application of these and other forecast tools. In that regard, we will refer to any such mapping with the generic nomenclature of AI/ML and not make any attempt to parse fine distinctions.

This report is organized as follows. Sections 2 and 3 present the findings of this review, organized into a brief overview of current activities (section 2) and existing obstacles to these efforts (section 3). Section 4 provides suggestions for transforming these experiences into a framework that will accelerate future advances. Also included are references, a list of acronyms, and a list of participants who contributed their understanding to this report.

2. CURRENT ACTIVITIES

A demonstration of the growth of AI/ML in the atmospheric sciences is provided by a 23 February 2022 search of all American Meteorological Society (AMS) publications with the words "machine learning" in the title – this search produced 55 abstracts – of which nearly two-thirds were published since 2020 (see also Fig. 1 from Chase et al. 2022 for longer term trends in AI/ML publications). Attendees of recent annual meetings of the AMS have qualitatively experienced a rapid growth in the number of papers presented involving some aspect of data science, and integrated throughout multiple sessions (Fig. 1).

NOAA's Center for Artificial Intelligence (NCAI) provides a count of data science projects across the agency, and this count reveals considerable activity by NWS and across NOAA's Line Offices and mission areas. This count revealed ~188 self-reported projects in 2020, and ~263 in 2022 (Rob Redmon, personal communication, 2022). In response to this growing interest, the AMS recently launched a new journal devoted to AI/ML for earth systems (https://www.ametsoc.org/index.cfm/ams/publications/journals/artificial-intelligence-for-the-earth-systems/). Although the idea of a "landing place" for AI/ML is a positive, one potentially unfortunate consequence of this effort may be a siloing of AI/ML. Previously, AI/ML research has been published in the context of a particular weather analysis or forecasting application, and in that way the scientific context was preserved. Whether such siloing materializes, of course, will be dependent on the kinds of papers that are directed to the new journal, and it may be that the journal will facilitate the ability of researchers to find a wide variety of papers touching upon AI/ML.

In discussing current projects with contributors to this report, a number were identified and are listed in Table 1. Please note that this list is not exhaustive and is provided merely to indicate a measure of the scope and breadth of this activity within the NWS. This mirrors the findings of the NAIS, which listed a number of existing AI/ML efforts within NOAA that pertain to the NWS, including (1) quality control of weather observations; (2) improving physical parameterization for weather, ocean, ice modeling, and improving the computational performance of numerical models; (3) aiding weather warning generation; (4) supporting partners in wildfire detection and movement; and (5) using machine learning for reliable and efficient processing, interpretation, and utilization of earth observations. We note that AI/ML tools, if implemented properly, can assist operations with the well-known problem of data overload, since trained users can deploy them to get the sense of large volumes of data and extract explicit information relatively quickly. Table 2 provides a sampling of future projects suggested by contributors to this report. Despite the extensive work that this list represents, it is only a subset of what is possible as AI/ML efforts continue to grow.

Several obstacles to successful operational implementation were noted by multiple contributors to this report and will be discussed in detail in section 3. At this stage, one can view NWS AI/ML activity as broad-based and growing, but uncoordinated. Again, from the NAIS: "Despite this notable progress, the true potential for AI to advance NOAA's mission has not been realized because all NOAA AI activity heretofore has originated within individual offices with no institutional support. Additionally, some development has been redundant because of a lack of awareness across the agency due to the absence of a coordinating directive or authority."

Our survey suggests the roots of this redundancy are a result of the lack of agency coordination as noted above, but also the need to perform any such coordination with a thoughtful inclusion of the expertise and operational requirements of specific entities [e.g., the needs of the Storm Prediction Center (SPC) are not identical to those of the National Hurricane Center (NHC) nor that of an individual Weather Forecast Office]. Accordingly, any proffered solutions must take this needed domain expertise and site-specific applications into account.

3. LIMITATIONS

Successful ML development depends on three pillars: (1) ample, quality controlled datasets; (2) technical skills for development; and (3) domain expertise – familiarity both with the forecast problems and the operational logistics of the setting where that problem is being considered. As mentioned in section 2, a number of roadblocks in the path from research to operational implementation exist in the AI/ML space, relevant to each of these pillars. These roadblocks include: extensive data requirements and efforts for data management, availability and coordination of fundamental datasets such as reforecasts and observations/analyses, lack of community benchmark datasets and software codes, limited operational computational resources, limited workforce training, the siloing of technical and domain experts, and the lack of reliable and coordinated funding for ML to bring in academic expertise. These are discussed in turn below.

First and foremost, the data requirements of AI/ML techniques are extensive. In order to train such algorithms, the best practice is to split these data (which consists of all the inputs and the desired outputs) into three segments: a training segment, a validation segment, and an independent test segment. The training segment is used for the model development, the validation segment is for hyperparameter tuning (e.g., weights, biases and activation functions in a neural network), and the test segment to evaluate the generalization of the results once training and tuning are completed. Datasets are necessarily large, since this process requires "exploration" of the n-dimensional variable space – if these data do not sufficiently fill this space, then the AI/ML scheme will not be able to produce a good mapping of inputs to outputs in that area, leading to potential performance errors. Secondly, since this mapping is nonlinear, multiple representations within this space will reduce the deleterious impacts of noisy data.

Secondly, it is a truism of this work that considerable time is spent simply managing datasets (often referred to as data wrangling). For example, in a survey conducted by the Earth Science Information Partners (ESIP) Data Readiness cluster, an open cross-sector collaboration that the NCAI contributes to, responses to the question "In your typical AI/ML application development, roughly what percentage of your time do you spend on finding, accessing, and preprocessing data?" showed that only 20% of the survey group spent a quarter or less of their time on that activity, and nearly half indicated that they spent the majority of their time on this task (Rob Redmon, personal communication, 2022). This is the case since the variety of needed inputs come in many different formats, from multiple sources, and may further need to be synced in space and time before being presented to an AI/ML scheme for training. These data must also be quality controlled to limit the amount of noise that is presented. This argues for some kind of data clearinghouse, perhaps including specialized test datasets, as well as modular software for different kinds of AI/ML applications (e.g. Hamill 2015; Fig. 2).

A critical inclusion in such a data library should be reforecast datasets. These datasets, although computationally costly to produce (see section 4), are extraordinarily valuable for validating particular weather events, addressing calibration issues, and general predictability studies. Larger member ensembles are valuable for providing proper baselines for probabilistic forecasting. However, owing to the cost of producing such ensembles, the number of members and the archived output has been restricted, making the most recent datasets less useful. This expense should be supported, and available variables should be increased for the purpose of post-processing in general and AI/ML work in particular. Concomitant with that effort should be the collation of relevant observations/analyses [an example of the latter is the need for long time series of quality high-resolution analyses in Alaska and Hawaii to improve the National Blend of Models (NBM)]. Further, convenient formatting of such datasets drastically improves efficiency of AI/ML/post-processing efforts (for example, chunked netCDF datasets for easy access and reduced data-wrangling time). The production of such benchmark datasets, organized according to agreed-upon standards and frameworks, would be a major step forward in facilitating AI/ML development efforts.

Computational resources are also a limitation. Space and compute needed for development of AI/ML tools can be substantial, owing to the size of the datasets and the data cycling needed for training those algorithms. Further, such training can be more effective when GPUs rather than CPUs are available. Currently, such a development system is lacking. Operational computational resources are also a limitation, since finding compute slots on the NWS operational system is not always possible. Without increased availability of these computational resources, development of AI/ML tools will be constrained and when developed, the transition of those technologies to operations will not occur.

The PWR-2021 report noted an important workforce challenge related to AI/ML, specifically "staying nimble requires a workforce with a broader and evolving range of technical skills and spectrum of talents. Future workforces will include meteorologists working with other experts in Earth sciences, HPC, artificial intelligence (AI) and machine learning (ML), observing, data assimilation, modeling technologies, social sciences, etc. Strategies to increase the workforce capacity will be essential given the increasing demands for these skills."

Currently, NWS employees tend to come from two main areas of academic training: meteorologists (GS-1340) and physical scientists (GS-1301). The former classification is quite strict in terms of the requirements, whereas the latter is more generic, but neither explicitly requires training in AI/ML. The GS-1340 requirements, in particular, leave little room for expansion to include this discipline and are now 30 years old. It is not necessarily the case that every NWS employee be an AI/ML expert with the ability to develop such models themselves, but at a minimum, they should be sufficiently aware of this domain to speak intelligibly with AI/ML developers and in particular to recognize both opportunities and pitfalls associated with application of these technologies to areas within their purview. Accordingly, the NOAA/NWS standards should be revised to promote greater flexibility and expansion of skills that are needed now and those that will be needed in the future. There are some in NOAA working to promote broadening of the future workforce to include more Data Science and Social Science federal service classifications (Rob Redmon, personal communication, 2022), and these efforts should be supported and continue.

This connects to the requirement for domain expertise. Since the technical expertise for AI/ML is substantial, often such experts come from computer science, and such individuals are typically not well-versed in the details of either the meteorology or operational logistics. While it is possible to come from such a background and gain additional knowledge through further academic training (e.g. coursework), deep understanding of operational needs and logistical challenges in addition to meteorological knowledge are required to build AI/ML tools that can be incorporated into routine use. This suggests that a balance between centralization and local expertise needs to be established for successful coordination of AI/ML activity across the NWS.

The solution is to focus on team-based approaches, where the individual members of those teams (or at least one such member) has the training sufficient to bridge the gap between the meteorological/operational domain and the AI/ML domain. Beyond the development stage, there is a need for users to be sufficiently aware of the strengths and weaknesses of AI/ML tools in order to be able to use them judiciously rather than simply as a black box. Equally as important, a lack of sophisticated understanding of these strengths and weaknesses likely reinforces resistance to change rather than an attitude of exploring possibilities. This connects to the active areas of interpretable ML and trustworthiness – in order for such tools to be employed, their credibility will be critical.

Thus, the proposed software and data library should be expanded to include AI/ML "librarians" (i.e., consultants) who can partner with domain experts across the NWS to facilitate projects housed within particular centers or forecast offices, where the domain expertise resides. This collaborative approach will help avoid the stove-piping of such activity that presently occurs, and by providing realistic guidance concerning what is and is not possible, will help to reduce psychological barriers to seeking new solutions to old problems. Another example of how team-based approaches add value is the issue of feature/predictor selection. Time and attention employing meteorological intuition is necessary to determine potentially useful features, and to limit the constraints imposed by the "curse of dimensionality" (the size of needed training data increases exponentially as the dimensions of the AI/ML problem expand, so efforts to select and reduce relevant features are important).

Finally, if resources allow, this partnering should extend across NOAA (for example, excellent training development materials are being developed by the NCAI) and also include collaboration with external, academic experts. In the latter case, the current structure of such collaborations presents several obstacles to effective academic partnering, a situation that was highlighted by numerous contributors to this report.

One such impediment is the need for fundamental exploratory AI/ML work, which contrasts with the readiness level (RL) criteria used in collaborative opportunities such as the Joint Technology Transfer Initiative (JTTI) program ["Readiness Level (RL) 4 or above, which means the concept has been already developed and validated in their own or another laboratory environment and is ready to be tested in a NOAA pseudo-operational environment."] In the AI/ML domain, owing to the rapid development in this field and the need to do considerable exploration of possible benefits of particular new approaches in an operational context, the limited opportunities for funding research at lower RL levels blocks innovation. This approach promotes incremental rather than the high risk – high reward work that is needed.

One academic contributor commented that there are insufficient dedicated resources to increase RLs on projects, for example, the difficulty of showing that the technique works well enough to gain review in an operational testbed. While OAR labs are intended to be the places where mid-RL level research is done within NOAA, from the academic perspective, there is a lack of coordination between efforts across NOAA. Another substantial obstacle is the inconsistent timeline between academic work and operations. Whether or not such tools are developed and have potential to be operationally useful, transitioning them to the operational computing system (such as NOAA's Weather and Climate Operational Supercomputing System [WCOSS] or a cloud system) is a further challenge, owing to availability of that resource (such as limited compute slots or funding for time on a cloud system) and the time of NOAA collaborators to effect implementation and support. At present, there is in this sense no centralized home or dedicated support for AI/ML development within the NWS.

More generally, there are relatively few funding opportunities for academic collaborators, and as such, even those academics inclined to pursue the difficult and time-consuming work of bridging the gap from research to operations (R2O) are often better able to succeed professionally by directing their efforts in more traditional ways, such as fundamental research through the National Science Foundation (NSF). It is important to recognize that this R2O "valley of death" is artificial and imposed entirely by our structures and incentives – this means that the valley can be crossed if efforts are undertaken to do so. In the next section, we suggest possible solutions to these problems.

4. SOLUTIONS AND FUTURE DIRECTIONS

In section 3, a number of obstacles to rapid progress in the use of AI/ML within the NWS were highlighted. Here, we provide some possible solutions, subject to the particulars regarding funding and staffing for which we do not have information at the time of this writing. We summarize these recommendations in Table 3, and discuss each below.

Given the need for workforce training and development in AI/ML, it would seem reasonable to identify the NOAA Center for Artificial Intelligence (NCAI) as an AI/ML training resource. This group has already begun efforts to develop example Jupyter notebooks and R materials organized in a "learning journey" style to ultimately encourage the broad AI community of practice to contribute materials to an NCAI curated library. Additional materials specific to NWS interests could be developed with collaboration from National Centers and other NWS entities [e.g., post-processing with the Meteorological Development Laboratory (MDL)], and NCAI staff have indicated an interest in undertaking that effort. In that regard, NCAI has requested approval for a public repository landing page (NOAA github), where they would stage NCAI created and contributed examples (e.g., rip current detection and others are in development) and expand from there with contributions across NOAA. Additionally, it is likely that a focus on hiring in the

NWS with scientific background in both meteorology and AI/ML will be needed in addition to enhancing the training of existing staff.

As noted in section 3, the current uncoordinated nature of such training should be augmented with a software, data, and consulting clearinghouse. These datasets should include a variety of standardized datasets that could be used to develop different types of AI/ML applications, depending on the need, and most importantly, as a reference against which to compare AI/ML applications. This library should include modular software to facilitate AI/ML application development and should extend, at the minimum, to include the variety of standard techniques currently in wide use, such as Random Forests (RF), multilayer perceptron neural networks (MLP-ANN), and convolutional neural networks (CNN). Since platforms such as Google Tensorflow are already in wide use, it would be sensible to leverage those capabilities in developing this library.

A key element of this clearinghouse is the need for AI/ML consultants who can facilitate the development and use of these tools by domain experts across the NWS. These consultants would be able to partner with NWS experts on specific projects of interest to those organizations without dispersing that expertise into the many existing silos. At the same time, this would allow for developing institutional knowledge concerning ongoing projects and reduce duplication of effort. This partnering will likely lead to the added, crucial development of in-house AI/ML expertise within those specific areas through the project basis of that activity. Where this clearinghouse is located within NWS is immaterial to the overall concept and should be driven by logistical considerations – one likely location for it might be the MDL, given the extensive experience with post-processing within that group. Notably, with the advent of the COVID pandemic and the remarkable success of virtual work across the NWS, it should be possible to establish a kind of hybrid organization for this clearinghouse, which in the competitive environment for AI/ML expertise will allow for less difficulty in staffing.

The above clearinghouse concept should work well for individual Weather Forecast Offices (WFOs) as well as the NWS Centers, provided that sufficient manpower is provided within the clearinghouse to work as collaborative development and implementation teams. This latter is obviously crucial as demand for such partnering is likely to be substantial (see Tables 1 and 2).

This raises the issue of partnering with academia. Currently, academic partnering occurs through a variety of mechanisms including JTTI, CSTAR, and Cooperative Institutes. Adjusting the RL funding opportunities (FO) to sync more realistically with academic research would allow the crucial exploratory AI/ML work that is needed but stunted by current RL FOs to develop. Either some funding vehicle specific to that concept needs to be developed or existing ones should be appropriately adjusted. The exploratory niche is clearly the most obvious place for academic work and fits well with NWS needs in the larger sense. For example, calibration works in opposition to the rare event nature of many of the forecasts of most interest – how to balance these considerations within a specific operational context necessitates coordinated, exploratory efforts along with input from operational experts.

However, there should be a more seamless, end-to-end pathway through which projects could pass from exploration/development to testbed to proving ground to operational implementation, along with the necessary personnel for ongoing support of those efforts. This process needs to be formalized but must entrain NWS personnel from the beginning in order to retain crucial operational domain expertise. Theme-based calls for such efforts (e.g., heavy rainfall, wildfires, etc.), connected to a process as detailed above would likely lead to more rapid progress than is currently possible. However, given the need for code to be open source and the potential intellectual property issues this entails for some academic and corporate partners, such a requirement must be made clear at the onset of any such collaboration. Given the public nature of much academic research, however, there is substantial room for such "open source" activity and intellectual property concerns are unlikely to substantially diminish that activity.

A costly[1], but valuable resource in AI/ML development work, and more broadly in predictability studies, is reforecast datasets. As stated by Hamill (Vannitsem et al. 2018, Chapter 7), "The training data should span a long period of time, thereby providing multiple samples of the range of possible future environmental conditions." The reforecast has the additional benefit of providing a stable (from a computational platform standpoint) dataset which will provide more robust weights in usual AI/ML tools such as ANNs, since the underlying data-generating mechanism (the computer model) is not changing. As noted previously, observational/analysis time series in convenient formats alongside the reforecasts are required. An example of the effective use of such datasets is the use of quantile mapping for precipitation forecasts within the NBM (Hamill et al. 2017; Hamill 2022, pers. comm.).

There are several possible workarounds to this problem. "Smart sub-sampling" of reforecasts is one approach. Kravtsov et al. (2022) built a high-dimensional empirical model of temperature and precipitation that was capable of producing a minimal subset of dates that provide representative sampling of local precipitation distributions across the contiguous US, both in training and independent test data. In order to generate this model, however, a long time series of (reanalysis) data was needed.

Adaptive AI/ML is another possibility. Roebber (2021) devised an adaptive, ensemble-based postprocessor employing neural networks which eliminates the need for retraining as data inputs (including model systems) change. Additionally, the dependence of this technique on training data size was shown to be comparable to multiple linear regression.

Using AI/ML to directly produce reforecasts (Weyn et al. 2021) is another possibility, where here the deep-learning model is again trained using reanalysis data. Weyn et al. (2021) discuss their application of such a trained model for producing highly computationally efficient weather forecasts and large number reforecast ensembles (in their case, 85,800 reforecasts were generated in a few hours on a single GPU). This model currently provides only a few output variables at 1.4° latitude-longitude grid spacing, and is not competitive with state-of-the-art forecast models at short-to-medium range, but is remarkably skillful nonetheless. Further, despite the lack of physics being directly incorporated into the model, it is able to learn physics-based phenomena directly from the data. It is straightforward to add additional variables to this model, as shown by

---

[1] EMC estimates that a 30-year reforecast using 20 fully coupled members at 0.25° grid spacing and 60 atmosphere only members at 0.50° grid spacing would cost approximately $15 million. A 30-year, 80 member at 0.25° grid spacing might cost $56 million (David Novak, 2022, personal communication).

Weyn et al. (2020; 2021), and based on these early results, it is likely that further skill improvements are possible. One limitation, perhaps most especially for long-run climate models where drift becomes important, is the lack of conservation laws. However, this can also be addressed in the learning process, and based on present results appears not to be a substantial problem for forecasts extending through sub-seasonal time scales.

We note that the ability of this approach to readily generate large member, multi-model ensembles that include initial condition uncertainty may provide an additional operational benefit. It is sometimes the case that prior to a major weather event, details concerning controlling factors for the event are poorly known. One such example was the 3 May 1999 tornado outbreak in Oklahoma and Kansas (Roebber et al. 2002). As noted by those authors, prior to this event "… although there was evidence to support severe convection, the prospects for convective initiation were mixed, the information supporting supercellular organization was ambiguous until late, and no observational, conceptual, or NWP model evidence existed to support an outbreak scenario." Their analysis, using "potential vorticity (PV) surgery," indicated that both the likelihood of an outbreak scenario and the location of that convection were highly sensitive to details concerning the arrival of a PV anomaly in the southern airflow. Other uncertainties related to convective initiation in the weakly forced environment, a situation that was also subject to details concerning the southern anomaly.

Many other examples exist in the literature. Ribeiro et al. (2022) used a 40-member ensemble to demonstrate the low short-range predictability of a derecho event, related to convection initiation, the organization of a dominant bow echo mesoscale convective system (MCS) and MCS maintenance. Cases with recurving Western North Pacific typhoons are often associated with large increases in ensemble spread. Aiyyer (2015) used an ensemble prediction system to show spread increases peak at approximately 4-5 days after recurvature, and propagate downstream in a wave packet in the midlatitude storm track.

Implementation of a cluster analysis tool that could quickly identify particular sets of initial conditions, allowing forecasters to select those ensemble members from the large pool, would allow forecasters to improve their situational awareness and improve operational forecast confidence. This connects directly to the ability to generate large ensembles so that a sufficiently dispersive set of scenarios could be examined and understood.

Finally, we note that an, as yet, insufficiently explored area of AI/ML/post-processing is agent-based modeling (e.g. Morss et al. 2017; Roebber and Crockett 2019; Harris et al. 2021). Agent-based models (ABMs) are a form of computer simulation in which a system is governed by the interaction of individual "agents" which follow a set of "local" rules. For example, traffic engineers can model traffic using differential equations, and the traffic flow is thereby considered as continuous. In reality, of course, traffic is composed of individual vehicles and can be modeled this way by considering each vehicle as an agent – the behavior of the system (the traffic flow) emerges from the collective behavior of the individual agents (the vehicles). Such models have a natural connection to human decision-making and the social sciences (e.g. Miller and Page 2007), and in the decision support context of the NWS, there is a need to consider this technique as a means for understanding the context in which users employ forecast information. For example, Harris et al. (2021) have employed coupled ABMs to consider the impact of tropical cyclone forecasts on evacuation effectiveness, where the collective response of individual agents can sometimes produce non-intuitive system responses.

This bottom-up interaction of multiple systems at multiple scales, with concomitant emergent properties, can allow for a deeper understanding of complex systems, such as the weather forecast and warning system employed by the NWS. As noted by Morss et al. (2017), "ABMs are being employed to represent interactions and feedbacks between natural and human systems (Parker et al. 2003; French 2010; Boone et al. 2011; Rounsevell et al. 2012; Farmer et al. 2015; Barton et al. 2016)." Morss et al. (2017) and Harris et al. (2021) extend those interactions to include the forecast as well as the actual weather, since the forecast, to the extent that it is acted upon, will influence the system response. With anticipated future investment by the NWS in social science research, agent-based modeling will need to be another tool to be judiciously employed in future NOAA work, and if done effectively, will assist in the IDSS goal of reducing loss of life and other negative impacts of weather events.

Finally, within the IDSS framework, scenarios present information to end-users in the most concise and actionable form. For example, many end-users are not interested or capable of ingesting the full probability distribution of a given forecast event, but rather prefer scenarios based on the most-likely outcome and the most-likely worst case outcome, specific to their risk factors. This requires that scenarios be core-partner specific. This is largely a generic NWS issue rather than specific to AI/ML, but AI/ML approaches often lend themselves well to these representations.

The many exploratory AI/ML approaches summarized in this section represent opportunities for groundbreaking advances in operations, but also highlight the disconnect that currently exists within NOAA/NWS between research and operations. As highlighted earlier in this section, there is a pressing need for support of such high-risk/high-reward research within all of NOAA.

ACKNOWLEDGMENTS

The author would like to thank the many individuals and their organizations who participated in the discussions leading to this report. These are listed in the Contributors section below, where the identified organization is that of the individual at the time of the discussion. I am also grateful to Dr. Stephan Smith for commissioning this report and to Dr. Judy Ghirardelli for her careful review of this manuscript. This paper is the responsibility of the author and does not necessarily represent a position of the National Weather Service or of any other government agency.

REFERENCES

Aiyyer, A., 2015. Recurving western North Pacific tropical cyclones and midlatitude predictability. *Geophysical Research Letters*, **42**, 7799-7807.

Barrett, L. F., 2021: This is how your brain makes your mind. *MIT Technology Review*, 9-11, August 25.

Barton, C. M., I. I. T. Ullah, S. M. Bergin, J. S. Sarjoughian, G. R. Mayer, J. E. Bernabeu-Auban, A. M. Heimsath, and M. F. Acevedo, 2016: Experimental socioecology: Integrative science for Anthropocene landscape dynamics. *Anthropocene*, **13**, 34–45.

Boone, R. B., K. A. Galvin, S. B. BurnSilver, P. K. Thornton, D. S. Ojima, and J. R. Jawson, 2011: Using coupled simulation models to link pastoral decision making and ecosystem services. *Ecol. Soc.*, **16**, 6.

Chase, R. J., D. R. Harrison, A. Burke, G. M. Lackmann, and A. McGovern, 2022: A machine learning tutorial for operational meteorology, Part I: Traditional machine learning. CoRR abs/2204.07492.

Farmer, J. D., C. Hepburn, P. Mealy, and A. Teytelboym, 2015: A third wave in the economics of climate change. *Environ. Resour. Econ.*, **62**, 329–357.

Foster, K. R., and H. Kokko, 2008: The evolution of superstitious and superstition-like behaviour. *Proc. of the Royal Soc. B*, doi.org/10.1098/rspb.2008.0981.

French, C., 2010: People, societies, and landscapes. *Science*, **328**, 443–444.

Glahn, H. R., and D. A. Lowry, 1972: The use of model output statistics (MOS) in objective weather forecasting. *J. Appl. Meteor*., **11**, 1203–1211.

Hamill, T. M., 2015: New Directions in Statistical Post-Processing. *Harry R. Glahn Symposium, 95th American Meteorological Society Annual Meeting*, 6 January 2015.

——————, E. Engle, D. Myrick, M. Peroutka, C. Finan, M. Scheuerer, 2017. The U.S. national blend of models for statistical postprocessing of probability of precipitation and deterministic precipitation amount. *Mon. Weather Rev*., **145**, 3441–3463.

Harris, A., P. J. Roebber, and R. Morss, 2021: An agent-based modeling framework for examining the dynamics of the hurricane-forecast-evacuation system. *International Journal of Disaster Risk Reduction*, **7**:102669, doi:10.1016/j.ijdrr.2021.102669.

Kravtsov, S., P. Roebber,T. M. Hamill, and J. Brown, 2022: Objective methods for thinning the frequency of reforecasts while meeting post-processing and model validation needs. *Wea. Forecasting.,* doi.org/10.1175/WAF-D-21-0162.1.

NOAA Science Advisory Board, 2021: A report on priorities for weather research. NOAA Science Advisory Board Report, 119 pp.

Ribeiro, B., Weiss, S. and L. Bosart, 2022: An Analysis of the 3 May 2020 low-predictability Derecho using a convection-allowing MPAS ensemble. *Wea. Forecasting*, **37**, 219-239.

Roebber, P.J., D.M. Schultz and R. Romero, 2002: Synoptic regulation of the 3 May 1999 tornado outbreak. *Wea. Forecasting*, **17**, 399-429.

Shermer, M., 2009: Patternicity. *Scientific American*. 299. 48. 10.1038/scientificamerican1208-48.

Weyn, J. A., D. R. Durran, and R. Caruana, 2020. Improving data-driven global weather prediction using deep convolutional neural networks on a cubed sphere. *Journal of Advances in Modeling Earth Systems*, ***12***, e2020MS002109. doi.org/10.1029/2020MS002109

——————, D. R. Durran, R. Caruana, and N. Cresswell-Clay, 2021: Sub-seasonal forecasting with a large ensemble of deep-learning weather prediction models. *Journal of Advances in Modeling Earth Systems*, **13**, e2021MS002502. doi.org/10.1029/2021MS002502.

Wilson, L. J., and M. Vallée, 2002: The Canadian Updateable Model Output Statistics (UMOS) system: Design and development tests. *Wea. Forecasting,* **17,** 206–222.

——————, and M. Vallée, 2003: The Canadian Updateable Model Output Statistics (UMOS) system: Validation against Perfect Prog. *Wea. Forecasting*, **18**, 288–302.

LIST OF ACRONYMS

ABM	Agent-based Model
AI	Artificial Intelligence
AMS	American Meteorological Society
ANN	Artificial Neural Network
AWC	Aviation Weather Center
CNN	Convolutional Neural Network
COVID	Coronavirus Disease
CPC	Climate Prediction Center
CSTAR	Collaborative Science, Technology, and Applied Research
CWA	County Warning Area
EMC	Environmental Modeling Center
HPC	High Performance Computing
IDSS	Impact-based Decision Support Services
JTTI	Joint Technology Transfer Initiative
MDL	Meteorological Development Laboratory
ML	Machine Learning
MLP	Multilayer Perceptron
MLR	Multiple Linear Regression
MOS	Model Output Statistics
NAIS	NOAA Artificial Intelligence Strategy
NBM	National Blend of Models
NCAI	NOAA Center for Artificial Intelligence

NESDIS	National Environmental Satellite, Data, and Information Service
NHC	National Hurricane Center
NOAA	National Oceanographic and Atmospheric Administration
NSF	National Science Foundation
NWC	National Water Center
NWP	Numerical Weather Prediction
NWS	National Weather Service
OPC	Ocean Prediction Center
OPG	Operations Proving Ground
OSTI	Office of Science and Technology Integration
PSL	Physical Sciences Laboratory
RF	Random Forest
RL	Readiness Level
R2O	Research-to-operations
SPC	Storm Prediction Center
SWPC	Space Weather Prediction Center
WCOSS	Weather and Climate Operational Supercomputing System
WFO	Weather Forecast Office
WPC	Weather Prediction Center

REPORT CONTRIBUTORS

Aviation Weather Center (AWC)	Joshua Scheck and team
Climate Prediction Center (CPC)	Dave Dewitt
Colorado State University	Russ Schumacher
Environmental Modeling Center (EMC)	Brian Gross Jason Levit Vladimir Krasnopolsky
Meteorological Development Laboratory (MDL)	Judy Ghirardelli Jeffrey Craven
National Hurricane Center (NHC)	Brian Zachry
National Water Center (NWC)	Edward Clark Fred Ogden
NESDIS	Mike Pavolonis Sid Boukabara
NOAA Center for Artificial Intelligence (NCAI)	Rob Redmon Jordan Watson

	Heather McCullough Douglas Rao
NWS DTW Weather Forecast Office	Greg Mann
NWS Western Region	Matt Jeglum
Ocean Prediction Center (OPC)	Joseph Sienkiewicz
Operational Proving Ground (OPG)	John (JJ) Brost
Physical Sciences Laboratory (PSL)	Tom Hamill
Science and Technology Integration (STI)	Hendrik Tolman
STI Modeling Program Team	Jamese Sims
Space Weather Prediction Center (SWPC)	Enrico Camporeale
Stony Brook University	Brian Colle
Storm Prediction Center (SPC)	Russell Schneider Israel Jirak
The Climate Corporation	Brad Colman
University of Oklahoma	Amy McGovern
University of Washington	Dale Durran
Weather Prediction Center (WPC)	David Novak Jim Nelson Kathryn Gilbert

Table 1. A non-exhaustive list of AI/ML projects currently in development or operations in the U.S. National Weather Service. Where known, acronyms in parentheses denote the type of mapping that is being used (RF=random forest; ANN= artificial neural network; CNN=convolutional neural network).

Group	Project
Aviation Weather Center (AWC)	Probability of turbulence
Climate Prediction Center (CPC)	Cool-season precipitation forecasts based on relationships with recent and past SSTs (Matt Switanek and Tom Hamill); Weeks 3-4 precipitation (ML post-processing of GEFSv12 reforecasts; Rochelle Worsnop)
Environmental Modeling Center (EMC)	QC, Data thinning, AQ transport, bias correction, model physics, ocean/lake waves, grid diagnostics, SSMI data retrieval, optimizing model parameters, radiation
National Hurricane Center (NHC)	Rapid intensification (Mark DeMaria; ANN)
NWS Western Region	Diagnosing fog from webcams; use METAR, GFS, reforecast ensemble to produce probabilistic weather element forecasts; flash flood forecasting in nowcast period (MRMS and RF)
National Water Center (NWC)	Colorado River basin reservoir predictions
Storm Prediction Center (SPC)	Ensemble post-processing (Amy McGovern; RF); Wind reports and damage (Bill Gallus); Convective outlooks (Russ Schumacher; RF)
Weather Prediction Center (WPC)	Excessive rainfall outlook (Russ Schumacher; RF); Frontal analysis (Amy McGovern); Precipitation type (Heather Reeves); Road temperature (Heather Reeves); Heavy rainfall nowcast (Paul Roebber; CNN).

Table 2. A non-exhaustive "wish list" of AI/ML applications that could be developed in support of the U.S. National Weather Service.

Group	Project
Climate Prediction Center (CPC)	Weeks 3-4 precipitation forecasts (ANN, CNN); Reservoir management for flood control and water supply; Week 2 extreme events; precipitation regime transitions (wet-to-dry and reverse)
DTW Weather Forecast Office	Scenario-based forecasts for IDSS
Environmental Modeling Center (EMC)	Genetic optimization of model parameters; Process-oriented diagnostics (e.g. dryline forecasts)
Meteorological Development Laboratory (MDL)	Grid and scenario-based post-processing; Probabilistic precipitation forecasts (quantile mapping and rank-weighted best-member dressing, Tom Hamill and Michael Scheuerer)
NWS Western Region	NBM improvements – handling extreme events (not black swans), reduce forecaster editing
National Water Center (NWC)	Data-driven ML reservoir predictions
Ocean Prediction Center (OPC)	Offshore thunderstorms; Waves in swell-dominated regimes
Operations Proving Ground (OPG)	IDSS engine (impact-based decision services)
Space Weather Prediction Center (SWPC)	Combined physics-based/ML space weather forecasts (boosted RF).
Storm Prediction Center (SPC)	Extracting information from NWP in computationally efficient way

Table 3. Recommendations for future directions in NWS AI/ML.

Recommendation	Description
Establish a NWS AI/ML clearinghouse.	Clearinghouse to be staffed in a hybrid format, and contain supporting data sets, baseline methods/software, verification statistics libraries, and consultants to partner on NWS AI/ML projects.
Partner with NCAI for NWS staff training.	NCAI is building training capacity; can partner with NWS to develop "learning journey" materials specific to NWS needs.
Focus on hiring NWS staff with scientific background in both meteorology and AI/ML	NWS is not currently staffed adequately to implement AI concepts, and as such, new hires in this area are needed in addition to staff training.
Adjust RL funding vehicles to account for needed exploratory AI/ML work, both within NWS and with academic partners.	Current RL funding opportunities do not allow for exploratory AI/ML work, which is essential at this stage of development.
Construct a theme-based, end-to-end project pathway from exploration to testbed to proving ground to operational implementation.	Current construct has many gaps and results in frequent one-off projects that do not lead to operationally useful products.
Develop agent-based modeling capability within the NWS.	IDSS connects naturally to agent-based modeling since such models can directly link natural and human systems, and can reveal non-intuitive, emergent properties of complex systems.
Sponsor continued production of reforecast datatsets, including comprehensive and accessible archives.	Reforecast datasets are computationally expensive, but are extremely valuable to AI/ML research; some workarounds are developing.
Format AI/ML forecast tools in a scenario-based framework, such as most-likely and most-likely worst-case scenarios, where these scenarios are core-partner specific.	Many forecast end-users find scenario-based forecasts linked to their risk factors to be the most actionable information. AI/ML forecast tools can easily be constructed in this way.

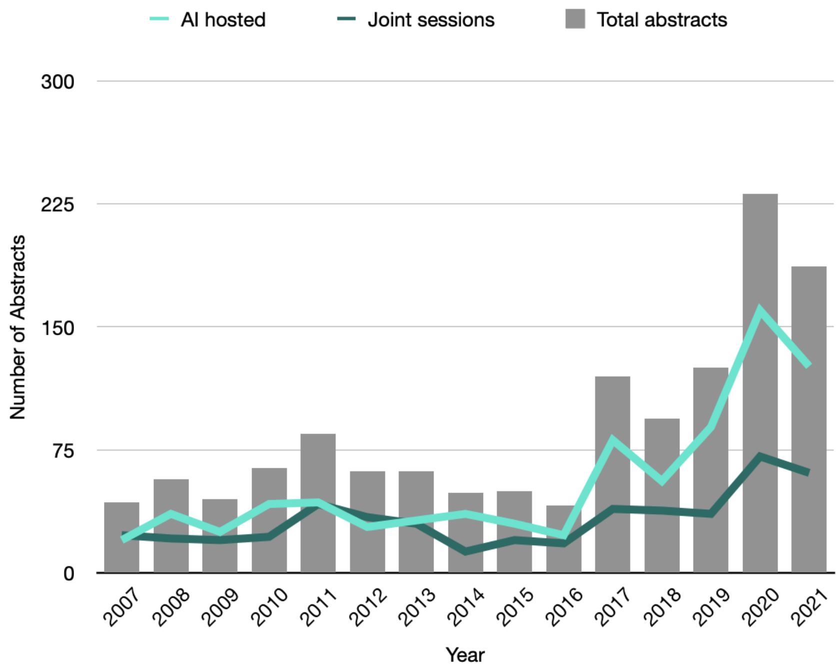

Figure 1. AI/ML activity at the AMS Annual Meeting (Amy McGovern, personal communication, 2022).

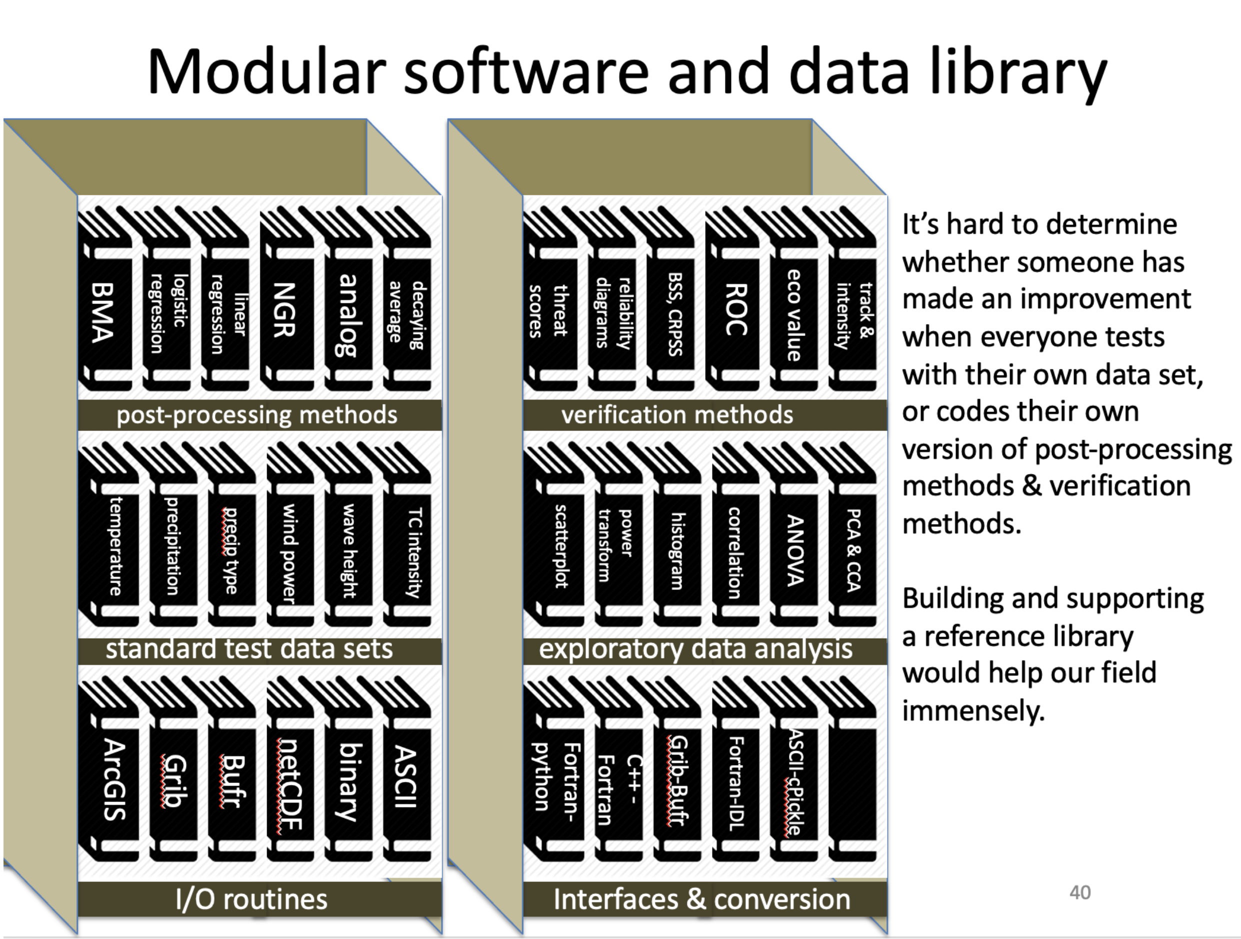

Figure 2. A proposed software and data library (from Hamill 2015).