HYPERION RESEARCH

HPC User Forum Update

Successes and Challenges of AI Integrated into Weather and Climate Forecasting

Thomas Sorensen and Earl Joseph
December 2024

IN THIS UPDATE

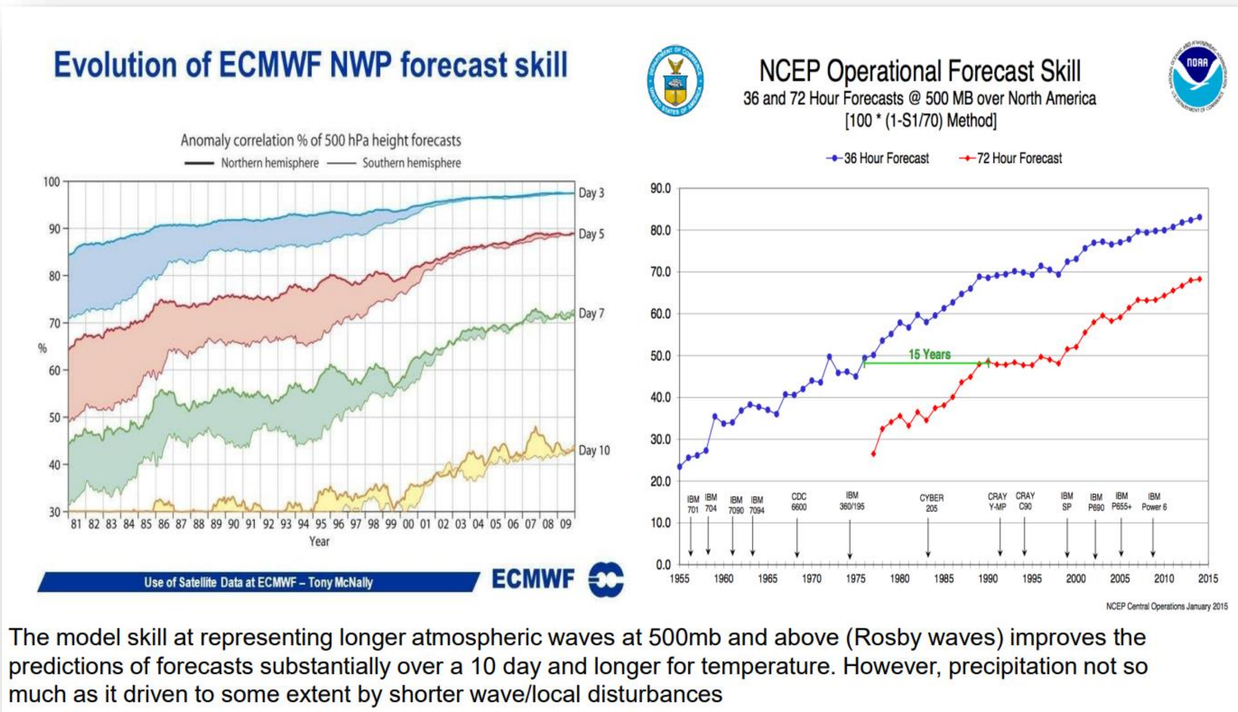
The HPC User Forum was established in 1999 to promote the health of the global HPC industry and address issues of common concern to users. In September 2024, the 85th HPC User Forum took place at Argonne National Laboratory. This update summarizes a presentation from that conference given by Rao Kotamarthi, Senior Scientist in the Environmental Science division at Argonne National Laboratory, where he also serves as a Chief Scientist. Additionally, he is the Science Director at the Center for Climate Resilience and Decision Science (CCRDS). Kotamarthi discussed ongoing activities in climate research leveraging AI/ML techniques, as well as the plans and challenges relating to the research.

Source: Rao Kotamarthi, 2024

PRESENTATION: ACCELERATING CLIMATE DISCOVERY SCIENCE USING AI/ML BY RAO KOTAMARTHI, SENIOR SCIENTIST, ENVIRONMENTAL SCIENCE DIVISION AT ARGONNE NATIONAL LABORATORY

For Kotamarthi, the pivotal question surrounding weather and climate science is regarding predictability. Climate, as a larger-scale system that is more challenging to observe than weather, demands a more complex set of models and methodologies to accurately predict. In the past two decades short-term weather forecasting has undergone an incredible positive shift in statistical accuracy with local 3-day forecasts near 100% accuracy. These advancements were made possible in no small part through the use of powerful compute resources unlocking higher resolution capability at faster speeds. However, with more fine resolution comes higher rates of predictability, scaling this method brings significant power demands.

FIGURE 1

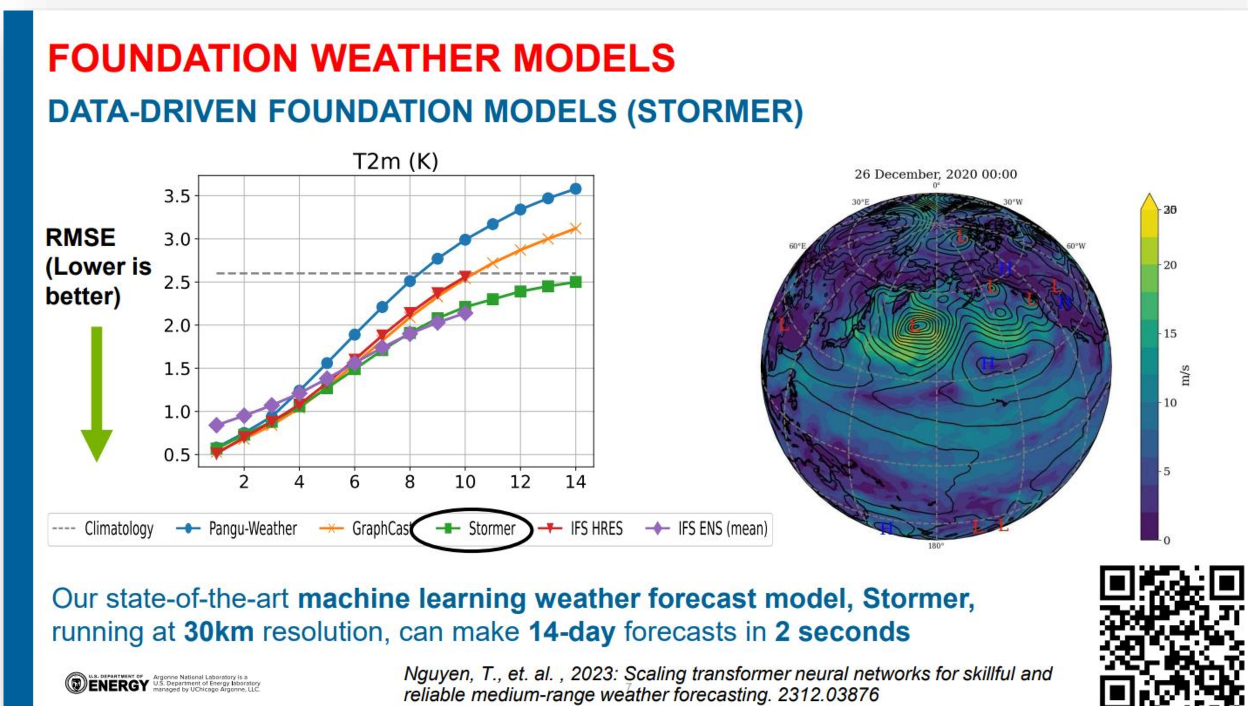


Source: Rao Kotamarthi, Argonne National Laboratory, 2024

Weather prediction computing is not solely based on observed environmental data. Environmental interactions that happen on a scale smaller than the resolution of the program are calculated via process models that are designed to account for processes such as cloud microphysics, aerosol dynamics, and radiative transfer. These physics-informed process emulators account for a considerable amount of their computing resources.

The first major innovation, according to Kotamarthi, was replacing these emulators, called “process models,” with machine learning surrogates. Replacing these “expensive” models in the larger weather model system would greatly increase efficiency and decrease time to solution by “throwing lots of the workload onto GPUs.” They started with radiative transfer and boundary layer physics, some of the most demanding elements, with plans to onboard more in a metered roll out into regional and global scale models. One major challenge faced during this process is maintaining and monitoring model stability over long integration times, where problems are expected to occur. Details of load balancing with new and increased hardware hybridity have not been fully explored and must be measured carefully.

FIGURE 2

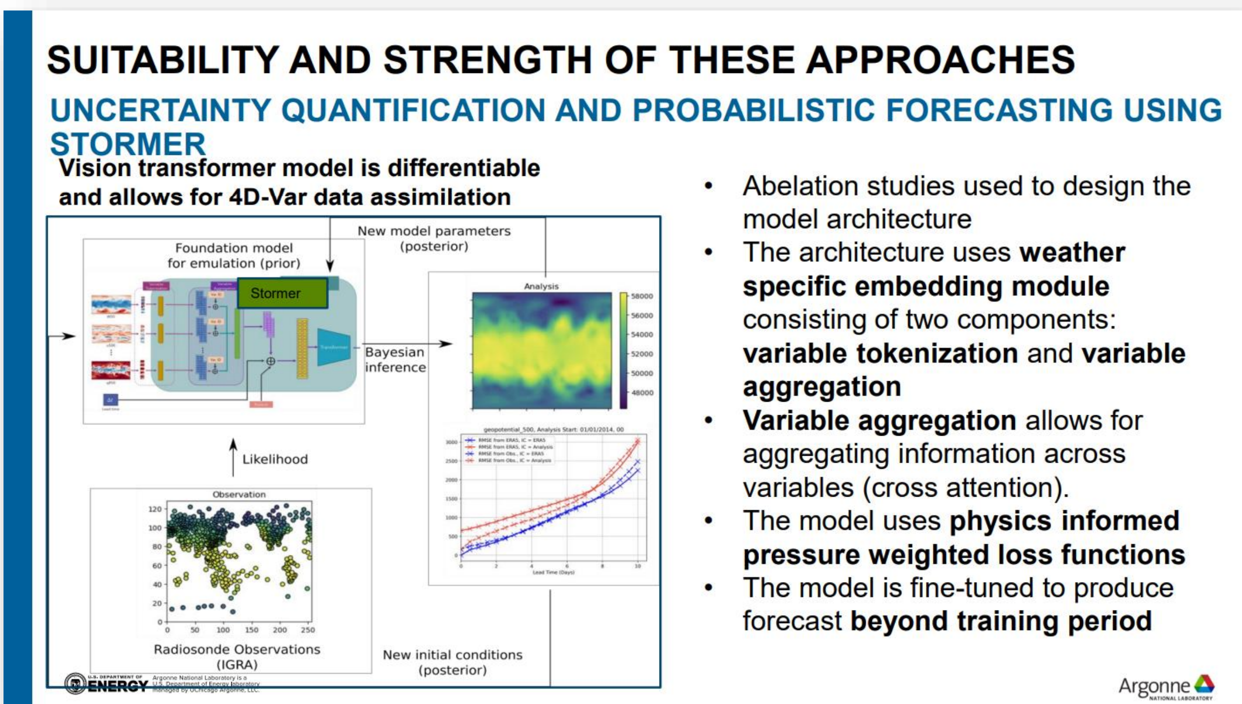


Source: Rao Kotamarthi, Argonne National Laboratory, 2024

Kotamarthi went into detail regarding physics-informed domain-aware RRTM emulators, the radiative transfer model which is one of the most complex interaction simulations involved in a weather model of this kind. This emulator seeks to predict the interactivity of the rising and setting sun's warmth cycles on atmospheric particles and moisture and its effect on weather phenomena. The results were promising upon replacement of this emulator with a deep learning algorithm, even “as good” as a typical simulation in an average climate model.

Another methodology that has been gaining efficacy rapidly over the past few years is foundation weather models. The idea is that a model is developed that is entirely data driven with observational data sets. Some of these AI models have achieved great success with the wholesale exclusion of calculations and simulation only leveraging historical data. In fact, the STORMER model has proven to be as effective as the traditional forecasting IFS Ensemble model with extended 14-day forecasts as opposed to the previous 10-day. Furthermore, the current inexpensive nature of inferences of this kind allows for significantly faster results to be provided for the same forecast window. However, models of this kind rely heavily on accurate and location-specific data, with forecast accuracy increasing in areas with a larger amount of quality data.

FIGURE 3

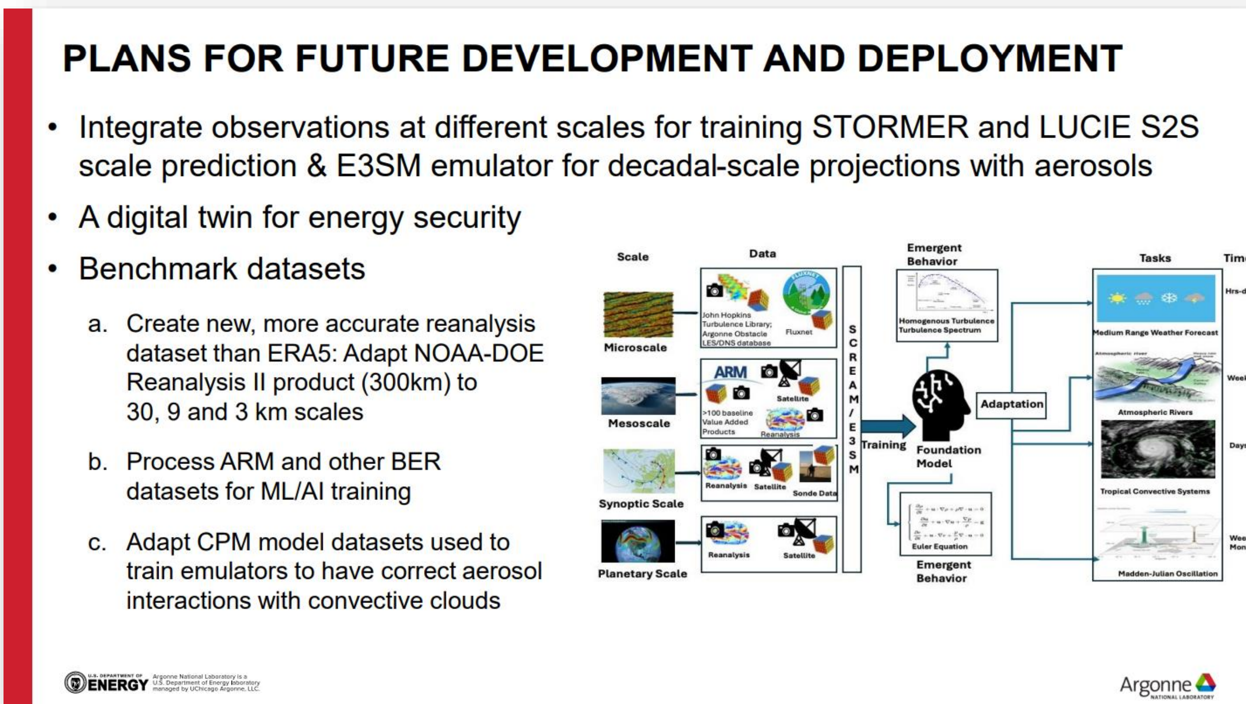


Source: Rao Kotamarthi, Argonne National Laboratory, 2024

These models, however accurate, are subject to comprehensive tests to assess their suitability and strength. The models are ameliorated through ablation studies used to design the architecture, including variable aggregation which allows for aggregating information across variables, and fine-tuning with data like atmospheric physics to produce beyond the training period.

Kotamarthi explained that for a stable weather model to inform or become an effective climate model, the scale of time demands a different approach. Climate models seek to create informed outputs not in the measure of days but years, possibly thousands of years. In an attempt to create a model in similar ways using climate data from the previous decade, the results were promising in terms of the measured drift and bias. One important element of the development of an effective climate model is informing the model that the climate does change and is changing. A climate model simply trained on historical data in this way will need fine tuning and other interventions to allow larger shifts to be considered and predicted.

FIGURE 4



Source: Rao Kotamarthi, 2024

As for plans for future deployments and developments, one major task is the continued integration of observations at different scales for training models like STORMER, specifically for decadal-scale projections with aerosols. Furthermore, a digital twin project is in the works for energy security, and an effort to benchmark datasets, including the creation of new, more accurate datasets.

Climate mapping and weather prediction are critical missions in modern science and technology as well as highly suitable candidates for AI/DL/ML integration. Weather forecasting has already seen a tremendous boon in deep learning models with significant developments being made in the climate counterpart. However, weather forecasting presents a much more stable problem than long-term climate prediction, meaning much more careful and innovative efforts must be made for increased success in the area. As is common with nearly all AI applications in scientific fields, larger amounts of new, high quality data are a must as well as creative design solutions to manage the technology.

For more information or to view this and other presentations given at HPC User Forums dating back to 2008, visit www.hpcuserforum.com.

About Hyperion Research, LLC

Hyperion Research provides data driven research, analysis and recommendations for technologies, applications, and markets in high performance computing and emerging technology areas to help organizations worldwide make effective decisions and seize growth opportunities. Research includes market sizing and forecasting, share tracking, segmentation, technology and related trend analysis, and both user & vendor analysis for multi-user technical server technology used for HPC and HPDA (high performance data analysis). We provide thought leadership and practical guidance for users, vendors and other members of the HPC community by focusing on key market and technology trends across government, industry, commerce, and academia.

Headquarters

365 Summit Avenue
St. Paul, MN 55102
USA
612.812.5798
www.HyperionResearch.com and www.hpcuserforum.com
