

High Performance Computing Strategic Plan 2015-2020 *(Final Draft)*



Office of the Chief Information Officer
National Oceanic and Atmospheric Administration
United States Department of Commerce



Introduction

Current State

NOAA manages and operates major investments for operational and research and development (R&D) high performance computing. These investments are managed as an integrated enterprise by the NOAA Office of the Chief Information Officer's High Performance Computing Program. NOAA's High Performance Computing Board provides strategic guidance and oversight for the execution of this enterprise and the allocation of HPC resources.

The NOAA Weather and Climate Operational Supercomputing System investment provides reliable HPC capabilities essential to run real-time numerical models generating millions of weather guidance products daily. This guidance is incorporated into public and private sector forecast processes to protect our nation's lives and livelihood. Since this investment is critical to the nation, NOAA operates geographically dispersed identical primary and backup systems in Reston, VA and Orlando, FL.

The NOAA R&D High Performance Computing System investment enables significant improvements in weather and climate research and modeling. NOAA operates development computing located in Fairmont, WV and Boulder, CO primarily supporting the development of weather and seasonal to interannual climate model predictions bound for implementation on the operational systems. NOAA's research computing operated by Oak Ridge National Laboratory primarily supports improvements in the skill, resolution, and complexity of models used for Earth System research, understanding its variations and changes, and for predictions, and projections.

NOAA leverages other U.S. government agency shared services to provision its enterprise high performance computing services. Model developers utilize dedicated computing allocations on multiple leadership class systems at Department of Energy's (DOE) Oak Ridge National Laboratory. NOAA competes for access to national high-performance computing facilities housing some of the world's most advanced supercomputers through agency programs such as the DOE's Innovative and Novel Computational Impact on Theory and Experiment (INCITE) and the National Science Foundation's (NSF) Extreme Science and Engineering Discovery Environment (XSEDE).

Vision for future state

NOAA depends on High Performance Computing (HPC) in one form or another to meet many of its missions. Comprehensive numerical modeling of the global system, which requires HPC, is the cornerstone of weather forecasting, weather and climate research, and understanding ecosystems and coastal issues. In FY 2013, for the first time, every NOAA Line Office received an allocation on NOAA's HPC systems, demonstrating that the demand for HPC is growing beyond NOAA's traditional climate and weather users. NOAA must move to an end-to-end integrated HPC for mission critical modeling and remove the distinction between operational and research HPC, creating a single delivery system for reliable, dependable HPC to meet NOAA's modeling requirements. Capacity and architecture requirements would be met through a mix of NOAA core HPC and larger-scale HPC shared with other agencies.



Science and Applications Drive NOAA’s Computing Requirements

New applications are continually under development to address significant scientific questions that can accelerate NOAA towards meeting its mission goals. Weather prediction based on numerical models supports an increasingly weather-vulnerable economy and society. Climate simulations, predictions, and projections inform science, service, and stewardship decisions and form the basis for assessments of current and future states of the climate system that identify causal factors, processes, potential impacts, and the attendant uncertainties. Operational prediction applications generate information to meet requirements for execution time, reliability and accuracy. The suite of NOAA applications incorporates scientific understanding and numerical prediction system development into improved beneficial products that society uses. Prediction and projection systems for use across NOAA mission goals are computationally expensive because they include the complex interactions of the atmosphere, ocean, land surface, cryosphere, chemically active atmospheric constituents, biogeochemical cycles on land and in the ocean, and terrestrial and oceanic ecosystems. The applications must address local to global spatial scales and timescales from seconds to centuries.

Prediction accuracy depends critically on the ability to add observed information on the initial state of the atmosphere and other domains (such as the ocean, land surface and ice regions) to the forecast system. In addition to augmenting observations, advanced data assimilation techniques increase computing requirements. Over the next decade, global and regional data assimilation and model capabilities and techniques will become more integrated into a single system capable of providing forecast data from less than one hour to at least one month with seasonal climate extensions to one year. NOAA’s research programs seek to extend this to interannual to decadal time scales.

A “Warn-on-Forecast” (WoF) system, composed of large-domain nests over CONUS, Alaska, Hawaii and Puerto-Rico together with high resolution nests over regions of anticipated severe weather and major airports, will provide required information on the location and intensity of severe weather and broad support for aviation weather operations; a similar system will improve hurricane intensity prediction. A WoF system is expected to push exascale computing capabilities by 2023.

With broader geographic coverage, global models are the key to accurate forecasting of major storms with oceanic origins, such as Hurricane Sandy, the nor’easters that hit New England in February 2013, the major snowstorm that hit Washington DC in 2010, and the devastating floods caused by “Atmospheric Rivers” hitting the US West Coast. Global models are also critical to NOAA’s success in preparing the public 3-8 days in advance for major tornado outbreaks. These advanced atmospheric models coupled to state-of-the-art ocean, biosphere, cryospheric, and chemistry including biogeochemistry and ecosystem models are needed to predict the likelihood of such significant weather and climate extremes and events at longer lead times.

Over the past decade, NOAA’s coastal ocean predictions have increasingly relied on modeling with HPC to provide information for coastal community challenges such as storm and climate scale inundation planning, resilience and response, safe and profitable marine transportation, and to predict and mitigate the impacts of hazards and toxins ranging from oil spills to harmful algal blooms. Sophisticated 4-D ocean information, produced quickly and reliably, is critical to ensuring the provision of forecasts on a



regionally and locally relevant scale. Complicated biological, chemical, physical and ecological interconnections and coupled models connecting riverine, estuary, and ocean systems require HPC, and are critical to predicting the full complement of impacts on the coast. Improved extratropical and tropical storm surge and sea level change information at the parcel scale is critical, as witnessed in a series of coastal storms including Sandy. The use of unstructured grids and ensemble approaches to storm surge to produce higher localized resolution, and realistically estimating the combined effects of surge, tides, waves and rivers will only increase the future HPC demand.

NOAA's ecosystem modeling community is a growing user of HPC. Scientific advice provided to living marine resource managers by NOAA includes results from simulations of various combinations of environmental conditions, ecological responses, and management decisions. Models examine the combined effects of physical, chemical, biological, and anthropogenic forcing on marine ecosystems through spatially-explicit simulations. Challenges include running multiple, multi-decadal and vertically-integrated models that include many affected species groups. To explore a broader range of scenarios (e.g. more climate scenarios, more configurations of assessment methods and management strategies) and conduct multiple runs of the simulation for statistical power, high performance computing resources are critical to timely and successful completion of the mission. As computing resources and approaches are developed the scientific information provided by NOAA becomes more robust for understanding and managing ecosystems in a changing environment.

The evolution of sampling for marine fauna includes optical technologies that require commensurate HPC resources. Recent advances in field and bench top instruments are helping move the enumeration and measurement of biological samples into the digital realm. Selection of in-focus video objects and machine identification of species require HPC platforms that can handle large video and still image data sets and aggressive algorithms to process them. Advances here would accelerate what is presently a relatively slow data stream into the analysis of ecosystem and fishery status.

NOAA's mission requirements will likely continue to expand as they have over the past decade, requiring additional HPC capacity. Increased numerical resolution, increasingly complex models that capture the realism of the Earth System processes and interactions, and the use of ensembles to better quantify uncertainty are needed for these requirements, and all require significantly enhanced HPC capabilities. These will also require new approaches in data management, transmission, and storage.

Increased Resolution

Greater accuracy and detail are achieved by increasing the horizontal and vertical resolution of the models used in NOAA's prediction systems. High-resolution (3-10 km) ocean models are critical for support of coastal inundation and marine safety, the assimilation of ocean observations for more accurate predictions with quantified uncertainty, the delivery of seasonal-to-decadal predictions, and for coastal climate applications such as vulnerabilities to storm surge in a changing climate. A WoF system, consisting of a 3 km large-domain nest and 1 km local nests increases computation requirements up to 200 times. High resolution nested coastal ocean models are needed to predict local oceanographic conditions to support marine transportation, ecological forecasting, and coastal management issues.



The next generation global weather prediction system is estimated to run at horizontal resolutions of 3-10 km, with more accurate representation of physical processes, non-hydrostatic dynamics, high resolution nests for local prediction and advanced data assimilation techniques. Such weather-resolving regional models embedded in coarse resolution (15km) global models will be used to examine the impact of climate variability and change on the distribution and frequency of severe weather and other extremes.

Extension of the atmospheric model top along with increased vertical resolution are needed to capture the intricacies of vertical processes from the surface through the boundary layer into the troposphere, stratosphere, and mesosphere. With these extensions, we can accurately capture processes, such as convection, turbulence and the Madden Julian Oscillation (MJO), and troposphere-stratosphere interactions related to the Quasi-Biennial Oscillation (QBO), North Atlantic Oscillation (NAO), and the Arctic Oscillation (AO), for more accurate intraseasonal to interannual climate predictions particularly for the North American sector.

These are computationally expensive propositions. Each doubling in horizontal resolution requires eight times the computational capacity to complete the model prediction of a given duration in the same amount of time. Extending models toward the upper boundary of the atmosphere with increased vertical resolution requires four to eight times the computational capacity needed for today's models

Capturing complexity in the Earth System

Increasingly sophisticated representations of Nature in NOAA models require additional computational capacity. For example, fully interactive chemistry and aerosols in high-resolution coupled models are needed to better understand and predict the regional impact of pollution, ocean acidification, the recovery of the stratospheric ozone layer and, in particular, the ozone holes, and the impact of aerosols on climate. Advanced numerical representations of aerosol and cloud physics, including aerosol-cloud interactions will enable better understanding of the interactions between cloud processes, convection, radiation, and dynamics and their impact on physical climate feedbacks and sensitivity. The addition of atmospheric chemistry and biogeochemistry including land-atmosphere interactions, aerosols, and cloud physics increases the computational requirement by about a factor of three.

Adding complete biogeochemical cycles (e.g., for nitrogen and phosphorous) in the land and ocean improves our understanding of the impacts of climate change on coastal ecosystems, fisheries, ocean acidification, and the risks of harmful algal blooms, as well as the ecosystem changes directly resulting from human activities (such as fires, land use, etc). Adding terrestrial ecosystems and oceanic biogeochemistry for Earth System prediction requires an increase in the computational capacity by a factor of three.

Coupled climate models using high resolution component models of the atmosphere, ocean, cryosphere, and chemistry, including biogeochemistry and ecosystem physics will accelerate the delivery of high-resolution regional predictions on time scales from weeks to decades, including sea-level rise, Arctic sea ice extent, and the localized risk of hurricanes, drought, and other crucial long-term heads up warnings of major events anywhere in the Nation. Coupling with wave forecasting is critically important for coastal and marine safety and a necessary complement to storm surge as demonstrated by the impact of Hurricane Sandy. Yet coupling atmosphere, ocean, wave, land, cryosphere, biogeochemistry and ecology in a



prediction system can require 3-4 times the computational capacity since all the model components must be integrated concurrently in their individual domains.

It is essential to plan for these types of mission expansions both in the personnel to develop and maintain the software and in the HPC capacity.

Quantifying Uncertainty

Specifying a measure of confidence (or, likewise, forecast uncertainty) for any model forecast is essential for completing the forecast information communicated to users. Confidence measures can inform the operational forecaster and other users of warnings and outlooks for high impact events and the area impacted. Ensemble-based systems provide the best means of providing confidence information. Ensemble systems have typically been run at lower resolution than the single “deterministic” forecast; in the future, in order to maximize the information to the users, it is highly desirable to execute the ensembles at the highest possible resolution with the highest number of members. Consequently, quantification of forecast uncertainty through an ensemble strategy requires a 10-50 times increase in compute capacity. Ensemble techniques are also used to assess the uncertainty in climate projections, requiring a similar HPC capacity increase.

Goals and Objectives

NOAA has established five Goals for High Performance Computing that will ensure that the agency is best positioned to use HPC to meet its missions. These goals provide a path forward to maintain NOAA’s success in providing HPC services, position NOAA to take advantage of new technologies, and to plan for growing user needs. These goals will enable NOAA to provide HPC to its mission users in more effective and more efficient ways. The Goals of the NOAA High Performance Computing Strategic Plan are to:

Goal 1: Provide enterprise HPC services to enable the agency’s mission.

Objective 1: Maintain core competency and HPC enterprise capability to provide reliable computing to NOAA’s research and operational missions and to other federal agencies.

Computing capacity, short and long-term storage, and scientific analysis are central to NOAA’s mission. NOAA will continue to operate and maintain its core enterprise capability and services for computing; post-processing and analysis; and long-term storage assets. In-house HPC expertise familiar with the mission applications, is a key asset essential to optimize, continuously modernize, and exploit the use of NOAA’s HPC capabilities for the maximum return.

NOAA will act as a shared services provider enabling other federal agencies to leverage NOAA’s core HPC enterprise capabilities in a cost reimbursable manner. This will enable other federal agencies to gain access to HPC in a cost effective manner by leveraging purchasing power and reducing the burden for these agencies to maintain separate HPC assets.



Objective 2: Leverage other federal agencies HPC shared services to increase NOAA's access to leadership-class computing and novel architectures

On the path to exascale computing, NOAA will leverage leadership-class facilities that are available throughout the Government. Access to these national assets will enable NOAA to scale their work further and leverage engineering and resources that would not be otherwise attainable. Access to leadership class computing will enable breakthrough research.

NOAA will gain access to these world-class facilities through a combination of dedicated and shared computing services at other agencies. For research requiring leadership-class computing over a multi-year duration, NOAA will develop specific interagency agreements to secure dedicated computing time providing scientists a consistent development platform for research. For research which stretches the scalable limits of its models, NOAA will explore national shared computing services through existing agency programs such as DOE INCITE and NSF XSEDE.

Objective 3: Provision and allocate HPC services consistent with requirements of NOAA's mission workflow.

Computational and storage capabilities are determined by model and workflow requirements. Working with agency partners, resources will be acquired and configured to achieve optimal results. Feedback will be provided through the allocation process to ensure the modeling workload is assigned to the most appropriate computational environments and consistent with NOAA mission objectives.

Goal 2: Improve linkage between NOAA mission requirements and HPC solutions

Objective 1: Establish integrated support of NOAA's applications to optimize HPC solutions and to guide model development based on changing technologies

As heterogeneous computing architectures and their associated programming environments mature, the interplay between hardware and software application design has become more tightly coupled. NOAA needs to continually adapt to a more complex set of emerging programming standards and wider array of heterogeneous computing solutions. For example, it is important to note that when increased resolution or additional ensemble members are required, more computational work can be farmed out to additional processors so that the overall pace of the system execution is generally unaffected. In contrast, increasingly complex models demand new approaches to concurrency, so that additional physical processes and variables can be assigned to the additional processors. NOAA will develop an integrated software engineering team, comprised of lab and center personnel, to enhance its software engineering discipline and expertise to achieve optimal code performance and scaling (while maintaining code portability), maximize the efficiency of transitioning research to operations, and enable effective collaborative model development with partners both internal and external to NOAA. In close collaboration with the modeling community, this integrated software team will provide the necessary



bridge for environmental modeling to work in tandem with the existing integrated management team, which currently manages the acquisition, provisioning, and day to day operation of HPC.

Objective 2: Develop scientific and software assurance methodologies to deal with increasing complexity of HPC systems

The mean time between failures for HPC will decrease dramatically as the number of processing cores and integrated components within the HPC increase by orders of magnitude. This increase in system complexity leads to more frequent job runtime interruption and potential data corruption requiring additional sophistication in the applications to achieve the current level of runtime reliability and data integrity. NOAA will need to invest in software automation techniques required for fault tolerant data movement, and error detection, correction, and handling to ensure mission applications run reliably. NOAA scientists will have to increasingly work together with software engineers to develop workflows that have fault tolerance built in from their inception. Interdisciplinary teams will become the norm for model development within NOAA.

Objective 3: Explore innovative methodologies and solutions to improve efficiency of data storage and effectiveness of data analytics

Within the next five years, NOAA's hierarchical storage management system, within its HPC program, will store hundreds of petabytes of data. While growth in computational demand continues at an exponential rate, storage performance and capacity is lagging behind creating an incrementally increasing cost for storage over time.

NOAA will refine its workflows to ensure appropriate trade-offs are made when storing data. Engagement with industry will continue to ensure that storage and data movement technologies can continue to scale to the required capability and reliability. Hardware and software innovations will be exploited to enhance scientific analysis.

Goal 3: Recognize and plan for emerging uses of HPC

Objective 1: Perform outreach and education to NOAA programs about HPC services.

The number of mission areas requiring HPC services has grown over the past few years and will continue to grow in the future. NOAA has incorporated new modeling capabilities, such as the National Ocean Service's 14 operational forecast systems, onto its HPC systems for coasts and estuaries. Migrating these applications from a workstation environment allows users to run more complex models at higher speeds and greater resolution. Aside from these benefits, users can also leverage access to high-resolution model output from NOAA's traditional HPC user communities. NOAA anticipates increased HPC use for ecosystem and coastal applications and other analytical applications.

Outreach and education will be provided to communities within NOAA which have not traditionally used HPC. The HPC program will help translate the complexities of using HPC environments involved with parallel programming, debugging, data movement and storage. Training will be provided by experts from



NOAA, our partners, and industry. Enterprise HPC wikis, help desk support and NOAA application analysts will assist this NOAA community in making efficient use of HPC resources.

Objective 2: Enhance support to accommodate new communities and technologies

Computational testbeds will be established to support initial “development” allocations. These testbeds will be associated with applications and technology liaisons to assist new users. The NOAA applications lead will engage new Principal Investigators to ensure that they can adapt their codes to new platforms and that they are aware of the latest programming standards.

As NOAA’s HPC program has evolved and matured, the user community has grown and diversified. To foster a strong relationship between the HPC program and its users, NOAA will establish an HPC User Committee with representation from the major user communities and some of the “pioneer” users identified by the NOAA Allocation Committee. The User Committee will bring the advice, guidance, and point of view of computer users to the attention of the HPC program; and to exchange information concerning effective utilization of the HPC resources available to NOAA.

Objective 3: Evolve funding model to account for emerging HPC users

Currently, NOAA’s HPC assets are funded primarily by OAR and NWS. (Operational supercomputing is augmented by NOS pursuant to an MOA with NWS on hydrodynamic modeling for oceans and coasts.) NOS, NESDIS, and NMFS, each receive small allocations on NOAA’s HPC. As these new users’ mission requirements grow, a new funding model will need to be established to recognize the cost of providing those services. Significant new computational and storage allocations will only be provided when programming, network, and HPC support are provided by the requestor. In the long term, emerging users will need to contribute funding and some personnel support towards the overall cost of the HPC enterprise and provided services.

Goal 4: Effectively adopt latest HPC technologies to drive efficiencies.

Objective 1: Access novel architectures

High Performance Computing is at the cutting edge of IT technology. HPC trends rapidly change through successive generations of technology and require that applications constantly adapt to ensure they can best exploit the computational platforms. Access to novel architectures enables developers to explore the opportunities and challenges associated with running new and existing applications on the next generation of computing.

NOAA will identify technology testbeds for novel architectures. The capabilities of other agencies, university partners, industry and our own assets will be leveraged. Early adopters will be able to evaluate the programming, I/O handling, software, and storage environments for NOAA’s mission workload.

Objective 2: Create training opportunities and workshops to enable exchange of knowledge between model developers and HPC solution providers



Understanding the programming, software, and technology limitations of next-generation architectures is beneficial when implementing new algorithms for numerical models. Advanced techniques ensure that applications can scale and exploit new platforms without having to be completely rewritten.

NOAA experiences will be communicated with industry and our partners to ensure that our concerns, “lessons learned”, and successes can be leveraged to improve the next-generation technologies. NOAA will benefit by ensuring that appropriate enhancements will be made to enable environmental codes to exploit these architectures. This will serve as a benefit not just for NOAA, but the HPC community as a whole - as the technology will be more stable, functional, and better performing.

Objective 3: Develop integrated enterprise-wide competency to recognize, plan, and evaluate emerging technologies for future acquisitions

The leading NOAA applications engineers will work with domain scientists to identify agency-wide benchmarks that are both relevant and computationally challenging. These applications will be cross-compiled on various architectures and made available to the entire modeling community and HPC vendors. As appropriate, performance numbers will be made publicly available. Since new software capabilities are typically frozen during operational transitions to new technologies, these transitions should be planned and resourced to occur with maximum reliability and minimum elapsed time.

Regular interactions with HPC vendors will be ongoing to ensure that both technology challenges and successes are communicated. Through these regular interactions, NOAA’s leading application engineers and community of users will remain informed of the latest software and technology trends by participating in various community forums.

Goal 5: Maximize effectiveness of HPC solutions

Objective 1: Provision HPC within a functional service framework

NOAA has a diverse set of application workload requirements that are best aligned with an equally diverse number of configurations of the HPC systems. Optimally locating these mission applications will require analysts to work with the user community to select among: high availability computing; large-scale, high I/O computing; extreme scale computing; emerging technology testbeds; or post-processing and analysis systems.

NOAA will provision HPC systems to align with these functional workload characteristics. High availability computing is designed to support cyclical, time-critical workloads characterized by its resiliency features comprised of three similar systems. Large-scale, high I/O computational systems will support workloads that are characterized by their data intensity, and massively parallel applications. NOAA will support next-generation computing technology by provisioning both extreme scale computing and emerging technology testbeds. The extreme scale computing supports workloads that require a scale that cannot be achieved without the use of leadership-class machines. Emerging technology testbeds will enable users to evaluate the programming, I/O handling, software and storage environments for NOAA’s mission workloads of the most cutting edge IT technology.



Objective 2: Provide secure HPCC enterprise services

The HPC Program will leverage the NOAA Security Operations Center and other NOAA enterprise security services for centralized logging, reporting, and incident awareness. Additionally, this will improve the program's integration with the NOAA Computer Incident Response Team (N-CIRT).

As NOAA's integrated modeling suite seeks to utilize more community based applications, the need for collaboration with users outside of NOAA in the university and international communities will increase. While the need for external collaboration is increasing, the need for more intense security is also increasing as the threats become more sophisticated. NOAA needs to explore innovative approaches to securing its HPC systems while enabling insightful collaboration.

Objective 3: Improve the user experience

NOAA will enhance tools and methods to provide users a consistent user experience across its integrated HPC resources. For example, NOAA will improve workload and software management practices, establish common reporting, and create common documentation and communication tools.

NOAA will institute improvements to its HPC governance model and the establishment of an HPC User Committee. All of NOAA's HPC will be acquired, managed and allocated as an integrated set of resources, based on the functional workload requirements of the modeling applications.