NOAA Data Assimilation Strategy

Strategic Plan
2024–2033

DOI: 10.25923/yz0v-my29

Signature Page

MORGAN.MIC
HAEL.COTTMA
N.1616131016

Digitally signed by
MORGAN.MICHAEL.C
OTTMAN.1616131016
Date: 2024.09.30
11:24:51 -05'00'

Dr. Michael Morgan,
Assistant Secretary of Commerce for Environmental Observation and Prediction

LE
BOEUF.NICOLE.R
ENE.1365870841

Digitally signed by LE
BOEUF.NICOLE.RENE.13658
70841
Date: 2024.09.24 15:19:26
-04'00'

Ms. Nicole LeBoeuf, Chair, Earth Systems Integration Board
Assistant Administrator, National Ocean Service (NOS)

SHONTZ.KATHR
YN.ANN.139237
5631

Digitally signed by
SHONTZ.KATHRYN.ANN.1
392375631
Date: 2024.12.02 14:02:09
-05'00'

Ms. Kathryn Shontz, Chair, NOAA Modeling Team
Deputy Director, Office of Common Services (OCS)

Table of Contents

SECTION 1

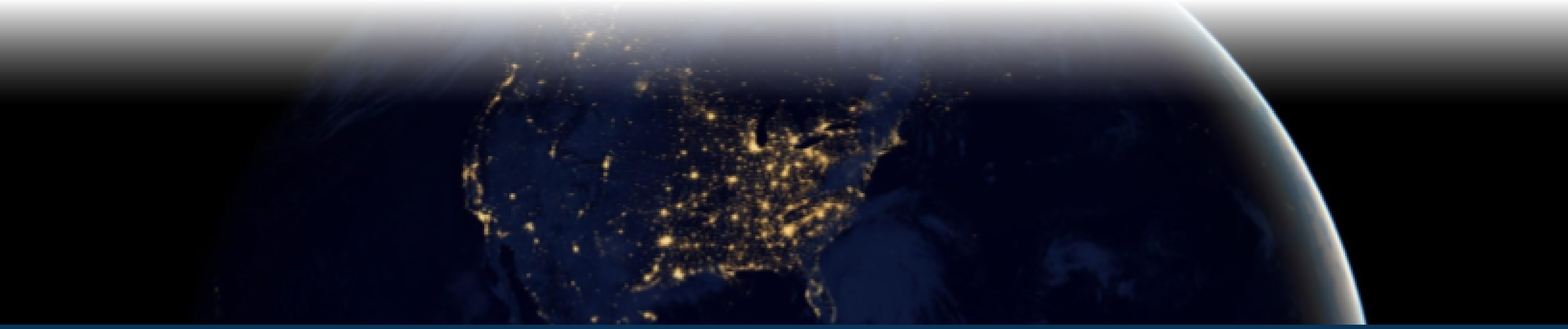
Introduction

The National Oceanic and Atmospheric Administration (NOAA) Data Assimilation (DA) Strategy addresses the grand challenges, opportunities, and key focus areas for NOAA to pursue over the next 10 years to fully realize the information potential from both current and future environmental observations in NOAA's DA applications, and continuous delivery of analyses to users and downstream applications. DA is generally described as the end-to-end process of combining environmental observations and model data to estimate the true state of the Earth system (i.e. the analysis) and its components of the atmosphere, ocean, land, cryosphere, and ecosystems. DA is a critical component of Earth system prediction, for its use of observations to inform initial conditions for model simulations across scales, for providing analyses of active or emerging environmental threats in near real-time to stakeholders and decision makers, and for producing reanalyses of historical conditions that are important for both Earth system monitoring as well as for training data-driven models. Therefore, NOAA's mission and service areas fundamentally depend on DA applications and their ability to utilize environmental observations.

Coinciding with the drafting of this strategy, NOAA is transitioning multiple DA systems to the Joint Effort for Data assimilation Integration (JEDI) framework along with other partners of the Joint Center for Satellite Data Assimilation. The JEDI infrastructure will unify and promote fully coupled DA across Earth system modeling components including for coastal and hydrologic applications, allow for more rapid advancement of DA algorithms and integration of new observation types, and provide a framework allowing NOAA to more easily adopt technologies such as Artificial Intelligence (AI).

Just as DA tools and methods are evolving to advance the state-of-the-art, so too are the observing systems evolving to include new, innovative in-situ and remotely sensed observations, as well as constellations with exceedingly large volumes of data due to increasing spectral, spatial and temporal resolutions. To meet current and emerging DA grand challenges, NOAA must adapt to the evolving technologies and observing systems and adopt a DA Strategy.

The DA strategy is commissioned by the NOAA Modeling Team (NMT) under the NOAA Earth Systems Integration Board (ESIB), and is an objective of NOAA's Modeling Strategy[1]. The DA strategy is responsive to and aligned with the NOAA Science Council's NOAA Artificial Intelligence[2] and NOAA Cloud Strategies[3]. The strategy is also responsive to many of the NOAA Science Advisory Board's Priorities for Weather Research (PWR)[4] study recommendations for Observations and Data Assimilation, Information Delivery, and Foundational Elements.

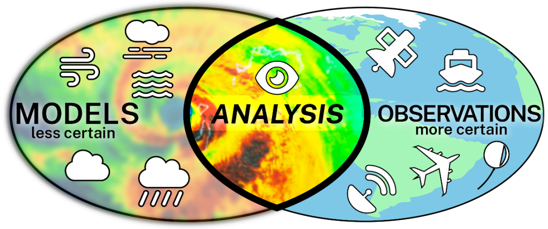


SECTION 2

VISION AND PURPOSE

Vision

To serve societal needs through fully-coupled and continuous Earth system data assimilation capabilities which maximize the impact of environmental observations.



The Data Assimilation Concept

Data assimilation is a (statistically) optimal combination of uncertain environmental models and observations. The result of assimilating observations is an analysis of the Earth system state that is more accurate and complete compared to either the model or observations alone. Data assimilation is practiced daily across NOAA, including as part of the numerical weather prediction enterprise, or for coupled ocean and coastal modeling. For example, the next Global Forecast System (GFSv17) will optimally combine coupled short-range forecasts of the atmosphere, ocean, ice, land, aerosols, and waves with a comprehensive set of Earth system observations obtained from a variety of observing platforms that include satellites, aircraft, weather balloons, surface stations, surface-based radars, and ocean buoys. The resulting analysis is then used routinely for a number of applications, including situational awareness, initialization of NOAA operational forecasts, experimental forecasts with data-driven models, and for understanding the state of the Earth system climate and variability (through reanalyses).

Data Assimilation Grand Challenges

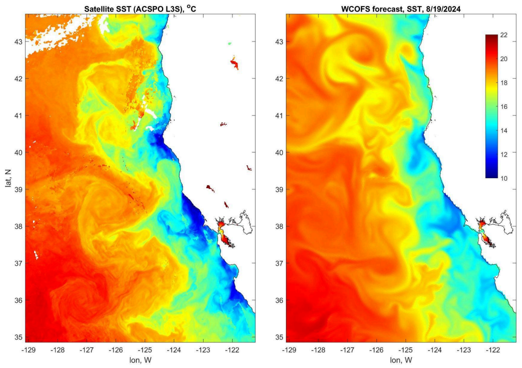
The NOAA Data Assimilation Strategy will address the following grand challenges facing data assimilation across the agency.

Current observations. Increasing the volume of observations assimilated in DA applications.

New Observations. Reducing timelines for integrating and transitioning new observations to operational data assimilation systems.

Observing systems. Minimizing gaps in the observing system, particularly for in-situ and sub-surface oceanic observations, as well as creating efficient processes to inform observing systems planning and assess observation impacts.

Nonlinear multiscale data assimilation. Developing impactful approaches to assimilate observations across spatial scales in highly nonlinear or non-Gaussian environments, for example satellite radiances affected by clouds and precipitation.



Nonlinear Data Assimilation

(LEFT) Level 3 Sea-surface Temperature (SST) retrieved from the Visible Infrared Imager Radiometer Suite (VIIRS) and **(RIGHT)** West Coast Operational Forecast System (WCOFS) forecast 09 SEP 2021.

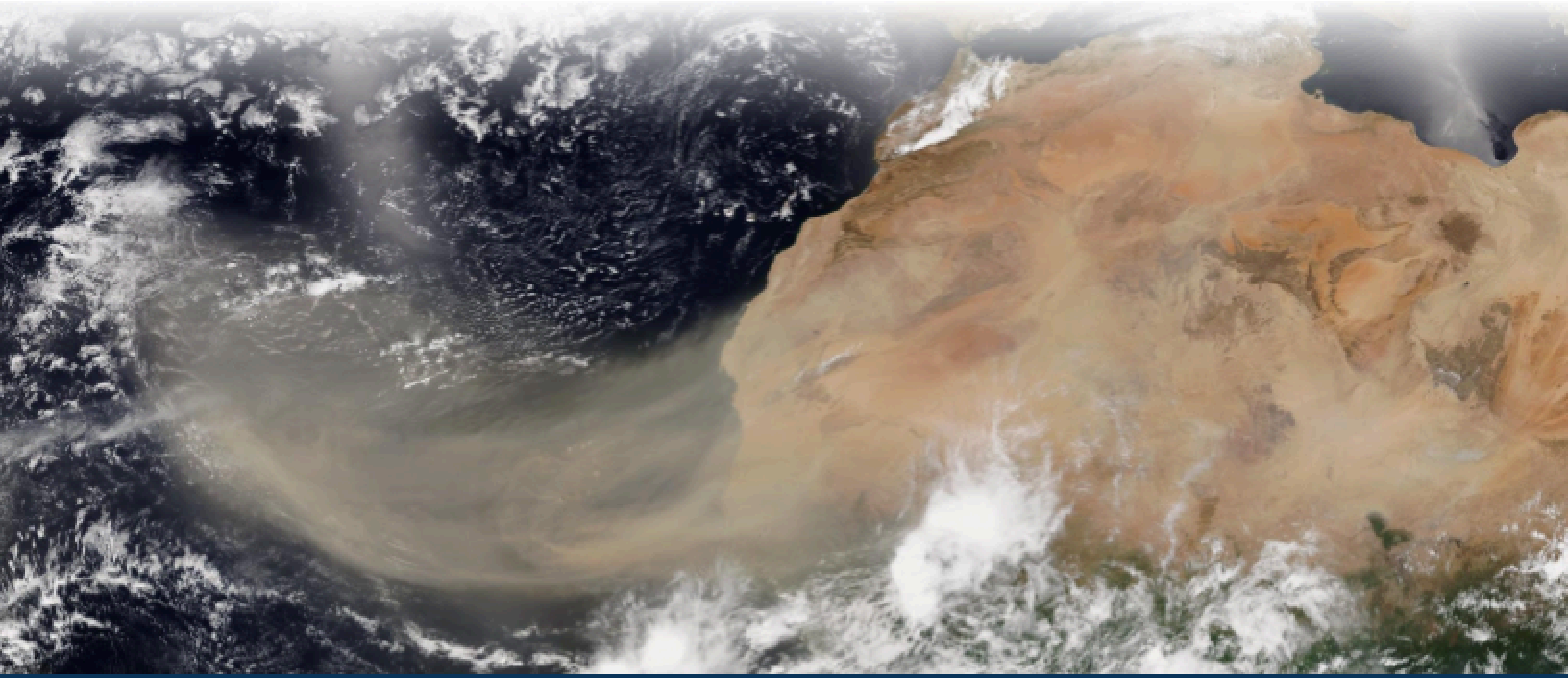
Strongly nonlinear dynamics results in limited predictability of coastal flows and requires advanced DA methods such as 4-D variational data assimilation (4DVAR). The WCOFS assimilates Sea-Surface Temperature (SST), altimetry and other data to constrain uncertainty and yield accurate ocean forecasts in support of navigation, search and rescue, environmental hazards response, fisheries, and tourism, for example.

Coupled data assimilation. Advancing the science for coupled Earth system data assimilation and the ability to effectively characterize constraints across components.

Efficiency. Reduce the computational cost to process and assimilate observations and the latency between the time of observation and availability for decision making.

Reanalysis. Maximizing impact from the historic observational record to provide the best reconstruction of historic weather, climate, and ecosystem conditions and translate this knowledge to improved predictions.

Workforce. Hiring and retention of qualified data assimilation experts, along with training the next generation data assimilation workforce with expanded skill sets, including data science.



Purpose

As the global observing system evolves and expands, information will be provided by an increasing variety of sources, and at volumes exponentially greater than today. Compounding the challenge is an increasing societal need for environmental information in real-time. To keep pace and meet demand, NOAA must adapt and develop enhanced capabilities toward meeting the grand challenges by processing data thoroughly but efficiently, preparing for and integrating new observations into operational DA applications more rapidly, and maximizing the impact of all relevant observations on analysis and reanalysis datasets.

The DA Strategy will enable NOAA to build and strengthen expertise not only in fundamental data assimilation techniques, but also foster development of the skills required for the next generation of data assimilation scientists in new observation types and tools such as data science, Artificial Intelligence (AI), software engineering, and cloud computing, with the potential to address some expertise gaps through innovative partnerships. These skill sets will be essential for implementing the DA Strategy and drive development of new techniques enabling continuous data assimilation where observations have an impact on the analysis within seconds or minutes after measurement. The development includes engineering new infrastructure to process and assimilate a continuous stream of data from various sources, as well as to perform quality assurance, dissemination, and archiving; new AI techniques to filter and extract the most impactful information, blurring the boundaries between observation and analysis; and new tools to rapidly assess the impact of observing systems on analysis quality and to inform future observing system requirements.

To meet the 10 year vision, the DA Strategy will provide the focus necessary to align research and development efforts with resource and operational needs across NOAA toward a flexible framework with the ability to continuously assimilate and fuse together observations across the Earth system.

SECTION 3

NOAA DATA ASSIMILATION STRATEGIC GOALS

Goal 1: Build required infrastructure for continuous development, integration, and data assimilation

The data assimilation end-to-end processes within the development, testing, and operational life-cycle requires availability of considerable infrastructure, storage, and compute resources. NOAA will benefit from seamless integration and deployment of DA applications, and seamless, continuous assimilation of environmental observations, promoting not only efficient use of resources, but also keeping pace with increasing data volumes and delivery of near-real time analyses. Fully transitioning to the JEDI framework will provide a cornerstone from which future innovations may be developed and integrated.

Objective 1.1: Develop capabilities for continuous data assimilation, automated optimization and integration of innovations in research and operational systems, including new observations and new data assimilation algorithms across Earth system components and time and space scales.

Objective 1.2: Advocate for robust compute and storage resources and capabilities for assessing impacts on NOAA's mission from new observation sources, including from non-NOAA sources, and to highlight gaps in the global observing system and potential solutions.

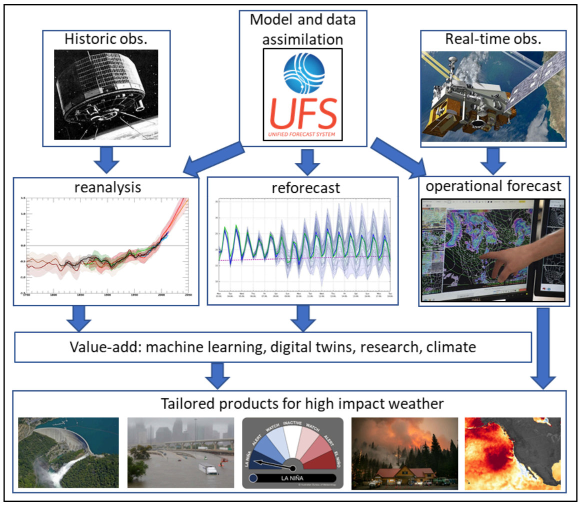
Objective 1.3: Adopt commercial cloud architectures to increase efficiencies in data management supporting research and operational data assimilation, add flexibility to scale compute resources to expedite high-impact studies, and accelerate transition of research to operations.

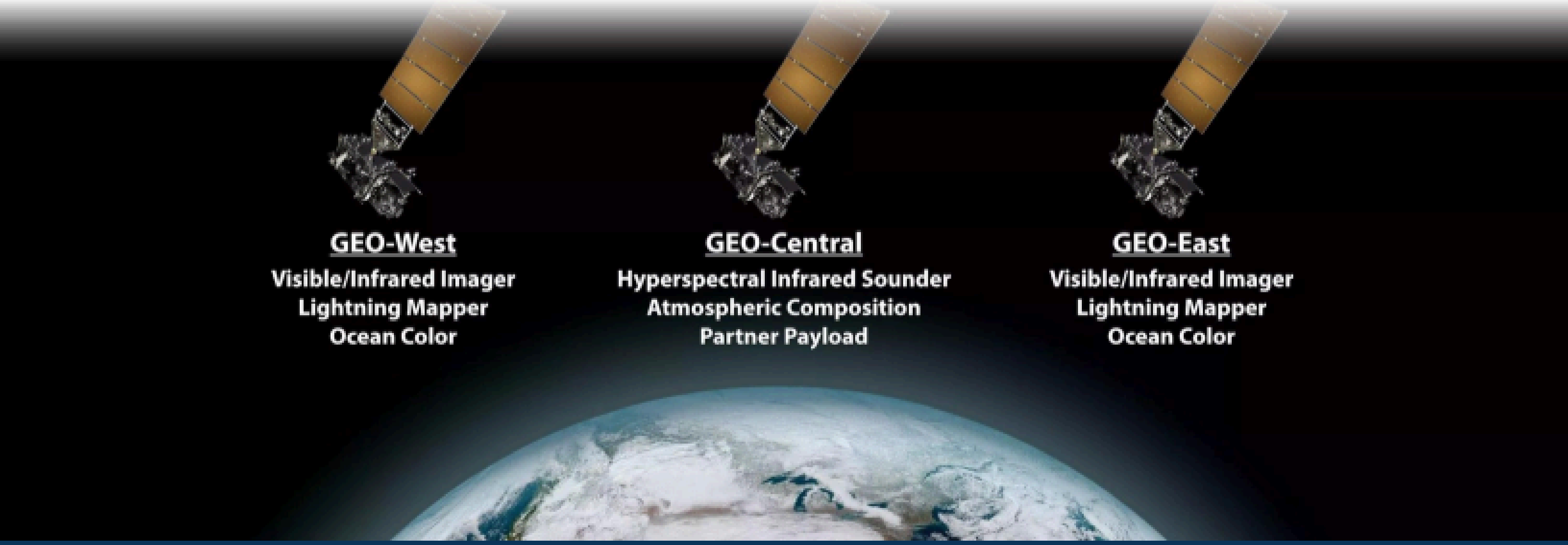
Objective 1.4: Expand software and hardware infrastructure to accommodate, and develop new tools using AI, to process and assimilate a growing volume of observations from: new high resolution (spectral, spatial, and temporal) data sources; emerging in-situ and remote sensing reanalysis platforms, such as uncrewed systems, smallsats/cubesats, non-NOAA systems; and existing but underused sampling systems such as sustained field-based biological sampling while reducing the latency between observation and ingest into applications.

Objective 1.5: Develop a sustained and continuous effort to provide the best possible Earth system reanalysis and reforecast datasets (global and regional) through improved DA techniques and increased use of observations, populating open data lakes, supporting model development and evaluation, as well as providing high quality, expansive data for developing AI algorithms.

The Importance of Reanalyses and Reforecasts

Earth system reanalyses and reforecasts are essential digital infrastructure that support NOAA's operational and research missions as well as the rapidly growing climate and weather enterprise. Reanalysis combines historical observations with modern Earth system models to generate a spatially and temporally complete history of the Earth system. Reforecasts (initialized from reanalyses) are retrospective forecasts using the operational forecast model that can be used to correct errors in today's forecast. Combined, they are fundamental to producing useful forecasts and understanding the predictability of high impact weather, water, and climate events (such as heat and cold extremes; hurricanes; fire weather; hydrological, solar, and wind droughts; winter storms; flooding; and marine heat waves). This is especially apparent with the emergence of machine learning and digital twin tools that use reanalysis and reforecast datasets to enhance the value of "raw" output from the NOAA operational models. The recent "Report on Priorities for Weather Research " (PWR report[4]) identified a sustained reanalysis and reforecast capability as one of the top 10 critical recommended actions.





Goal 2: Maximize the use, value, and impact of Earth system observations

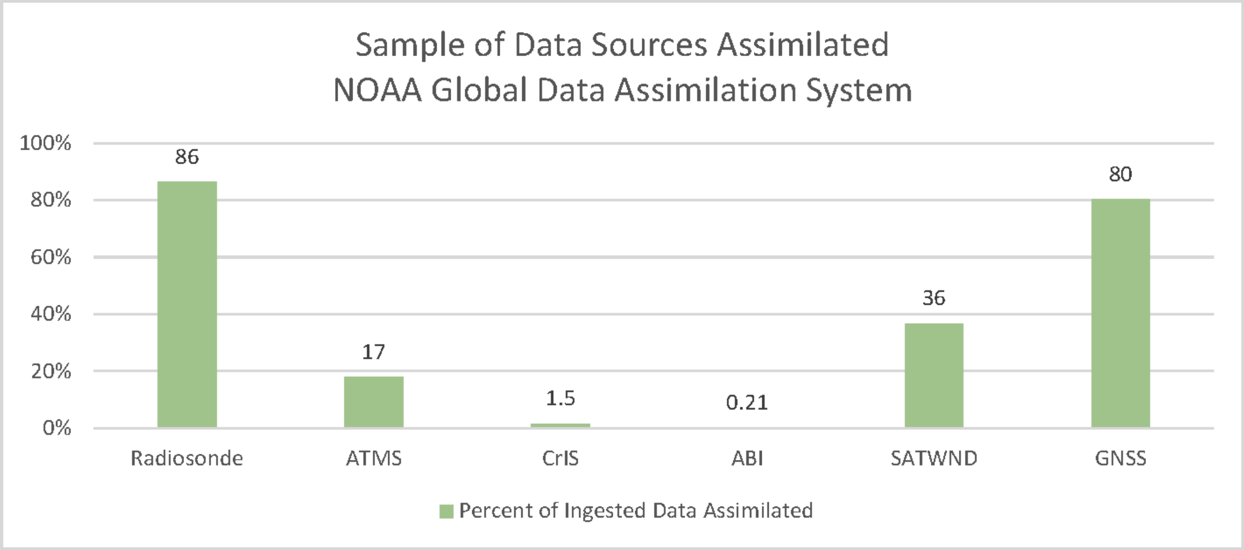
Data assimilation requires accurate estimates of the model state uncertainty and observation error, including the uncertainty in representing the model state in observation space. Additionally DA approaches may require assumptions that introduce further constraints which limit the impact of observations on the analysis. NOAA will pursue a holistic approach to maximize the use of observations in DA, from identifying where future value in observations can be gained, to exploring the use of technologies such as AI to improve observation use and impact.

Objective 2.1: Develop a fully coupled, Earth system data assimilation system through leveraging and improving components and subcomponents, to produce a physically consistent, complete, and accurate representation of the Earth system state.

Objective 2.2: Maximize the use of historical and contemporary observations in analysis and reanalysis by optimizing observation error characterization, bias correction, sensor calibration/validation, quality control, data selection, and observation operators, including leveraging AI methods.

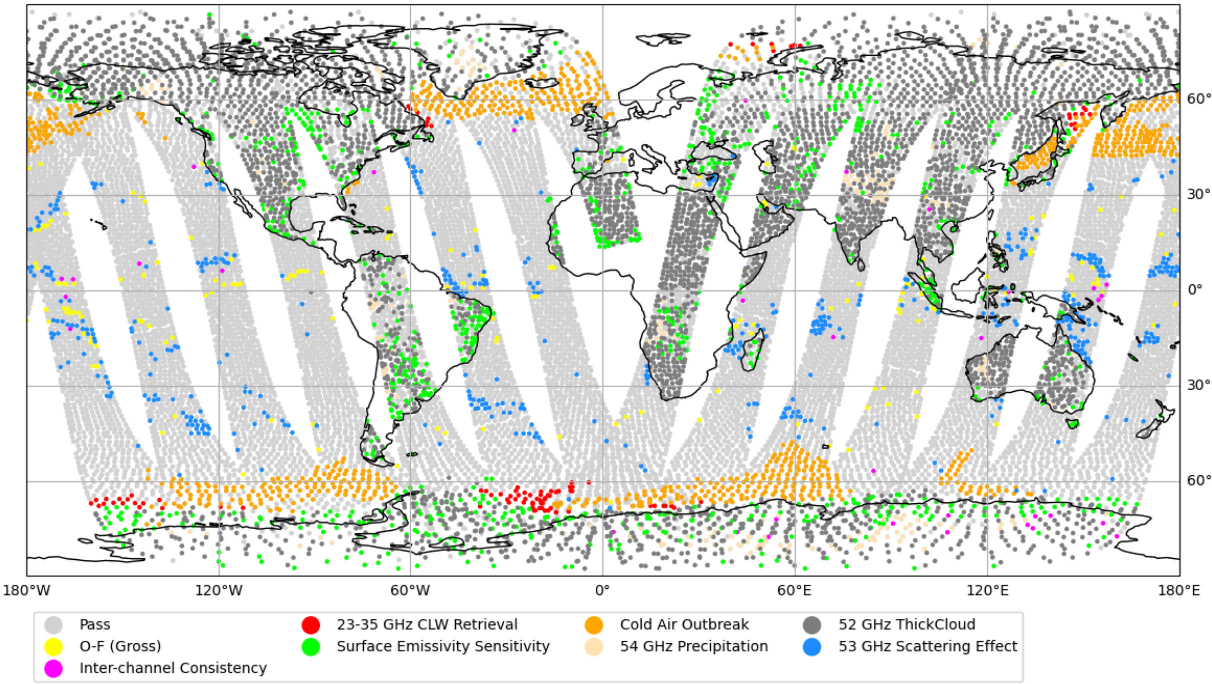
Objective 2.3: Develop an efficient, standing capability to quantitatively assess the impact of individual or multiple components of the global observing system on NOAA applications, with well-defined metrics, and to accelerate their readiness and transition to operations, including both observations operationally assimilated as well as missions of opportunity.

Objective 2.4: Document data gaps in the global observing system, including from under-utilized observations and non-NOAA sources, particularly with respect to the representation of the Earth system, and develop plans and priorities for addressing gaps.



Observation Utilization Rate

Numerous in-situ and remote sensing observing systems provide critical information that, when combined, describe a nearly complete picture of the current Earth system state. For data assimilation applications, the observations are used to derive more accurate initial conditions for environmental prediction systems. Effective and efficient use of observations, however, are determined by a number of factors including the data quality, representativeness in the model state and the ability to model the observation, as well as computational constraints. The figure above illustrates the percentage of data assimilated relative to the number of observations ingested (for the case of ATMS, ~1.8M) for one cycle of NOAA's Global Data Assimilation System (GDAS), for a selection of observing systems used. As illustrated, the utilization rate varies significantly by data source. Low utilization for Advanced Technology Microwave Sounder (ATMS) is primarily due to spatial thinning and challenges assimilating observations impacted by light scattering from precipitation hydrometeors and a large uncertainty in surface emissivities; for the Cross-track Infrared Sounder (CrIS), spectral thinning is performed to remove redundant information or information not relevant to the model state; and for the Advanced Baseline Imager (ABI), high resolution data are thinned spatially and temporally to reduce computational needs. The NOAA DA Strategy will support approaches to maximize the use of observing system data with an emphasis on extracting the most impactful information from all observations.



Quality Control of Observations

Data assimilation requires strict quality control to ensure that the observations assimilated reflect the true state of the environment with minimal uncertainty, to be effective at correcting errors in the model state background and leading to an accurate analysis of the Earth system state. Observations will be quality controlled either because the measurement is taken with too high degree of uncertainty (e.g. noise), or because the model state is not able to represent the phenomenology of the observation (e.g. cloud microphysics through observation operator functions). The above plot illustrates the quality control flags of the Advanced Microwave Sounding Unit (AMSU-A) for a 12-h assimilation window for the Global Data Assimilation System (GDAS). Light gray points show where observations passed quality control checks, compared with all other colors illustrating the various criteria in which the observations failed quality control. In general, observations fail quality control due to the inability of the model state to properly characterize the observation, including over land and sea ice due to emissivity, or over the ocean due to the scattering of frozen precipitation. New techniques should enable improved characterization of these scenes in both the model state and observation operators.

Goal 3: Transform and build a workforce that synergizes skill sets needed for the advancement of fundamental data assimilation science, emerging technologies, and observing systems to achieve operational success

The current availability of DA expertise is insufficient to meet the needs both internal and external to NOAA. In addition, the required skills for developing next generation DA capabilities are evolving and broadening. NOAA will focus on creative and diverse approaches to build and grow the next generation DA workforce, promoting use of NOAA's state-of-the-art DA tools such as JEDI in academia and the private sector for training and research, and pursuing new partnerships towards sustaining a workforce proficient not only in data assimilation science, but data science, AI, software engineering, and observation technology.

Objective 3.1: Establish NOAA as a premiere center of excellence for Earth system modeling and data assimilation, promoting professional and technical development of its existing workforce through innovative work assignments and training, to address data assimilation grand challenges through highly-esteemed, well-resourced, dynamic, and inclusive teams.

Objective 3.2: Strengthen the education-to-career pipeline to grow the data assimilation workforce. Leverage and diversify Science, Technology, Engineering and Math (STEM) initiatives to support internships, visiting scientists, and exchange programs through university consortiums, including the Cooperative Institutes and the NOAA Cooperative Science Centers.

Objective 3.3: Partner with academia and the private sector, and integrate highly-specialized subject matter experts with new skill sets, particularly in emerging areas of advanced computing architectures, cloud computing, AI, and new observing systems, on multidisciplinary teams for improving efficiency and performance toward continuous, coupled DA.

Objective 3.4: Broaden involvement of the NOAA data assimilation workforce in the international research community.

Objective 3.5: Foster collaboration and workforce development by adopting an open-science framework to include open-source, open-data along with data stewardship and dissemination, and open system configurations, toward enabling the research, academic, and private industry communities to train and grow DA expertise needed for NOAA's mission.

Goal 4: Optimize efficiencies in research, development, and operational transition of new capabilities across the data assimilation enterprise

Data assimilation activities across NOAA will be well coordinated, and will require identifying where current gaps in research and operational capabilities exist, along with the corresponding opportunities to improve current data assimilation systems and increase readiness for future observing systems with the purpose of maximizing value and to reducing duplication of effort.

Objective 4.1: Establish a NOAA Data Assimilation Team to fully coordinate efforts across Line Offices, and improve efficiencies in planning, funding, developing, and implementing new capabilities.

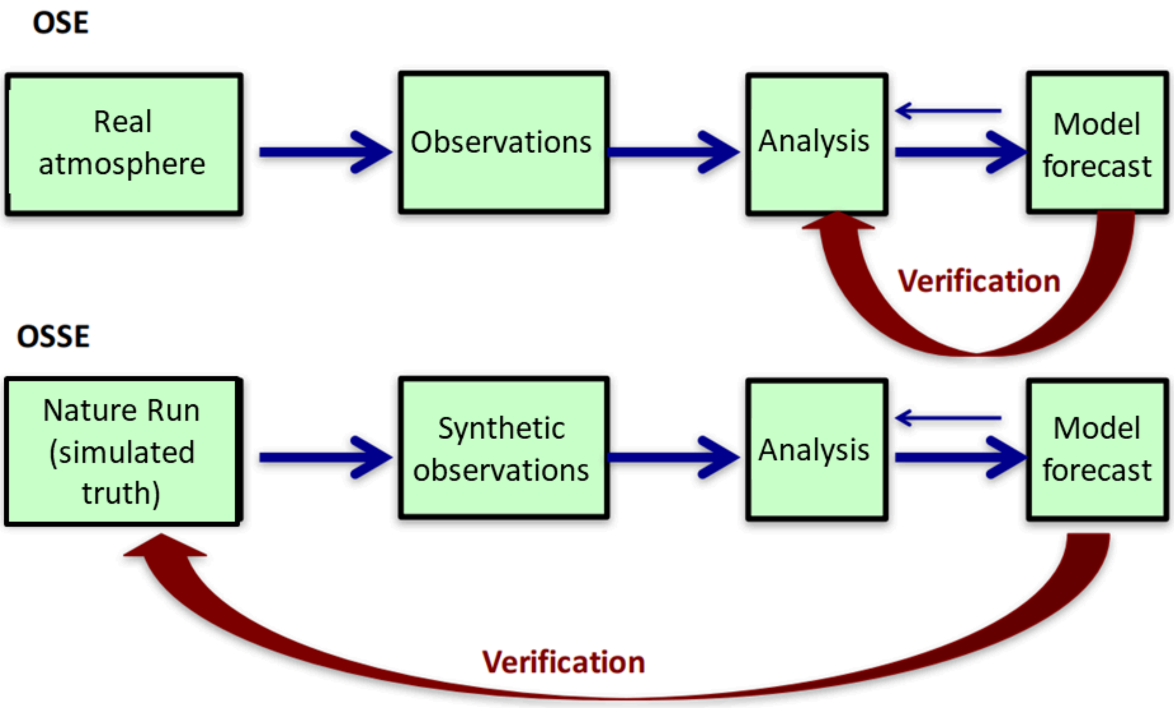
Objective 4.2: Develop a roadmap of future observing system capabilities, and derive end-to-end action plans for integration of planned observational capabilities into NOAA research and operations.

Objective 4.3: Define a broad set of Earth system analysis and modeling performance metrics that can be used in evidence-based prioritization of investments in the end-to-end data assimilation pipeline.

Objective 4.4: Invest in research that will enable continuous data assimilation across scales and embrace AI technologies to increase efficiency targeting specific DA processes or to fully emulate traditional DA systems.

Objective 4.5: Align High Performance Computing capacity, availability, and accessibility with research investments in data assimilation, including algorithm development, Observing System Experiments, Observing System Simulation Experiments, and data assimilation emulators using AI.

Objective 4.6: Build frameworks such as JEDI and processes which promote collaboration between research and operational teams to accelerate research-to-operations and operations-to-research (R2O2R) of new data assimilation capabilities, and create seamless boundaries between cutting-edge research and operational applications.



Exploring the Impact of Observations on Environmental Prediction

Observing system experiments (OSEs) and observing system simulation experiments (OSSEs) are tools used to conduct quantitative evaluations of the value of observing systems. OSEs address today's modeling and observing systems capabilities. These assessments use real data (not simulated) that allow quantification of the value of existing data. For example, an evaluation may run one numerical weather forecast experiment based on current observations and compare it with a forecast where one type of data is removed from the experiment. The difference in the skill of the forecast between the two experiments indicates the impact of the data that was removed. This type of evaluation helps optimize the use of current observations in our existing modeling systems (i.e., enhanced data assimilation strategies, more realistic characterization of observations, and leveraging of existing observations not currently used). However, this methodology cannot be used to analyze the value of enhancing the current global observing system with additional observations that do not yet exist. To quantify the impact of proposed observing capabilities, observing system simulation experiments (OSSEs) are needed. OSSEs are modeling experiments used to perform an objective evaluation of the potential benefits of proposed instruments. These studies differ from OSEs in that all the observations (current and proposed) are simulated. Similar to OSEs, experiments are compared by including and then excluding a particular data type. A numerical weather prediction forecast which provides a complete representation of the assumed true state of the atmosphere is used to simulate all the observations and to quantify the accuracy of the predictions from the experiments.

References

1. National Oceanic and Atmospheric Administration Earth System Integration Board's Modeling Team (2022). NOAA Modeling Team Charter and Terms of Reference. Available at: https://vlab.noaa.gov/web/noaa-modeling-team/document-library

2. National Oceanic and Atmospheric Administration (2020). NOAA Artificial Intelligence Strategy. Available at: https://sciencecouncil.noaa.gov/wp-content/uploads/2023/04/2020-AI-Strategy.pdf

3. National Oceanic and Atmospheric Administration (2020). NOAA Cloud Strategy. Available at: https://sciencecouncil.noaa.gov/wp-content/uploads/2022/08/2020-Cloud-Strategy.pdf

4. National Oceanic and Atmospheric Administration Science Advisory Board (2021). A Report on Priorities for Weather Research (PWR). Available at: https://sab.noaa.gov/wp-content/uploads/2021/12/PWR-Report_Final_12-9-21.pdf

Acknowledgements

The NOAA Data Assimilation Strategy was led by Dr. Michael Morgan, Assistant Secretary of Commerce for Environmental Observations and Prediction (ASEOP). This strategy was directed by Dorothy Koch (OAR) and Brian Gross (NWS), Former Chair and Former Vice-Chair of the NOAA Modeling Team (NMT), and by Kathryn Shontz (NESDIS) and Curtis Alexander (OAR), current Chair and Vice-Chair of the NMT, under the Earth Systems Integration Board (ESIB). The NOAA Data Assimilation Strategy Tiger Team was chaired by Kevin Garrett (NWS) and co-chaired by Sergey Frolov (OAR). Significant contributions to the Strategy were provided by Curtis Alexander, Lidia Cucurull, Hyun-Sook Kim (OAR); Daryl Kleist and Vijay Tallapragada (NWS); Eric Bayler (NESDIS); James Thorson (NMFS); Jiangtao Xu and Alexander Kurapov (NOS). We would also like to thank Mridula Srinivasan (NMFS, and Special Advisor to the ASEOP) for providing valuable review and feedback of the Strategy, and Shira Francis (NWS) for providing feedback and logistical support to the Tiger Team. Finally we would like to thank the NMT members, and all ESIB teams for their review and feedback on the Strategy.

Appendices

Definitions

Artificial Intelligence: From NOAA's Artificial Intelligence Strategy, Artificial Intelligence (AI) 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 (ML) refers to mathematical models able to perform a specific task without using explicit instructions, instead relying on patterns and inference. Deep learning (DL) is a subset of ML that utilizes artificial neural networks capable of learning from unstructured or newly added data. The use of labeled training data can further improve the AI predictive capability through supervised ML.

Continuous Data Assimilation: The concept of running data assimilation in real-time, in order to assimilate observations as they become available, in contrast to traditional data assimilation systems which are executed on a predefined cadence, with observations covering a window around the analysis time (e.g. +/- 3 hours).

Data Assimilation: The process to optimally combine model and observational data to produce a best estimate of the state, for example, of the Earth system and its components.

Data Lake: A centralized repository that ingests, stores, and allows for processing of large volumes of data in its original form.

Earth System: Constructed from a comprehensive set of components including the atmosphere, land, ocean, cryosphere, and associated ecosystems, which interact through complex physical processes that both reflect the state of, and drive the evolution of, weather, water, and climate.

Photo Credits

Front Cover

- A Saildrone collects data beneath the Golden Gate Bridge alongside a NOAA research vessel. Image credit: NOAA, Courtesy of Saildrone Inc.
- One of NOAA's Unmanned Aerial Vehicles (UAV). Image credit: NOAA Photo Library.
- Rendering of GOES-U spacecraft with Earth and sun. Image credit: NOAA/Lockheed Martin.
- NOAA's supercomputer Dogwood, located in Manassas, Virginia (2023). Image credit: NOAA.

Page 3 - View of North America on Globe. Image credit: NOAA Weather Program Office.

Page 5 - NOAA-20 satellite captured this image of an expansive plume of dust from the Sahara Desert, also known as the Saharan Air Layer, traveling westward across the Atlantic Ocean. Image credit: NOAA Satellite Flickr.

Page 6 - NOAA's Satellite Operations Facility in Suitland, Maryland, during sunset. Image credit: NOAA NESDIS.

Page 7 - Corridor of computer servers. Image credit: NOAA (iStock).

Page 8 - Conceptualization of the GeoXO constellation. Image credit: NOAA NESDIS.

Page 9 - NOAA P-3 with brand new color scheme. Image credit: NOAA/OMAO/AOC | Lieutenant Kevin Doremus.

Page 11

- Dr. Anna Shlyaeva presenting her work at the 3rd JEDI Academy in Boulder, Colorado. The academy consists of both lecture and practical sessions. Image credit: UCAR/K. Shanahan.
- 3rd JEDI Academy participants working together on practicums in the afternoon in a computer room in NCAR Foothills Lab 2. Instructors stand in a circle discussing until someone has a question. Image credit: UCAR/K. Shanahan.
- Buoy deployment. Image credit: NOAA.
- Mussie Kebede was one of the students supported by NOAA's Educational Partnership Program with Minority Serving Institutions Cooperative Science Centers. He now works as a meteorologist for the National Weather Service, stationed in the Forecast Operations Branch of the Hydrometeorological Prediction Center. Image credit: NOAA/Mussie Kebede.