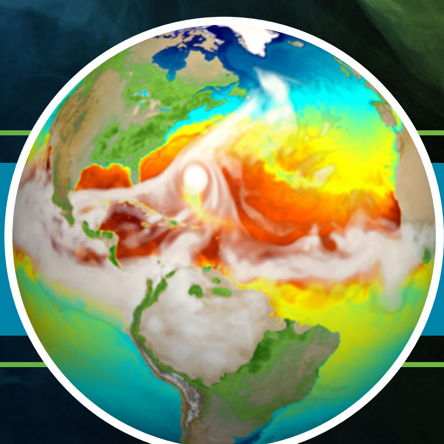
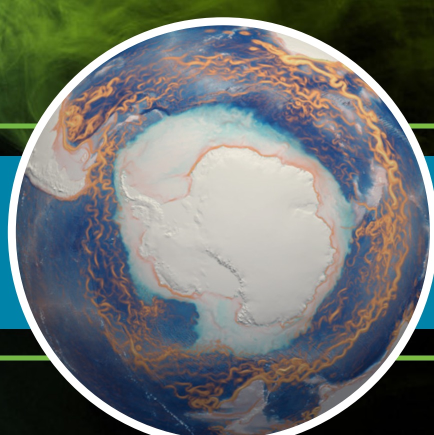
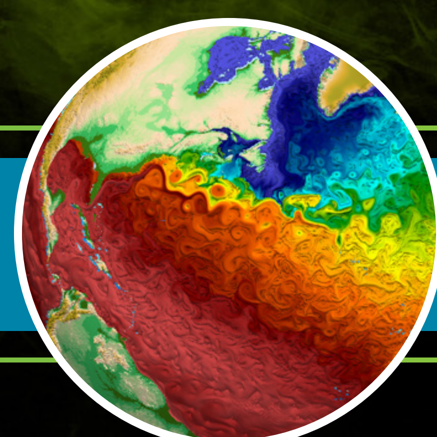
U.S. DEPARTMENT OF ENERGY | Office of Science

National Oceanic and Atmospheric Administration
U.S. Department of Commerce

Workshop on the Initialization of High-Resolution Earth System Models

APRIL 9 – 10, 2018

DOE/SC-0195 ▪ NOAA Technical Report OAR CPO-6 ▪ doi:10.25923/20NE-2B12

About the Energy Exascale Earth System Model (E3SM) outputs on the cover:

From left to right:

Left – In high-resolution (1/8 degree), E3SM shows the column of water vapor with a hurricane in the Atlantic.

Middle – E3SM shows in high-resolution the ocean (gold shows kinetic energy), sea-ice extent (white area around Antarctica), and circumpolar deep water (white currents meandering away from Antarctica).

Right – This E3SM high-resolution ocean modeling output shows ocean surface temperature.

Workshop on the Initialization of High-Resolution Earth System Models

Held April 9 – 10, 2018, Rockville, Maryland
Published October 2018

Convened by
U.S. Department of Energy
Office of Science
Office of Biological and Environmental Research
U.S. Department of Commerce
National Oceanic and Atmospheric Administration
Climate Program Office

Dorothy Koch
DOE Earth and Environmental System Modeling Development

Dan Barrie and Annarita Mariotti
NOAA Climate Program Office
Modeling, Analysis, Predictions and Projections

Workshop Chairs

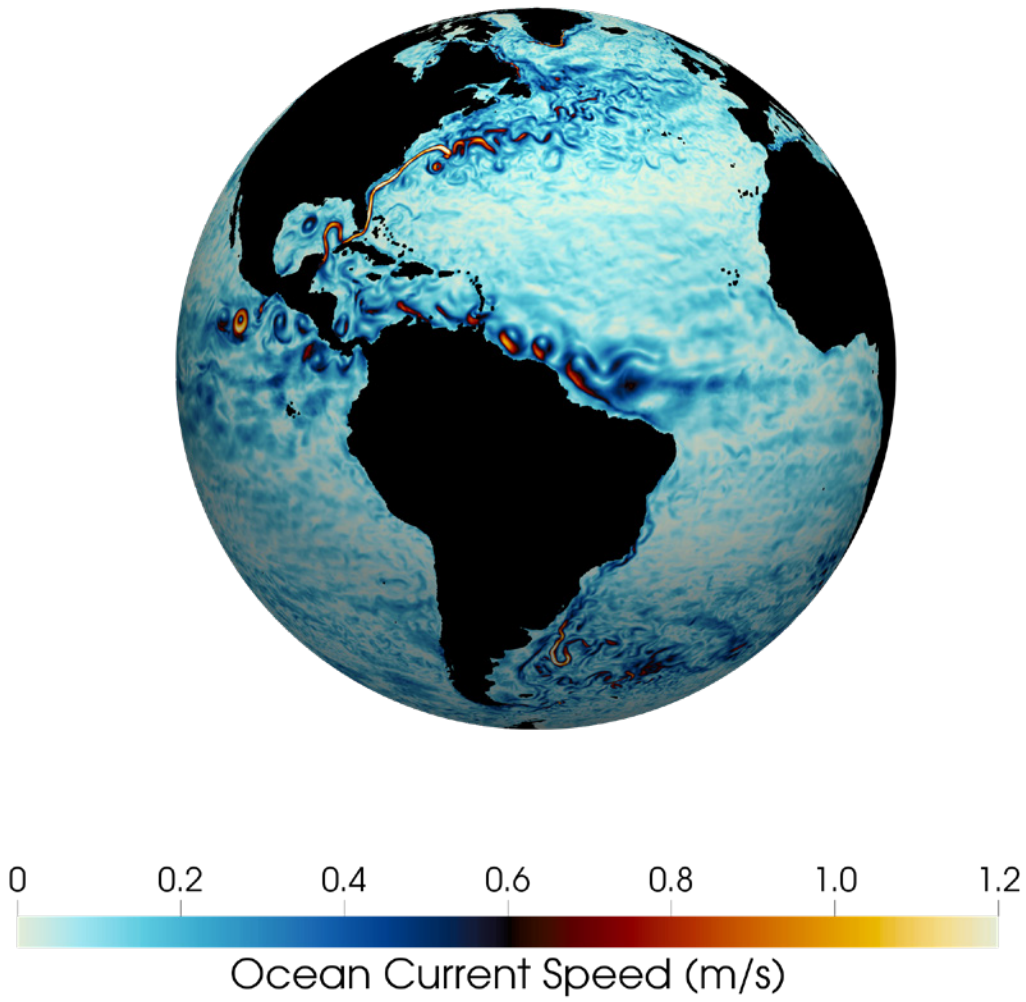
Steve Penny, University of Maryland, National Centers for Environmental Prediction
Todd Ringler, Los Alamos National Laboratory

Workshop Facilitators and Rapporteurs

Alistair Adcroft, Princeton University
Santha Akella, National Aeronautics and Space Administration
Craig Bishop, U.S. Naval Research Laboratory-Monterey
Gokhan Danabasoglu, National Center for Atmospheric Research
Ben Kirtman, University of Miami, Rosenstiel School of Marine and Atmospheric Science
Patrick Laloyaux, European Centre for Medium-range Weather Forecasts
Steve Penny, University of Maryland, National Centers for Environmental Prediction
Todd Ringler, Los Alamos National Laboratory
Matthew Mazloff, Scripps Institution of Oceanography
Luke Van Roekel, Los Alamos National Laboratory
Anthony Rosati, NOAA Geophysical Fluid Dynamics Laboratory
K Chad Sockwell, Florida State University
Shu Wu, University of Wisconsin

U.S. DEPARTMENT OF ENERGY | Office of Science

National Oceanic and Atmospheric Administration
U.S. Department of Commerce


0
0.2
0.4
0.6
0.8
1.0
1.2
Ocean Current Speed (m/s)

Executive Summary

From weather to climate timescales, weeks to decades, societal needs require detailed predictions of the atmosphere, ocean, and land-surface to facilitate informed planning. Central to this scientific endeavor is the ability to accurately initialize the earth system models (ESMs), since frequently the skill in forecasts, predictions, and projections depends on the quality of the initialization. However, initialization of high-resolution ESMs remains an outstanding challenge, particularly for the climate modeling community where initialization has received relatively less attention and where longer-term information is needed. This need, after a review of previous workshop outcomes, led the U.S. Department of Energy (DOE) and National Oceanic and Atmospheric Administration (NOAA) to host a workshop that focused exclusively on the opportunities that data assimilation (DA) and, specifically, coupled data assimilation (CDA) might provide to the ESM enterprise with particular attention toward initialization.

The workshop was held April 9-10, 2018, at the Rockville Hilton in Rockville, Maryland, with 41 participants. It was composed of seven sessions, each with a distinct topical theme that related to the broad challenge of initializing high-resolution ESMs. The first session was used to frame the workshop challenge. Two of the sessions (2 and 7) focused the current state of the art of CDA within laboratory and operational centers at both the short-term, subseasonal and long-term, climate time scales. Three of the sessions (3, 4, and 5) addressed new algorithmic approaches to CDA at high resolution. Session 6 focused on the computational science solutions for the management of large input/output streams required to carry out the assimilation procedure.

The total body of research surrounding initialization of high-resolution ESMs is not large enough to assert with confidence which research pathways will prove most fruitful. With this "research risk" in mind, CDA is a particularly good choice for focused research activities since CDA can be leveraged by the ESM community to address problems beyond initialization, such as studying the causes of model drift, remedying model biases, and optimizing the use of earth system observational platforms. As a result, seven research priorities were identified during the workshop:

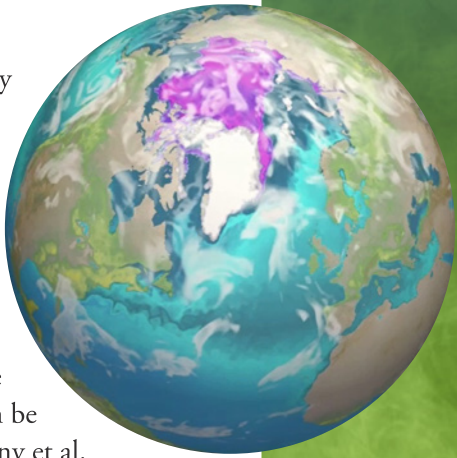
1. Continued foundational research on algorithmic approaches to CDA.
2. Continued exploration of the benefits of strongly coupled DA versus weakly coupled or individual-component DA.
3. Implementation of extensible CDA frameworks for use within high-resolution ESMs, such as the DOE Energy Exascale Earth System Model (E3SM) high-resolution configuration.
4. Computational, algorithmic, and middleware solutions to link forward models to CDA system in ESM grand-challenge configurations.
5. Development of testbeds and metrics to measure the fidelity of various initializing approaches.
6. Development of holistic CDA approaches/frameworks that support the diversity of CDA applications.
7. Growth of CDA capability and expertise in the context of E3SM.

U.S. DEPARTMENT OF ENERGY | Office of Science
National Oceanic and Atmospheric Adminstration
U.S. Department of Commerce

Contents

Introduction

From weather to climate timescales, societal needs require detailed predictions of the atmosphere, ocean, and land-surface to facilitate informed planning. Central to this scientific endeavor is the ability to accurately initialize the earth system models (ESMs), since frequently the skill in forecasts, predictions, and projections depends on the quality of the initialization. However, initialization of high-resolution ESMs remains an outstanding challenge, particularly for the climate modeling community where initialization has received relatively less attention and where longer-term information is needed. For this community, "high-resolution" currently implies an ocean resolution that permits the emergence of mesoscale eddies (usually ¼-degree or finer) and an atmosphere grid spacing of approximately 25 km or less. The community has yet to identify any single approach to high-resolution ESM initialization that could be considered a "best practice." For the foreseeable future, the procedure to initialize high-resolution ESMs will depend on the scientific questions to be addressed with the simulation. International modeling centers use high-resolution ESMs for diverse scientific problems, ranging from subseasonal to decadal predictability to measuring the forced-transient response to greenhouse gas emissions. It remains unclear what aspects of the earth system become more predictable as model resolution increases and what improvements in the representation of physical processes are necessary for this additional predictability to be realized. While it has long been recognized that simulations of the "forward model" are grand-challenge-scale computations, less attention has been directed toward the equally large grand challenge of robust initialization of the high-resolution ESMs.



Even at longer time scales, estimating the state of the earth system has the opportunity to play a critical role in initializing earth system models because much of the system predictability is contained in the phase of dominant modes of variability, such as El Niño-Southern Oscillation (ENSO), Pacific Decadal Oscillation (PDO), and Atlantic Meridional Overturning Circulation (AMOC) which often span multiple components of the earth system. But to obtain an optimal estimate of the state vector spanning the entire ESM, some form of coupled data assimilation (CDA) is needed. CDA provides, among other services, the ability to incorporate observational data measured across the entire earth system onto a single, fully gridded state estimate by using the coupled ESM dynamics. The resulting state estimate can then be used to initialize forecasts made with this coupled model. Previous workshops (Penny et al. 2017, Penny and Hamill 2017, Anderson et al. 2017) also identified CDA as an important technique for the initialization of high-resolution ESMs, though it was found that the complexity and robustness of the CDA procedure vary between implementations. At high resolutions, the implementation of CDA has open scientific, algorithmic, and computational questions that remain to be answered.

Our scoping of previous workshop outcomes and the sentiment of likely participants led the U.S. Department of Energy (DOE) and National Oceanic and Atmospheric Administration (NOAA) to host a workshop that focused exclusively on the opportunities that data assimilation and, specifically, CDA might provide to the earth system modeling enterprise with particular attention toward initialization.

Summary of Workshop

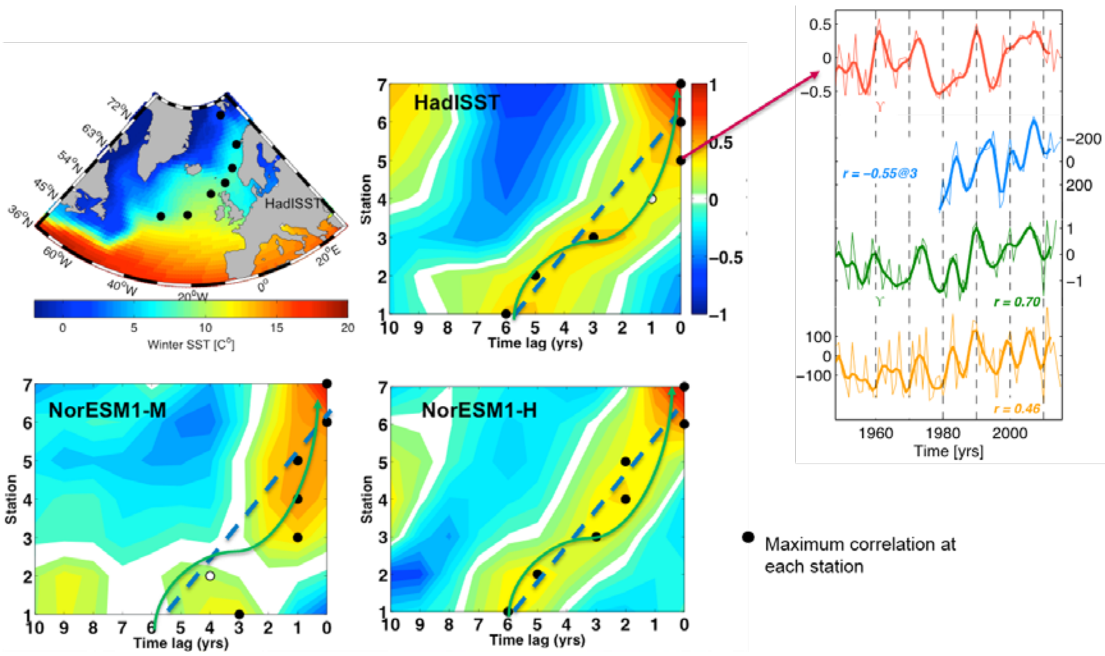
The workshop was held April 9-10, 2018 at the Rockville Hilton in Rockville, Maryland, with 41 participants (see Appendix A). The agenda is listed in Appendix B. The workshop was composed of seven sessions, each with a distinct topical theme that related to the broad challenge of initializing high-resolution ESMs. The first session was used to frame the workshop challenge. Two of the sessions (2 and 7) focused the current state of the art of CDA within laboratory and operational centers at both the short-term, subseasonal and long-term, climate time scales. Three of the sessions (3, 4, and 5) addressed new algorithmic approaches to CDA at high resolution. Finally, Session 6 focused on the computational science solutions for the management of large input/output (I/O) streams required to carry out the assimilation procedure. Each session started with two or three presentations and was followed by a facilitated discussion based, in part, on integrating questions presented in the agenda. Each session of the workshop is summarized immediately below.

1

Session 1: Challenges of Initializing High-Resolution ESMs with a Small Ensemble

The session focused on methods currently used to initialize high-resolution coupled climate simulations. Long-range predictions often gain skill by resorting to an ensemble prediction approach. Due to the computational expense, options are needed where ensemble size is restricted to either a few or in some cases a single member. Definitions of high resolution are to some degree linked to computational resources, but here are defined as 0.1-degree resolution in at least one large basin in the ocean and an atmospheric model resolution that is near the hydrostatic limit (approximately 0.25°). The DOE ESM, called Energy Exascale Earth System Model (E3SM), presently utilizes a modified HighResMIP protocol (Haarsma et al. 2016), which uses a single brief control run for model initialization. This approach is a slightly modified and truncated version of the typical (multi-century) climate model spin-up approach: due to the expense of the high-resolution simulation, the coupled system is forced with perpetual 1950s conditions for only 50 simulated years before initiating the late-20th-century transient simulation. While limits on computational resources require such an approach, the effectiveness of the initialization method remains unclear, at least for the E3SM.

Experience with the Norwegian Climate Prediction Model (NorCPM) demonstrates improved fidelity in the representation of the North Atlantic system when moving from a 1.0-degree ocean-ice system to a 0.25-degree ocean-ice system (Langehaug et al. 2018). Following Hewitt et al. (2017), the improved fidelity is due to the better representation of topographic steering, a resolution permitting mesoscale eddies, and the ability to support sharper ocean temperature and salinity fronts. But at 0.25-degree resolution, the NorCPM cannot afford an ensemble size sufficiently large to produce the dynamically evolving forecast error covariance field that is the foundation of the Ensemble Kalman Filter (EnKF) data assimilation method. Since the EnKF approach approximates the covariance matrix by analyzing the relationship of the ensemble members, the accuracy of the assimilation procedure depends on having a sufficiently large ensemble size (Evensen 1994). To mitigate the impact of using a small number of ensemble members, NorCPM is testing a hybrid data assimilation method following Hamill and Snyder (2000) that blends a static error covariance matrix, estimated using an ensemble drawn from an



Assimilation provides improved skill of the observed ocean anomaly from the sub-polar gyre into the high-resolution Norwegian Earth System Model (NorESM1-H) (Årthun et al. 2017 and Langehaug et al. 2018).

existing high-resolution model integration, with a dynamical (flow-dependent) error covariance matrix estimated utilizing an ensemble of low-resolution online model integrations.

The notion of using multiple scales in the data assimilation algorithm was a theme that the workshop returned to frequently. In most Numerical Weather Prediction (NWP) applications, the range of resolved scales is relatively limited. For this reason, most data assimilation (DA) solution methods do not explicitly account for the presence of multiple spatiotemporal scales. However, for 'seamless prediction,' in which the same system is used for both weather and climate timescales, the ability to isolate the long-timescale climate modes provides an opportunity for extended long-range prediction by filtering out small-scale variability that acts as noise in the system. See the Session 4 summary below for more details.

Session 2: Climate Modeling Centers' Roadmaps for Coupled Data Assimilation

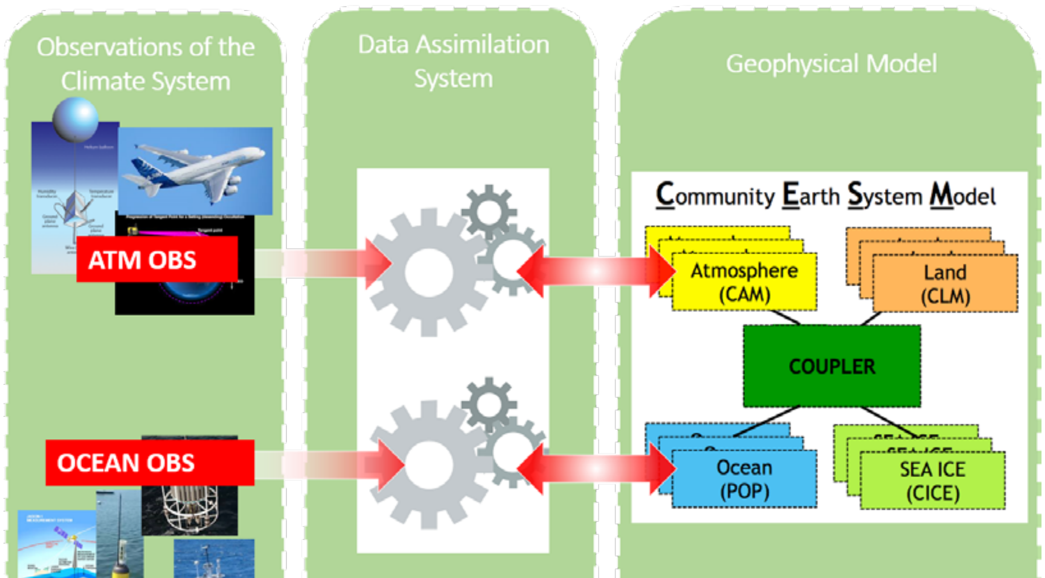
With the challenge to be addressed by the workshop framed in Session 1, the next step is to better understand the strategies and roadmaps from two modeling centers: the National Center for Atmospheric Research (NCAR) and the Geophysical Fluid Dynamics Laboratory (GFDL).

NCAR uses two techniques for ESM initialization. The first is a 300-year pre-industrial, hindcast spin-up, followed by an integration from 1948-2007 using CORE-II forcing (Large and Yeager 2004). Decadal predictions are then initialized from staggered states. The second approach applies data assimilation with a 48-member EnKF implemented in the Data Assimilation Research Testbed (DART, https://www.image.ucar.edu/DAReS/DART/) using a weakly coupled data assimilation (WCDA) scheme in which the atmosphere and ocean are assimilated independently. Both the hindcast spin-up and data assimilation approaches are compared using sea surface temperature (SST) decadal prediction skill and the magnitude of the AMOC. Of particular concern for the initialization of the ocean is that multiple ocean reanalyses produced by the international community have more variability of the AMOC reanalysis than found in forward model runs. This finding indicates large systematic errors in the forward model runs and a high level of uncertainty in the accuracy of ocean reanalyses in the Atlantic basin.

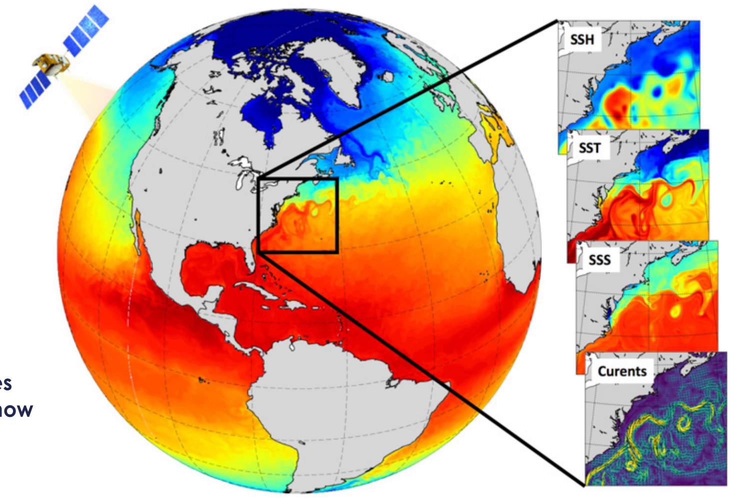
A reanalysis experiment was performed using a WCDA atmosphere and ocean system covering 1970-1982. The data assimilation was implemented with DART, using a 30-member EnKF. However, even at low resolutions, this analysis was too expensive to extend to the present day. NCAR is currently testing 84 members along with an improved DART system, and the initial results are promising. NCAR is now proposing to run a 2005-2016 ocean hindcast, with strongly coupled data assimilation (SCDA) in which the cross-domain error covariance is used to permit each observation to have an impact on the entire coupled system at the analysis time.

GFDL emphasized the importance of all three components in the CDA system: the data assimilation scheme, observing network, and the model. Weaknesses in any one of these primary ingredients can block advancements. GFDL currently uses the Ensemble Coupled Data Assimilation (ECDA; Chang et al. 2013) system. They have found that coupled modeling is vital to prevent initialization shocks and adjustments that can degrade the prediction skill of ENSO. GFDL's decadal predictions show skill in the NINO3 variability and Arctic sea ice extent. The simulated AMOC will likely yield some decadal predictability.

Dramatic improvements in the observing of the global ocean have occurred due to the deployment of the Argo float network (Roemmich et al. 2009) measuring in-situ temperature and salinity profiles starting in the early 2000s and reaching its targeted coverage around 2005. The Argo float network allows for significantly more accurate estimates of the upper 2000 m ocean state. Furthermore, CDA has the potential to leverage much of this ocean-state information to improve state estimates of



Weakly coupled data analysis, which is cutting edge for climate models such as the Community Earth System Model (CESM), assimilated data into the ocean and the atmosphere separately.



An active area of research is on how best to initialize high-resolution features like ocean-eddies, what data to use, how frequently, etc.

other components of the earth system. The opportunities for observations in one component to influence and improve state estimates in other components extends well beyond this Argo example. For instance, Laloyaux et al. (2016) showed hat atmospheric scatterometer observations can improve ocean temperatures in tropical cyclones.

Overall, both the NCAR and GDFL roadmaps point toward the use of assimilation methods that are increasingly coupled in nature, with the strength of the system coupling and the algorithmic approaches still very much a topic of basic research.

3 Session 3: Approaches for the Initialization of Eddying Ocean Configurations

Initializing high-resolution models with an eddying ocean configuration has been shown to benefit forecasts and climate predictions (Kirtman et al. 2012). Some of the motivations for using eddying ocean initializations for high-resolution models are the following: (1) high-resolution initializations have shown improvements in rainfall along SST fronts, (2) the forcing on the atmosphere by the ocean is seen to be stronger in eddy-rich regions, and (3) the presence of eddies affects forecast quality. There is an unresolved debate as to whether using an initialization that obtains the eddy's imprint on the large scales, as opposed to initializing specific eddies, can improve the skill of predictions at various timescales. In particular, it was found that the negative skill in strongly eddying, western boundary current regions was brought closer to zero when using an eddying configuration (Kirtman et al. 2017). This reflects the fact that the climatological structure of the boundary current was seen to improve with high-resolution initializations.

Although advantages can be found in an eddying ocean initialization, the high computational cost of these model resolutions still poses a problem. At eddying resolutions the variability in the SST front can force changes in the position of atmospheric jet stream that, in turn, lead to changes in rainfall patterns. A concern is what minimum resolution is required, for example, in order to adequately resolve these differences with relation to the Gulf Stream. Within the ocean, the strong vertical coherence of ocean eddies provides opportunities to improve predictability down to at least the depth of the thermocline.

Other important discussion questions include what aspects of ocean mesoscale features are essential for prediction of various physical phenomena and at what timescales do resolving individual eddies remain imperative. It is agreed that resolving the small scales will have a positive impact on short-term, high-resolution initializations and forecasts. However, the magnitude of these benefits must be weighed against available computational resources. For seasonal to decadal prediction, accurate initialization of the large scales is essential; thus a cost-benefit analysis to achieve this goal is required. It seems that there must be an investigation into the benefits of resolving certain aspects of ocean mesoscale in the initialization process, and to what degree this improves predictions at all scales. Simultaneously, an initialization strategy must be developed to mitigate costs in order to optimize the cost-benefit relationship.

There is some belief that for longer timescale prediction it is not essential to initialize specific eddies, but rather to obtain a large-scale imprint of climate modes through an appropriate statistical representation of the eddies. However, it is clear that low-resolution models are not merely coarsened versions of the high-resolution models due to the introduction of significant systematic errors that compound into long-term model biases. For example, low-resolution models typically have poor representation of coastal phenomena. Skill is not claimed by high-resolution models in eddy-rich regions, but negative skill is brought closer to zero. Although it is likely that the atmosphere is affected directly by a higher-resolution ocean model, it can be difficult to attribute the cause to the ocean rather than to other atmospheric phenomena.

Given the expense of the large modeling systems, it may be that collaborative comparisons of various DA methods are required. At that point, it may also be necessary to develop standardized metrics. Even if eddies are not a primary concern, DA of an eddying ocean can improve other aspects of the simulation such as the boundary currents.

Session 4: Exploration of Data Assimilation Methods

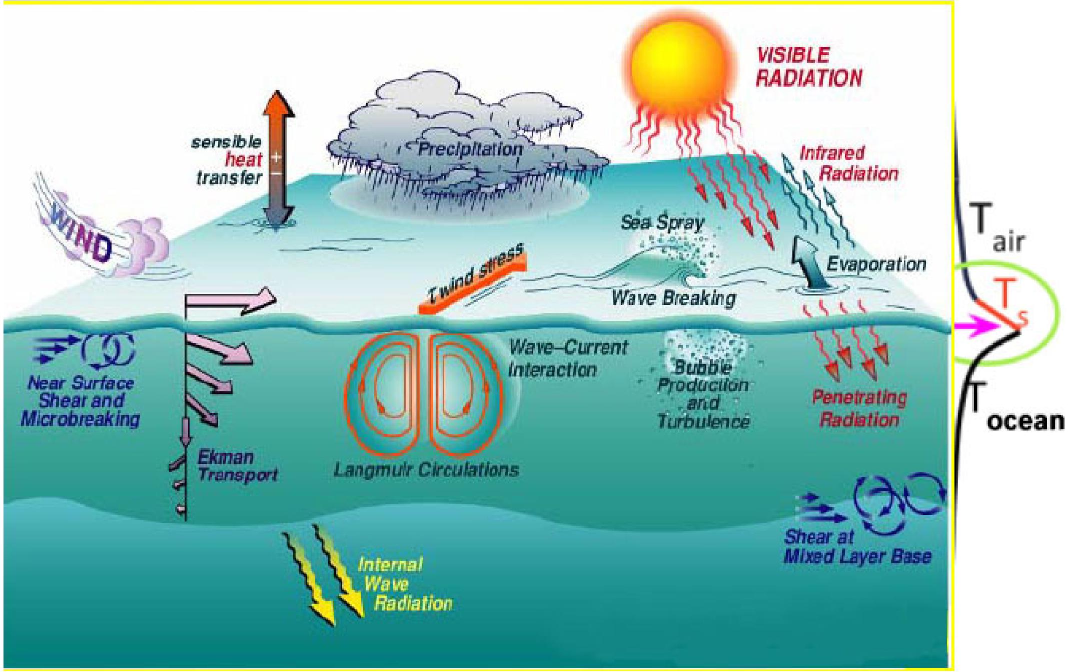
Coupled data assimilation presents new challenges that have not yet been resolved by the DA community. At present, the primary challenges being addressed by the major modeling centers are technical, such as how to extend single-component DA methods to operate successfully with coupled ESM configurations. The leading DA methods applied in the atmosphere and ocean are variational methods (e.g., 3D-Var or 4D-Var) and ensemble methods (e.g., the EnKF). Variational methods typically require a static background error covariance matrix to be defined—in a SCDA application this implies that a full climatological background error covariance must be defined for the full coupled ESM. From a mathematical perspective, these matrices are ill conditioned due to the presence of disparate spatiotemporal scales. For 4D-Var, which requires a linearized model and adjoint, the prospect of developing these for an ESM application is a significant challenge. Alternatively, the relative simplicity of the ensemble methods has allowed for a more straightforward transition to coupled model environments. However, the presence of disparate timescales requires larger ensembles and more frequent analyses, so the computational costs of ensemble methods at high resolution can be prohibitive.

Two methods were presented to manage challenges arising from CDA applied to ESMs. A temporal averaging approach was applied to the background error covariance matrix to separate scales in the assimilation step. The use of climatological information is critical for high-resolution cases in which ensemble solutions are not viable. As an alternative to developing a software-based tangent linear model (TLM) and adjoint, an ensemble-based approximation of the TLM has been proposed.

Session 5: Data Assimilation for Shorter Time-Scale Prediction

The focus of this session is primarily on operational prediction and knowledge that can be transferred to the initialization of ESMs. A highlighted application is the assimilation of satellite radiance measurements to estimate SST. Experiences from the National Centers for Environmental Prediction (NCEP) global forecast system demonstrate that both the analysis and forecast errors are affected by the amount of observational data, details such as model grid resolution and choice of parameterizations, and data assimilation schemes. These experiences gained from the NWP community could be beneficial for the broader climate sciences community—from S2S to beyond decadal timescales.

Observational data includes in-situ observations such as radiosondes and Argo profiling floats, as well as satellite measurements covering the atmosphere and ocean surface. Using such observations to initialize an ESM requires a significant level of understanding of the various observational sources, including their errors and biases. The ESMs also have biases, for example, critical physical processes such as the diurnal cycle of SST have not been widely incorporated. The SST biases in ESMs often appear relatively quickly after initialization, e.g., in typically less than six months of a free integration. The National Aeronautics and Space Administration (NASA) Global Modeling and Assimilation Office (GMAO) is currently developing methods to assimilate satellite radiances directly.



Observational data includes in-situ observations such as radiosondes and Argo profiling floats as well as satellite measurements covering the atmosphere and ocean surface. (Edson et al. 2007, doi:10.1175/BAMS-88-3-341)

One project that has taken advantage of using observations to identify the sensitivity of model parameters, help to tune those parameters, and reduce ESM biases is the Cloud-Associated Parameterizations Testbed (CAPT). However, in CAPT experiments, different initialization schemes lead to different sensitivities, implying a proper data assimilation scheme for initialization is still vital. Once observations, models, and DA are implemented together, differentiating the impact of each component becomes more difficult.

Session 6: Middleware Solutions to Remove the O/I Barriers in Data Assimilation

Since the majority of forward models and assimilation models communicate through data written to disk, I/O has the potential to be the bottleneck controlling throughput. As model resolution increases and more frequent assimilation increments are desired, the I/O barrier will become a more pressing concern. This will be increasingly likely in the years ahead as the I/O system transfer rates remain relatively flat as compared to overall computer system processing power. Currently, there is disagreement as to whether or not today's I/O systems produce a bottleneck for CDA. This is complicated by the fact that at operational centers the I/O systems are typically handled by a dedicated team that works separately from the scientists.

Regardless, there are ways in which bottlenecks arising from I/O can be mitigated. One such technique is "in situ" or "online" data assimilation. This strategy avoids reading and writing to disk by distributing the ensemble members and posterior vector among processes as opposed to writing a restart file at the end of the DA procedure. Although additional memory is required, efficiency is ensured due to sufficient scalability of the method. Significant speedups can be obtained by avoiding reading and writing to hard memory. This method has greater potential in high-resolution simulations where single restart files are hundreds of Gb in size. The benefits of this method come at the cost of a more complicated programming model.

A potentially flexible and efficient approach is an online ECDA hybrid system that combines an online EnKF with an Ensemble Optimal Interpolation (EnOI) in the same framework. EnOI is applied to a single (or a few members of) a high-resolution model simulation capturing the static part of the covariance and the EnKF on an ensemble of low-resolution model capturing the temporal evolution of the covariance. This system should be flexible in its application to a high-resolution model under a given computational constraint. This system would also make full use of both versions of e.g., E3SM in high and low resolutions.

The trade-off between performance and programming effort, especially for an online system, motivates the use of high-level software interfaces. Also known as middleware, these packages lay the groundwork for online DA, thereby removing much of the coding effort. Two such middleware packages were discussed for this purpose.

The first middleware, Decaf, provides a workflow for creating interfacing layers that lower the barriers to adoption by external applications (Dreher and Peterka 2017). Decaf has seen success in other disciplines and is part of the Coupling Approaches for Next-Generation Architectures (CANGA) effort supported by DOE's Scientific Discovery through Advanced Computing (SciDAC). The second middleware approach, Adaptable IO System (ADIOS), has been used with great success in improving high-performance computing (HPC) I/O applications on large

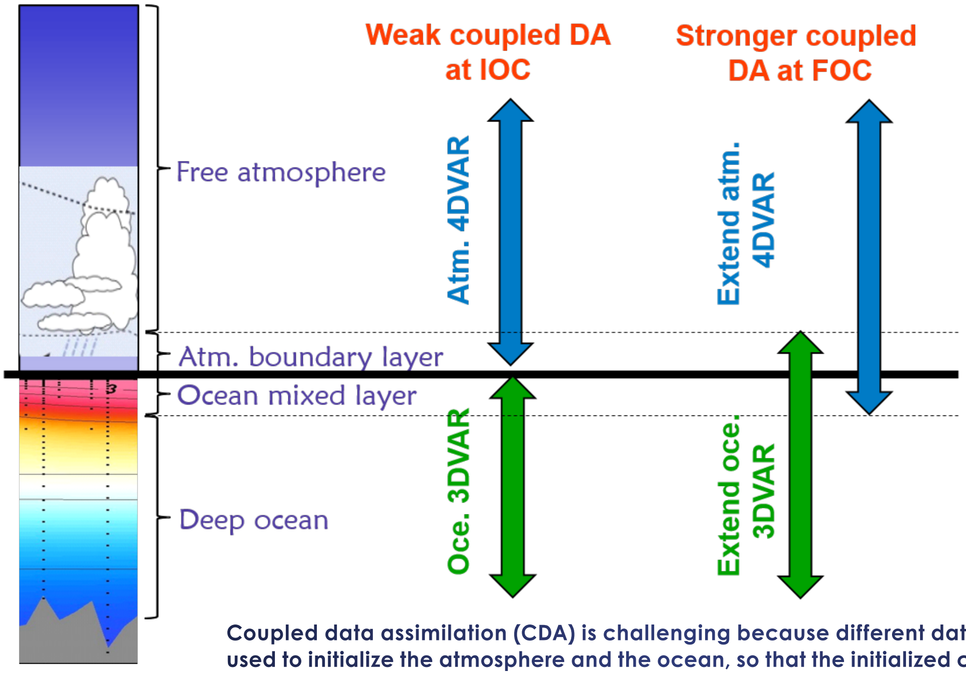
HPC machines such as Titan (Lofstead et al. 2008). This software is also being used in extremely data-heavy applications such as particle collisions and molecular dynamics. ADIOS provides freedom in the construction of data files and online analysis.

Questions about portability of the packages still exist since the DA community uses a wide variety of architectures. It seems that stronger communication between the I/O package developers and the dedicated I/O support teams for each DA effort would be beneficial in addressing the possible use of these packages and how I/O transfer rate limitations will be addressed in the future. The fundamental questions to be considered are: what will the future for I/O look like as resolution increases, and will I/O be a bottleneck in the future? Does online DA produce a large enough benefit to motivate its use? What advantages do Decaf and ADIOS provide to entice the adoption if these middleware interfaces? More specifically, how much work will be required (or saved) to make the packages run in an optimal manner on a particular machine?

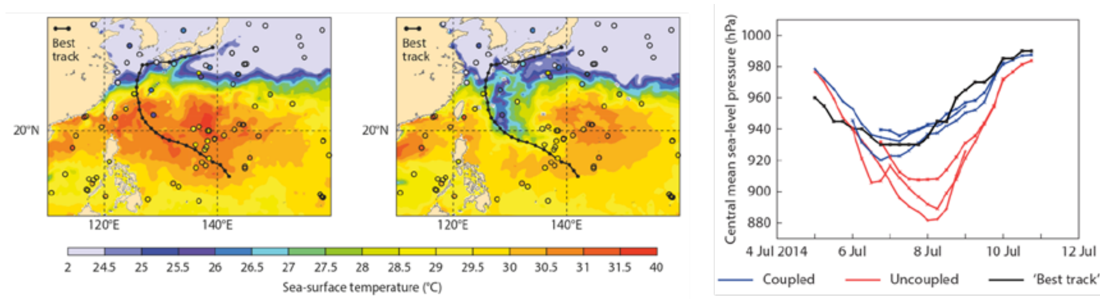
Session 7: Data Assimilation Research in Support of Operational Centers

The last session of the workshop focused on data assimilation research in support of operational centers. The discussion covered community-developed DA software to accelerate implementation, inter-agency coordination and collaboration on prediction capability in need of DA, and a perspective of coupled DA from the vantage point of operational European Centre for Medium-Range Weather Forecasts (ECMWF) needs and objectives.

On the topic of community development, the group discussed the Joint Effort for Data assimilation Integration (JEDI). JEDI is an interagency plan for the development of a unified community DA system for research and operations. The intention is that the software will be sufficiently modular and portable to allow for widespread use, application, and development across the DA community.



Coupled data assimilation (CDA) is challenging because different data are typically used to initialize the atmosphere and the ocean, so that the initialized coupled ocean and atmosphere is subject to coupling "shock." Next-generation systems are under development to systematically improve initialization of the coupled system.



Increased skill for modeling tropical cyclone Neoguri is demonstrated, including better sea-level pressure, for a model that includes ocean as well as atmosphere (K Mogensen et al. 2012).

The intention is also to use community development activities, such as JEDI, to promote collaboration among different US national modeling systems. It is an evolving system which, at present, supports a few models. As users' contributions are integrated, other models/ components are added and revised.

One effort to coordinate research on earth system predictability is the inter-agency Earth System Predictability Capability (ESPC) initiative. At the national level, the ESPC vision is to establish a global physical earth system analysis and prediction system to provide seamless predictions covering hours to decadal timescales including the atmosphere, ocean, land, cryosphere, and space. Such an effort requires state-of-the-art data assimilation to be successful. ESPC is instantiated with various US agencies, including the Navy. Over the next five years, the Navy's ESPC (NESPC) will evolve from a collection of component-centric (e.g., atmosphere, ocean, and sea-ice) prediction models into a fully coupled ESM. As a result, the associated DA method will also evolve from its present form, a weakly coupled DA using 4DVar and 3DVar in the atmosphere and ocean, respectively, toward a strongly coupled DA approach. The strongly coupled DA algorithm (Frolov et al. 2016) will extend the atmosphere's analysis to include the ocean boundary layer and the ocean's analysis to include the atmospheric boundary layer, thereby coupling the atmosphere and ocean data assimilation procedure through an overlapping interface solver.

The perspective provided by P. Laloyaux from ECMWF was in many ways consistent with NESPC's roadmap, in the sense that while stronger coupling is needed in the assimilation procedure, we do not require a direct-solve of the entire coupled system. One compelling example to demonstrate the importance of developing stronger atmosphere-ocean data assimilation is tropical cyclone-ocean SST interaction. ECMWF has clearly shown that incorporating cold SST wakes behind tropical cyclones improves forecast skill by reducing errors in sea-level pressure. The cold wake inhibits the growth of the cyclone and, in turn, reduces the propensity of the forecast to predict excessively low sea-level pressure at the cyclone core. While the atmosphere-ocean coupling is critical to more accurate tropical cyclone prediction, the required "tightness" of the coupling is still a matter of research.

Summary of Data Assimilation Methods in Use Today, by Whom, and for What Purpose

A variety of data assimilation methods are currently used to initialize ESMs. The approaches used typically differ between the climate modeling community and the operational prediction centers. Operational prediction centers have generally been the leaders in pursuing a data assimilation strategy for initializing ESMs, due in large part to their longstanding success in using data assimilation to initialize numerical weather prediction models. A need to extend forecast skill beyond the 2-week barrier has led to a search for longer timescale signals in the coupled earth system that can be predictable with the appropriate initialization. The data assimilation methods used in practice range from relatively simple statistical interpolation techniques to quite sophisticated hybrid methods. This spectrum is characterized by wide ranges in algorithmic and coding complexity, and computational cost.

Optimal interpolation (OI) is a statistical weighted averaging method that combines model background information with observational data. An alternative scheme uses the calculus of variations to find a minimum of some cost function that is typically composed of terms associated with the model background and observations at a given time (called 3D-Var), but may also include other terms to enforce physical balance constraints. These methods are still commonly used for ocean applications—OI, for example, by the Simple Ocean Data Assimilation (SODA) system (Carton and Giese 2008) and 3D-Var NEMOVAR by ECMWF (Mogensen et al. 2012) and the Global Ocean Data Assimilation System (GODAS) at NCEP (Penny et al. 2015).

As a more sophisticated variational method that includes the temporal information of the observations, 4D-Var uses the climatological structure functions described above but uses a linearized version of the model to propagate misfits to the observations into the appropriate corrections at the model initialization time. An alternative formulation called 'weak constraint' 4D-Var also allows for some uncertainty in the model during the optimization process. The 4D-Var methods are used for atmospheric applications by operational centers such as ECMWF, the UK MetOffice, and in some applications by the US Navy.

The next advancement in the design of structure functions that allowed broader spatial impact of observations was to identify that a time-dependent structure could be estimated with Monte Carlo methods. Use of real-time ensemble forecasts or historical databases of ensemble information allows for the estimation of the 'errors of the day' and the climatological errors, respectively.

The Ensemble Kalman Filter (EnKF) is perhaps the best-known Monte Carlo type method used in the earth sciences. The primary limitations of the EnKF are the significant computational costs in running multiple simultaneous instances of the full nonlinear model and its sensitivity to systematic model errors. To address computational costs, DA methods, such as the EnOI, have been used to replace a real-time running ensemble with a database of precomputed ensemble members. The EnOI approach has been used by the Australian Bureau of Meteorology for ocean applications.

A variety of new hybrid methods were developed to take advantage of the strengths of climatological and time-dependent solution methods, as well as the strengths of direct and iterative solution methods. The first hybrid methods focused on the construction of a combined climatological and dynamic estimation of the model background structure functions to upgrade the solutions computed via variational methods. The current hybrid approach at NOAA/NCEP uses a 3D-Var with a model run at higher resolution with ensemble information provided via a second EnKF data assimilation system using a model run at lower resolution.

The most straightforward ESM initialization strategy may be the direct or 'brute force' insertion (via interpolation to the target model grid) of a more sophisticated reanalysis product produced by an operational center. However, it should be understood that any use of such a product will necessarily contain biases exhibited in the numerical models used to construct the reanalysis. The reanalysis product should not be mistakenly characterized as synonymous with observation.

Operational centers started moving toward CDA upon facing the seasonal and climate prediction problems. Early pioneering efforts were the Climate Forecast System at NCEP and the ECDA system at GFDL. More recently, CDA has been developed at ECMWF and implemented in a 20th-century reanalysis. Today, both ECMWF and NCEP are moving toward coupled forecasts at multiple scales, which will require initialization from CDA.

In this workshop, we have seen a variety of methods implemented for the initialization of ESMs. These include: insertion, 4D-Var (NPS, ECMWF), 4D-Var using the local ensemble tangent linear model (NPS), EnKF-offline (DART), EnKF-online (GFDL, Fast Ocean Atmosphere Model [FOAM]), multiscale EnOI (Qingdao National Laboratory for Marine Science and Technology [QNLM]), and multigrid DA (Massachusetts Institute of Technology General Circulation Model [MITGCM]).

Scientific Opportunities and Research Priorities

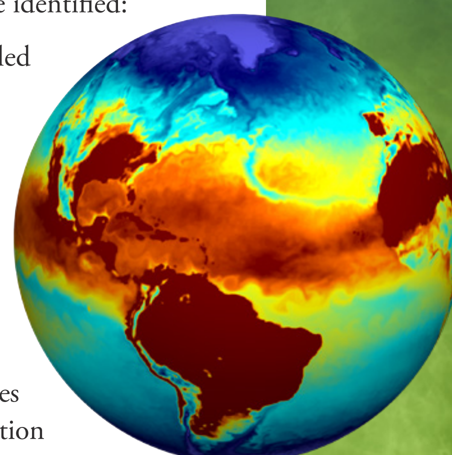
As highlighted in the introduction, how to initialization ESMs remains a modeling grand challenge, especially at high resolution. Many fundamental issues of ESM initialization remain unresolved, such as the relative importance of system equilibration and accuracy in determining the initial model state, and how this trade-off varies with simulation timescale. The total body of research surrounding initialization of high-resolution ESMs is not large enough to assert with confidence which research pathways will prove most fruitful. With this "research risk" in mind, CDA is a particularly good choice for focused research activities since CDA can be leveraged by the ESM community to address problems beyond initialization, such as studying the causes of model drift, remedying model biases, and optimizing the use of earth system observational platforms.

Since many of the long-standing biases in ESM begin to form within six months of the start of integrations initialized using data assimilation, CDA offers a new approach to identifying the causes of model drift (Zadra et al. 2017). CDA provides a way to understand the emergence of bias by analyzing the processes that cause a model to drift rapidly away from an observed quasi-balanced state. Closely related to the study of bias emergence is evaluating the impact of improved representation of physical processes. By assimilating new observational data that improves the representation of some physical process, e.g., tropical convection, we can quantitatively measure the importance of specific physical processes, both locally and remotely, on the fidelity of the simulation (Slivinski et al. 2018, submitted).

At longer research time horizons, growing the ESM capability to include CDA is viewed as an essential step in the maturity of the climate modeling activity. As the time and space boundaries between weather prediction and climate modeling continue to blur, the communities are appropriately becoming less distinct. On the ESM side of this relationship, adoption of CDA and its associated capabilities is a critical step in merging with the weather prediction community. The benefit bringing these communities closer is the potential for seamless modeling and prediction across the entire time-space continuum that encompasses weather prediction and climate modeling. Adoption of CDA by the ESM community will also allow for stronger collaboration with the observational communities. While observational data plays a vital role in the ESM community, the vast majority of observational data is used in the post-processing and evaluation of ESM simulations. This after-the-fact application of observational data is not conducive to building a healthy relationship between the communities. Since, by design, CDA brings observational and modeling data sets into a unified setting, it offers the opportunity to create a more vibrant and equitable relationship between these communities. One example of this is the climate observing system of the future as envisioned by Weatherhead et al. (2018). The methodology for designing this observing system depends on the ability to conduct Climate Observing System Simulation Experiments (COSSE) that, in turn, depends on ESMs having the ability to assimilation observational datastreams.

With these near- and long-term opportunities in mind, seven research priorities are identified:

1. Continued foundational research on algorithmic approaches to CDA: Coupled data assimilation is, and will remain for the foreseeable future, an important topic of basic research. By definition, CDA is composed of a collection of dynamically connected systems. The research priorities should include both methods for “component-wise” assimilation—such as the atmosphere and ocean—as well as methods for incorporating system coupling into the assimilation process. Concerning core algorithmic approaches to assimilation, leading methods such as 4D-Var, the EnKF, and their respective variants are all viewed as competitive paths forward. Within the ESM community, the burden of maintaining the TLM and adjoint software has inhibited adoption of variational methods. Research priorities should include the development of easy-to-adopt algorithms for on-the-fly construction of an approximate TLM and adjoint. Within the context of the EnKF, research priorities should include methods to minimize the number of required ensemble members, blending of static and dynamic error covariance information, and exploitation of multiple resolutions to inform the assimilation procedure at the highest resolution while mitigating computational costs.



2. Continued exploration of the benefits of strongly coupled DA versus weakly coupled or individual-component DA: While, in theory, the earth system could be assimilated as one very large, multi-scale, multi-physics system, the approach to CDA has been more fragmented in current applications—due in large part to the historically separated nature of the model components and model development, but also because of complexities that arise because of the presence of multiple

U.S. DEPARTMENT OF ENERGY Office of Science

National Oceanic and Atmospheric Adminstration U.S. Department of Commerce spatiotemporal scales. The impact of coupling strength (i.e., the degree of representation of the cross-domain error information) used to propagate the impact and influence of observational data from one system into other systems has to be more fully characterized.

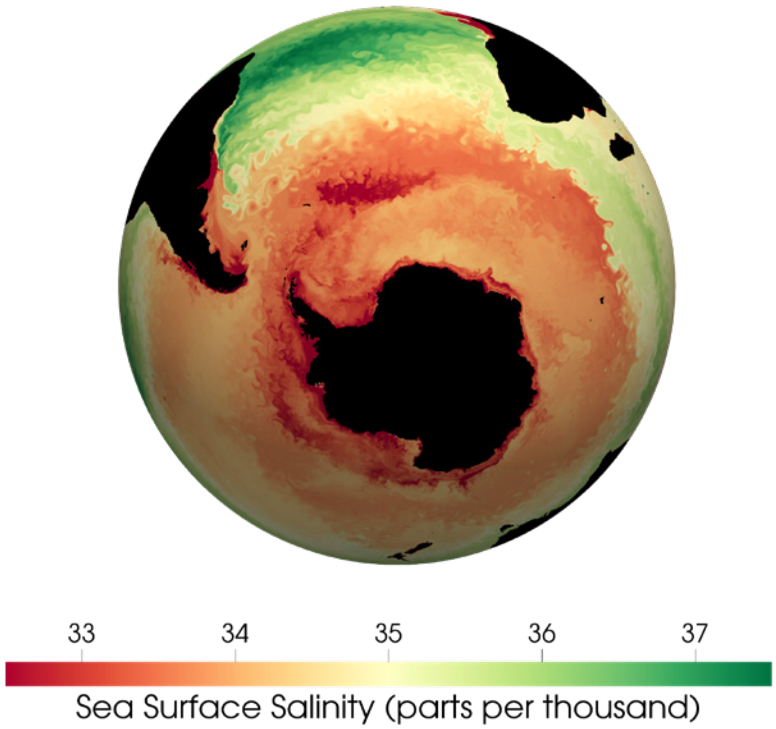
3. Implementation of extensible CDA frameworks for use within high-resolution ESMs, such as the DOE E3SM high-resolution configuration: Building a complete and performant CDA system is a significant technological undertaking. Given the improvement in software practices, revision control, and regression testing that have taken place over the last decade throughout the ESM community, the notion of a community-developed-and-supported CDA "toolbox" has merit. While each CDA system will differ in its mission and application space, many of the underlying methods, algorithms, and software packages can be shared across the community. Research priorities should include the identification of aspects within CDA that can be easily leveraged across the community.

4. Computational, algorithmic, and middleware solutions to link forward models to CDA system in ESM grand-challenge configurations: Data assimilation and forward models have traditionally been connected through the I/O system, i.e., the data assimilation system writes the "increment" to disk, and then the increment is read by the forward model and used to produce a forecast that is then written to disk to be ingested by the DA system for the next assimilation cycle. As computer architectures continue to evolve, the chip processing and associated interconnect network speeds continue to outstrip the speed of the I/O system, thus making I/O increasingly expensive, at least relative to other parts of the system. Since avoiding costly I/O is a broad issue with high-performance computing, the continued development of "middleware" solutions that allow the read/write to be replaced with direct memory get/put should be a research priority. The importance of removing the I/O bottleneck grows rapidly as the model resolution grows and, as a result, could be particularly important for applications like E3SM that emphasize grand-challenge simulations.

5. Development of testbeds and metrics to measure the fidelity of various initializing approaches: Given that the underlying methodology for implementing CDA is still a topic of research, it is important that the community has available a set of testbeds and metrics for the evaluation of different approaches. The testbeds should be community-shared configurations, ranging from idealized to real world, that each initialization system is encouraged to undertake. The metrics are the shared quantitative measures of testbed configurations that allow the different CDA systems to be compared and contrasted. The definition and adoption of these testbeds and metrics before the maturity of numerous CDA systems will, in all likelihood, accelerate progress across both ESM and CDA communities.

6. Development of holistic CDA approaches/frameworks that support the diversity of CDA applications: Research priorities should emphasize approaches that exploit the varied applications of CDA. CDA has the potential to expand the scientific scope of ESMs to provide high-fidelity earth system state estimates for model initialization, as a method for the study of causes of model bias, as a means of evaluating and constraining model parameters, and as a tool for evaluating the impact of novel observing systems. But these diverse applications will not be realized by happenstance. Rather, each of these powerful applications of CDA will need to be incorporated into the design, development, and maintenance of a CDA implementation for ESMs. Research priorities should include the implementation of CDA methods that promote the wide-ranging applications of this technique.

7. Growth of CDA capability and expertise in the context of E3SM: Data assimilation has long been a research capability within organizations that conduct weather forecasting and research. Since weather forecasting and research has not been a focus within DOE, data assimilation expertise in the context of earth system modeling is lacking. Yet, the development of the underlying methods and algorithms needed to build a robust data assimilation system (e.g., variational methods and EnKF) has long been a core capability with DOE's Office of Science Advanced Scientific Computing Research (ASCR) portfolio. Priority should be given to research pathways that "marry" the data assimilation capabilities already available within ASCR to the E3SM application with sufficient strength and durability to ensure that scientific use of CDA becomes routine with DOE's earth science's community.



References

Anderson, J., Bell, M., Hacker, J., Marzouk, Y., Sandu, A., and Snyder, C., 2016. Report from a Workshop An Examination of Data Assimilation Algorithms, Observations, and Applications in the Context of Next Generation Computing.

Årthun, M., Eldevik, T., Viste, E., Drange, H., Furevik, T., Johnson, H., and Keenlyside, N., 2017. "Skillful prediction of northern climate provided by the ocean." *Nature Communications 8*(15875). https://doi.org/10.1038/ncomms15875

Carton, J.A., and Giese, B.S., 2008. "A reanalysis of ocean climate using Simple Ocean Data Assimilation (SODA)." *Monthly Weather Review 136*(8): 2999-3017.

Chang, Y.S., Zhang, S., Rosati, A., Delworth, T.L., and Stern, W.F., 2013. "An assessment of oceanic variability for 1960–2010 from the GFDL ensemble coupled data assimilation." *Climate Dynamics 40*(3-4): 775-803.

Dreher, M., and Peterka, T., 2017. *Decaf: Decoupled dataflows for in situ high-performance workflows* (No. ANL/MCS-TM-371). Argonne National Laboratory (ANL), Argonne, Illinois.

Edson, J., Crawford, T., Crescenti, J., Farrar, T., Frew, N., Gerbi, G., Helmis, C., Hristov, T., Khelif, D., Jessup, A., Jonsson, H., Li, M., Mahrt, L., McGillis, W., Plueddemann, A., Shen, L., Skyllingstad, E., Stanton, T., Sullivan, P., Sun, J., Trowbridge, J., Vickers, D., Wang, S., Wang, Q., Weller, R., Wilkin, J., Williams III, A., Yue, D.K.P., and Zappa, C., 2007. "The coupled boundary layers and air–sea transfer experiment in low winds." *Bulletin of the American Meteorological Society 88*(341–356). https://doi.org/10.1175/BAMS-88-3-341

Evensen, G., 1994. "Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics." *Journal of Geophysical Research: Oceans 99*(C5): 10143-10162.

Frolov, S., Bishop, C.H., Holt, T., Cummings, J., and Kuhl, D., 2016. "Facilitating strongly coupled ocean–atmosphere data assimilation with an interface solver." *Monthly Weather Review 144*(1): 3-20.

Haarsma, R.J., Roberts, M.J., Vidale, P.L., Senior, C.A., Bellucci, A., Bao, Q., Chang, P., Corti, S., Fučkar, N.S., Guemas, V., and von Hardenberg, J., 2016. "High resolution model intercomparison project (HighResMIP v1. 0) for CMIP6." *Geoscientific Model Development 9*(1): 4185-4208.

Hamill, T.M., and Snyder, C., 2000. "A hybrid ensemble Kalman filter–3D variational analysis scheme." *Monthly Weather Review 128*(8): 2905-2919.

Hewitt, H.T., Bell, M.J., Chassignet, E.P., Czaja, A., Ferreira, D., Griffies, S.M., Hyder, P., McClean, J.L., New, A.L., and Roberts, M.J., 2017. "Will high-resolution global ocean models benefit coupled predictions on short-range to climate timescales?" *Ocean Modelling 120:* 120-136.

Kirtman, B.P., Bitz, C., Bryan, F., Collins, W., Dennis, J., Hearn, N., Kinter, J.L., Loft, R., Rousset, C., Siqueira, L., and Stan, C., 2012. "Impact of ocean model resolution on CCSM climate simulations." *Climate dynamics 39*(6):1303-1328.

Kirtman, B.P., Perlin, N., and Siqueira, L., 2017. "Ocean eddies and climate predictability." *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 27(12): 126902.

Laloyaux, P., Thépaut, J.N., and Dee, D., 2016. "Impact of scatterometer surface wind data in the ECMWF coupled assimilation system." *Monthly Weather Review 144*(3): 1203-1217.

Langehaug, H.R., Sandø, A.B., Årthun, M., and Ilıcak, M., 2018. "Variability along the Atlantic water pathway in the forced Norwegian Earth System Model." *Climate Dynamics:* 1-20.

Large, W. G., and S. Yeager, 2004. *Diurnal to decadal global forcing for ocean and sea-ice models: the data sets and flux climatologies*. NCAR Technical Note, NCAR/TN-460+STR, CGD Division of the National Center for Atmospheric Research.

Lofstead, J.F., Klasky, S., Schwan, K., Podhorszki, N., and Jin, C., 2008. "Flexible IO and integration for scientific codes through the adaptable IO system (ADIOS)." *In Proceedings of the 6th international workshop on Challenges of large applications in distributed environments* (pp. 15-24). ACM.

Mogensen, K., Balmaseda, M.A., and Weaver, A., 2012. *The NEMOVAR ocean data assimilation system as implemented in the ECMWF ocean analysis for System 4*. European Centre for Medium-Range Weather Forecasts.

Penny, S.G., Behringer, D.W., Carton, J.A., and Kalnay, E., 2015. "A hybrid global ocean data assimilation system at NCEP. *Monthly Weather Review 143*(11): 4660-4677.

Penny, S.G., and Hamill, T.M., 2017. "Coupled data assimilation for integrated earth system analysis and prediction." *Bulletin of the American Meteorological Society 97*(7): ES169-ES172.

Penny, S.G., and coauthors, 2017. Coupled Data Assimilation for Integrated Earth System Analysis and Prediction: Goals, Challenges and Recommendations. *World Meteorological Organization*. Available online (last accessed 3 April 2018): https://www.wmo.int/pages/prog/arep/wwrp/new/documents/Final_WWRP_2017_3_27_July.pdf

Slivinski, L., G.P. Compo, J.S. Whitaker, P.D. Sardeshmukh, J.-W. A. Wang, K. Friedman, and C. McColl, 2018. "What is the impact of additional tropical observations on a modern data assimilation system?" *Monthly Weather Review,* submitted

Roemmich, D., Johnson, G.C., Riser, S., Davis, R., Gilson, J., Owens, W.B., Garzoli, S.L., Schmid, C., and Ignaszewski, M., 2009. "The Argo Program: Observing the global ocean with profiling floats." *Oceanography 22*(2): 34-43.

Weatherhead, E.C., Wielicki, B.A., Ramaswamy, V., Abbott, M., Ackerman, T.P., Atlas, R., Brasseur, G., Bruhwiler, L., Busalacchi, A.J., Butler, J.H., and Clack, C., 2018. "Designing the Climate Observing System of the Future." *Earth's Future 6*(1): 80-102.

Vannitsem, S., and V. Lucarini, 2016. "Statistical and dynamical properties of covariant lyapunov vectors in a coupled atmosphere-ocean model—multiscale effects, geometric degeneracy, and error dynamics." *Journal of Physics A: Mathematical and Theoretical 49*(22): 224001, doi:10.1088/1751-8113/49/22/224001.

Zadra, A., Williams, K., Frassoni, A., Rixen, M., Adames, Á.F., Berner, J., Bouyssel, F., Casati, B., Christensen, H., Ek, M.B., and Flato, G., 2018. "Systematic Errors in Weather and Climate Models: Nature, Origins, and Ways Forward." *Bulletin of the American Meteorological Society 99*(4): ES67-ES70.

Appendix A: Participants

Last Name	First Name	Organization	Email
Adcroft	Alistair	Princeton University	alistair.adcroft@noaa.gov
Akella	Santha	NASA	santha.akella@nasa.gov
Bader	David	LLNL	bader2@llnl.gov
Barrie	Dan	NOAA	daniel.barrie@noaa.gov
Bishop	Craig	NRL-Monterey	craighuntlybishop@gmail.com
Biven	Laura	DOE	laura.biven@science.doe.gov
Compo	Gil	CIRES	gilbert.p.compo@noaa.gov
Counillion	Francois	NERSC	francois.counillon@nersc.no
Danabasoglu	Gokhan	NCAR	gokhan@ucar.edu
Draper	Clara	NOAA	clara.draper@noaa.gov
el Akkraoui	Amal	GMAO	Amal.ElAkkraoui@nasa.gov
Geernaert	Gary	DOE	gerald.geernaert@science.doe.gov
Higdon	David	Virginia Tech	dhigdon@vt.edu
Hnilo	Justin	DOE	justin.hnilo@science.doe.gov
Hogan	Pat	NRL-Stennis	patrick.hogan@nrlssc.navy.mil
Huang	Jin	NOAA	jin.huang@noaa.gov
Johnson	Brad	NOAA	bradford.johnson@noaa.gov
Joseph	Renu	DOE	renu.joseph@science.doe.gov
Kirtman	Ben	University of Miami Rosenstiel School	bkirtman@rsmas.miami.edu
Klein	Steve	LLNL	klein21@llnl.gov

Last Name	First Name	Organization	Email
Koch	Dorothy	DOE	dorothy.koch@science.doe.gov
Laloyaux	Patrick	ECMWF	patrick.laloyaux@ecmwf.int
Laviolette	Randall	DOE	randall.laviolette@science.doe.gov
Leung	Ruby	PNNL	Ruby.Leung@pnnl.gov
Lu	Feiyu	NOAA	feiyu.lu@noaa.gov
Mazloff	Matt	Scripps	mmazloff@ucsd.edu
McClean	Julie	Scripps	jmcclean@ucsd.edu
Penny	Steve	University of Maryland	Steve.Penny@noaa.gov
Peterka	Tom	ANL	tpeterka@mcs.anl.gov
Podhorszki	Norbert	ORNL	pnorbert@ornl.gov
Rahul	Mahajan	NOAA	rahul.mahajan@noaa.gov
Ringler	Todd	LANL	todd.ringler@science.doe.gov
Rosati	Anthony	GFDL	tony.rosati@noaa.gov
Sluka	Travis	NCEP/CPC, UMD	travis.sluka@noaa.gov
Sockwell	Chad	Florida State University	kcs12j@my.fsu.edu
Van Roekel	Luke	LANL	lvanroekel@lanl.gov
Vernieres	Guillaume	JCSDA	guillaume.vernieres@noaa.gov
Wu	Shu	University of Wisconsin	swu33@wisc.edu
Zhang	Shaoqing	Qingdao	szhang@ouc.edu.cn
Zhengyu	Liu	Ohio State	liu.7022@osu.edu

Appendix B: Agenda

Monday, April 9, 2018

07:30–08:30		**Registration**
08:30		**Convene**
08:30–08:40		**Dorothy Koch and Dan Barrie** DOE interests in high-resolution initialization NOAA interests in high-resolution initialization
08:40–09:00		**Steve Penny and Todd Ringler –** Welcome and workshop introduction
09:00–10:00	**Session 1: Small-ensemble, high-res initialization**	
	20 min	**Ruby Leung** E3SM's approach for initialization of v1 high-resolution climate simulations
	20 min	**Francois Counillon** Can we use flow-dependent assimilation with a high-resolution earth system model?
	20 min	**Facilitator: Gokhan Danabasoglu** Discussion questions: The high-resolution ESM activity has been a leader in term of HPC-compatible algorithms, design, and implementation. Should/can this same activity take on the task of evaluating the various "scalable" approaches for coupled-system initialization? **Session rapporteur: Luke Van Roekel** 1. Small-ensemble, high-res initialization report
10:00–10:30		**Break**
10:30–12:00	**Session 2: Coupled data assimilation roadmaps**	
	20 min	**Steve Penny** Opportunities for leveraging coupled data assimilation capabilities for initialization of high-resolution ESMs
	20 min	**Gokhan Danabasoglu** NCAR's approach, experience, and roadmap for coupled model data assimilation
	20 min	**Anthony Rosati** GFDL's approach, experience, and roadmap for coupled model data assimilation
	30 min	**Facilitator: Ben Kirtman** Discussion questions: Are there any "lesson learned" from past prediction and projection activities that are applicable to the initialization of high-resolution ESMs? ESMs continue to push to longer prediction time scales, at least for certain phenomena. Will this trend eventually grow to contain the CMIP-relevant time scales?

		Session rapporteur: Matthew Mazloff
		2. Coupled data assimilation roadmaps report
12:00–13:00		**Lunch break**
13:00–14:30	**Session 3: Initialization of eddying ocean configurations**	
	20 min	**Ben Kirtman**
		Initialized prediction with a global coupled ocean eddy-resolving model
	20 min	**Julie McClean**
		Initialization sensitivity of fully-coupled E3SM v0 approximate present-day transient climate simulations
	20 min	**Matthew Mazloff**
		Multigrid 4D-var optimization for initialization of high-resolution ocean models: decoupling the forward and adjoint models
	30 min	**Facilitator: Alistair Adcroft**
		Discussion questions:
		Is there an opportunity to coordinate the testing and evaluation of different approaches to initialization of eddying ocean models?
		Should we consider defining metrics for this evaluation?
		Does the overall importance of including an eddying ocean model vary with the time scale of the coupled system prediction?
		Session rapporteur: K Chad Sockwell
		3. Initialization of eddying ocean configurations report
14:30–15:30	**Session 4: Data assimilation methods**	
	20 min	**Shaoqing Zhang**
		A high-efficiency approximation of EnKF for coupled model data assimilation
	20 min	**Craig Bishop**
		The local ensemble tangent linear model with dynamical order reduction—an enabler for coupled model 4DVar
	20 min	**Facilitator: Steve Penny**
		Discussion questions:
		Do we have observations to test the coupled error covariance models?
		Session rapporteur: Todd Ringler
		4. EnKF within a coupled data assimilation setting report
15:30–16:00		**Break**

Monday, April 9, 2018 (continued)

16:00–17:30 **Session 5: Data assimilation for shorter time-scale prediction**

20 min **Gil Compo**

Sensitivities of the NCEP global forecast system to observations, initializations, and model formulation

20 min **Santha Akella**

Role of interface variables—SST, SSS, ice surface temperature—in coupled DA

20 min **Steve Klein**

The emerging coupled-CAPT effort within DOE

30 min **Facilitator: Anthony Rosati**

Discussion questions:

Are there opportunities for the broader climate sciences community—from S2S to beyond decadal timescales—to benefit and build from the capabilities developed within the "NWP" community?

How are observations currently used in the initialization of climate simulations?

Session rapporteur: Shu Wu

5. Data assimilation for shorter-time-scale prediction report

Tuesday, April 10, 2018

07:30–08:30		**Registration**
08:30		**Convene**
08:30		**Steve Penny and Todd Ringler –** Welcome
08:30–10:00	**Session 6: Possible solutions to the DA I/O barrier**	
	20 min	**Shu Wu and Zhengyu Liu** Online implementation of data assimilation schemes in ESM and potential optimizations
	20 min	**Tom Peterka** Improving the efficiency of loosely coupled separate parallel programs that communicate over the HPC interconnect
	20 min	**Norbert Podhorszki** The Adaptable IO System (ADIOS)
	30 min	**Facilitator: Patrick Laloyaux** Discussion questions: Does production-like coupled data assimilation depend on solving the I/O problem? Are software solutions like Decaf, ADIOS, and maybe others useful to the data assimilation community? If so, who in the community is positioned to explore these new approaches? **Session rapporteur: K Chad Sockwell** 6. Possible solutions to the DA I/O barrier report
10:00–10:30		**Break**
10:30–12:00	**Session 7: Operational centers and data assimilation**	
	20 min	**Guillaume Vernieres** JEDI and the initialization of coupled models
	20 min	**Pat Hogan** Navy Earth System Prediction Capability: Progress and plans
	20 min	**Patrick Laloyaux** How much coupling does ECMWF need: implicit and explicit cross-correlations
	30 min	**Facilitator: Craig Bishop** Discussion questions: What is the appropriate weighting between individual groups exploring one-off approaches and coordinated, community-based efforts? **Session rapporteur: Santha Akella** 7. community activities report
12:00–12:30	30 min	**Closing discussion** Discussion questions: Gaps and research priorities Near-term opportunities for coordination, evaluation, testing? Where to from here?

U.S. Department of Energy | Office of Science
National Oceanic and Atmospheric Adminstration
U.S. Department of Commerce

Appendix C: Abbreviations and Acronyms

ADIOS	Adaptable Input-Output System
AMOC	Atlantic Meridional Overturning Circulation
ASCR	Advanced Scientific Computing Research
CANGA	Coupling Approaches for Next-Generation Architectures
CAPT	Cloud-Associated Parameterizations Testbed
CDA	Coupled Data Assimilation
CORE	Co-ordinated Ocean-Ice Reference Experiments
COSSE	Climate Observing System Simulation Experiments
DA	data assimilation
DART	Data Assimilation Research Testbed
DOE	U.S. Department of Energy
E3SM	Energy Exascale Earth System Model
ECDA	Ensemble Coupled Data Assimilation
ECMWF	European Centre for Medium-Range Weather Forecasts
EnKF	Ensemble Kalman Filter
EnOI	Ensemble Optimal Interpolation
ENSO	El Niño-Southern Oscillation
ESM	Earth System Model
ESPC	Earth System Predictability Capability
FOAM	Fast Ocean Atmosphere Model
GFDL	Geophysical Fluid Dynamics Laboratory
GMAO	Global Modeling and Assimilation Office

GODAS	Global Ocean Data Assimilation System
HPC	high-performance computing
I/O	input/output
JEDI	Joint Effort for Data assimilation Integration
MITGCM	Massachusetts Institute of Technology General Circulation Model
NCAR	National Center for Atmospheric Research
NCEP	National Centers for Environmental Prediction
NESPC	Navy Earth System Predictability Capability
NorCPM	Norwegian Climate Prediction Model
NOAA	National Oceanic and Atmospheric Administration
NWP	Numerical Weather Prediction
OI	optimal interpolation
PDO	Pacific Decadal Oscillation
QNLM	Qingdao National Laboratory for Marine Science and Technology
SCDA	strongly coupled data assimilation
SciDAC	Scientific Discovery through Advanced Computing
SODA	Simple Ocean Data Assimilation
TLM	tangent linear model
UK	United Kingdom
WCDA	weakly coupled data assimilation

For More Information

Dorothy Koch, dorothy.koch@science.doe.gov

DOE Earth and Environmental System Modeling

https://science.energy.gov/ber/research/cesd/earth-and-environmental-system-modeling/

Annarita Mariotti, annarita.mariotti@noaa.gov

NOAA Modeling, Analysis, Predictions and Projections Program

https://cpo.noaa.gov/Meet-the-Divisions/Earth-System-Science-and-Modeling/MAPP