Naval Research Laboratory

Washington, DC 20375-5320

NRL/7530/FR--2021/1

NRL Atmospheric Data Assimilation and ONR Code 31: Workshop Proceedings

ELIZABETH SATTERFIELD,[1] NANCY BAKER,[1] CHRIS JONES,[2] REZA MALEK-MADANI,[3] MICHAL BRANICKI,[4,5] WILLIAM CAMPBELL,[1] ALBERTO CARRASSI,[6,7] SUE CHEN,[1] PEDRAM HASSANZADEH,[8] DANIEL HODYSS,[9] SARAH KING,[1] MATTHIAS MORZFELD,[10] HANS NGODOCK,[11] NANCY NICHOLS,[12] CHRISTIAN SAMPSON,[2] ELAINE SPILLER,[13] ANDREW STUART,[14] PETER JAN VAN LEEUWEN,[15] KEVIN VINER,[1] ERIK VAN VLECK,[16] AND MAX YAREMCHUK[2]

[1]*U.S. Naval Research Laboratory, Marine Meteorology Division*
[2]*University of North Carolina, Chapel Hill*
[3]*Office of Naval Research, Code 31*
[4]*Department of Mathematics, University of Edinburgh, UK*
[5]*The Alan Turing Institute for Data Science, UK*
[6]*School of Mathematical, Physical and Computational Sciences and National Centre for Earth Observations, University of Reading, UK*
[7]*Mathematical Institute, University of Utrecht, the Netherlands*
[8]*Rice University*
[9] *U.S. Naval Research Laboratory, Remote Sensing Division*
[10]*Scripps Institution of Oceanography, University of California, San Diego*
[11]*Ocean and Atmospheric Science and Technology Directorate, Oceanography Division*
[12]*School of Mathematical, Physical and Computational Sciences, University of Reading, UK*
[13]*Department of Mathematical and Statistical Sciences, Marquette University*
[14]*Computing and Mathematical Sciences, Division of Engineering and Applied Science, California Institute of Technology*
[15]*Department of Atmospheric Science, Colorado State University*
[16]*Department of Mathematics, University of Kansas*

June 14, 2021

REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 0704-0188*

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
14-06-2021	NRL Formal Report	Nov 5-19, 2020

4. TITLE AND SUBTITLE NRL Atmospheric Data Assimilation and ONR Code 31: Workshop Proceedings	**5a. CONTRACT NUMBER** N0001421WX00025
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER 0601153N
6. AUTHOR(S) Elizabeth Satterfield,[1] Nancy Baker,[1] Chris Jones,[2] Reza Malek-Madani,[3] Michal Branicki,[4,5] William Campbell,[1] Alberto Carrassi,[6,7] Sue Chen,[1] Pedram Hassanzadeh,[8] Daniel Hodyss,[9] Sarah King,[1] Matthias Morzfeld,[1], Hans Ngodock,[11] Nancy Nichols,[12] Christian Sampson,[2] Elaine Spiller,[13] Andrew Stuart,[14] Peter Jan van Leeuwen,[15] Kevin Viner,[1] Erik Van Vleck,[16] and Max Yaremchuk[2]	**5d. PROJECT NUMBER** 100001643639.0010
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER 75-1L35-0-1-5

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Research Laboratory Marine Meteorology Division 7 Grace Hopper Ave. Mail-stop 2 Monterey, CA 94934	**8. PERFORMING ORGANIZATION REPORT NUMBER** NRL/7530/FR--2021/1
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) NRL Base 1	**10. SPONSOR/MONITOR'S ACRONYM(S)**
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited

13. SUPPLEMENTARY NOTES
[1]U.S. Naval Research Laboratory, Marine Meteorology Division; [2]University of North Carolina, Chapel Hill; [3]Office of Naval Research, Code 31; [4]Department of Mathematics, University of Edinburgh, UK; [5]The Alan Turing Institute for Data Science, UK; [6]School of Mathematical, Physical and Computational Sciences and National Centre for Earth Observations, University of Reading, UK; [7]Mathematical Institute, University of Utrecht, the Netherlands; [8]Rice University; [9]U.S. Naval Research Laboratory, Remote Sensing Division; [10]Scripps Institution of Oceanography, University of California, San Diego; [11]Ocean and Atmospheric Science and Technology Directorate, Oceanography Division; [12]School of Mathematical, Physical and Computational Sciences, University of Reading, UK; [13]Department of Mathematical and Statistical Sciences, Marquette University; [14]Computing and Mathematical Sciences, Division of Engineering and Applied Science, California Institute of Technology; [15]Department of Atmospheric Science, Colorado State University; [16]Department of Mathematics, University of Kansas

14. ABSTRACT
Between November 5 and 19, 2020, we held a virtual workshop bringing together the Naval Research Laboratory (NRL) Data Assimilation groups, Office of Naval Research (ONR) program managers and both the national and international university community. The focus of the workshop was to identify the main areas of collaboration, to discuss collaborations in terms of the timelines for transition for Navy operational use (e.g. short-range transitions ~3 years, mid-range transitions ~3-6 years, long-range transitions ~10 years), and to identify what is needed for a productive collaboration (e.g., interactions, funding types).

15. SUBJECT TERMS Workshop proceedings

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			Dr. Elizabeth Satterfield
					19b. TELEPHONE NUMBER *(Include area code)*
UU	UU	UU	UU	22	(301) 351-5999

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

This page intentionally left blank

CONTENTS

This page intentionally left blank

EXECUTIVE SUMMARY

Between November 5-19, 2020, we held a virtual workshop bringing together the Naval Research Laboratory (NRL) Data Assimilation groups, Office of Naval Research (ONR) program managers and both the national and the international university communities. The focus of the workshop was to identify the main areas of collaboration, to discuss collaborations in terms of the timelines for transition for Navy operational use (e.g., short-range transitions ~3 years, midrange transitions ~3-6 years, long-range transitions ~10 years), and to identify what is needed for a productive collaboration (e.g., interactions, funding types). The workshop was structured into six main sessions as follows:

1. Overview of Data Assimilation at NRL
2. Recent Collaborations
3. Short–Mid Range Collaborations: Coupled Data Assimilation, Model Error and Weak Constraint
4. Next Generation DA: Nonlinear, Non-Gaussian, Multi-scale
5. Next generation DA: Machine Learning
6. DA for Next Generation Models: Lagrangian Issues and Dynamic Meshes

Each session had several presentations and extended discussions. The workshop concluded with breakout groups in four main categories: 1) Model Error and 4DVar 2) Coupled Data Assimilation 3) Machine Learning in Data Assimilation and 4) Nonlinearity and Non-Gaussianity in Next Generation Data Assimilation. Each working group was tasked with writing a high-level topical white paper outlining the state of the science and key issues for future research and development.

This page intentionally left blank.

NRL ATMOSPHERIC DATA ASSIMILATION AND ONR CODE 31: WORKSHOP PROCEEDINGS

1. INTRODUCTION

Between November 5 and 19, 2020, we held a virtual workshop bringing together the Naval Research Laboratory (NRL) Data Assimilation groups, Office of Naval Research (ONR) program managers and both the national and the international university communities. The focus of the workshop was to identify the main areas of collaboration, to discuss collaborations in terms of the timelines for transition for Navy operational use (e.g., short-range transitions ~3 years, midrange transitions ~3-6 years, long-range transitions ~10 years), and to identify what is needed for a productive collaboration (e.g., interactions, funding types). The workshop was structured into six main sessions as follows:

1. Overview of Data Assimilation at NRL
2. Recent Collaborations
3. Short –Mid Range Collaborations: Coupled Data Assimilation, Model Error and Weak Constraint
4. Next Generation DA: Nonlinear, Non-Gaussian, Multi-scale
5. Next Generation DA: Machine Learning
6. DA for Next Generation Models: Lagrangian Issues and Dynamic Meshes

Each session had several presentations and extended discussions. The workshop concluded with breakout groups in four main categories: 1) Model Error and 4DVar 2) Coupled Data Assimilation 3) Machine Learning in Data Assimilation and 4) Nonlinearity and Non-Gaussianity in Next Generation Data Assimilation. Each working group was tasked with writing a high-level topical white paper outlining the state of the science and key issues for future research and development.

In what follows, Section 2 summarizes the workshop presentations and discussions, and Section 3 provides the reports from each working group.

2. OVERVIEW OF WORKSHOP SESSIONS

2.1 Opening Remarks and Overview of NRL Data Assimilation Efforts across Divisions

The workshop began with opening remarks by Chris Jones, who encouraged all attendees to consider collaboration in terms of short-range (~3 years), midrange (3-6 years), and long-range (~10 years) transitions. For longer-range transitions, attendees were encouraged to think about how Navy models and observing systems would look in 10 years and beyond. The introductory remarks were continued by Dr. Reza Malek-Madani, who discussed the five ONR codes and the 3 divisions within ONR Code 31, which is primarily focused on information science. He next discussed his portfolio, Applied Computational Analysis, and highlighted overlaps between the remit of his portfolio and systematic techniques for using empirical data to inform decisions. Dr. Malek-Madani specified that he was seeking guidance in aligning this portfolio with NRL research directions.

The introductory session continued with overview talks by each NRL division in attendance. Dr. Nancy Baker presented first for NRL 7500 Marine Meteorology Division, and provided an overview of the NAVDAS-Accelerated Representer (AR) data-assimilation system and current research-and-development efforts. Dr. Baker discussed some key challenges, including: higher resolution, nonlinearities, convective-permitting models, longer-range predictability and Earth system modeling,

Manuscript approved June 14, 2021.

high-quality ensembles for hybrid DA, representing model error, and the movement toward community software such as JEDI (JCSDA) and MET (DTC). The Navy relevance for this work was framed in terms of the direct link to the customer, with diagnostics and verification geared toward naval needs, and uncertainty quantification.

Dr. Hans Ngodock next provided an overview for NRL 7300 Ocean Sciences data assimilation. He described the NCODA system and the Global (HYCOM) model. He discussed the NCODA 3DVar and Augmented State AS-EnKF, as well as the regional (NCOM) 4DVar representer method, Wave-4DVar, Acoustic 4DVar, a variety of coupled applications (Ocean-Wave, Ocean-Acoustics, Ocean-Atmosphere), and the generalization of NCOM-4DVar. Dr. Ngodock provided several examples for potential future directions, including image assimilation, multi-scale, adjointless 4DVar, and ice assimilation.

Dr. Dan Hodyss gave an overview of data-assimilation efforts within the NRL 7200 Remote Sensing and 7600 Space Science Divisions. He briefly discussed the divisions' efforts for aerosol assimilation, and provided details on their efforts with neutral variable assimilation in the stratosphere, the mesosphere and the thermosphere. Those efforts include LETKF development, data assimilation in an observation-limited environment and the handling of the long-distance correlations that arise in the variables of the ionosphere.

After the introductory briefs, the initial discussion session focused on NRL funding types (e.g., 6.1, 6.2 and 6.4), NRL funding sources (primarily Navy, NOAA, and NASA), Technical Readiness Levels, and research to operations and operations to research. Some challenges were discussed in terms of communicating why data assimilation matters, why we use certain DA approaches, and what the key challenges are. Because DA is interdisciplinary, it can be challenging to create a narrative of relevance for proposal evaluation panels with primarily physical science or applied mathematics expertise. There was also some discussion of using DA to improve physics and feedback into modeling, as well as discovery-based machine learning.

2.2 Recent Collaborations

Our next session focused on recent and ongoing collaborations.

Dr. Nancy Nichols provided an overview of correlated observation error and error of representation, which are topics that she has collaborated on with several NRL scientists including Elizabeth Satterfield, Dan Hodyss and William Campbell. She presented an overview of the sources of observation error, including representation error, and methods by which observation error covariances can be diagnosed. Dr. Nichols discussed approaches to incorporate correlated observation error in the assimilation and the resulting improvements in the analysis, as well as issues of convergence and conditioning. She also presented detailed techniques for reconditioning the observation error covariance matrix, specifically using ridge regression and eigenvalue modification.

Dr. Matti Morzfeld gave an overview of his work on particle filtering. He detailed the importance of particle filters (PFs) for nonlinearity and discussed localized PFs. He specified cases in which a PF would be useful (e.g. nonlinearity in prior and posterior) and what was needed to make PF algorithms computationally feasible. Dr. Morzfeld next detailed ongoing issues, including: the effects of resolution and nonlinear observation operators on nonlinearity, the need to develop a mathematical theory for localization and multi-scale localization, the importance of standardized DA test problems, "Effective Dimension", and the optimal window for DA.

Dr. Nancy Baker overviewed other external collaborations and discussed vehicles for collaborations including contracts, grants, cooperative agreements, and sponsored research. She detailed some recent sponsored research, including: LETKF development (I. Szunyogh, TAMU), Adjoint-Based Sensitivity Diagnostics (D. Daescu, PSU), Observation Impact of pseudo-radiosondes (B. Hoover, UW/CIMSS), Holm Transform and Improvements to Background Error Covariances (S. Fletcher, CSU/CIRA). Dr. Baker emphasized that collaborative research with universities and NOAA Cooperative Institutes can be pathways to interns, post-docs, and federal employees. She also detailed some issues with NRL directly providing funding for partnerships with academia, including that the NRL funding tends to be linked to operational transitions and is uncertain from year to year. Funding delays and expiration dates can make planning difficult, especially for academia that seek funding timelines more aligned with doctoral student dissertations. She proposed that long term partnerships and internal funding mechanisms would be needed to optimize these collaborations.

The second discussion session focused on emerging short-mid range collaborations. There was a productive discussion on what level of sophistication is needed for simplified models to provide meaningful results for operational applications, and what the dependence is between the problem being investigated and the level of sophistication of the simplified model. This was an ongoing theme and discussion throughout the workshop, and is discussed further in section 2.7.

2.3 Short- to Mid-Range Collaborations: Coupled Data Assimilation, Model Error and Weak Constraint

Dr. Alberto Carrassi gave an overview of coupled DA. He discussed coupling of fast and slow scales and showed the mechanism of error propagation across scales with the aid of linear approximations. He then numerically experimented and corroborated his findings using the coupled atmosphere-ocean coupled model (MAOOAM; De Cruz, 2016). He showed that cross-compartments effects are stronger in the slow-to-fast direction, but intercompartmental effects are stronger for the fast compartment (Tondeur et al., 2020). This behavior indicates that fast scale error must be controlled for effective prediction. Furthermore, using covariant Lyapunov vectors, he demonstrated that quasi-neutral modes, and the quasi-degeneracy of the Lyapunov spectrum, are a result of the coupling process. All these modes must be monitored and controlled adequately by the data assimilation, implying the need to enlarge the ensemble properly beyond what is needed in the uncoupled scenario (Carrassi et al., 2021). The importance of coupled DA also was put forward in the context of a combined DA-ML approach, whereby a hybrid, physics-based-plus-data-driven model, is constructed to approximate unresolved scales. The use of coupled DA was seen to be crucial to propagate the information effectively across scales before being used by the ML step (Brajard et al., 2021). Finally, Dr. Carrassi showed some experiments using 4DVar in a single-column atmosphere-ocean model that indicated that strongly coupled 4DVar is needed to transfer information fully across compartments and to reduce imbalance (Smith et al., 2015, 2020).

Dr. Peter Jan Van Leeuwen gave a presentation on Weak Constraint 4DVar. He posed questions regarding where the analysis should be done within the DA time window, and pointed out that while 4DVar solves for the mode rather than the mean, we do not really understand what happens once this solution is propagated forward with a nonlinear model. He indicated that weak-constraint (En)4DVar is likely more useful for prediction than strong-constraint. He also discussed issues with localization in ensemble (or hybrid) smoothers, but noted that these problems are likely easier to solve in a weak constraint formulation. He examined methods for estimating model errors, noting that these model errors are likely to be time-correlated, and further, that misspecification of model error decorrelation time scales can lead to large analysis errors. Estimating decorrelation time scales can be done, but the convergence tends to be slow and non-uniform. Dr. Van Leeuwen detailed methods to use weak-constraint DA to estimate both structural and random components of the model errors, as errors can be estimated at every time step at every grid point for every variable. The structural components are a direct route to model improvement and learning new physics.

The third discussion session focused mainly on further refining the topics for break out groups and identifying participants. These topics will be covered in depth in Section 3, so they will be listed only briefly here:

1. Weak Constraint: The unique dual form of NAVDAS-AR allows for a true weak constraint formulation. To capitalize on this ability, we need ways to specify model error. Weak constraint permits long window DA, and so we need to understand the optimal window length and the application dependence. We also require improved methods for treatment of pre-conditioning and convergence.
2. Coupled DA: Strongly coupled approaches, coupled covariances, Hybrid ML-DA approaches for the coupled problem, issues of time scale separation and sensitivity analysis.
3. Machine Learning: DA can benefit from ML by using ML for modeling unresolved scales (parameterization), uncertainty quantification, data selection and quality control, or to emulate a model (data-driven models).
4. New Solver Approaches to handle nonlinearity and non-Gaussianity: particle filters, localization, hybrid ensemble-PF, particle flows, MCMC framework to identify sampling techniques.
5. Next Generation Models: the NEPTUNE system will have the capability of high-resolution nested domains within the global system; However, we do not currently have the capability to tackle multi-scale or nested DA.

2.4 Next Generation DA: Nonlinear, Non-Gaussian

Dr. Elaine Spiller discussed challenges in PFs. The "curse of dimensionality" limits the size of the problem that PFs can be applied to, and for practical application, demands dimension reduction of some type. Various recently developed strategies for localization with PFs show promise for earth system data-assimilation applications. This localization leads to the issue of "stitching back together" locally applied PF analyses, which is complicated further by resampling. Another challenge for PFs is the typical paucity of model runs, and hence particles. Surrogates — statistical, ML, or physical — are a promising strategy to overcome this challenge. Dr. Spiller described a recent proof-of-concept simulation experiment in which utilizing a Gaussian process emulator fit with 100 particles within a PF performs on par or better with 10^5 member PF on a nonlinear/non-Gaussian test case.

The subsequent discussion session on nonlinear and non-Gaussian aspects of ensemble data assimilation was led by Dr. Matti Morzfeld. The key discussion topics/questions were as follows:

1. To what extent does nonlinearity in the model or observation operator interact with or cause non-Gaussianity of prior and posterior distributions?
2. To what extent are outer loops (var/hybrid DA) effective in addressing nonlinearity and non-Gaussianity in the analysis and forecast?
3. How do we initialize the forecasts in view of nonlinearity and in view of non-Gaussian posteriors?
4. What is the best use of PF hybrids, e.g., traditional ensemble DA for nearly linear parts of the problem, PFs for nonlinear parts?
5. Theory for localization
6. New algorithms

2.5 Next-generation DA: Machine Learning

Dr. Pedram Hassanzadeh talked about recent advances in using machine learning for data-driven spatiotemporal forecasting, showing examples from a multi-scale Lorenz 96 (Chattopadhyay et al, 2020a) system and ERA5 reanalysis. It was shown that some variants of recurrent and convolutional neural networks (RNNs and CNNs) can yield accurate predictions, and that adding physical consistency to the neural networks, for example, through equivariance-preserving components, can improve the forecast accuracy (Chattopadhyay et al, 2020b). The power of transfer learning for improving the generalization (i.e., extrapolation) capabilities of data-driven forecast models was demonstrated (Chattopadhyay et al, 2020c) and some of the recent work in adding data assimilation (an ensemble Kalman filter) to the forecast model was shown. At the conclusion of his talk, Dr. Hassanzadeh highlighted the benefits of such data-driven forecast models (e.g., to generate thousands of ensemble members inexpensively, avoiding the need for localization) and some of the outstanding challenges (e.g., generalization, learning the atmosphere's full state vector).

Dr. Andrew Stuart discussed ML in high dimensions, motivated by the recent successes of various neural network architectures and deep learning in addressing the supervised learning problem to find the function **F** mapping between spaces **X** and **Y**. The classical settings are ones where **X** is a finite dimensional Euclidean space and where **Y** is either a finite dimensional Euclidean space (regression) or a set of finite cardinality (classification). He discussed algorithms that address the problem for spaces of functions **X** and **Y** in a general infinite-dimensional setting. The main reason behind such an approach is driven by the need to derive ML techniques whose performance is dimension-independent (or it at least scales well with increasing dimension). The resulting algorithms have the potential to speed-up large-scale computational tasks arising in science and engineering in which the map **F** must be evaluated many times. Dr. Stuart introduced a new, overarching approach, and described a number of distinct methodologies that are built from this approach. The key concept is to design methods resulting in learned approximate models that can be transferred between different grid resolutions and different discretization methods. Once learned, the approximate model may be reused in many different situations. Basic theoretical results were explained and numerical results were illustrated for solution operators arising from elliptic PDEs, from the Burgers equation and from the Navier-Stokes equation. Potential for generalization of such an approach to data assimilation was discussed; in particular, the use of such techniques for learning the equivalent of the Kalman gain (or the control operator) was mentioned

Dr. Alberto Carrassi led a discussion session on machine learning and DA. The key discussion topics/questions were as follows:

1. How can ML contribute to DA workflow, "What ML can do for DA?"
2. How can we best combine ML and DA, "What DA can do for ML?"
3. How can we unify DA/ML from a theoretical point of view?

2.6 DA for Next Generation Models: Lagrangian Issues and Dynamic Meshes

The final session focused on longer-range collaborations aimed at next-generation models. Dr. Michal Branicki discussed theoretical connections and differences between two distinct notions of Lagrangian DA, and the Eulerian DA aimed at improving predictions of scalar and vector fields (e.g., temperature, velocity, etc.). Lagrangian DA can involve the use of trajectory data to improve the Eulerian predictions, or it can be aimed at improving predictions of tracer trajectories using both the Eulerian and Lagrangian data/observations. Dr. Branicki stressed that while the Eulerian and Lagrangian DA can be considered and analyzed in a unified abstract framework, the accuracy and performance of DA approaches have to be assessed differently depending on whether the outcome is aimed at estimating field-based or trajectory-based quantities. This is due to the fact arising from dynamical systems theory/analysis that even small errors in DA estimates of Eulerian fields (e.g., velocity) are likely to lead to large errors in trajectory predictions. Some probabilistic, information-theoretic metrics for assessing the accuracy of DA for Lagrangian predictions were outlined, as were potential approaches for tuning the models used in Lagrangian DA to improve the trajectory-based predictions. Other issues highlighted in the presentation included the potential for introducing biases and artifacts in the Eulerian (velocity) estimates in oceanographic predictions based on the Lagrangian observations fed into DA algorithms were confined to the top layer of the water column. Finally, Dr. Branicki outlined his ongoing work on the analysis of convergence of various (Bayesian, time-sequential) DA algorithms incorporating both the Eulerian and Lagrangian observations, which apply to moving meshes and time-varying number of observed trajectories.

Dr. Christian Sampson discussed DA on adaptive meshes in sea ice models. Numerical solvers using adaptive meshes can focus computational power on important regions of a model domain capturing important or unresolved physics. The adaptation can be informed by the model state or external information, or can be made to depend on the model physics. In this latter case, one can consider the mesh configuration as part of the model state. If observational data is to be assimilated into the model, the question of updating the mesh configuration with the physical values arises. Adaptive meshes present significant challenges when using popular ensemble data assimilation (DA) methods. Dr. Sampson and colleagues developed a novel strategy for ensemble-based DA for which the adaptive mesh is updated along with the physical values. This involves including the node locations as a part of the model state, itself, allowing them to be updated automatically at the analysis step. This poses a number of challenges that they resolved to produce an effective approach that promises to apply with some generality. They evaluated their strategy with two testbed models in 1D compared with a strategy that was developed previously that did not update the mesh configuration. They found that updating the mesh improves the fidelity and convergence of the filter. Ongoing work on extensions to 2D are ongoing, but present further challenges that need to be overcome. The key to implementing a joint physics and mesh update is the configuration of an augmented vector that includes variables characterizing the underlying numerical mesh. In 1D, there is little ambiguity, as the natural approach is to just add the grid points of the underlying mesh. But in 2D, the characterization of the “grid” in terms of such a vector is not as straightforward when, for instance, using a finite element method. In that case, physical values may be stored at the centers of the elements, whereas the elements, themselves, are determined by their vertices. Considerations of the mesh type (triangular, hexagonal, square, etc.) also add complications.

Dr. Erik Van Vleck discussed dynamically adapting meshes and DA. For ensemble-based data-assimilation techniques, the use of adaptive spatial meshes present some challenges. These include the development of computational models with flexible adaptive discretization, the combination of time-evolving spatial meshes that support ensemble members with observations and their potentially time dependent locations, and a desire (for both accuracy and efficiency) to minimize interpolation and extrapolation. There are two general classes of adaptive spatial meshing techniques. In the quasi-Lagrange approach, the mesh moves continuously with time, and the discretized PDE is supplemented with a convective term to reflect mesh movement. With rezoning approaches, the mesh is updated at each time level using mesh equations or via mesh generation codes, the PDE solution is interpolated from the previous mesh to the new mesh, and it is often necessary to use conservative interpolation schemes that preserve invariants. With both approaches, mesh refinement (adding or subtracting mesh points) is often necessary when there is change in the complexity of solutions. Dr. Van Vleck surveyed augmented variable and common mesh approaches and discussed the use of metric tensors to generate and combine meshes. The remainder of his talk included the application of these types of techniques to model problems (1D and 2D inviscid Burgers, and 2D Shallow Water Equation) and ideas for future development based on the use of metric tensors, variational approaches, and machine learning approaches.

2.7 Idealized Models and Hierarchies of Testbed Models

The success and advancement of numerical techniques for data assimilation are, at least in part, due to the availability of useful idealized models that are accepted within the community. An idealized model represents some aspects of a "real," or operational model, but its solution can be obtained easily on a laptop or a desktop machine. Prototyping DA techniques on idealized problems is naturally faster than running a full-blown test in an operational setting. Provided that the results obtained with an idealized problem are interpreted carefully, numerical prototyping within an idealized framework is valuable for true scientific application. Perhaps it is fair to say that successful application of a new DA technique in a number of idealized problems is a gateway to possible application in an operational system (with several steps in between, of course).

Prominent examples of idealized models are the Lorenz models, which mimic geofluids' chaotic behavior and conservation quantities (Lorenz, 1963; 1996). These models allow for adapting the model degree of nonlinearity as well as their dimension and are thus very powerful tools when studying, e.g., ensemble DA techniques. Other examples of idealized models include the Kuramoto-Sivashinsky equation (Kuramoto & Tsuzuki 1975; Sivashinsky, 1977), or the Burgers equation (Burgers, 1948), displaying respectively spatially extended chaos and shock wave propagation. Models of two- and three-body problems have also been used effectively to examine the application of energy constraints in data assimilation (Watkinson et al, 2007). Other simplified test problems are simplifications of the Navier-Stokes equations and include the dry and moist quasi-geostrophic models (see, e.g., Evensen 1994, Lapeyre & Held 2004), the primitive equations (see, e.g., Ades and van Leeuwen, 2015), the shallow-water equations (see, e.g., Holm et al. 2020), and a barotropic vorticity model (see, e.g., Browne 2016). Further idealized models include an extension of the Lorenz model to include thermodynamics (Vissio & Lucarini 2020), an extension of the shallow-water model to include a rain component (Wursch & Craig, 2014), the Modular Arbitrary-Order Ocean-Atmosphere Model (MAOOAM) (De Cruz et al. 2016), the atmospheric "SPEEDY" model (Molteni 2003, Kucharski et al. 2006) and its atmosphere-ocean-land-sea-ice coupled version "SPEEDO" (Severijns and Hazeleger, 2010), as well as the model put forward by Majda and coworkers (Majda et al. 1997). These models are used routinely to test new DA algorithms or new extensions of existing DA schemes. Typically, the procedure is to present a new idea for a DA technique and then to test it on one or more of the idealized models and to compare the performance to existing DA technology. If the new DA techniques lead to improved analysis and forecasting in the idealized models, then the technique is deemed promising for more realistic or real application. In addition to being useful for (rapid) prototyping of DA methods, idealized models are invaluable tools in the classroom and to familiarize students with DA techniques. Further, the Joint Effort for Data assimilation Integration (JEDI) contains quasi-geostrophic and Lorenz models, previously mentioned, as part of the Object Oriented Prediction System (OOPS) package, which is publicly available via the JCSDA GitHub.

Often, available idealized models are not representative of the current challenges in DA. For example, most of the idealized problems do not account satisfactorily for the effects of increasing model resolution, either because the models are too simple even to increase resolution (ordinary differential equation models, such as the Lorenz models), or because the effects of a refined resolution are very different from the relevant Navier-Stokes equations (e.g., increasing resolution in a quasi-geostrophic or shallow-water model is very different from increasing resolution in a Navier-Stokes model). Similarly, current and future DA systems must be able to account for a large number of spatial and temporal scales, which is commonly referred to as a "multi-scale" problem. While there are several multi-scale test problems available in the literature (see, e.g., Lorenz 2005), most of them are not appropriate because errors are dominated by a single scale (calling into question the actual multi-scale behavior of the idealized problem). Lastly, there is an interest in coupled ocean-atmosphere models and coupled DA, and how idealized models can emulate the difficulties and sub-compartment interactions of a coupled DA system is currently under investigation (Smith et al, 2015, 2017, 2018, 2020). Developing idealized models that use adaptive (potentially non-conservative) meshes is another area of important and desired developments. Some examples exist whereby “classical” idealized models (the Burger and the Kuramoto equations) have been equipped with such a moving mesh numerical scheme in 1D (Aydogdu et al, 2019), and data-assimilation techniques have been developed for moving mesh models of shallow ice sheets (Bonan et al, 2016, 2017), but further work is needed to port those in 2D and 3D.

In summary, there is a need to produce a new generation of an idealized model hierarchy that faithfully represents the challenges of modern and next-generation DA. Creating this new generation of idealized models may be a community effort, but we envision NRL playing a leading role in the development and deployment of a new generation of idealized problems. This new generation of idealized models will pave the way to successful application of advanced, nonlinear, non-Gaussian, multi-scale and coupled DA techniques. NRL can take a leading role because of its strong connections to operational forecasting, which can enable the verification of the applicability and usefulness of the idealized models. Additionally, NRL’s upcoming model NEPTUNE is configurable into idealized test cases such as a splitting supercell, various mountain wave configurations, or a baroclinic wave.

3. REPORTS FROM WORKING GROUPS

The workshop concluded with breakout groups in four main categories: 1) Model Error and 4DVar, 2) Coupled Data Assimilation, 3) Machine Learning in Data Assimilation, and 4) Nonlinearity and Non-Gaussianity in Next Generation Data Assimilation. The working groups were tasked with identifying main areas of collaboration, transition potential and lead times, and writing a high-level topical white paper outlining the state of the science and key issues for future research and development. The out-briefs from each working group follow.

3.1 Model Error and 4DVar

Variational data assimilation, especially 4DVar (four-dimensional variational data assimilation), is the workhorse of most modern operational global data-assimilation systems. 4DVar finds the best state of the model at the start of a time window, utilizing observations distributed throughout the specified time window, a prior forecast, and a mathematical model of background and observational error. Most operational implementations of 4DVar apply the strong-constraint formulation, which explicitly assumes that the model error can be neglected over the assimilation window. However, it has been recognized for some time now that errors in the model equations can be substantial, and much better forecasts can be expected if these model errors would be accounted for in the data assimilation. This weak-constraint formulation of 4DVar (i.e., including the model error) confers several advantages, including 1) a better fit of the initial conditions to the observations, 2) the ability to use longer assimilation time windows, such that a more accurate solution can be obtained, and 3) the capability to start the model forecast at the end of the assimilation window (versus the beginning), because the analysis is a best trajectory over the time window, thereby gaining time and accuracy.

The advantages of weak-constraint 4DVar have been realized by the operational community, although the complexity of implementation and additional memory requirements have stalled progress. However, NRL is unique in that the global atmospheric and regional ocean data-assimilation systems solve the problems in observation space using the dual-space formulation. This approach differs from all other operational centers including ECMWF and allows for very efficient implementation of representation of the model errors as part of the DA solver with almost no additional memory requirements. For these reasons, this is an appropriate time to exploit this unique NRL capability.

To reach this potential, further research is needed to determine how 1) to specify the model error covariance, 2) to integrate the model error into the data-assimilation code, exploring existing infrastructure, 3) to optimize flow-dependent information, 4) to design and implement localization, and 5) to decide the optimal window length. Although this work is close to operational, it does contain a substantial fundamental research component, too. The model error covariance will have to be estimated iteratively and fundamental research is needed on the most efficient way to do this. Preconditioning of a weak-constraint 4DVar in dual space, in connection with outer loops and more extensive correlations in the observational covariance matrix, is a largely open question. Optimal ways to do the hybridization with ensemble methods need exploration, and flexible hybridization options are essential to this. Research into these fundamental questions is needed because fast convergence of the whole new algorithm is essential, as the Navy operational DA systems are subject to real-time constraints.

3.2 Coupled DA and DA for Adaptive Moving Mesh Models

As the complexity of Earth system models increases, the Navy's operational DA schemes will need to adapt. Information will flow increasingly between different identifiable Earth system subcomponents, and it will be crucial to understand how to optimize the DA approach in order to make the most of the information coming from observational data, in particular from new satellite-based observations that often sense multiple Earth system components. This issue is particularly evident with subsystems that historically were uncoupled, but are becoming progressively more integrated. The most notable case is the coupling of oceans, surface waves and the atmosphere, but other cases such as the atmosphere with land or sea ice, and the oceans with sea ice, have significant naval relevance.

Two key characteristics that make coupled DA challenging are the separations of scales between components and their temporal forcing-feedback nonlinearity interaction. The ocean-atmosphere coupling suffers from a strong time (and space) scale separation, with a slow ocean and fast atmosphere or fast surface waves and land. How the information propagates across these scales affects the efficacy of the adopted DA scheme. Many operational centers have made significant advances in the practical implementation of coupling ocean, surface wave, land and atmosphere; however, these developments generally have not been accompanied by corresponding advances in the theoretical understanding of the problem. Variants of the Lyapunov spectrum form the basis of an understanding of the impact of time-scale separation through revealing how the error propagation is driven by the coupling processes (Tondeur et al., 2020). This provides not only coupling information between different system components but also between combinations of variables, as well as a measure of the strength of the coupling and information flow. This understanding can be used as a basis for observational design. It is, however, not known how to handle efficiently and effectively the combined time- and space-scale separation that is inevitable in situations of naval interest. We will need to understand error evolution across a spectrum of spatial and temporal scales in order to make medium- to long-range forecasts. Development of a hierarchy of coupled problems that capture the key features of differing time and space scales with increasing complexity will be instrumental to our understanding and eventual capabilities.

Regional models can be coupled through nesting. The logical extension of this strategy is to improve the approximation of localized and/or anisotropic features in computational models through the use of adaptive meshes. We anticipate such configurations becoming a key component in future modeling as they will be needed to capture regions with steep gradients in the geophysical field of interest. This is particularly important when observation locations may occur in such regions, thereby challenging the propagation of data content. A smart relocation of mesh nodes to properly describe areas of sharp gradient is desirable. Ensemble methods are crucial here, as numerical approximations in the variational method can destabilize under the same conditions of steep gradients. Adaptive meshes, however, pose specific challenges to ensemble DA methods, as the underlying mesh may change across ensemble members. Each ensemble member may be represented on a different mesh, whereby differences include the position of the individual mesh nodes but also their total number. Significant progress has been made to model problems in one spatial dimension (Aydogdu et al., 2019), but the appropriate technology for two and three dimensions remains relatively undeveloped, although variational adaptive meshing techniques provide a unified framework that has been extended successfully to higher space dimensions. Because the mesh points contain physical information and can be cast as part of the state variable (Sampson et al., 2021), the issue has similarities to the coupled DA problem, with information flow between different parts of the state (here the physical variables and the mesh) that needs to be unveiled and used properly in the DA process. The use of adaptive spatial meshing allows for the design of ensemble meshes to maximize the skill of the DA process by adjusting meshes to balance the approximation of ensemble members and observations.

Our discussions centered on the overarching goal of understanding the evolution of error and flow of information from observations between the components of the coupled systems. The ultimate challenge would be to construct a single coupled cycling system for global and limited-area applications that accounts for the differences in observing networks and error-growth characteristics of the ocean, the atmosphere, the cryosphere and the land.

We also advocated for an intentional approach to developing a hierarchy of dedicated models in which to develop and test new approaches to exploring these research questions. This hierarchy could start with idealized models based on simplified geometries and/or low mode approximations, could extend to intermediate-complexity models, and ultimately could lead to a full-scale operational model. While each given issue may necessitate the formulation of a dedicated model hierarchy, this approach of building and systematic use of these model hierarchies ultimately will afford faster development of ideas and their transfer to the operational level.

The following problems were posed as relevant for coupled DA; many of these problems and questions could be explored using a hierarchy of testbed models varying in complexity, and as such would be ideal for collaborations with universities:

1. When and where is strongly coupled data assimilation of the most benefit to the Navy? What are the relevant time and spatial scales for the Navy?
2. What are the relevant data-assimilation windows for each component of the coupled system? How should the observation processing for coupled DA change with respect to the different space/time scales of the different components? How should the spatial and temporal thinning differ?
3. Given that the error of representation likely will be a large contributor to the observation-error covariances, how should the observation-error covariances differ within each component of the coupled system?
4. What is needed for localization in space and time to form ensemble and/or hybrid covariances? How does the observation processing affect the coupled covariances and localization?
5. While limited-area models can be coupled through nesting, a logical extension of this strategy is to improve the approximation of localized and/or anisotropic features in computational models through the use of adaptive meshes. How does the use of adaptive spatial meshes impact the flow of information and evolution of error in simple coupled DA systems, and how does this translate to operational systems of naval relevance?

The outcomes of these short- to long-term research efforts (1-5) are expected to enhance NRL's coupled global and regional analysis/reanalysis and forecasting capabilities by accelerating the research leading ultimately to operational transitions. Goals of items 1-4 (transitional and applied) are expected to be achievable in two to five years by building on ongoing DA technology, while item 5 (basic research) is aimed at moving the state of the art for adaptive meshes and is in the higher-risk/higher-reward category. The overarching goal falls within the five- to 10-year range. This modeling and data-assimilation research and development also will be of substantial interest to the broader data-assimilation, numerical weather and ocean prediction, and climate-modeling communities.

3.3 Machine Learning

A number of areas where machine learning (ML) could improve the DA workflow or where DA could contribute to ML applications were identified. In the first category, the power of different types of ML methods (e.g., deep neural networks) for user-specified feature detection in multi-scale data, learning distribution of high-dimensional data, and emulation of nonlinear dynamics is utilized. For the short- to mid-term, the following problems are of particular interest to the NRL and academic partners: 1) computing tangent linear models (TLM) and adjoints from ensemble data for use in operational DA, 2) customizable pre-processing of observational data (e.g., quality control, filling missing data, adaptive data sampling; e.g., Kadow et al., 2020) and automated identification of specific features (e.g., sea ice boundaries and leads/polynyas for use in DA), and 3) approximating spatially correlated and/or situation-dependent observation error covariances and sparsifying their inverses while minimizing the information loss (e.g., for upcoming interferometry of the ocean surface delivered by the SWOT satellite). These issues are of immediate importance for the operational DA community. NRL is uniquely qualified for introducing such innovations, particularly given its proven track record of bringing basic research all the way through to transition to computation-intensive operations.

For mid-term efforts, we recognized the following as problems of mutual interest and practical importance: 4) building data-driven (and interpretable) models to predict and correct errors in the forecast models, 5) estimating model error covariances for use in weak-constraint 4DVar, and 6) coupling situation-dependent observation error covariance matrices with 4DVar control variables during the minimization. Due to the formulation of NRL's DA systems in observation space, NRL is particularly interested in research on weak-constraint 4DVar with operational atmospheric and ocean models.

For long-term efforts, we identified these problems as having practical importance but requiring further theoretical development and close collaboration between NRL and academic partners: 7) developing data-driven forecast models for high-dimensional nonlinear systems (e.g., Chattopadhyay et al., 2020a), for example, to provide optimized large-ensemble forecasts improving the Eulerian and/or Lagrangian predictions, and 8) formulating a Bayesian framework for combining ML and DA (e.g., through Bayesian neural networks) with the aim to quantify uncertainty systematically in the DA estimates, and to assimilate nonlinear observations efficiently, including the information obtained from Lagrangian tracers, AUVs, acoustic tomography, etc.

In turn, DA can be used to improve the quality of the training data for ML models, e.g., in applications involving learning data-driven or data-informed subgrid-scale parameterization schemes from noisy and sparse or spatially localized observations (Brajard et al., 2021). The NRL is uniquely positioned to lead such efforts, given its access to a vast amount of (sometimes exclusive) in-situ and remote-sensing data from the atmosphere and the ocean combined with extensive DA expertise.

In problems 1-8 listed above, we also identified the following challenges regarding the use of ML methods that need to be addressed using novel mathematical frameworks and a hierarchy of well-designed test cases: i) interpretability, ii) generalization (e.g., Chattopadhyay et al., 2020c), iii) robustness, and iv) incorporating physics and other prior constraints.

The outcomes of the short- to long-term efforts (1-8) are expected to enhance NRL's forecasting capabilities by accelerating the operational DA framework and/or improving its accuracy, reliability, and the potential for a rapid deployment in new environments. Goals of items 1-3 (transitional and applied) are expected to be achievable in two to three years by building on existing ML methodologies and DA technology, while items 7 and 8 (basic research) are aimed at moving the state of the art and are in the high-risk/high-reward category. The developed models and frameworks also will be of substantial interest to the broader DA, weather prediction, climate modeling, and ML communities.

3.4 Nonlinearity and Non-Gaussianity in Next-Generation Ensemble Data Assimilation

We expect that as numerical models increasingly produce high-resolution, realistic simulations of the real Earth system, DA systems necessarily will become more nonlinear and more non-Gaussian in the future. The reason is that an increase in computational power typically results in users' increasing the resolution of the numerical models and/or embedding detailed process models, which, in turn, is likely to cause more significant nonlinear phenomena. This nonlinearity of the model directly leads to non-Gaussian prior and posterior distributions, which are at the core of statistically sampled or ensemble-based DA (see, e.g., Morzfeld and Hodyss 2019). We thus expect that the desire to use non-Gaussian methods will increase. We further expect that nonlinear and non-Gaussian characteristics of the DA system will be amplified by new observations that are obtained via nonlinear observation operators (similar to current all-sky observations, for example) and possibly non-Gaussian model or observation error statistics. Current DA systems struggle, or simply fail, in a highly nonlinear and non-Gaussian regime and there is a need to design a next generation of ensemble DA techniques that are applicable and efficient under nonlinear and non-Gaussian conditions (see, e.g., Morzfeld and Hodyss 2019).

We have identified three specific research objectives, relevant to Navy and DoD objectives, to create the mathematical and computational foundations for the design of efficient nonlinear and non-Gaussian ensemble DA:

• **Research Objective 1:** Metrics. There exists an urgent need to understand the factors that control the non-Gaussian characteristics of the PDFs we are trying to describe using Monte Carlo methods. We suggest the development of a variety of metrics that help elucidate how and what factors may lead to an increase in nonlinearity and non-Gaussianity, as well as how factors such as increases in the number and types of specific observations and model improvements may produce more Gaussian impacts. We further study the delicate interplay of nonlinearity and non-Gaussianity with respect to resolution of the numerical model, the number of observations being assimilated, and the nonlinearity of observation operators. Such a theory is indeed required to study the effects of nonlinearity and non-Gaussianity on the forecast ability of ensemble DA under nonlinear and non-Gaussian conditions.

• **Research Objective 2:** Hybrid methods. We plan to design efficient hybrid DA schemes by first identifying and then separating nonlinear/non-Gaussian characteristics from nearly linear/Gaussian ones. This separation allows us to target specifically the nonlinearity/non-Gaussianity, while using more traditional (and computationally less expensive) techniques that exploit nearly linear/Gaussian characteristics, e.g., on larger spatial or temporal scales (see, e.g., Slivinski et al. 2015).

• **Research Objective 3:** Future methods. In the long term, we believe that there will be sufficient computational resources available that (i) the models will be run at extremely high resolutions invoking extreme nonlinearities, and (ii) highly non-Gaussian techniques could be used. Of course, because of the properties of geophysical fluids and their tendency to result in intrinsically high-dimensional systems, the dimension of this problem will be enormous, and any plausible techniques will have to allow for both non-Gaussianity and sampling issues in extremely high dimensions. Contemporary techniques that appear to be heading toward this goal are localized particle filters (Morzfeld et al. 2018, Poterjoy et al. 2019, Potthast et al. 2019), particle flows (Pulido and van Leeuwen, 2019) and optimization-based Monte Carlo samplers (Bardsley et al., 2014). A second class of techniques that may become feasible for ensemble DA are Markov chain Monte Carlo (MCMC) methods, some of which recently have been shown to scale well with high dimensions, and are well-known to be able to handle strong nonlinearities (Morzfeld et al. 2019, Tong et al. 2020).

We envision that this research will guide the design of new and efficient numerical techniques for data assimilation under highly nonlinear and non-Gaussian conditions. We will test specifically the potential of algorithms that rely on hybrid techniques (see Research Objective 2 above), and merge traditional ensemble DA, such as the ensemble Kalman filter (EnKF) or ensemble 4DVar (EnVar), with nonlinear techniques, such as particle filters, particle flows, or optimization-based sampling techniques. Interestingly, it has been shown that there is a very natural way to embed an EnKF or variational methods in particle filters and particle flows (see, e.g., the recent review Van Leeuwen et al., 2019). A second critical component of nonlinear DA will be the wide use of emulators (via Gaussian processes or other machine learning inspired techniques, (see, e.g., Bayarri et al. 2015, Spiller et al. 2014) that will help ameliorate deficiencies due to a limited ensemble size, which we expect to be significant under nonlinear and non-Gaussian conditions.

We view these three research objectives as medium-range efforts, with Objective 1 being the most "open ended" research task, which may be viewed as a long-term effort (but continuously progressing in 3- to 6-year increments). All three objectives would benefit from involving (graduate) students or postdocs, because the objectives generate research projects that are academic and of practical relevance.

4. SUMMARY AND CONCLUDING REMARKS

Between November 5 and 19, 2020, we held a virtual workshop bringing together the Naval Research Laboratory (NRL) Data Assimilation groups, Office of Naval Research (ONR) program managers and both the national and international university community. The focus of the workshop was to identify the main areas of collaboration, to discuss collaborations in terms of the timelines for transition for Navy operational use, and to identify what is needed for a productive collaboration. This workshop helped to establish relationships between NRL scientists and the university community, paving the way for research to operations. Conversely, discussions at this meeting helped inform the university community of future directions that would be beneficial to NRL.

We identified the following key areas of collaboration:

1) Model Error and 4DVar
2) Coupled Data Assimilation
3) Machine Learning in Data Assimilation
4) Nonlinearity and Non-Gaussianity in Next Generation Data Assimilation.

Below, we will summarize our recommendations briefly.

Model Error and 4DVar: NRL's unique dual-form 4DVar differs from all other operational centers including ECMWF, and allows for very efficient implementation of representation of the model errors as part of the DA solver, with almost no additional memory requirements. For these reasons, this is an appropriate time to exploit this unique NRL capability. To reach this potential, further research is needed to determine how 1) to specify the model error covariance, 2) to integrate the model error into the data-assimilation code, exploring existing infrastructure, 3) to optimize flow dependent information, 4) to design and implement localization, and 5) to decide the optimal window length. Although this work is close to being operational, it does contain a substantial fundamental research component. We view this work as being a short- to mid-range effort, with potential to have an immediate impact on forecast skill.

Coupled Data Assimilation: As the complexity of Earth system models increases, the Navy's operational DA schemes will need to adapt. Information (and errors) increasingly will flow between different Earth system subcomponents, and it will be crucial to optimize the DA approach in order to make the most of the information coming from observational data. Improvements to the coupled data assimilation can be achieved through research focused on 1) knowing when and where strongly coupled DA is of the most benefit to the Navy, 2) optimizing the assimilation of observations within coupled systems, particularly from new, satellite-based observations that often sense multiple Earth system components, 3) developing effective space/time localization for ensemble and/or hybrid error covariances, and 4) developing adaptive meshes for limited area applications. The outcomes of this short- to long-term research efforts are expected to enhance NRL's coupled global and regional analysis/reanalysis and forecasting capabilities by accelerating the research leading ultimately to operational transitions.

Machine Learning in Data Assimilation: DA is uniquely positioned to integrate ML because the two fields share a common theoretical foundation. Improvements to the DA workflow can be made through 1) improvements to the TLM and adjoint models 2) Improvements to quality control and 3) model error estimation and correlation. Data-driven models can be used to improve the forecast model and its uncertainty quantification, e.g., through improved parameterizations. These issues are of immediate importance for the operational DA community and aspects project onto short- to long-range efforts.

Nonlinearity and Non-Gaussianity in Next-Generation Data Assimilation: We expect that as numerical models increasingly produce high-resolution, realistic simulations of the real Earth system, DA systems necessarily will become more nonlinear and more non-Gaussian in the future. Three specific research objectives, relevant to Navy and DoD objectives, were identified: 1) metrics 2) hybrid methods, and 3) future methods. These objectives aim to create the mathematical and computational foundations for the design of efficient nonlinear and non-Gaussian ensemble DA.

We found this workshop both informative and beneficial. We encourage continued collaboration between ONR, NRL, and the university community for the purposed of 1) identifying short- to mid-range transitions that may follow from basic research 2) providing guidance for long-range needs and 3) better positioning NRL to be able to take full benefit from basic research being performed at academic institutions both nationally and internationally.

REFERENCES

1. Ades, M., van Leeuwen, P.J., 2015. The Effect of the Equivalent-Weights Particle Filter on Dynamical Balance in a Primitive Equation Model. Monthly Weather Review, 143(2), 581-596.

2. Aydogdu, A., Carrassi, A., Guider, C.T., Jones C.K.R.T. and Rampal P., 2019. Data assimilation using adaptive, non-conservative, moving mesh models. Nonlinear Processes in Geophysics, 26(3):175–193, 2019.

3. Bardsley, J.M., Solonen, A., Haario, H., Laine, M., 2014. Randomize-then-optimize: A method for sampling from posterior distributions in nonlinear inverse problems. SIAM Journal on Scientific Computing 36 (4), A1895-A1910

4. Bayarri, M. J., Berger, J., Calder, E., Patra, A., Pitman, E. B., Spiller, E., Wolpert, R., 2015. Probabilistic Quantification of Hazards: A Methodology Using Small Ensembles of Physics Based Simulations and Statistical Surrogates. International Journal for Uncertainty Quantification 5(4), 297-325

5. B. Bonan, M.J. Baines, N.K. Nichols and Partridge D. A moving-point approach to model shallow ice sheets: a study case with radially symmetrical ice sheets. The Cryosphere, 10, 2016, 1-14, doi:10.5194/tc-10-1-2016. http://www.the-cryosphere.net/10/1/2016/tc-10-1-2016.pdf .

6. B. Bonan, N. K. Nichols, M. J. Baines, and D. Partridge. Data assimilation for moving mesh methods with an application to ice sheet modelling. Nonlinear Processes in Geophysics, 24, 515 – 534, 2017. https://doi.org/10.5194/npg-24-515-2017

7. Brajard J, A Carrassi, M Bocquet and L Bertino. Combining data assimilation and machine learning to infer unresolved scale parameterization. Phil. Trans A of the Roy Soc., 2021, In press. Available at arXiv:2009.04318.

8. Browne, P.A., 2016. A comparison of the equivalent weights particle filter and the local ensemble transform Kalman filter in application to the barotropic vorticity equation. Tellus A: Dynamic Meteorology and Oceanography, 68:1, 30466.

9. Burgers, J. M., 1948. A mathematical model illustrating the theory of turbulence. Advances in applied mechanics 1, 171-199.

10. Carrassi A, M Bocquet, J Demayer, C Grudzien, P Raanes, and S Vannitsem. Data assimilation for chaotic dynamics. In press Springer Data Assimilation Book (Vol. IV), 2021. Available at arXiv:2010.07063.

11. Chattopadhyay, A., Hassanzadeh, P. and Subramanian, D., 2020a. Data-driven predictions of a multiscale Lorenz 96 chaotic system using machine-learning methods: reservoir computing, artificial neural network, and long short-term memory network. Nonlinear Processes in Geophysics, 27(3), pp.373-389.

12. Chattopadhyay, A., Nabizadeh, E. and Hassanzadeh, P., 2020b. Analog Forecasting of Extreme-Causing Weather Patterns Using Deep Learning. Journal of Advances in Modeling Earth Systems, 12(2), p.e2019MS001958.

13. Chattopadhyay, A., Subel, A. and Hassanzadeh, P., 2020c. Data-driven super-parameterization using deep learning: Experimentation with multi-scale Lorenz 96 systems and transfer-learning. Journal of Advances in Modeling Earth Systems, p.e2020MS002084.

14. De Cruz, L., Demaeyer, J., and Vannitsem, S., 2016. The Modular Arbitrary-Order Ocean-Atmosphere Model: MAOOAM v1.0. Geoscientific Model Development 9, 2793–2808.

15. Evensen, G., 1994. Sequential data assimilation with a nonlinear quasigeostrophic model using Monte Carlo methods to forecast error statistics. Journal of Geophysical Research 99(10), 10143–10 162.

16. Holm, H., Sætra, M., van Leeuwen, P.J., 2020. Massively parallel implicit equal-weights particle filter for ocean drift trajectory forecasting. Journal of Computational Physics: X 6(2590–0552), 100053.

17. Kadow, C., Hall, D.M. & Ulbrich, U., 2020. Artificial intelligence reconstructs missing climate information. Nature Geoscience 13, 408–413

18. Kucharski, F., Molteni, F., Bracco, A., 2006. Decadal interactions between the western tropical Pacific and the North Atlantic Oscillation. Climate Dynamics 26, 79-91.

19. Kuramoto, Y., Tsuzuki, T., 1975. On the formation of dissipative structures in reaction-diffusion systems. Progress of Theoretical and Experimental Physics 54, 687–699.

20. Lapeyre, G., Held, I. M., 2004. The Role of Moisture in the Dynamics and Energetics of Turbulent Baroclinic Eddies. Journal of the Atmospheric Sciences 61(14), 1693-1710.

21. Lorenz, E.N., 1963. Deterministic nonperiodic flow, Journal of the Atmospheric Sciences 20(2), 130-141.

22. Lorenz, E.N., 1996. Predictability - a problem partly solved. Seminar on Predictability Vo. I, ECMWF.

23. Lorenz, E.N., 2005. Designing chaotic models. Journal of the Atmospheric Sciences, 62(5):1574–1587.

24. Majda, A.J., McLaughlin, D.W., Tabak, E.G, 1997. A one-dimensional model for dispersive wave turbulence. Journal of Nonlinear Science 7, 9–44.

25. Molteni, F., 2003. Atmospheric simulations using a GCM with simplified physical parametrizations. I. Model climatology and variability in multi-decadal experiments. Climate Dynamics 20, 175-191.

26. Morzfeld, M, Hodyss, D., 2019. Gaussian approximations in filters and smoothers for data assimilation. Tellus A: Dynamic Meteorology and Oceanography 71(1), 1-27.

27. Morzfeld, M., Tong, X.T., Marzouk, Y.M., 2019. Localization for MCMC: sampling high-dimensional posterior distributions with local structure, Journal of Computational Physics, 380, 1–28.

28. Morzfeld, M, Hodyss, D., Poterjoy, J. 2018., Variational particle smoothers and their localization. Quarterly Journal of the Royal Meteorological Society, 144, 806-825.

29. Poterjoy, J., Wicker, L. J., Buehner, M., 2019. Progress in the development of a localized particle filter for data assimilation in high-dimensional geophysical systems, Monthly Weather Review 147, 1107 – 1126.

30. Potthast R., Walter, A., Rhodin, A., 2019. A localized adaptive particle filter within an operational NWP framework. Monthly Weather Review 147 (1), 345-362

31. Pulido, M., P.J. van Leeuwen, Sequential Monte Carlo with kernel embedded mappings: The mapping particle filter. Journal of Computational Physics 396, 400-415.

32. Steffen, W., Richardson, K., Rockström, J. et al. The emergence and evolution of Earth System Science. Nat Rev Earth Environ 1, 54–63 (2020). https://doi.org/10.1038/s43017-019-0005-6

33. Sampson, C., Carrassi, A., Aydogdu A., and Jones, C.K.T, 2021. Ensemble Kalman Filter for non-conservative moving mesh solvers with a joint physics and mesh location update. (under review). arXiv:2009.13670v1.

34. Severijns, C. A. and Hazeleger, W.: The efficient global primitive equation climate model SPEEDO V2.0, Geosci. Model Dev., 3, 105–122, https://doi.org/10.5194/gmd-3-105-2010, 2010.

35. Sivashinsky, G., 1977. Nonlinear analysis of hydrodynamic instability in laminar flames, Part I. Derivation of basic equations. Acta Astronautica 4, 1177–1206.

36. Slivinski, L., Spiller, E., Apte, A., Sandstede, B., 2015. A Hybrid Particle–Ensemble Kalman Filter for Lagrangian Data Assimilation. Monthly Weather Review 143(1), 195–211.

37. Smith, P.J. , A.S. Lawless and N.K. Nichols. 2020. The role of cross-domain error correlations in strongly coupled 4D-Var atmosphere-ocean data assimilation. Quarterly Journal of the Royal Meteorological Society, 146, 2450–2465. http://dx.doi.org/10.1002/qj.3802

38. Smith, P.J. , A.S. Lawless and N.K. Nichols. 2018. Treating sample covariances for use in strongly coupled atmosphere-ocean data assimilation. Geophysical Research Letters, 45, 445-454. https://doi.org/10.1002/2017GL075534.

39. Smith, P.J., A.S. Lawless and N.K. Nichols. 2017. Estimating forecast error covariances for strongly coupled atmosphere-ocean 4D-Var data assimilation. Monthly Weather Review, 145, 4011 – 4035. doi: 10.1175/MWR-D-16-0284.1

40. Smith, P.J., A.M. Fowler, and A.S. Lawless. 2015. Exploring strategies for coupled 4D-Var data assimilation using an idealised atmosphere–ocean model. Tellus A: Dynamic Meteorology and Oceanography, 67, 1, doi: 10.3402/tellusa.v67.27025

41. Spiller, E.T., Bayarri, M.J., Berger, J.O., Calder, E.S., Patra, A.K., Pitman, E.B., Wolpert, R.L., 2014. Automating Emulator Construction for Geophysical Hazard Maps. SIAM/ASA Journal on Uncertainty Quantification 2, 126–152.

42. Tondeur, M, A. Carrassi, S. Vannitsem and M. Bocquet. On temporal scale separation in coupled data assimilation with the ensemble Kalman filter. J. Stat. Phys., 44. 101171, 2020

43. Tong, X.T., Morzfeld, M., Marzouk, Y.M., 2020. MALA-within-Gibbs samplers for high-dimensional distributions with sparse conditional structure. SIAM Journal of Scientific Computing 42(3), A1765–A1788.

44. Van Leeuwen, P.J., Nerger, L., Potthast. R., Reich. S., and Kunsch, H.R. (2019). A review of Particle Filters for Geoscience applications. Q. J. Royal Meteorol. Soc., doi: 10.1002/qj.3551

45. Vissio, G., Lucarini, V., 2020. Mechanics and thermodynamics of a new minimal model of the atmosphere. The European Physical Journal Plus 135, 807

46. Watkinson, L.R., A.S. Lawless, N.K. Nichols and I. Roulstone, Weak Constraints in Four Dimensional Variational Data Assimilation. Meteorologische Zeitschrift, 16, 2007, 767 - 776.

47. Wursch, M., and G. Craig. 201564. A simple dynamical model of cumulus convection for data assimilation research. Meteorologische Zeitschrift. 23, 483–490.