Exploring a New Computationally Efficient Data Assimilation Algorithm For Ocean Models

Elizabeth Carlson[1], Luke P. Van Roekel[2], Humberto C Godinez[2], Mark R. Petersen[2], and Adam Larios[1]

[1]University of Nebraska - Lincoln
[2]Los Alamos National Laboratory (DOE)

November 22, 2022

Abstract

We present a new data assimilation algorithm known as the Continuous Data Assimilation (CDA) algorithm that has been tested extensively in the mathematical literature and, most recently, in a downscaling simulation in the atmospheric literature. Unlike more common data assimilation methods, the CDA algorithm has an exponential convergence rate and is computationally efficient. This work is the first attempt to demonstrate the viability of the data assimilation algorithm in large-scale ocean models. We implement the CDA algorithm in the Model for Prediction Across Scales - Ocean in an idealized mesoscale eddy test case, demonstrating the ability of the data assimilation algorithm to capture the net effects of unresolved processes in low-resolution models.

Exploring a New Computationally Efficient Data Assimilation Algorithm For Ocean Models

Elizabeth Carlson[1,2], **Luke Van Roekel**[2], **Humberto Godinez**[3], **Mark Petersen**[4], **Adam Larios**[1]

[1]University of Nebraska - Lincoln, Department of Mathematics
[2]Los Alamos National Laboratory, T3: Fluid Dynamics and Solid Mechanics
[3]Los Alamos National Laboratory, T5: Applied Mathematics and Plasma Physics
[4]Los Alamos National Laboratory ,CCS-2: Computational Physics and Methods
[1]1400 R St, Lincoln, NE, 68588
[2]Bikini Atoll Rd., SM 30, Los Alamos, NM 87545
[3]Bikini Atoll Rd., SM 30, Los Alamos, NM 87545
[4]Bikini Atoll Rd., SM 30, Los Alamos, NM 87545

Key Points:

- Explore a new computationally efficient data assimilation method with strong theoretical support for exponential convergence to true state.
- This is the first application of new data assimilation algorithm to a large-scale ocean model.
- The integrated impacts of mesoscale eddies from a high resolution simulation are captured in a low resolution data assimilation simulation.

Corresponding author: Elizabeth Carlson, `elizabeth.carlson@huskers.unl.edu`

Abstract

We present a new data assimilation algorithm known as the Continuous Data Assimilation (CDA) algorithm that has been tested extensively in the mathematical literature and, most recently, in a downscaling simulation in the atmospheric literature. Unlike more common data assimilation methods, the CDA algorithm has an exponential convergence rate and is computationally efficient. This work is the first attempt to demonstrate the viability of the data assimilation algorithm in large-scale ocean models. We implement the CDA algorithm in the Model for Prediction Across Scales - Ocean in an idealized mesoscale eddy test case, demonstrating the ability of the data assimilation algorithm to capture the net effects of unresolved processes in low-resolution models.

Plain Language Summary

Data assimilation describes a set of methods that are used to incorporate observations into models to improve their representation of the current climate. One of the main difficulties with data assimilation for climate models is that measurements are often sparse in space, difficult to obtain (especially in the ocean), affected by instrument error, and not all variables in a system can be measured directly. Many data assimilation methods have been developed, the most popular of which are statistical in nature. However, these algorithms are computationally intensive and nontrivial to implement. In this paper, we test a novel data assimilation algorithm derived from a continuous framework that has been mathematically proven to converge exponentially fast and is simple to implement into existing models. We demonstrate in an idealized climate model that the novel data assimilation algorithm is able to very accurately capture the effects of a high resolution simulation in a low resolution simulation.

1 Introduction

Global ocean models are chaotic and highly sensitive to model inputs, and numerical approximations of critical processes present their own challenges to resolving the fine details of a flow field. Specifically, accurate ocean projections are difficult to obtain due to having incomplete initial conditions and computational limitations that restrict the ability to fully resolve the variety of different length and time scales present in the ocean. One way to mitigate these biases is to use data assimilation to incorporate observed data into the model (see, e.g., Dee (2005); He et al. (2014), and references thereof and therein). Currently the most popular data assimilation techniques are statistical in nature, the most common being the ensemble Kalman filter (EnKF) (see, e.g., Evensen (1997)) and 4DVAR (see, e.g., Trémolet (2007)). These techniques are popular because they exactly minimize the statistical error of linear systems, i.e. they are statistically optimal for these systems. However, these methods are subject to notable difficulties and assumptions, including 1) errors due to the linearization of nonlinear models, 2) difficulty in implementation due to requirement of adjoint model for derivatives, 3) convergence rates that are highly sensitive to the choice of initial conditions, and 4) specialized and highly nontrivial implementation. EnKF is not subject to (1) and (2), but is more likely to diverge (i.e., discount the influence of observations entirely) when there is not a large enough ensemble, when the probability distribution of the ensemble is not Gaussian, or if the model is strongly nonlinear (which distorts the Gaussian distribution of the ensemble), as is the case in standard ocean models (see, e.g., Houtekamer et al. (2014); Trémolet (2007); Evensen (1997); Lawson and Hansen (2004)). Even though for some of these difficulties methods have been devised to get around them, in the present work, we examine a new algorithm for data assimilation in ocean models proposed in (Azouani et al., 2014; Azouani & Titi, 2014) which gets around these difficulties naturally. This new method, known in the literature as the Azouani-Olson-Titi (AOT) or Continuous Data Assimilation (CDA) algorithm, avoids all of the difficulties mentioned above, and has demonstrated robust convergence in a variety of idealized test cases discussed below. Herein, we examine the performance of this algorithm in the context of a real ocean model.

History of a New Data Assimilation Algorithm

The CDA algorithm is based on the idea of feedback-control at the partial differential equation level. It bears a superficial resemblance to the so-called nudging or Newtonian relaxation methods introduced in (Anthes, 1974; Hoke & Anthes, 1976), with a crucial difference being the feedback control term is formed by spatial interpolation of the observed error, allowing for observations that are sparse in space. It is not prone to any of the difficulties associated with standard data assimilation methods. CDA is mathematically proven to converge exponentially fast in time to the data for a variety of fluid equations and in multiple settings, subject to only two conditions, making it very robust and versatile. Furthermore, it is very inexpensive to implement computationally. (See Section 2.2 for a more detailed explanation of these properties.) In this section, we will focus on the robustness of this method in a variety of idealized cases, highlighting which ones most directly relate to the needs of large-scale ocean modeling. In the following section, we will focus on how data assimilation is generally used to improve ocean modeling and how the CDA algorithm may also be used in these settings.

The CDA algorithm has been adapted to a wide variety of equations, including the 3D Navier-Stokes equations (NSE) (Biswas & Price, 2020; Clark Di Leoni et al., 2020), the 3D primitive equations (Pei, 2019), Bénard convection (Farhat et al., 2020, 2016b; Altaf et al., 2017; Farhat et al., 2015, 2017; Farhat, Johnston, et al., 2018), magnetohydrodynamic equations, (Hudson & Jolly, 2019), surface quasi-geostrophic equations, (Jolly et al., 2019, 2017), the Kuramoto-Sivashinsky equations, (Lunasin & Titi, 2017), the Brinkman-Forchheimer-Darcy model (Markowich et al., 2016), reaction-diffusion equations, (Azouani & Titi, 2014; Larios & Victor, 2021), and the Weather Research and Forecasting (WRF) model (Desamsetti et al., 2019). The CDA algorithm is also quite robust. Specifically, the exponential rate of convergence either to 0 or to a controllable error has been proven to hold in the the context of: sparse-in-time observations (Foias et al., 2016; Celik et al., 2019), statistical solutions (Biswas et al., 2018), systems assimilating time-averaged data (Jolly et al., 2019), systems with noisy observations (Bessaih et al., 2015), time-averaged data (Jolly et al., 2019), measurements in only certain components of the flow field (Farhat et al., 2015, 2016c, 2016a, 2016b, 2017; Farhat, Johnston, et al., 2018), approximations with reduced-order-models (Zerfas et al., 2019; Clark Di Leoni et al., 2020; García-Archilla, Novo, & Rubino, 2020), approximations with regularized models (D. A. F. Albanez & Benvenutti, 2018; D. A. Albanez et al., 2016; Farhat, Lunasin, & Titi, 2018; Larios & Pei, 2018), Leray weak solutions of 3D NSE (Biswas & Price, 2020), moving observers (Larios & Victor, 2021; Biswas et al., 2020), finite element methods (Larios et al., 2018; García-Archilla, Novo, & Titi, 2020; Gardner et al., 2020), spectral Galerkin discretization (Ibdah et al., 2019), post-processing Galerkin methods (Mondaini & Titi, 2018), systems with incorrect parameters (Carlson et al., 2020; Farhat et al., 2020), and systems with data given on a subdomain (Biswas et al., 2020).The algorithm has also seen several modifications (e.g., nonlinear feedback control) aiming at improved convergence rates (Larios & Pei, 2017; Rebholz & Zerfas, 2018).

The robustness of the algorithm demonstrated by the body of work is significant for the ocean community. For example, since not all state variables can be measured directly (specifically, temperature and salinity are measured, while velocities are not). Some idealized computational studies demonstrate that assimilating only temperature measurements may lead to failure in recovering the solution (Altaf et al., 2017), but this depends on the model used (Farhat et al., 2016c). However, convergence can be obtained with data that is blurred in time, i.e. averaged over a small time interval, (Jolly et al., 2019), a common problem that arises when taking measurements. Exponential convergence can also be proven when assimilating only two-dimensional surface data into the three-dimensional surface quasi-geostrophic equations (Jolly et al., 2017). This is extremely important in climate modeling as often, for the ocean especially, surface data is usually the most abundant observed data. Another difficulty that is present in obtaining measurements is obtaining sufficiently many observations over the full domain, and in (Biswas et al., 2020) it was recently proven (although in the context of the 2D NSE) exponential convergence can be obtained for observations only given on a subdomain.

How the CDA Algorithm May Address Data Assimilation Needs in Real-World Climate Modeling

In summary, the CDA algorithm has demonstrated significant versatility in idealized settings. The main drawback of this algorithm is that although it is statistically robust it is not necessarily statistically optimal; specifically, when assimilating noisy data, the error is bounded in terms of the trace of the covariance matrix of the assimilated variables (Bessaih et al., 2015). However, most statistical data assimilation algorithms (with the exception of EnKF) lack a mathematical theory to guarantee a statistically optimal convergence of the output for nonlinear models due to the necessary linearization of nonlinear models. Thus, the demonstrated advantages of the CDA algorithm in a variety of different settings make CDA a potentially viable alternative to current data assimilation methods for ocean modeling. Hence, since for climate modeling data assimilation is most often used 1) regionally in a downscaling setting, 2) in conjunction with data to initialize a simulation, and 3) to generate more accurate ensembles for climate projections, we describe how the CDA algorithm can address each of these aspects.

First, the CDA algorithm is most directly applicable in the downscaling setting, and it outperforms the more simple and popular downscaling method of nudging. Specifically, to the best of the authors' knowledge, the first, and currently only, real data study done utilizing CDA was in the recently published Desamsetti et al. (2019), which compared the CDA algorithm with grid and spectral nudging methods implemented in the Weather Researching and Forecasting model for downscaling. They conclusively demonstrated that the CDA algorithm performed better than or comparable to both grid and spectral nudging. It was only comparable to a spectral nudging method with a well-chosen cut-off wave number, and was considerably less expensive due to the lack of Fourier transforms.

Data assimilation is also a tool used to identify and reduce model bias. This is done either (1) via construction of a more balanced initial climate model state for long-term climate projections or (2) by improving estimates of uncertain parameters in model parameterizations. The CDA algorithm is potentially applicable for both methods as it provides convergence up to a quantifiable error in the data, discretization, and/or model. This is demonstrated in Carlson et al. (2020), Farhat et al. (2020), and Larios and Pei (2018). The paper Carlson et al. (2020) demonstrates that given an incorrect parameter the error between the true and the CDA solutions is controlled by the error in the parameter. The paper Farhat et al. (2020) demonstrates a slightly more complex case, proving that the Rayleigh-Bénard system with finite Prandtl number can be used to assimilate data into a system with infinite Prandtl number, with exponential convergence up to error controlled by the Prandtl number. In a slightly different vein, the paper Larios and Pei (2018) considers the data from the 2D NSE and assimilates it into the approximating model of the Navier-Stokes Voigt (NSV) equations (a regularized version of the NSE) using the CDA algorithm. The solution to the approximate data assimilation system converges exponentially fast to the true solution of the NSE up to an error determined solely by the parameter α that regularizes the system. Each of these works demonstrates how the CDA algorithm can be used to converge up to model bias and thus making model bias identifiable. Furthermore, model bias that is due to incorrect parameters can be corrected via parameter recovery, on which the CDA algorithm has some nascent literature in the ideal setting. In Carlson et al. (2020), the CDA algorithm was applied to the 2D incompressible Navier-Stokes Equations assuming the modeler has incorrect knowledge of the Reynolds number. Additionally, CDA is amenable to the existing method of applying statistical post-processing techniques that compare the model to the observed data to correct bias for parameter calibration (Durai & Bhradwaj, 2014; Glahn & Lowry, 1972; Lawson & Hansen, 2004; Woodcock & Engel, 2005; Gneiting et al., 2005; Delle Monache et al., 2006; Raftery et al., 2005; Bakhshaii & Stull, 2009; Du & Zhou, 2011; Cui et al., 2012; Satterfield & Bishop, 2014).

All of the works mentioned in the previous paragraph (excluding Desamsetti et al. (2019)) are either theoretical or have been tested in computationally idealized settings, making this work a timely application to more complex models. The goal of this paper is to determine how using the CDA algorithm in low resolution simulations allows the simulation to capture the net effects of resolved mesoscale eddies in a high resolution simulation. Specifically, we implement the CDA algorithm in the unstructured Model for Prediction Across Scales-Ocean (MPAS-O), and use the model output of a high resolution, eddy resolving simulation as a proxy for real data. The CDA algorithm is used in a low resolution simulation. We demonstrate the effectiveness of the algorithm in recovering net effects of unresolved features present in the high resolution simulation on the low resolution grid, consistent with Carlson et al. (2020).

The remainder of the paper is organized as follows, in Section 2, we describe the CDA algorithm, the idealized test case that is the basis for this study, and compare CDA to nudging. In Section 3, we present the results of the CDA simulation, briefly compare it to corresponding nudging simulations, and discuss the different metrics by which we quantify the effect of the CDA algorithm on the low resolution simulation. Finally, we summarize and discuss extensions of this study in Section 4.

2 Setup and Background

In this section, we describe the CDA algorithm, provide a brief description of the difference between the CDA algorithm and nudging, and present the setup of the test case.

2.1 CDA Algorithm

To describe the CDA algorithm as originally posed in Azouani et al. (2014), suppose a physical system is perfectly modeled by the differential equation

$$\begin{aligned}\frac{d\mathbf{u}}{dt} &= F(\mathbf{u})\\ \mathbf{u}(0) &= \mathbf{u}_0\end{aligned}$$

where F represents the physics evolving the system in time and $\mathbf{u}_0$ represents an unknown initial condition. For this system, we obtain a discrete set of operations which we interpolate in space, $I_\delta(\mathbf{u})$, where, in our case, δ represents the spatial distance between data points. In the theoretical works on CDA, I_δ is assumed to be a linear operator satisfying a particular bound; examples of these types of linear operators include nodal interpolation, (bi)linear interpolation, volume interpolation, and Fourier truncation, among others. The interpolated data is incorporated into the model using a feedback control term

$$\begin{aligned}\frac{d\mathbf{v}}{dt} &= F(\mathbf{v}) + \mu(I_\delta(\mathbf{u}) - I_\delta(\mathbf{v}))\\ \mathbf{v}(0) &= \mathbf{v}_0\end{aligned}$$

where $\mu > 0$ is a positive relaxation parameter and $\mathbf{v}_0$ is any (sufficiently smooth) initial condition (in many of the idealized computational tests of this data assimilation method, this initial condition is often taken to be 0). So long as μ is sufficiently large and δ is sufficiently small, the solution of the assimilated system converges exponentially fast to the original system once the solution to the original system is sufficiently close to the attractor. In other words, if the reference field, represented by $\mathbf{u}$, is sufficiently developed so that our observations are taken from a turbulent system, then the solution $\mathbf{v}$ to the assimilated system of equations converges exponentially fast to $\mathbf{u}$, the reference field.

Note that the only requirements for convergence of the CDA algorithm are a sufficiently large positive relaxation parameter μ and sufficiently many evenly spaced observations. These restrictions are mathematically very strong but in practice the CDA algorithm converges even when the restrictions on μ and the number of observations is significantly less than what is required in the mathematical theory; for example, in Gesho (2013), the 2D incompressible Navier-Stokes equations were simulated in the turbulent regime with only 49 observation points and $\mu = 24$, multiple orders of magnitude less than required by the theory. For other examples, see computational studies in, e.g., Altaf et al. (2017); Carlson et al. (2020); Desamsetti et al. (2019); Farhat et al. (2020); Farhat, Johnston, et al. (2018); García-Archilla, Novo, and Titi (2020); García-Archilla, Novo, and Rubino (2020); Gardner et al. (2020); Hudson and Jolly (2019); Larios and Pei (2017); Lunasin and Titi (2017); Zerfas (2019); Zerfas et al. (2019).

2.2 Interpolation: The Difference between the CDA Algorithm and Nudging

The CDA algorithm deceptively looks like the nudging algorithm, but the main difference lies in the way the feedback term is constructed. One of the most popular methods of nudging was presented by Stauffer and Seaman (1990), where the observational dataset is interpolated to the entire model grid (Bullock Jr. et al., 2018; Zheng & Weisberg, 2012; Zhang et al., 2016; Weisberg et al., 2009; Baptista et al., 2005; Fortunato et al., 2014; Robinson et al., 2011; Kilgren, 2006; F. Ye, 2017; Abbasi et al., 2018; Ding et al., 2012; X. Ye et al., 2020; Ge et al., 2020; Pringle, 2006; Fujisaki-Manome et al., 2017; Cazenave et al., 2018; Wei et al., 2014; Peng et al., 2014; Ge et al., 2013; Cowles et al., 2008; Fennel et al., 2016; Foreman et al., 2009). We note that Desamsetti et al. (2019) used this particular version of grid nudging for their paper, and their results demonstrated that the CDA algorithm was superior, and here we briefly explain at least one reason why. Recall that the CDA algorithm compares the interpolation of the observations as well as the interpolation of the model data, with system of equations given by

$$
\begin{aligned}
\mathbf{v}_t &= F(\mathbf{v}) + \mu(I_\delta(\mathbf{u}) - I_\delta(\mathbf{v})) \\
\mathbf{v}(0) &= \mathbf{v}_0.
\end{aligned}
$$

Using this same framework, the method of nudging in Stauffer and Seaman (1990) can be written as

$$
\begin{aligned}
\mathbf{v}_t &= F(\mathbf{v}) + \mu(I_\delta(\mathbf{u}) - \mathbf{v}) \\
\mathbf{v}(0) &= \mathbf{v}_0.
\end{aligned}
$$

This implies one is using the observations to inform the model everywhere in space, but it does so inconsistently. In particular, if at some time t_0 one actually obtained $\mathbf{u}(t_0) = \mathbf{v}(t_0)$, the true physical evolution equations would not be the same as the nudged evolution equations, which contain the an error term $\mu(I_\delta(\mathbf{u}) - \mathbf{v}) \neq 0$, errors which increase with increasing values of μ. In contrast, the CDA algorithm requires that the CDA model data must be interpolated *to the same grid* as the observations in order to make an accurate comparison, so that when $\mathbf{u}(t_0) = \mathbf{v}(t_0)$, the true physical evolution equations and the CDA evolution equations are identical.

Implementing the CDA algorithm becomes more technical in the setting where the observed grid is not a subset of the model grid. This difference is important yet subtle, and a more detailed explanation is provided in Appendix A

2.3 SOMA: Simulation Ocean Mesoscale Activity Test Case Setup and Numerics

The test case we consider is the Simulating Ocean Mesoscale Activity (SOMA) test case (Wolfram et al., 2015), which simulates a wind-driven double gyre similar to the subtropical and subpolar North Atlantic. The purpose of this test case is to provide a simplified analog of the mesoscale eddy rich Atlantic ocean. The original test case is described in detail in Appendices A and B of Wolfram et al. (2015) and utilizes the Model for Prediction Across Scales - Ocean (MPAS-O), the ocean component of the Energy Exascale Earth System Model (E3SM), created by the US Department of Energy (E3SM Project, 2018; Golaz et al., 2019; Petersen et al., 2019).

MPAS-O models the primitive equations on a finite volume spatial discretization on an unstructured mesh, which is built using spherical centroidal Voronoi tesselations. Temporal discretization is done via the split-explicit method. Unlike Wolfram et al. (2015), the simulations in this paper are run in an isopycnal configuration with vertical layers specified by the density, to decrease computational cost and more easily visualize isopycnal diffusivity. The simulations use 10 vertical layers with a density range of $1026.36\frac{\text{kg}}{m^3}$ to $1039.95\frac{\text{kg}}{m^3}$.

For reference, we write the CDA algorithm applied to the MPAS equations (Petersen et al., 2015; Ringler et al., 2013):

$$\frac{\partial \mathbf{u}}{\partial t} + \eta \mathbf{k} \times \mathbf{u} + w\frac{\partial \mathbf{u}}{\partial z} = -\frac{1}{\rho_0}\nabla p - \frac{\rho g}{\rho_0}\nabla z^{\text{mid}} - \nabla K + \mathbf{D}_h^u + \mathbf{D}_v^u + \mathcal{F}^u + \mu_{\mathbf{u}}(I_\delta(\mathbf{u}_{\text{ref}}) - I_\delta(\mathbf{u})) \tag{1}$$

$$\frac{\partial h}{\partial t} + \nabla \cdot (h\overline{\mathbf{u}}^z) + w|_{z=s^{\text{top}}} - w|_{z=s^{\text{bot}}} = 0 \tag{2}$$

$$\frac{\partial}{\partial t}(h\overline{\phi}^z) + \phi w|_{z=s^{\text{top}}} - \phi w|_{z=s^{\text{bot}}} = D_h^\phi + D_v^\phi + \mathcal{F}^\phi + h\mu_{\overline{\phi}^z}(I_\delta(\overline{\phi}^z_{\text{ref}}) - I_\delta(\overline{\phi}^z)). \tag{3}$$

We briefly explain the terms in the above equation.

- Equation (1) represents the evolution for horizontal normal velocity, equation (2) represents the layer thickness evolution, and equation (3) represents the tracer evolution.
- The diffusive terms $\mathbf{D}_h^u, \mathbf{D}_v^u, D_h^\phi$, and D_v^ϕ can be chosen so that the model can employ different types of diffusion, each serving specific purposes (for details, see (Ringler et al., 2013)).
  - For horizontal diffusion for the tracer term, this is 0 since we are assuming that diffusion in the advection scheme provides the horizontal diffusion. For horizontal diffusion on the velocity, harmonic diffusion suppresses eddies and diffuses jets and thus biharmonic diffusion is employed (for more details, see (Hecht et al., 2008)). The term is given explicitly by

    $$\mathbf{D}_h^u = \nabla \cdot \left(\frac{\nu_h}{\rho_m^{3/4}} \nabla[\nabla \cdot (\nabla \mathbf{u})] \right),$$

    where ν_h is the horizontal viscosity and ρ_m is the mesh density.
  - Vertical mixing follows the scheme outlined in (Pacanowski & Philander, 1981). For more details on the choice of viscosities, see (Wolfram et al., 2015).
- The term K represents the kinetic energy.
- The term $\overline{(\cdot)}^z$ represents a vertical average.
- The term $\mu_{\mathbf{u}} > 0$ and $\mu_{\overline{\phi}^z} > 0$ represent the CDA positive relaxation parameters.
- For a description of the boundary conditions employed, see Wolfram et al. (2015).

In this simulation we assimilate the horizontal velocity components, temperature, and salinity. Note that generally the CDA algorithm is implemented in the corresponding evolutionary equation, but the tracers are evolved via a system of equations involving the layer thickness h. Thus, using Equations (2) and (3) in conjunction with the product rule, we obtain the term for the assimilation as presented in Equation (3).

The goal of this study is to determine how using the CDA algorithm in low resolution simulations allows the simulation to capture the net effects of resolved mesoscale eddies in a high resolution simulation. Since the Rossby radius of deformation, the quantity approximating the length scale of the mesoscale eddies, is approximately 30 km (Wolfram et al., 2015) in this test case, the high resolution reference simulation is run at 8 km horizontal resolution (base mesh cell count: 122,807, culled mesh cell count: 88,056) to fully resolve the mesoscale eddies, and the low resolution simulations are run at 32 km horizontal resolution (base mesh cell count: 8,652, culled mesh cell count: 6,021), which precludes mesoscale eddies. At 32 km horizontal resolution, we run a control simulation and a CDA simulation. Observations are taken from the reference simulation and coarsened to a $0.5° \times 0.5°$ rectangular grid and also to a $1° \times 1°$ rectangular grid. We chose the $0.5° \times 0.5°$ grid as this is close to the resolution of the low resolution simulation, and the $1° \times 1°$ grid as it is common for observational data to be provided on this grid. In the low resolution data assimilation simulations, we do not expect to recapture the full mesoscale eddy field when the simulations are performed at low resolution, but the goal of this project is to determine how well the net effects of the mesoscale eddies can be captured in a low resolution simulation by assimilating the high-resolution data.

Given that MPAS-O utilizes a horizontally unstructured mesh, we implement the CDA algorithm following the method outlined in Section 2.2 and the nudging method of Stauffer and Seaman (1990) for a brief comparison in the context of an unstructured grid.

The reference and control simulations are run 5 1/2 simulated years to reach a near steady state (determined to be when the kinetic energy begins to equilibrate) from zero initial velocity. The data assimilation systems of equations are also initialized with zero initial velocity, but we begin assimilating the reference data from halfway through the fifth simulated year. This causes the CDA simulations to equilibrate within four simulation weeks, and to ensure this we run the simulations for eight simulation weeks. We chose the coefficients so that $\mu_{\overline{\phi}^z} = \mu_{\mathbf{u}} = \mu$. We drive the simulation using three different positive relaxation coefficients: $\mu = 2 \times 10^{-5}, 1 \times 10^{-5}$, and 2×10^{-6}, which correspond to forcing on time scales of approximately 14 hours, 27 hours, and 5 1/2 days. These particular positive relaxation coefficients were chosen to explore the range of effective coefficients: $\mu > 2 \times 10^{-5}$ did not significantly change the root mean square error, and $\mu \gtrsim 1 \times 10^{-4}$ caused blow-up in the magnitude of the velocities. Coefficients below 2×10^{-6} demonstrated in preliminary tests to be qualitatively ineffective in forcing the simulation to the desired quantities, which was reasonable considering the decorrelation time scale was computed to be approximately 10 days (Wolfram et al., 2015). All time averages used in the analyses are taken over the last two weeks of the eight-week simulations.

3 Results & Analysis

In this section, we demonstrate the convergence of the CDA algorithm by analyzing velocity magnitudes, variance of the velocity difference, and the root mean square error (RMSE). Figures 1 and 2 present the horizontal velocity magnitudes along the four least dense isopycnal surfaces (columns). All data is interpolated to the $0.5° \times 0.5°$ (Figure 1) and $1.0° \times 1.0°$ (Figure 2) observation grids for comparison. We note that, for all chosen nudging coefficients, the magnitude of the velocity in the CDA simulations is qualitatively indistinguishable from the reference simulation as represented on the $0.5° \times$

0.5° and $1.0^\circ \times 1.0^\circ$ observational grids. Figures 3 and 4 provide a clearer picture of the differences between the CDA and reference simulations by plotting the variance of the difference between the velocities in the CDA simulations and the reference simulation on the $0.5^\circ \times 0.5^\circ$ and $1.0^\circ \times 1.0^\circ$ grids, respectively. In these figures, we note that the stronger the positive relaxation coefficient μ, the better the convergence. Furthermore, confirming the results of Desamsetti et al. (2019), CDA performs as well as or better than nudging, with the most significant differences seen when the observations are given on the coarser $1.0^\circ \times 1.0^\circ$ grid. Table 1 shows the exact relative improvement of the CDA algorithm RMSE with respect to the nudging RMSE. The RMSE for the CDA algorithm is approximately 20% better than the RMSE for nudging with the observations given on the $0.5^\circ \times 0.5^\circ$ grid, and it is approximately 30% better than the RMSE for nudging with the observations given on the $1.0^\circ \times 1.0^\circ$ grid.

Remark. In our simulations nudging is not only less accurate but it is also less robust than the CDA algorithm. Specifically the simulation fails with the temperature variable becoming NaN in the case where $\mu = 1 \times 10^{-4}$ with observations given on the $1.0^\circ \times 1.0^\circ$ grid. The CDA algorithm tested with the same choices for μ given the same observations did not have this issue, but it also did not significantly improve the RMSE compared to the case where $\mu = 2 \times 10^{-5}$.

Remark. Future work should explore the optimality of mixed choices of $\mu_{\mathbf{u}}$ and $\mu_{\overline{\phi}^z}$. Analytically, the proof of convergence indicates that the stronger both μ's are taken, the faster it converges, hence it is not clear that a mixed choice of μ's would improve the convergence rate. However, the choice of numerical discretizations may affect the convergence. A preliminary test choosing $\mu_{\mathbf{u}} = 5 \times 10^{-5}$ and $\mu_{\overline{\phi}^z} = 1 \times 10^{-4}$ with observations given on the $0.5^\circ \times 0.5^\circ$ grid improves the velocity RMSE and passive tracer diffusion along isopycnals.

μ	Grid	% Improvement
2×10^{-5}	$0.5^\circ \times 0.5^\circ$	+23%
1×10^{-5}	$0.5^\circ \times 0.5^\circ$	+19%
2×10^{-6}	$0.5^\circ \times 0.5^\circ$	+5%
2×10^{-5}	$1.0^\circ \times 1.0^\circ$	+30%
1×10^{-5}	$1.0^\circ \times 1.0^\circ$	+29%
2×10^{-6}	$1.0^\circ \times 1.0^\circ$	+11%

Table 1. Relative Improvement of the CDA Algorithm Over Nudging

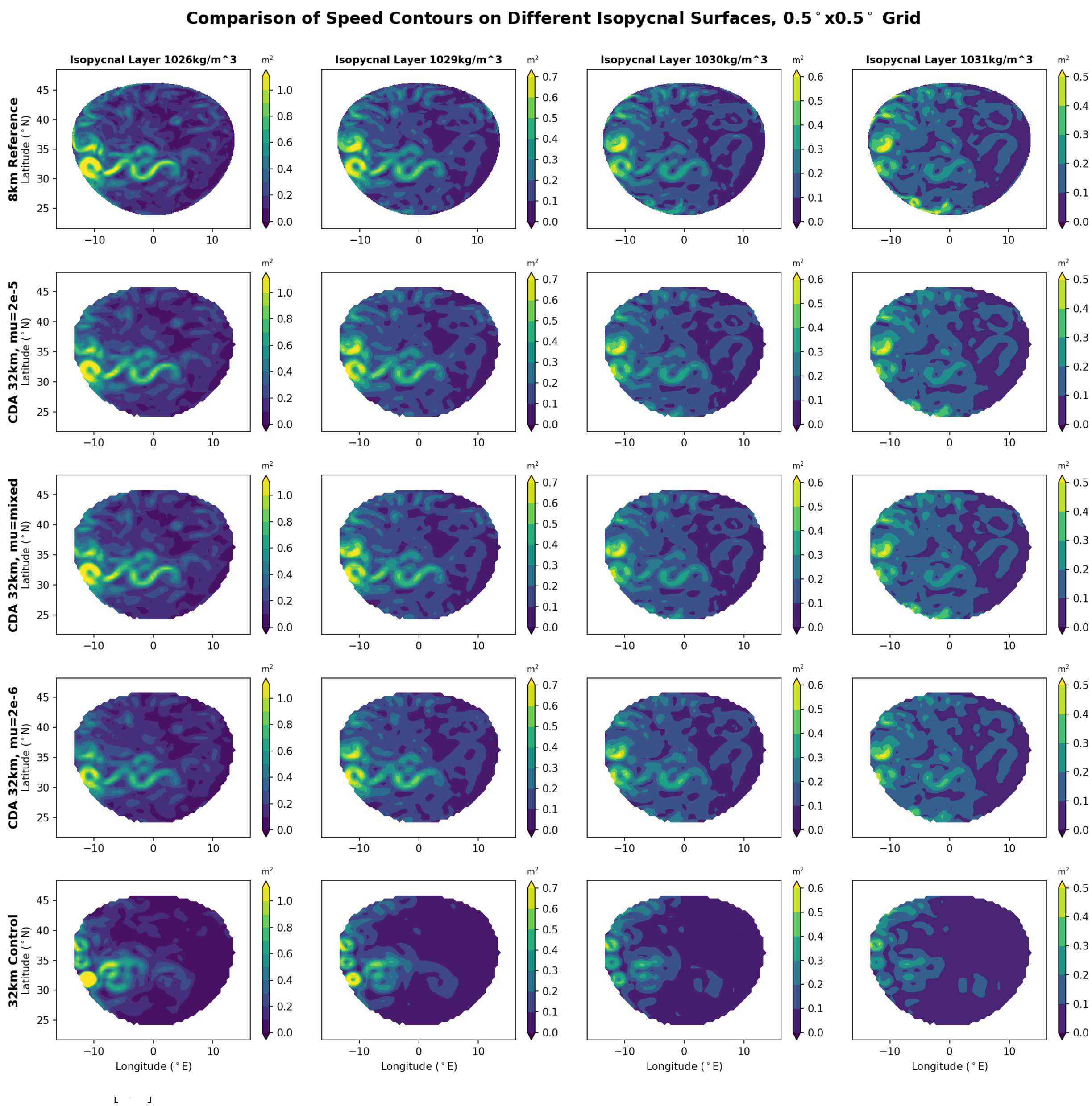


Figure 1. Comparison of Current Speed, Averaged over 2 weeks: Top Row– 8 km Reference Simulation, 0.5° × 0.5° resolution, Middle Rows– 32 km CDA Simulation, Bottom Row – 32 km Control Simulation

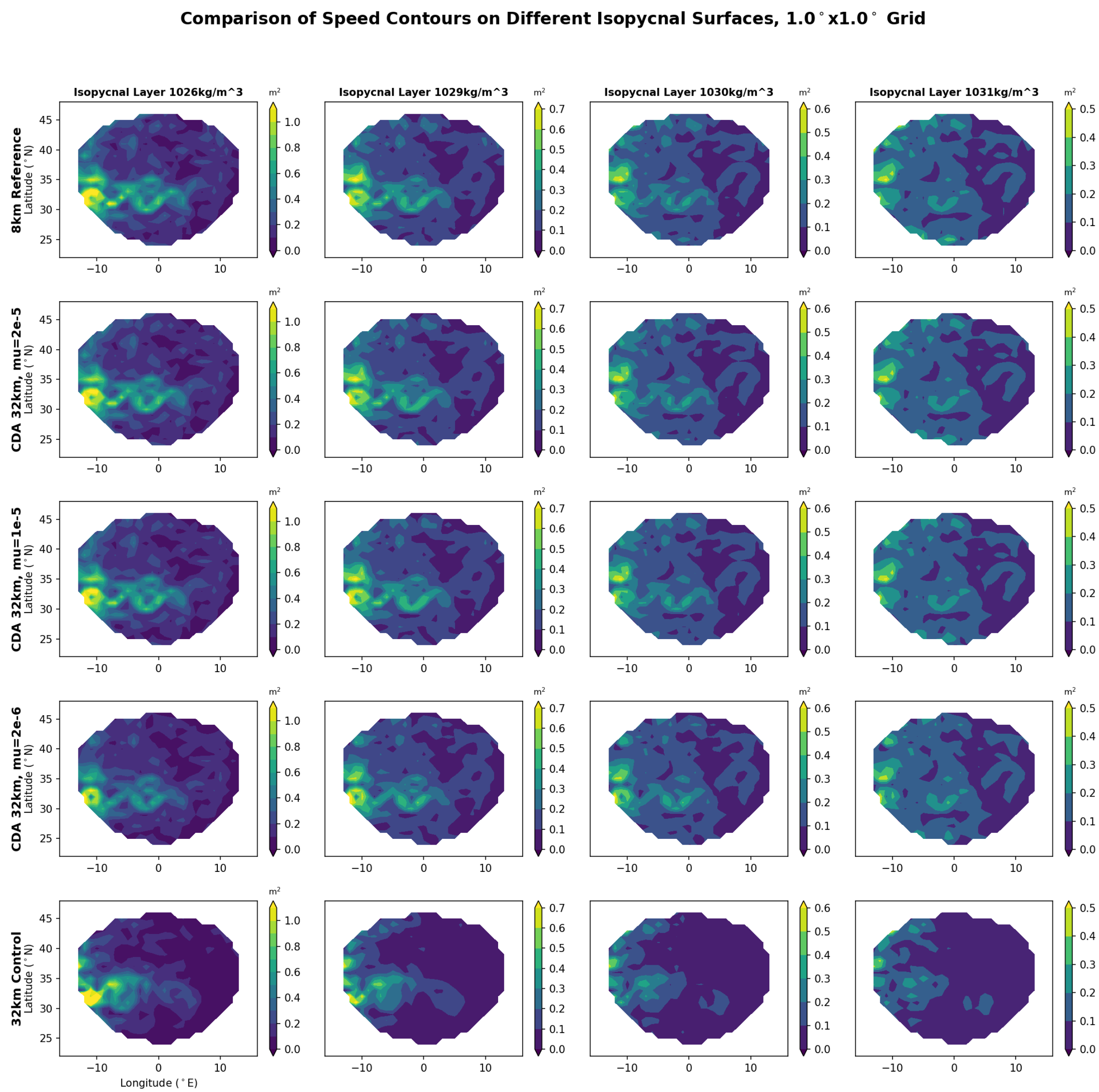
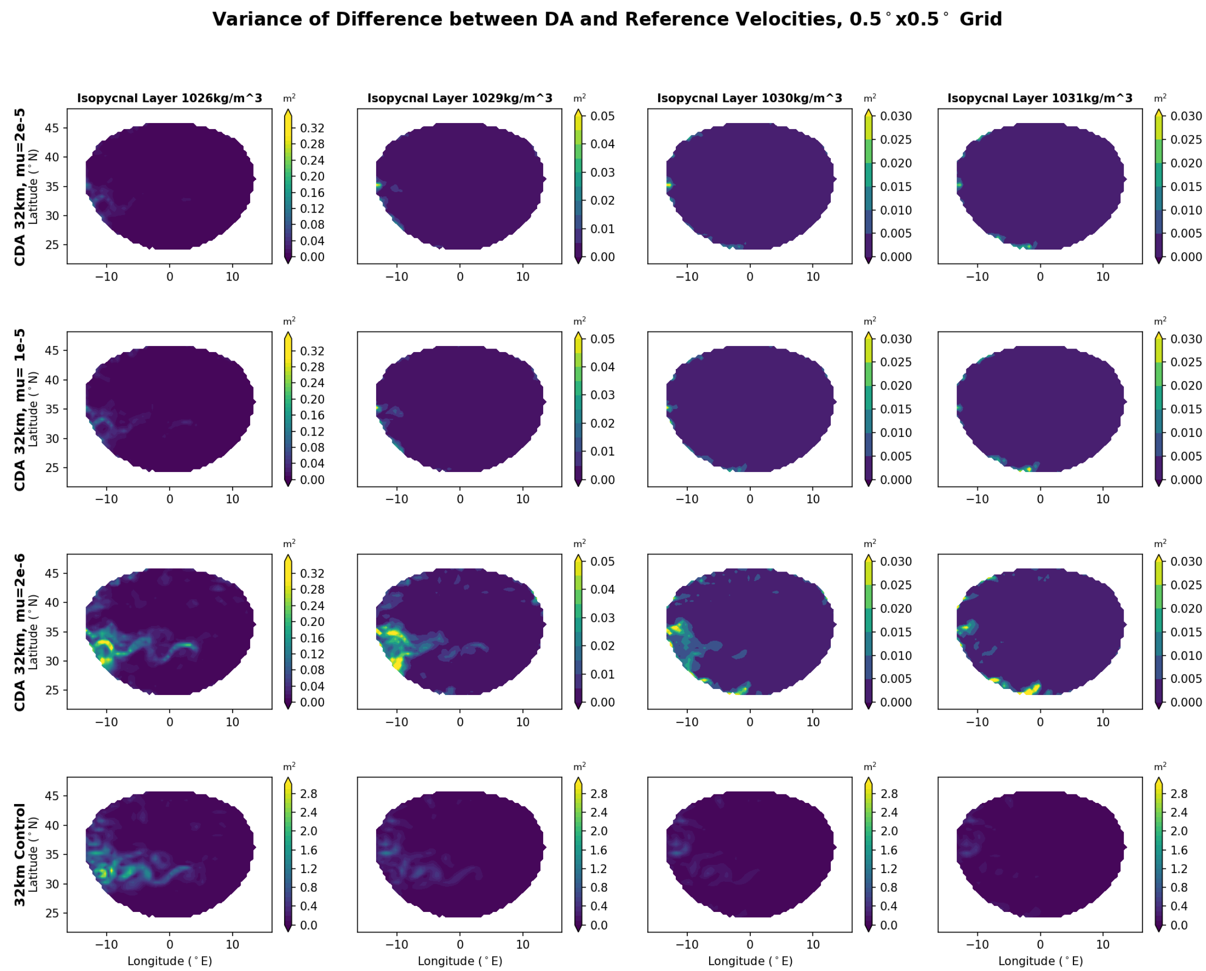
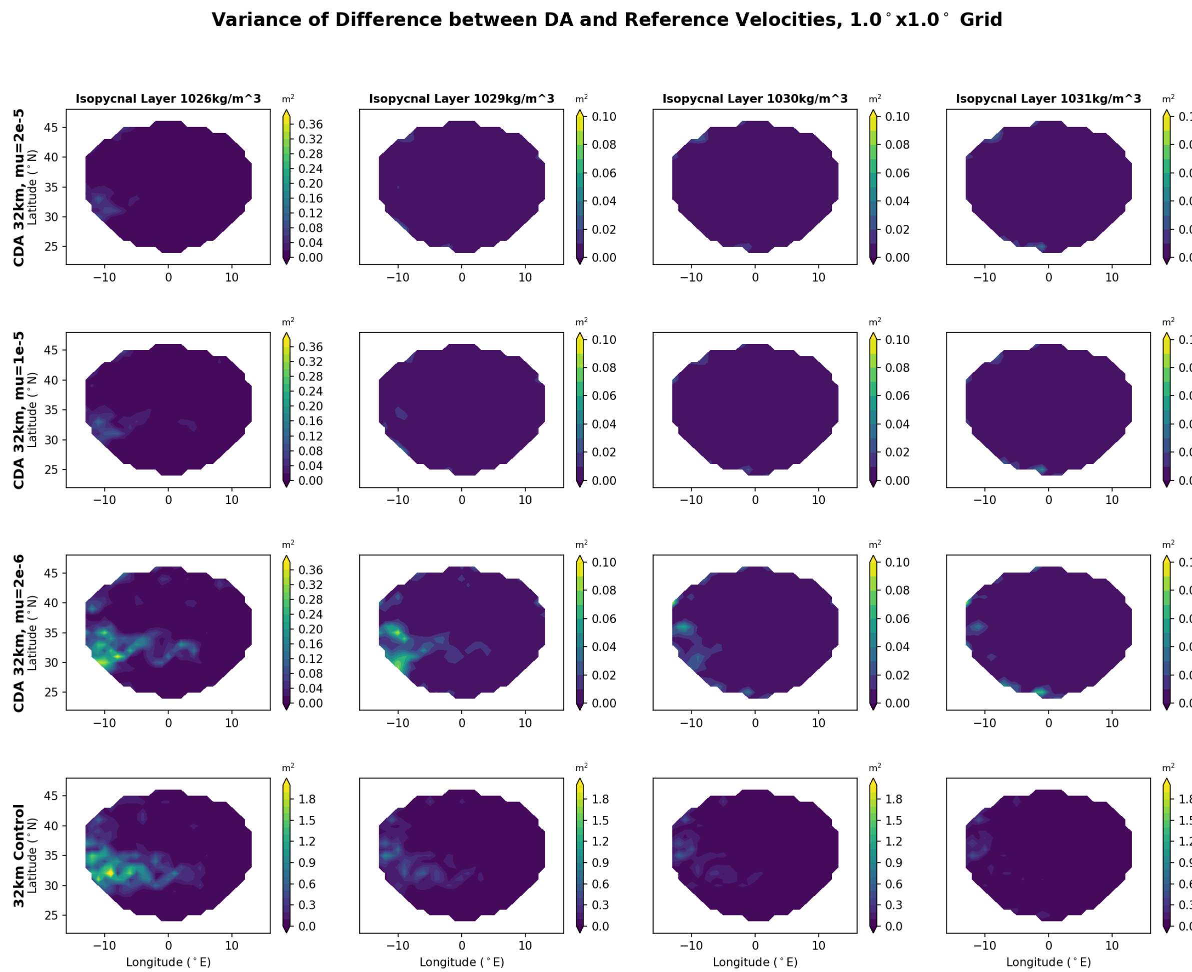


Figure 2. Comparison of Current Speed, Averaged over 2 weeks: Top Row– 8 km Reference Simulation, 1.0° × 1.0° resolution, Middle Rows– 32 km CDA Simulation, Bottom Row – 32 km Control Simulation



[!th]

Figure 3. Plots of the square of the magnitude of the difference of the horizontal velocities of the CDA data assimilation simulation and the reference simulation, assimilating observations on the regular $0.5^\circ \times 0.5^\circ$ grid. Note the difference in the colorbar range between the CDA and control simulations.



[!th]

Figure 4. Plots of the square of the magnitude of the difference of the horizontal velocities of the CDA data assimilation simulation and the reference simulation, assimilating observations on the regular $1.0^\circ \times 1.0^\circ$ grid. Note the difference in the colorbar range between the CDA and control simulations.

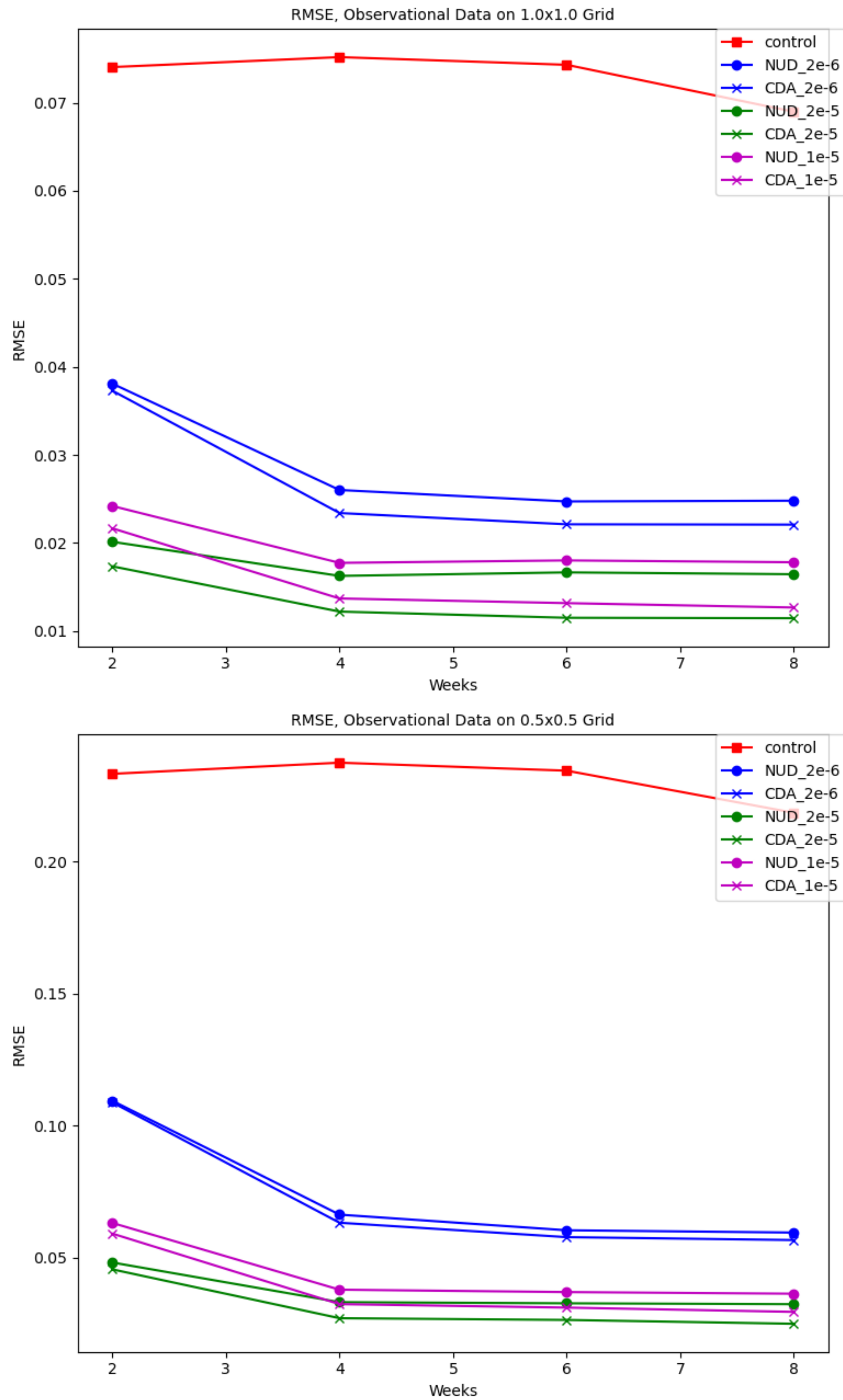


Figure 5. RMSE of horizontal velocity, averaged every 2 weeks from the start of the data assimilation simulations, with observational data given on a $1.0^\circ \times 1.0^\circ$ grid and $0.5^\circ \times 0.5^\circ$ grid, respectively.

Via the CDA algorithm, we expect to find the net effects of mesoscale eddies to be improved in the data assimilation simulation. Hence, we analyze how the CDA algorithm improves the effects of eddies via analysis of isopycnal diffusivity and eddy kinetic energy (EKE), defined by

$$\mathrm{EKE} = \frac{1}{2}\left(\overline{(u-\overline{u})^2} + \overline{(v-\overline{v})^2}\right),$$

where $\mathbf{u} = (u, v)$ is the horizontal velocity vector and the averaging is a time average over the final two weeks of the CDA simulation. We also analyze mass transport north and south of the jet to demonstrate the effects of resolution on the behavior of the double gyre and the CDA algorithm's ability to capture the correct mass transport. Overall, the CDA algorithm is determined to outperform the control simulation and produces results that are qualitatively similar to that of the high resolution reference simulation on the observation grids.

In Figures 6 and 7, which show the simulated EKE, we note that the the accuracy as well as the distribution of EKE is significantly improved in the CDA simulation. We particularly note the spatial representation of the EKE in the data assimilation simulations is more representative of the reference simulation than the control simulation. This is confirmed quantitatively in Table 2, which presents the RMSE for the EKE, where CDA yields an improvement up to two orders of magnitude relative to the control simulation (an approximate 100% relative improvement).

μ	Grid	RMSE	% Improvement
2×10^{-5}	$0.5° \times 0.5°$	5.37×10^{-6}	+96%
1×10^{-5}	$0.5° \times 0.5°$	8.18×10^{-6}	+94%
2×10^{-6}	$0.5° \times 0.5°$	3.70×10^{-5}	+73%
control	$0.5° \times 0.5°$	1.36×10^{-4}	–
2×10^{-5}	$1.0° \times 1.0°$	9.01×10^{-6}	+93%
1×10^{-5}	$1.0° \times 1.0°$	1.39×10^{-5}	+89%
2×10^{-6}	$1.0° \times 1.0°$	5.48×10^{-5}	+57%
control	$1.0° \times 1.0°$	1.29×10^{-4}	–

Table 2. EKE RMSE with respect to the observations, with % improvement relative to the control simulation.

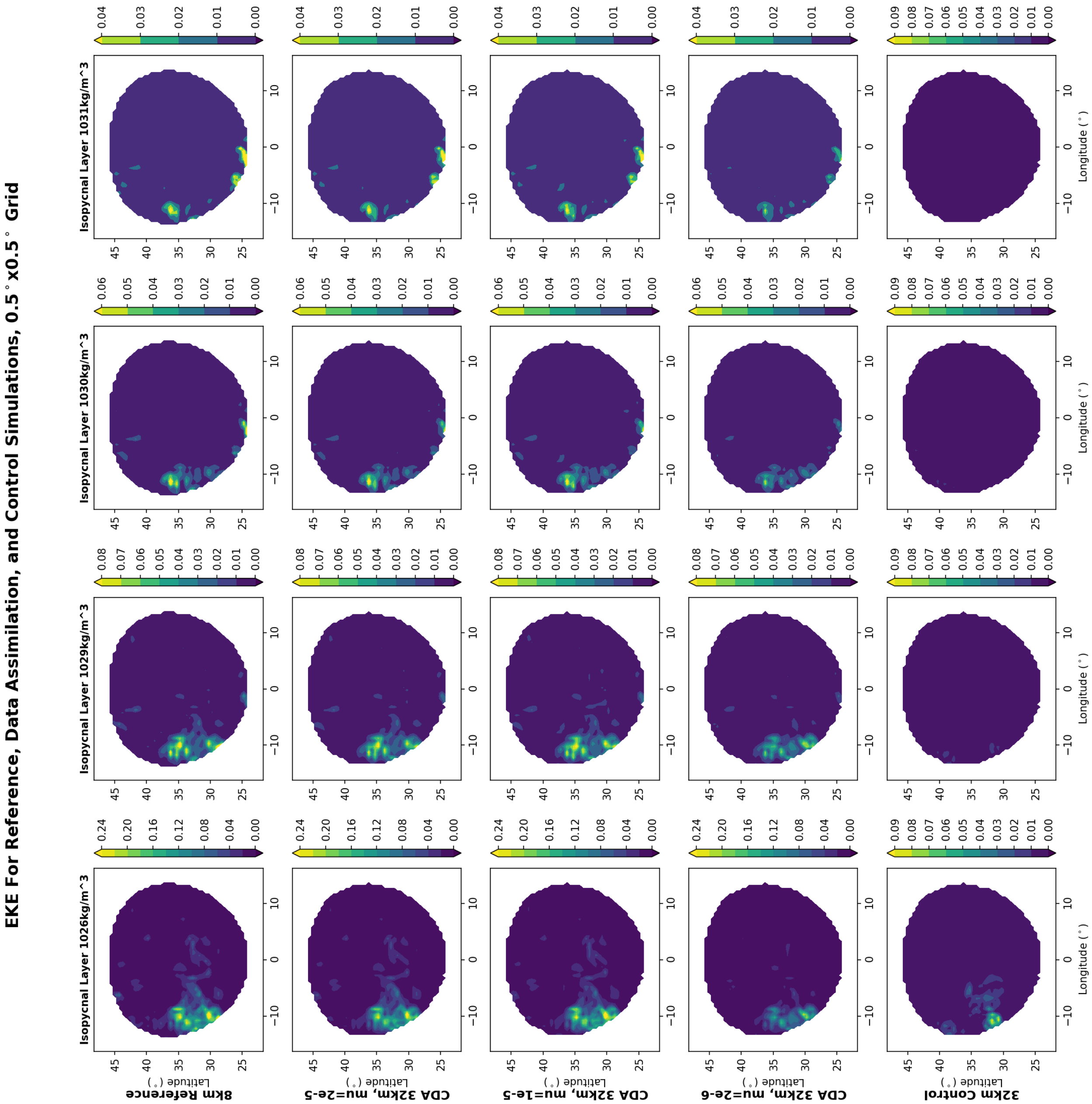


Figure 6. Plot of EKE on $0.5^\circ \times 0.5^\circ$ grid

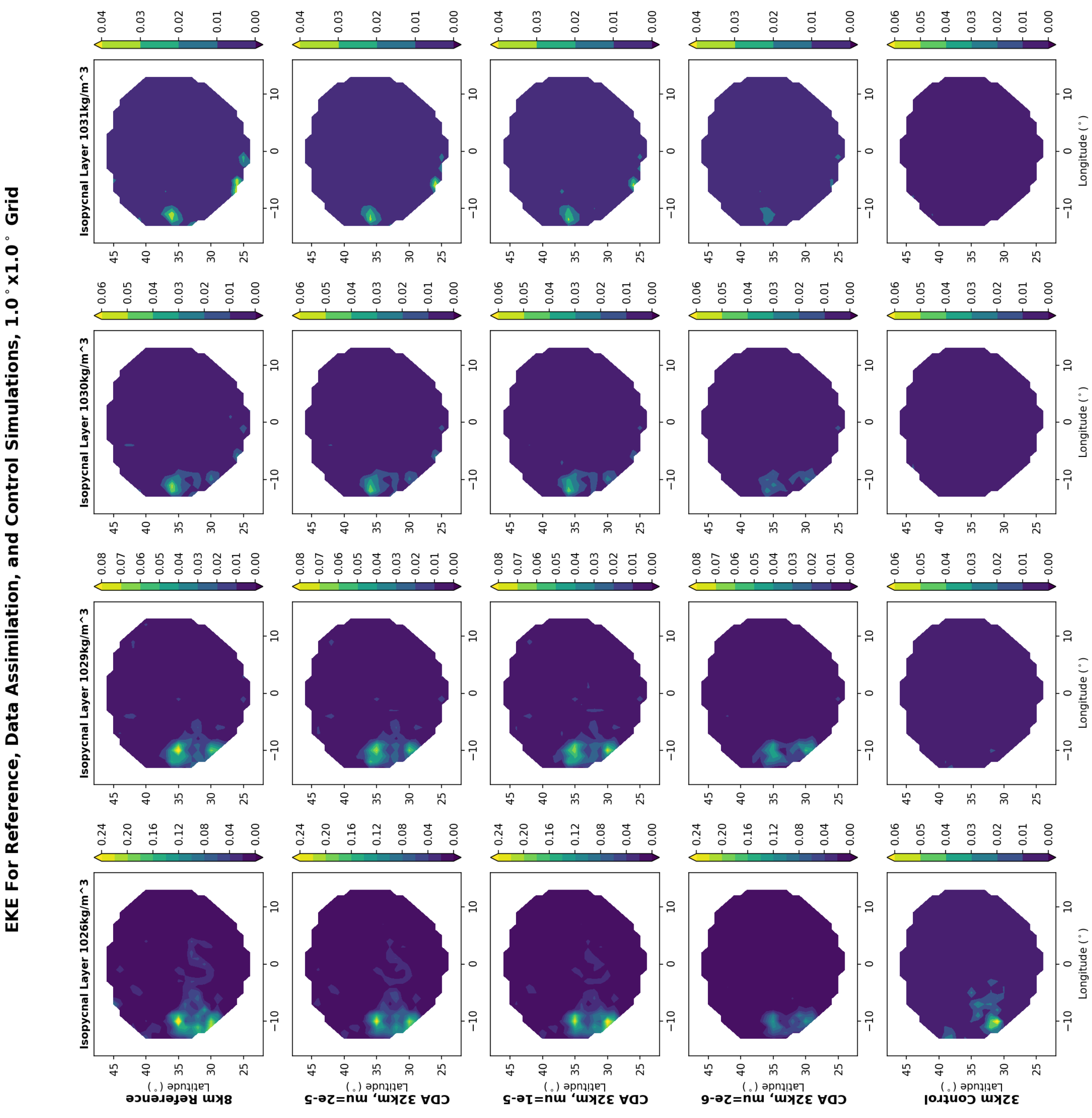
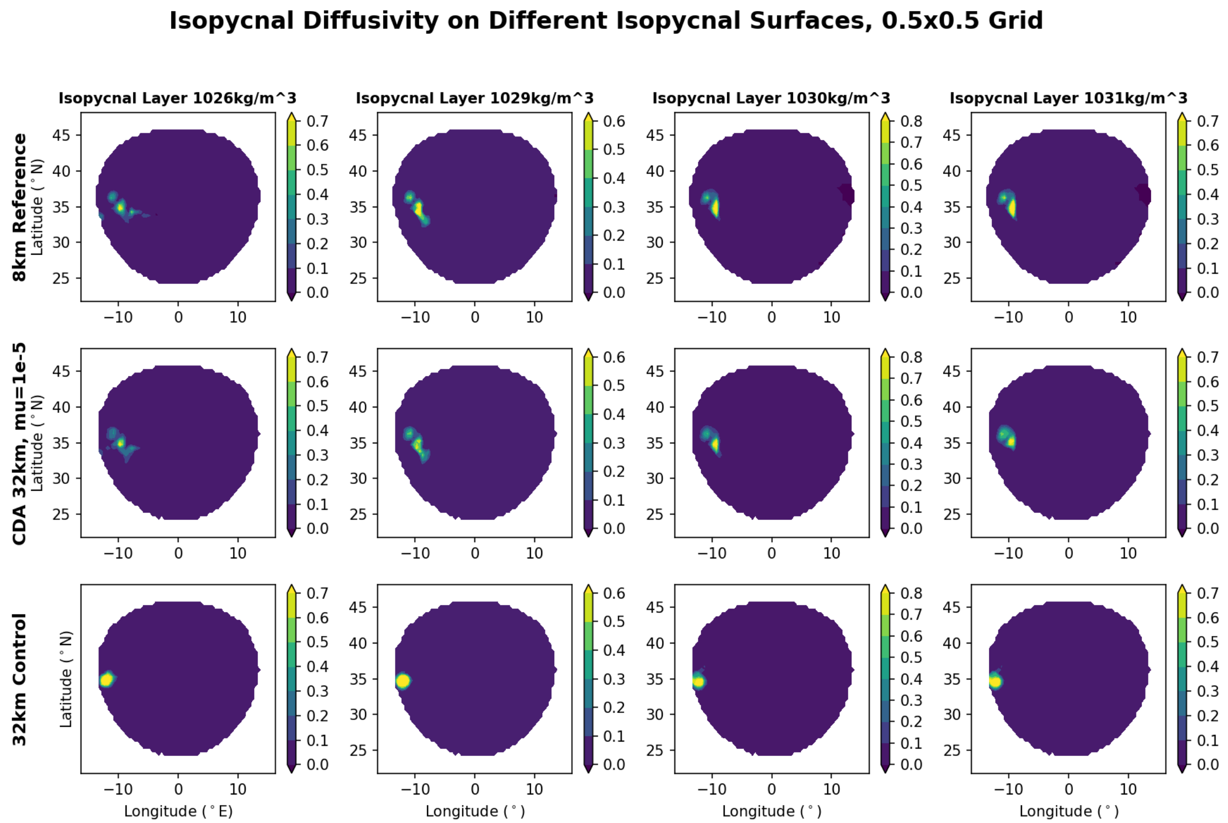


Figure 7. Plot of EKE on $1.0^\circ \times 1.0^\circ$ grid

Next we analyze the impact of the CDA algorithm on isopycnal diffusivity. A passive tracer was placed in a column in the middle of the jet near the western boundary, concentrated within a 100 km radius, and the tracer was evolved over two simulated weeks. Figures 8 and 9 show the averaged spread of the passive tracer over approximately 6 days starting 5 days into the simulation. Due to the qualitative similarity of the results for different μ's, only the isopycnal diffusivity for $\mu = 1\times10^{-5}$ is shown. The CDA simulations yields a diffusion more consistent with the reference simulation as compared to the control simulation. Table 3 presents the RMSE for the average tracer concentration, showing an improvement of an order of magnitude from the RMSE of the control simulation. The percent improvement with respect to the control simulation ranges from about 60-80%.

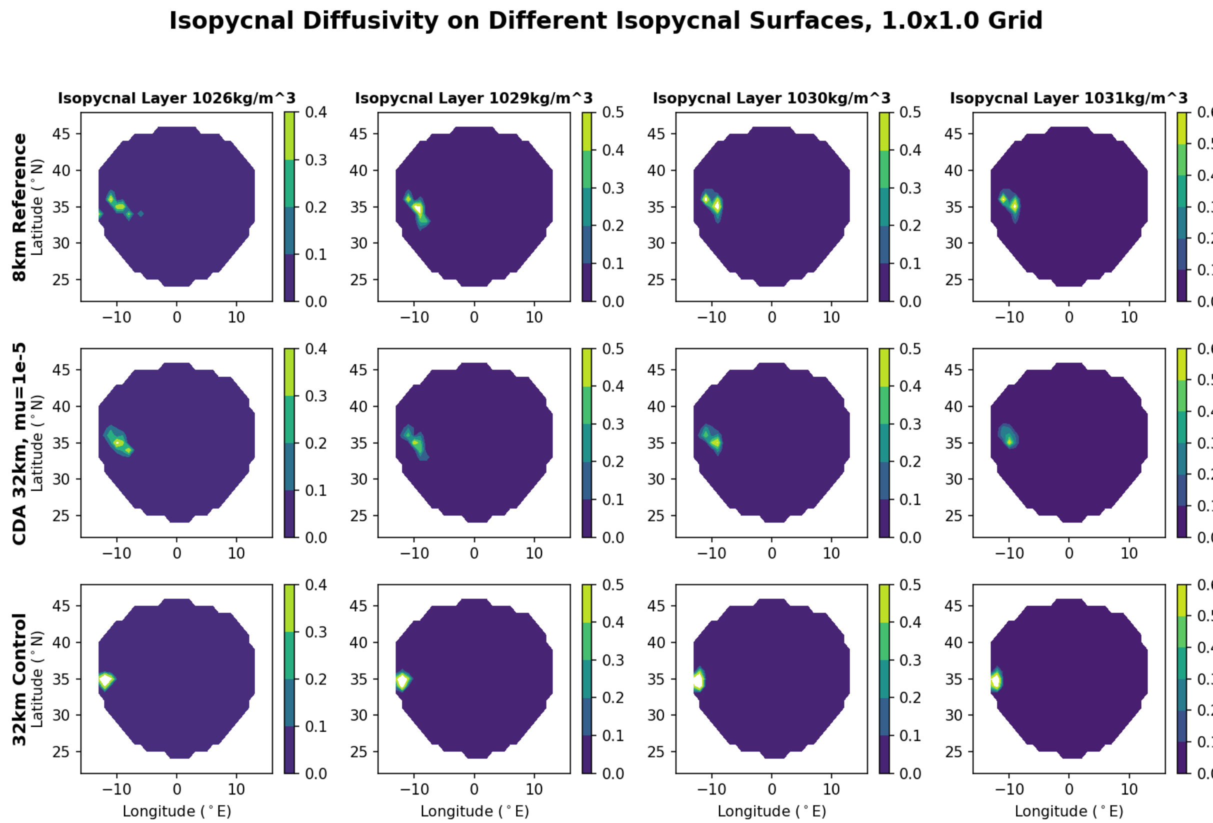
μ	Grid	RMSE	% Improvement
2×10^{-5}	$0.5^\circ\times0.5^\circ$	0.048	82%
1×10^{-5}	$0.5^\circ\times0.5^\circ$	0.050	81%
2×10^{-6}	$0.5^\circ\times0.5^\circ$	0.073	72%
control	$0.5^\circ\times0.5^\circ$	0.263	–
2×10^{-5}	$1.0^\circ\times1.0^\circ$	0.021	75%
1×10^{-5}	$1.0^\circ\times1.0^\circ$	0.021	75%
2×10^{-6}	$1.0^\circ\times1.0^\circ$	0.034	60%
control	$1.0^\circ\times1.0^\circ$	0.084	–

Table 3. RMSE of tracer concentration with respect to the observations, with % improvement of the CDA simulation relative to the control simulation.



[!th]

Figure 8. Figures plotting isopycnal diffusivity during the same time period on a $0.5° \times 0.5°$ grid, averaged over a approximately one week. Top row: reference simulation, Middle Row: CDA simulation with observations given on a $0.5° \times 0.5°$ grid, Bottom Row: Control simulation



[!th]

Figure 9. Figures plotting isopycnal diffusivity during the same time period on a $1.0^\circ \times 1.0^\circ$ grid, averaged over a approximately one week. Top row: reference simulation, Middle Row: CDA simulation with observations given on a $1.0^\circ \times 1.0^\circ$ grid, Bottom Row: Control simulation

Next, we look at how the CDA algorithm improves the mass transport north and south of the jet. Averaging the meridional velocity over approximately two weeks, we plot the approximate mass transport across two latitudes north and south of the jet, 41° N and 26° N, respectively. The mass transport through the jet was highly variable, causing analyses to be highly sensitive to the choice of latitude for mass transport computation. These latitudes were chosen far enough away from the jet so the mass transport would be less variable, and plots of mass transport at nearby latitudes were checked to verify the sensitivity to the choice of latitude. The fine scale features of the mass transport are better captured in all CDA simulations as compared to the control simulation, as demonstrated in Figures 10 – 13.

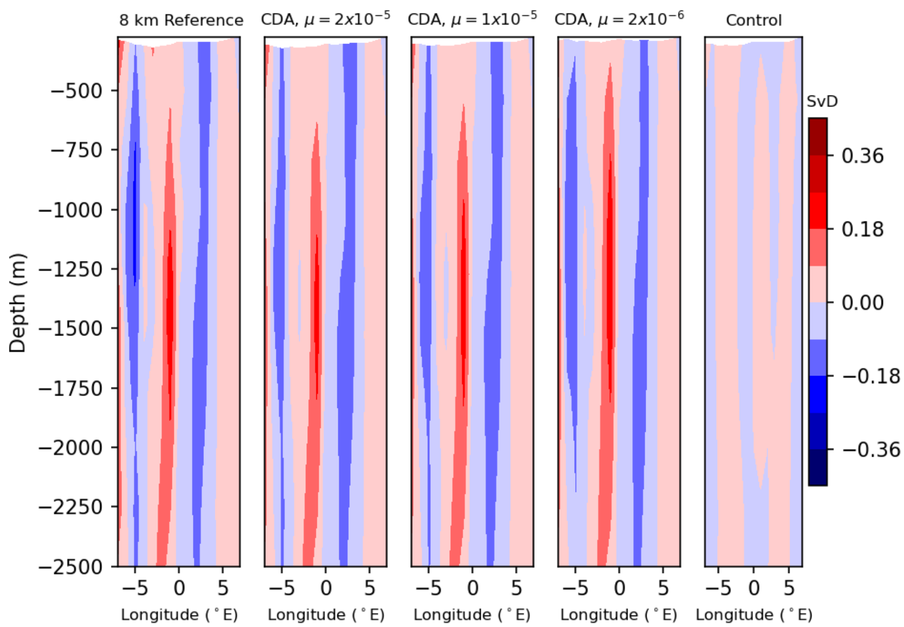


Figure 10. Mass Transport (SvD) at 26° latitude, $1.0^\circ \times 1.0^\circ$: The CDA algorithm better captures the finer-scale features as compared to the control simulation.

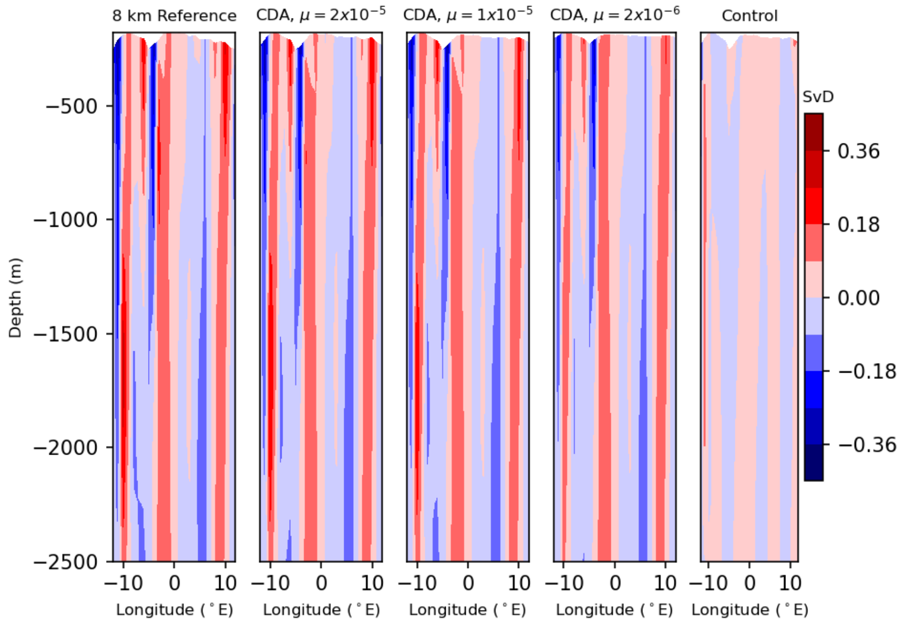


Figure 11. Mass Transport (SvD) at 41° latitude, $1.0° \times 1.0°$: The CDA algorithm better captures the finer-scale features as compared to the control simluation.

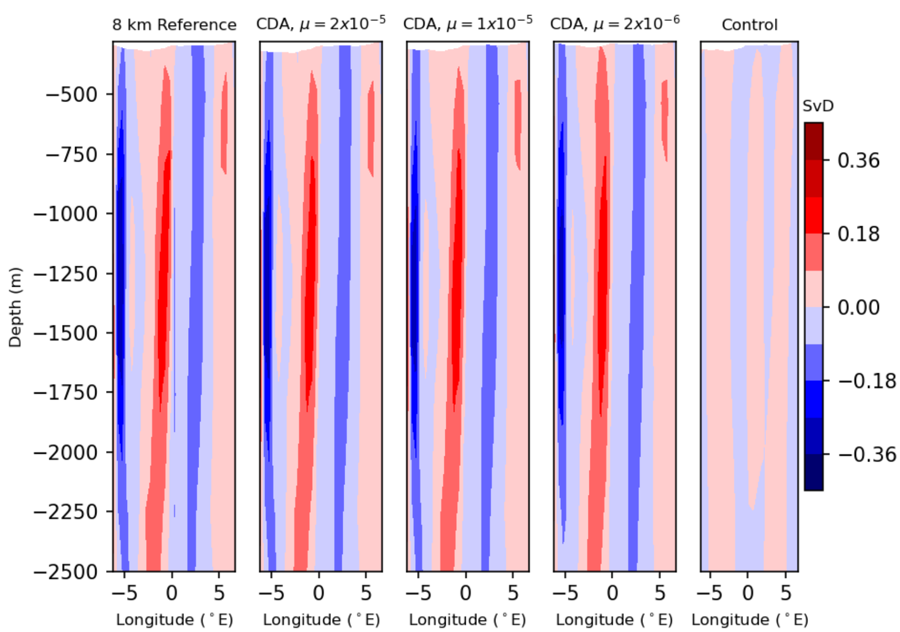


Figure 12. Mass Transport (SvD) at 26° latitude, $0.5° \times 0.5°$: The CDA algorithm better captures the finer-scale features as compared to the control simulation.

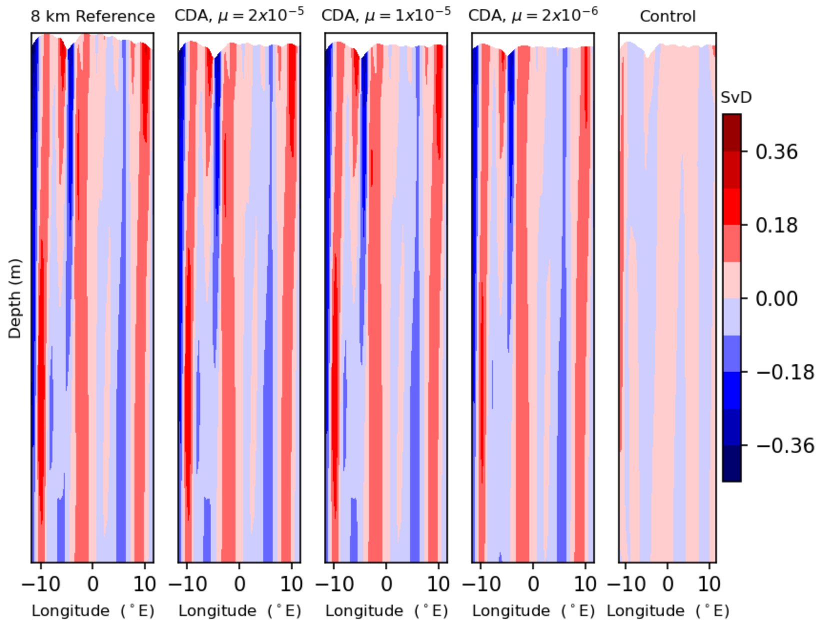


Figure 13. Mass Transport (SvD) at 41° latitude, $0.5^\circ \times 0.5^\circ$: The CDA algorithm better captures the finer-scale features as compared to the control simulation.

Finally, we compare the computational cost of running the CDA and nudging algorithms. Table 4 shows the approximate amount of computational runtime in seconds it takes to run the CDA and nudging simulations for approximately two simulation weeks on a large-scale HPC platform (specifically, the runs were done on 2 nodes with 36 Intel Xeon Broadwell E5-2695 processors and 128 GB of memory per node). The runtime is approximated on the order of 10^4, as fluctuations in runtime are large and highly dependent on how many other jobs are running. We note that, of course, CDA is slower than nudging due to the extra interpolation to the external grid. However, the consistent improvement in simulation fidelity of CDA over nudging is significant, with the improvement more marked in the setting where fewer observations are used. Furthermore, the computational time for CDA is not currently optimal, and could be improved by parallelizing the operation of interpolation of the model data to the observed grid and back. We also note that it only took four simulation weeks to equilibrate the data assimilation simulations, whereas it took 5 1/2 simulation years to equilibrate the control simulation. Hence, if one has observable data, assimilating said data using the CDA algorithm is faster and more accurate than running a model without data assimilation.

DA Method	Grid	Appx. Runtime (s)
CDA	$0.5^\circ \times 0.5^\circ$	4000
Nudging	$0.5^\circ \times 0.5^\circ$	1000
CDA	$1.0^\circ \times 1.0^\circ$	3000
Nudging	$1.0^\circ \times 1.0^\circ$	1000
Control	–	100

Table 4. Approximate Computational Cost for Two Simulation Weeks

4 Conclusions and Future Work

In this work, we presented the new CDA algorithm, its theoretical and computational advantages, and its advantages in comparison to the related data assimilation scheme of nudging. The CDA algorithm was demonstrated to qualitatively recapture the net effects of the eddy-resolving resolution flow field at coarser resolution, and improvement was seen in all measures of eddy effects as compared to the low resolution simulation. The horizontal velocity RMSE for the CDA algorithm was an order of magnitude better than the control simulation, and demonstrated a $10\% - 30\%$ improvement over the nudging simulation (see Figure 3 and Table 1)). The eddy kinetic energy, isopycnal diffusivity, and mass transport were also qualitatively identical to that of the high resolution simulation on the observation grids, with improvements in error between 60% and 100% relative to the control simulation. In summary, the CDA algorithm improved all metrics in the lower resolution simulation in very little simulated time. The theoretical support, computational recovery of mesoscale eddy effects, and computational speed of the algorithm support the conclusion that the CDA algorithm is a potentially viable algorithm for ocean models and should be explored further.

Future Directions

Since the CDA algorithm is a new data assimilation method, a number of factors need to be explored. For example, the choice of the positive relaxation coefficient μ is constant in this study, but it is common practice for the nudging coefficient to vary in space. To the best of the authors' knowledge, only three versions of the CDA algorithm with a varying μ have been computationally explored in Desamsetti et al. (2019), in Zerfas (2019), and in Larios and Pei (2017); Hudson and Jolly (2019). Another modification that can be explored would follow from Foias et al. (2016), where the observations at time t_n are assimilated over a time window $[t_n, t_{n+m})$, $m > 1$, addressing the fact that most datasets have data recorded less frequently than a typical timestep for a simulation.

Another avenue of further research would be to explore assimilating some instead of all state variables. This has been explored theoretically in, e.g., Farhat et al. (2015, 2016a, 2016b, 2016c, 2017) and computationally in, e.g., Altaf et al. (2017). Assimilating variables such as only temperature and salinity are practical when assimilating raw data instead of reanalysis data, and while the CDA algorithm has shown promise in this direction, theoretical and computational studies need to be completed to show the effectiveness of the CDA algorithm with respect to other data assimilation techniques. The authors have a publication in progress testing this in an idealized setting.

To address the most important data assimilation needs in the climate community, the CDA needs to be tested in more real data settings and against other more prominent methods of data assimilation. Finally, comparisons of the CDA algorithm to more commonly used statistical data assimilation algorithms needs to be explored in both ideal and real data settings.

Appendix A CDA vs Nudging: Discrete Setting with Mismatched Observation and Model Grids

In the continuous setting, the implementation of the CDA algorithm is straightforward. However, approximating and evolving models on discrete numerical grids presents a new yet subtle challenge. When the observational grid is a subset of the model grid, the explanation given in Section 2.2 is straightforward as the transition from understanding the concept in continuous space can be directly translated to discrete space. Complications arise in the setting where the observational grid is not a subset of the model grid. As a caricature, consider the following one-dimensional illustration. We are given observations on the grid Ω_{obs} (blue) and the simulation of the model is carried out on

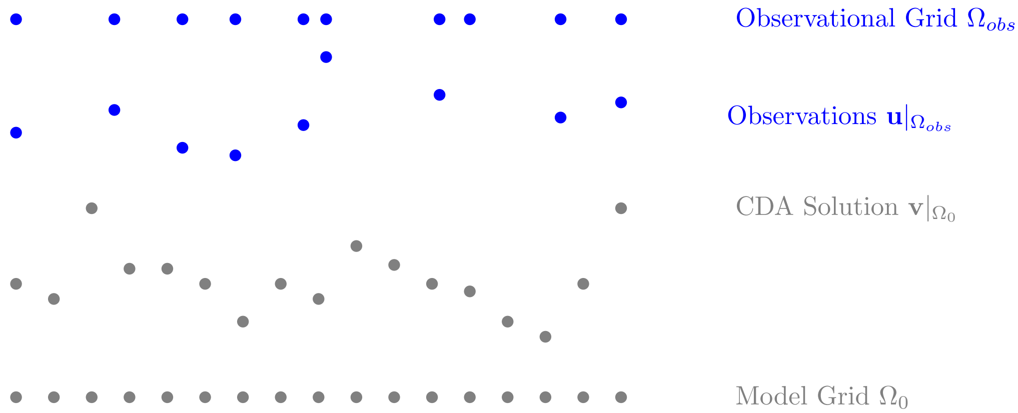


Figure A1. 1D Illustration of observations on observational grid Ω_{obs} (blue) and model solution as given on the model grid Ω_0 (gray).

the grid Ω_0 (gray), with spacing dx, for example. As we evolve the model, we only have the solution $\mathbf{v}|_{\Omega_0}$. Hence, we have the following information presented in Figure 1.

For the CDA algorithm, we need to have $I_\delta(\mathbf{u})$ and $I_\delta(\mathbf{v})$, the interpolation of these functions on Ω_{obs}. It is easy enough to compute $I_\delta(\mathbf{u})$. The temptation here to save computational time is to assimilate $I_\delta(\mathbf{u})|_{\Omega_0}$ directly. However, this is simply nudging, putting us back in the setting of modeling $\mathbf{v}_t = F(\mathbf{v}) + \mu(I_\delta(\mathbf{u}) - \mathbf{v})$. We need to know $\mathbf{v}$ on the *observational grid* in order to determine $I_\delta(\mathbf{v})$. However, since we only have $\mathbf{v}|_{\Omega_0}$ and $\Omega_{obs} \not\subseteq \Omega_0$, we need approximate values of $\mathbf{v}$ on Ω_{obs} in order to obtain an approximation of $I_\delta(\mathbf{v})$. In general, the observational grid is more sparse than the model grid, and hence we assume the interpolation $I_{dx}(\mathbf{v})$ of $\mathbf{v}|_{\Omega_0}$ to be a sufficiently good approximation of $\mathbf{v}$ in full space. This is illustrated in Figure 2.

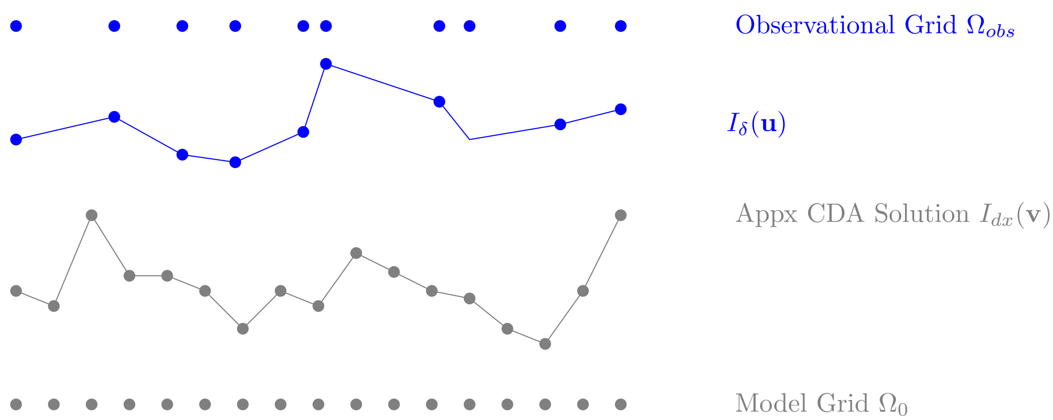


Figure A2. 1D Illustration of interpolation of observations to full space and approximation of CDA solution on all of space, since it is only known on the model grid Ω_0.

Now that we have an approximation of $\mathbf{v}$, we take the points of $I_{dx}(\mathbf{v})|_{\Omega_{obs}}$, and interpolate to obtain $I_\delta(I_{dx}(\mathbf{v}))$, an approximation of $I_h(\mathbf{v})$. This is illustrated in Figures 3 and 4.

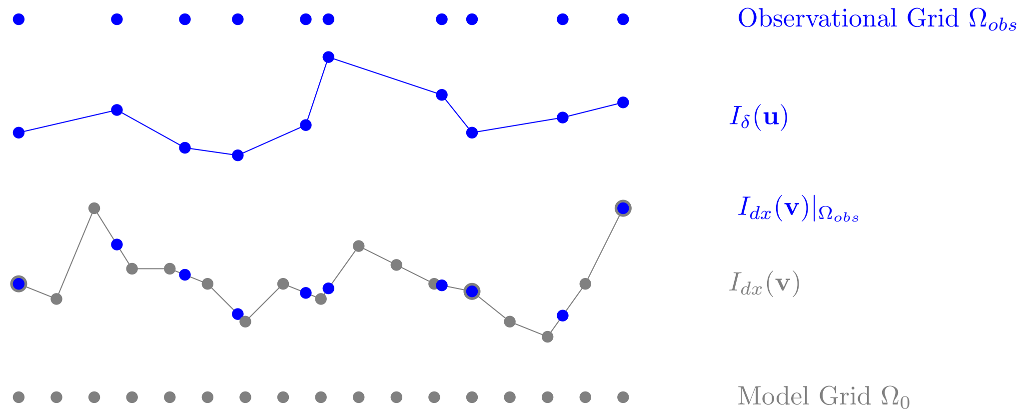


Figure A3. 1D Illustration of $I_{dx}(\mathbf{v})|_{\Omega_{obs}}$, the values of $I_{dx}(\mathbf{v})$ on the observational grid.

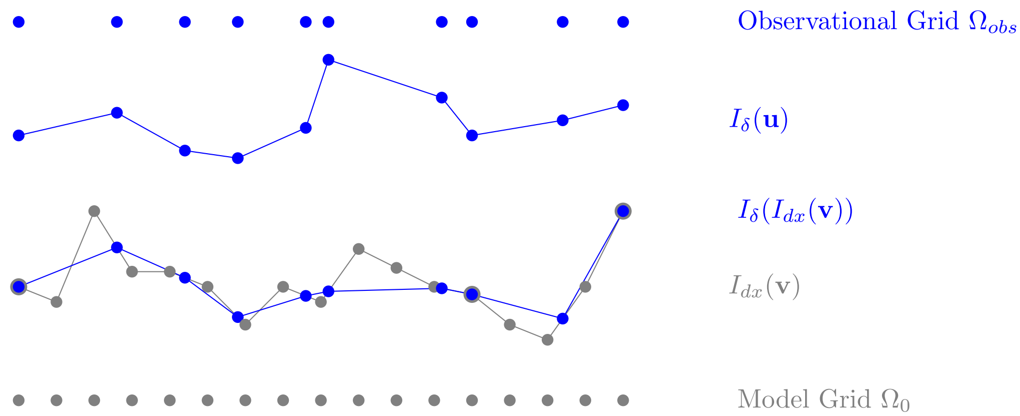


Figure A4. 1D Illustration of $I_{dx}(\mathbf{v})|_{\Omega_{obs}}$, the values of $I_{dx}(\mathbf{v})$ on the observational grid, and the interpolating those values to obtain $I_\delta(I_{dx}(\mathbf{v}))$.

Now that we have $I_\delta(\mathbf{u})$ and the approximation of $I_\delta(\mathbf{v})$ given by $I_\delta(I_{dx}(\mathbf{v}))$. These functions are assimilated into the model on the model grid; these are the gray values in Figure 5.

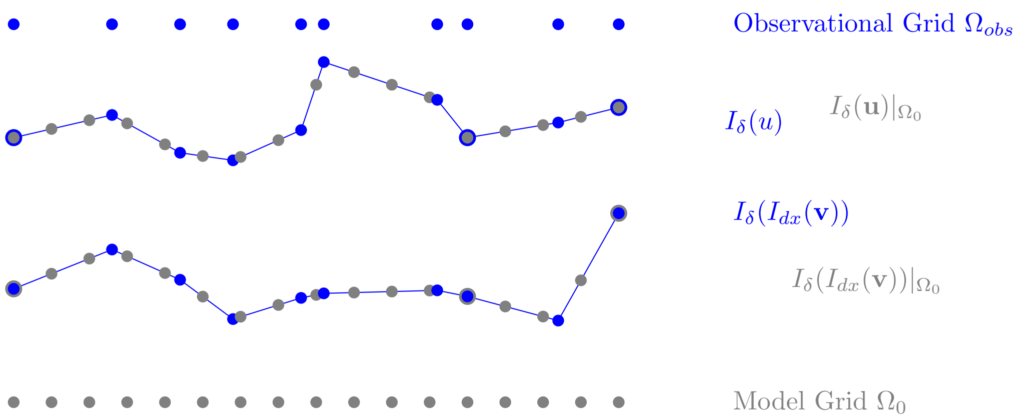


Figure A5. 1D Illustration of values of $I_\delta(\mathbf{u})|_{\Omega_0}$ and $I_\delta(I_{dx}(\mathbf{v}))|_{\Omega_0}$.

Acknowledgments

The research of author Elizabeth Carlson was supported in part by the NSF GRFP grant no. 1610400. This research was supported as part of the Energy Exascale Earth System Model (E3SM) project, funded by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research. Luke Van Roekel acknowledges the support of DOE Cooperative Agreement DE-SC0019340. This research used resources provided by the Los Alamos National Laboratory Institutional Computing Program, which is supported by the U.S. Department of Energy National Nuclear Security Administration under Contract No. 89233218CNA000001. The research from author Humberto Godinez presented in this article was supported by the Laboratory Directed Research and Development program of Los Alamos National Laboratory under project number 20160599ECR. Author Adam Larios was partially supported by NSF grants DMS-1716801 and CMMI-1953346. Elizabeth Carlson would like to thank the authors and the rest of the COSIM team at LANL for their hospitality and mentorship. The authors give special thanks to Dr. Philip Wolfram and Dr. Xylar Asay-Davis for their additional help with the physics and code.

The data that support the findings of this study are openly available in Zenodo at https://doi.org/10.5281/zenodo.4767741. The code used for this project is also openly available and can be found at https://github.com/MPAS-Dev/MPAS-Model.

This research was assigned the LA-UR number LA-UR-21-21353.

References

Abbasi, M. R., Chegini, V., Sadrinasab, M., & Sladatmousavi, S. M. (2018). Optimization of the modeled surface temperature by assimilation of SST data over the Persian Gulf. *IJMS*, *47(09)*, 1803-1808. Retrieved from `http://nopr.niscair.res.in/handle/123456789/44918`

Albanez, D. A., Nussenzveig Lopes, H. J., & Titi, E. S. (2016). Continuous data assimilation for the three-dimensional Navier–Stokes-α model. *Asymptotic Anal.*, *97*(1-2), 139–164.

Albanez, D. A. F., & Benvenutti, M. J. (2018). Continuous data assimilation algorithm for simplified Bardina model. *Evol. Equ. Control Theory*, *7*(1), 33–52. Retrieved from `https://doi.org/10.3934/eect.2018002` doi: 10.3934/eect.2018002

Altaf, M. U., Titi, E. S., Knio, O. M., Zhao, L., McCabe, M. F., & Hoteit, I. (2017).

Downscaling the 2D Benard convection equations using continuous data assimilation. *Comput. Geosci*, *21*(3), 393–410.

Anthes, R. A. (1974). Data assimilation and initialization of hurricane prediction models. *J. Atmos. Sci.*, *31*, 702–719.

Azouani, A., Olson, E., & Titi, E. S. (2014). Continuous data assimilation using general interpolant observables. *J. Nonlinear Sci.*, *24*(2), 277–304. Retrieved from `http://dx.doi.org/10.1007/s00332-013-9189-y` doi: 10.1007/s00332-013-9189-y

Azouani, A., & Titi, E. S. (2014). Feedback control of nonlinear dissipative systems by finite determining parameters—a reaction-diffusion paradigm. *Evol. Equ. Control Theory*, *3*(4), 579–594. Retrieved from `http://dx.doi.org/10.3934/eect.2014.3.579` doi: 10.3934/eect.2014.3.579

Bakhshaii, A., & Stull, R. (2009, 10). Deterministic ensemble forecasts using gene-expression programming. *Weather and Forecasting*, *24*(5), 1431-1451. Retrieved from `https://doi.org/10.1175/2009WAF2222192.1` doi: 10.1175/2009WAF2222192.1

Baptista, A., Zhang, Y.-L., Chawla, A., Zulauf, M., Seaton, C., Myers, E., ... Turner, P. (2005). A cross-scale model for 3d baroclinic circulation in estuary-plume-shelf systems: Ii. application to the Columbia River. *Continent Shelf Research*, *25*, 850–863.

Bessaih, H., Olson, E., & Titi, E. S. (2015). Continuous data assimilation with stochastically noisy data. *Nonlinearity*, *28*(3), 729–753. Retrieved from `http://dx.doi.org/10.1088/0951-7715/28/3/729` doi: 10.1088/0951-7715/28/3/729

Biswas, A., Bradshaw, Z., & Jolly, M. S. (2020). Data assimilation for the Navier-Stokes equations using local observables. *arXiv: Analysis of PDEs*. Retrieved from `https://arxiv.org/pdf/2008.06949.pdf`

Biswas, A., Foias, C., Mondaini, C. F., & Titi, E. S. (2018). Downscaling data assimilation algorithm with applications to statistical solutions of the Navier–Stokes equations. In *Annales de l'institut henri poincaré c, analyse non linéaire.*

Biswas, A., & Price, R. (2020). Continuous data assimilation for the three dimensional Navier–Stokes equations. *arXiv:Mathematics:Analysis of PDEs*. Retrieved from `https://arxiv.org/pdf/2003.01329.pdf`

Bullock Jr., O. R., Foroutan, H., Gilliam, R., & Herwehe, J. (2018). Adding four-dimensional data assimilation by analysis nudging to the Model for Prediction Across Scales – Atmosphere (version 4.0). *Geosci. Model Dev.*, *11*, 2897-2922. Retrieved from `https://doi.org/10.5194/gmd-11-2897-2018` doi: 10.5194/gmd-11-2897-2018

Carlson, E., Hudson, J., & Larios, A. (2020). Parameter recovery for the 2D Navier–Stokes equations via continuous data assimilation. *SIAM J. of Sci. Comput.*, *42(1)*, A250–A270. Retrieved from `https://doi.org/10.1137/19M1248583` doi: 10.1137/19M1248583

Cazenave, P., Torres, R., Blackford, J., Artioli, Y., & Bruggeman, J. (2018). Regional modelling to inform the design of sub-sea CO2 storage monitoring networks. *14th Greenhouse Gas Control Technologies Conference Melbourne*, 21-26. Retrieved from `https://ssrn.com/abstract=3366246`

Celik, E., Olson, E., & Titi, E. S. (2019). Spectral filtering of interpolant observables for a discrete-in-time downscaling data assimilation algorithm. *SIAM J. Appl. Dyn. Syst.*, *18*(2), 1118–1142. Retrieved from `https://doi.org/10.1137/18M1218480` doi: 10.1137/18M1218480

Clark Di Leoni, P., Mazzino, A., & Biferale, L. (2020, Feb). Synchronization to big data: Nudging the Navier-Stokes equations for data assimilation of turbulent flows. *Phys. Rev. X*, *10*, 011023. Retrieved from `https://link.aps.org/doi/10.1103/PhysRevX.10.011023` doi: 10.1103/PhysRevX.10.011023

Cowles, G. W., Lentz, S. J., Chen, C., Xu, Q., & Beardsley, R. C. (2008). Comparison of observed and model–computed low frequency circulation and hydrography on the New England Shelf. *Journal of Geophysical Research: Oceans*, *113(C9)*. Retrieved from https://doi-org.lanl.idm.oclc.org/10.1029/2007JC004394 doi: 10.1029/2007JC004394

Cui, B., Toth, Z., Zhu, Y., & Hou, D. (2012, 04). Bias correction for global ensemble forecast. *Weather and Forecasting*, *27*(2), 396-410. Retrieved from https://doi.org/10.1175/WAF-D-11-00011.1 doi: 10.1175/WAF-D-11-00011.1

Dee, D. P. (2005). Bias and data assimilation. *Quarterly Journal of the Royal Meteorological Society*, *131*(613), 3323-3343. Retrieved from https://rmets.onlinelibrary.wiley.com/doi/abs/10.1256/qj.05.137 doi: https://doi.org/10.1256/qj.05.137

Delle Monache, L., Nipen, T., Deng, X., Zhou, Y., & Stull, R. (2006). Ozone ensemble forecasts: 2. a Kalman filter predictor bias correction. *Journal of Geophysical Research: Atmospheres*, *111*(D5). Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2005JD006311 doi: 10.1029/2005JD006311

Desamsetti, S., Dasari, H., Langodan, S., Knio, O., Hoteit, I., & Titi, E. S. (2019). Efficient dynamical downscaling of general circulation models using continuous data assimilation. *Quarterly Journal of the Royal Meteorological Society*. Retrieved from http://hdl.handle.net/10754/656325 doi: 10.1002/qj.3612

Ding, Y., Bao, X., Yu, H., & Kuang, L. (2012). A numerical study of the barotropic tides and tidal energy distribution in the Indonesian seas with the assimilated finite volume coastal ocean model. *Ocean Dynamics*, *62*, 515-532. doi: 0.1007/s10236-011-0518-0

Du, J., & Zhou, B. (2011, 10). A dynamical performance-ranking method for predicting individual ensemble member performance and its application to ensemble averaging. *Monthly Weather Review*, *139*(10), 3284-3303. Retrieved from https://doi.org/10.1175/MWR-D-10-05007.1 doi: 10.1175/MWR-D-10-05007.1

Durai, V., & Bhradwaj, R. (2014). Evaluation of statistical bias correction methods for numerical weather prediction model forecasts of maximum and minimum temperatures. *Nat. Hazards*, *73*, 1229–1254. doi: 10.1007/s11069-014-1136-1

E3SM Project. (2018, April). *Energy Exascale Earth System Model (E3SM).* [Computer Software] https://dx.doi.org/10.11578/E3SM/dc.20180418.36. Retrieved from https://dx.doi.org/10.11578/E3SM/dc.20180418.36 doi: 10.11578/E3SM/dc.20180418.36

Evensen, G. (1997). Advanced data assimilation for strongly nonlinear dynamics. *Mon. Wea. Rev.*, *125*, 1342–1354.

Farhat, A., Glatt-Holtz, N., Martinez, V. R., McQuarrie, S. A., & Whitehead, J. P. (2020). Data assimilation in large Prandtl Rayleigh–Bénard convection from thermal measurements. *SIAM J. Applied Dynamical Systems*, *19*, 510-540.

Farhat, A., Johnston, H., Jolly, M., & Titi, E. S. (2018, 01). Assimilation of nearly turbulent Rayleigh–Bénard flow through vorticity or local circulation measurements: A computational study. *Journal of Scientific Computing*, *77*(3), 1519–1533. Retrieved from https://doi.org/10.1007/s10915-018-0686-x doi: 10.1007/s10915-018-0686-x

Farhat, A., Jolly, M. S., & Titi, E. S. (2015). Continuous data assimilation for the 2D Bénard convection through velocity measurements alone. *Phys. D*, *303*, 59–66. Retrieved from http://dx.doi.org/10.1016/j.physd.2015.03.011 doi: 10.1016/j.physd.2015.03.011

Farhat, A., Lunasin, E., & Titi, E. S. (2016a). Abridged continuous data assimilation for the 2D Navier–Stokes equations utilizing measurements of only one component of the velocity field. *J. Math. Fluid Mech.*, *18*(1), 1–23.

Farhat, A., Lunasin, E., & Titi, E. S. (2016b). Data assimilation algorithm for

3D Bénard convection in porous media employing only temperature measurements. *J. Math. Anal. Appl.*, *438*(1), 492–506.

Farhat, A., Lunasin, E., & Titi, E. S. (2016c). On the Charney conjecture of data assimilation employing temperature measurements alone: the paradigm of 3D planetary geostrophic model. *Mathematics of Climate and Weather Forecasting*, *2*(1).

Farhat, A., Lunasin, E., & Titi, E. S. (2017). Continuous data assimilation for a 2D Bénard convection system through horizontal velocity measurements alone. *J. Nonlinear Sci.*, 1–23. Retrieved from `http://dx.doi.org/10.1007/s00332-017-9360-y` doi: 10.1007/s00332-017-9360-y

Farhat, A., Lunasin, E., & Titi, E. S. (2018). A data assimilation algorithm: the paradigm of the 3D Leray-α model of turbulence. *Nonlinear Partial Differential Equations Arising from Geometry and Physics*. ((to appear))

Fennel, K., Laurent, A., Hetland, R., Justić, D., Ko, D. S., Lehrter, J., … Zhang, W. (2016). Effects of model physics on hypoxia simulations for the northern Gulf of Mexico: A model intercomparison. *Journal of Geophysical Research: Oceans*, *121*(8), 5731-5750. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2015JC011577` doi: 10.1002/2015JC011577

Foias, C., Mondaini, C. F., & Titi, E. S. (2016). A discrete data assimilation scheme for the solutions of the two-dimensional Navier–Stokes equations and their statistics. *SIAM J. Appl. Dyn. Syst.*, *15*(4), 2109–2142. Retrieved from `http://dx.doi.org/10.1137/16M1076526` doi: 10.1137/16M1076526

Foreman, M., Czajko, P., Stucchi, D., & Guo, M. (2009). A finite volume model simulation for the Broughton Archipelago, Canada. *Ocean Modelling*, *30*(1), 29 - 47. Retrieved from `http://www.sciencedirect.com/science/article/pii/S1463500309001085` doi: https://doi.org/10.1016/j.ocemod.2009.05.009

Fortunato, A. B., Nahon, A., Dodet, G., Pires, A. R., Freitas, M. C., Bruneau, N., … Oliveira, A. (2014). Morphological evolution of an ephemeral tidal inlet from opening to closure: the Albufeira inlet, Portugal. *Continental Shelf Research*, *73*, 49-63. doi: 10.1016/j.csr.2013.11.005

Fujisaki-Manome, A., Fitzpatrick, L. E., Gronewold, A. D., Anderson, E. J., Lofgren, B. M., Spence, C., … Xiao, C. (2017). Turbulent heat fluxes during an extreme lake-effect snow event. *Journal of Hydrometeorology*, *18*(12), 3145-3163. Retrieved from `https://doi.org/10.1175/JHM-D-17-0062.1` doi: 10.1175/JHM-D-17-0062.1

García-Archilla, B., Novo, J., & Rubino, S. (2020). Error analysis of proper orthogonal decomposition data assimilation schemes for the Navier–Stokes equations. *arXiv: Mathematics: Numerical Analysis*. Retrieved from `https://arxiv.org/pdf/2004.09127.pdf`

García-Archilla, B., Novo, J., & Titi, E. S. (2020). Uniform in time error estimates for a finite element method applied to a downscaling data assimilation algorithm. *SIAM J. of Numer. Anal.*, *58*(1), 410–429. Retrieved from `https://doi.org/10.1137/19M1246845` doi: 10.1137/19M1246845

Gardner, M., Larios, A., Rebholz, L. G., Vargun, D., & Zerfas, C. (2020). Continuous data assimilation applied to a velocity-vorticity formulation of the 2d Navier-Stokes equations. *arXiv: Numerical Analysis*. Retrieved from `https://arxiv.org/pdf/2006.07295.pdf` ((accepted for publication in AIMS Electronic Research Archive))

Ge, J., Ding, P., Chen, C., Hu, S., Fu, G., & Wu, L. (2013). An integrated East China Sea—Changjiang Estuary model system with aim at resolving multi–scale regional—shelf–estuarine dynamics. *Ocean Dynamics*, *63*, 881–900. doi: 10.1007/s10236-013-0631-3

Ge, J., Torres, R., Chen, C., Liu, J., Xu, Y., Bellerby, R., … Ding, P. (2020). Influence of suspended sediment front on nutrients and phytoplankton dynamics off the Changjiang Estuary: A FVCOM–ERSEM coupled model experiment. *Journal of Marine Systems*, *204*, 103292. Retrieved from http://www.sciencedirect.com/science/article/pii/S0924796319304294 doi: https://doi.org/10.1016/j.jmarsys.2019.103292

Gesho, M. (2013). *A numerical study of continuous data assimilation for the 2D–Navier Stokes equations using nodal points* (Unpublished master's thesis). University of Nevada – Reno.

Glahn, H. R., & Lowry, D. A. (1972). The use of model output statistics (MOS) in objective weather forecasting. *Journal of Applied Meteorology (1962-1982)*, *11*(8), 1203–1211. Retrieved from http://www.jstor.org/stable/26176961

Gneiting, T., Raftery, A. E., Westveld, I., Anton H., & Goldman, T. (2005, 05). Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. *Monthly Weather Review*, *133*(5), 1098-1118. Retrieved from https://doi.org/10.1175/MWR2904.1 doi: 10.1175/MWR2904.1

Golaz, J.-C., Caldwell, P. M., Van Roekel, L. P., Petersen, M. R., Tang, Q., Wolfe, J. D., … Zhu, Q. (2019). The DOE E3SM Coupled Model Version 1: Overview and evaluation at standard resolution. *Journal of Advances in Modeling Earth Systems*, *11*(7), 2089-2129. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018MS001603 doi: 10.1029/2018MS001603

He, Z., Thompson, K. R., Ritchie, H., Lu, Y., & Dupont, F. (2014). Reducing drift and bias of a global ocean model by frequency-dependent nudging. *Atmosphere-Ocean*, *52*(3), 242-255. Retrieved from https://doi.org/10.1080/07055900.2014.922240 doi: 10.1080/07055900.2014.922240

Hecht, M. W., Hunke, E., Maltrud, M. E., Petersen, M. R., & Wingate, B. A. (2008). Lateral mixing in the eddying regime and a new broad-ranging formulation. *Geophysical Monograph*, *177*, 339–352.

Hoke, J. E., & Anthes, R. A. (1976). The initialization of numerical models by a dynamics intialization technique. *Mon. Wea. Rev.*, *104*, 1551–1556.

Houtekamer, P., Deng, X., Mitchell, H., Baek, S.-J., & Gagnon, N. (2014). Higher resolution in an operational ensemble Kalman filter. *Mon. Wea. Rev.*, *142(3)*, 1143–1162. Retrieved from https://doi.org/10.1175/MWR-D-13-00138.1 doi: 10.1175/MWR-D-13-00138.1

Hudson, J., & Jolly, M. (2019). Numerical efficacy study of data assimilation for the 2D magnetohydrodynamic equations. *Journal of Computational Dynamics*, *6*(2158-2491_2019_1_131), 131. Retrieved from http://aimsciences.org//article/id/3857d52b-2b80-454e-87bb-f43b4dc394d8 doi: 10.3934/jcd.2019006

Ibdah, H. A., Mondaini, C. F., & Titi, E. S. (2019). Uniform in time error estimates for fully discrete numerical schemes of a data assimilation algorithm. *IMA J. Numer. Anal.*. Retrieved from https://doi.org/10.1093/imanum/drz043 doi: 10.1093/imanum/drz043

Jolly, M. S., Martinez, V. R., Olson, E. J., & Titi, E. S. (2019). Continuous data assimilation with blurred-in-time measurements of the surface quasi-geostrophic equation. *Chin. Ann. Math. Ser. B*, *40*, 721-764. Retrieved from https://doi.org/10.1007/s11401-019-0158-0 doi: 10.1007/s11401-019-0158-0

Jolly, M. S., Martinez, V. R., & Titi, E. S. (2017). A data assimilation algorithm for the subcritical Surface Quasi-Geostrophic equation. *Adv. Nonlinear Stud.*, *17*(1), 167–192. Retrieved from http://0-dx.doi.org.library.unl.edu/10.1515/ans-2016-6019 doi: 10.1515/ans-2016-6019

Kilgren, R. (2006). *Numerical simulations of large river plumes in the Pacific Northwest* (Unpublished master's thesis). Oregon Graduate Institute School of Science & Engineering.

Larios, A., & Pei, Y. (2017). Nonlinear continuous data assimilation.

((submitted) arXiv:1703.03546)

Larios, A., & Pei, Y. (2018). Approximate continuous data assimilation of the 2D Navier–Stokes equations via the Voigt-regularization with observable data. ((submitted) arXiv:1810.10616)

Larios, A., Rebholz, L., & Zerfas, C. (2018). Global in time stability and accuracy of IMEX–FEM data assimilation schemes for Navier–Stokes equations. *Computer Methods in Applied Mechanics and Engineering*, *345*, 1077-1093.

Larios, A., & Victor, C. (2021). Continuous data assimilation with a moving cluster of data points for a reaction diffusion equation: A computational study. *Commun. Comput. Phys.*, *29*, 1273–1298. Retrieved from https://arxiv.org/pdf/1812.01686.pdf

Lawson, W., & Hansen, J. (2004). Implications of stochastic and deterministic filters as ensemble-based data assimilation methods in varying regimes of error growth. *Mon. Wea. Rev.*, *132*, 1966–1981.

Lunasin, E., & Titi, E. S. (2017). Finite determining parameters feedback control for distributed nonlinear dissipative systems – a computational study. *Evolution Equations & Control Theory*, *6*(2163-2480_2017_4_535), 535. Retrieved from http://aimsciences.org//article/id/6e525aaa-1627-4790-bdb3-a9788ff15d4a doi: 10.3934/eect.2017027

Markowich, P. A., Titi, E. S., & Trabelsi, S. (2016). Continuous data assimilation for the three-dimensional Brinkman-Forchheimer-extended Darcy model. *Nonlinearity*, *29*(4), 1292–1328. Retrieved from http://dx.doi.org/10.1088/0951-7715/29/4/1292 doi: 10.1088/0951-7715/29/4/1292

Mondaini, C. F., & Titi, E. S. (2018). Uniform-in-time error estimates for the post-processing Galerkin method applied to a data assimilation algorithm. *SIAM J. Numer. Anal.*, *56*(1), 78–110. Retrieved from https://doi.org/10.1137/16M110962X

Pacanowski, R., & Philander, S. (1981). Parameterization of vertical mixing in numerical models of tropical oceans. *J. Phys. Oceanogr.*, *11*, 1443–1451. doi: 10.1175/1520-0485(1981)011,1443:POVMIN.2.0.CO;2

Pei, Y. (2019). Continuous data assimilation for the 3d Primitive equations of the ocean. *Communications on Pure & Applied Analysis*, *18*(1534-0392_2019_2_643), 643. Retrieved from http://aimsciences.org//article/id/5521e342-722d-417a-ada4-3e65f087b5de doi: 10.3934/cpaa.2019032

Peng, M., Schmalz Jr., R., Zhang, A., & Aikman III, F. (2014). Towards the development of the National Ocean Service San Francisco Bay Operational Forecast System. *J. Mar. Sci. Eng.*, *2(1)*, 247-286. Retrieved from https://doi.org/10.3390/jmse2010247 doi: 10.3390/jmse201047

Petersen, M. R., Asay-Davis, X. S., Berres, A. S., Chen, Q., Feige, N., Hoffman, M. J., . . . Woodring, J. L. (2019). An evaluation of the ocean and sea ice climate of E3SM using MPAS and interannual CORE-II forcing. *Journal of Advances in Modeling Earth Systems*, *11*(5), 1438-1458. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018MS001373 doi: 10.1029/2018MS001373

Petersen, M. R., Jacobsen, D. W., Ringler, T. D., Hecht, M. W., & Maltrud, M. E. (2015). Evaluation of the arbitrary Lagrangian-Eulerian vertical coordinate method in the MPAS-Ocean model. *Ocean Modelling*, *86*(0), 93 - 113. Retrieved from http://www.sciencedirect.com/science/article/pii/S1463500314001796 doi: http://dx.doi.org/10.1016/j.ocemod.2014.12.004

Pringle, J. M. (2006). Sources of variability in Gulf of Maine circulation, and the observations needed to model it. *Deep Sea Research Part II: Topical Studies in Oceanography*, *53*(23), 2457 - 2476. Retrieved from http://www.sciencedirect.com/science/article/pii/S096706450600230X (Dynamics of Plankton and Larval Fish Populations on Georges Bank, the North Atlantic U.S. GLOBEC Study Site) doi: https://doi.org/10.1016/

j.dsr2.2006.08.015

Raftery, A. E., Gneiting, T., Balabdaoui, F., & Polakowski, M. (2005). Using Bayesian model averaging to calibrate forecast ensembles. *Monthly Weather Review*, *133*(5), 1155-1174. Retrieved from `https://doi.org/10.1175/MWR2906.1` doi: 10.1175/MWR2906.1

Rebholz, L. G., & Zerfas, C. (2018). Simple and efficient continuous data assimilation of evolution equations via algebraic nudging.

Ringler, T., Petersen, M., Higdon, R., Jacobsen, D., Jones, P., & Maltrud, M. (2013). A multi-resolution approach to global ocean modeling. *Ocean Modelling*, *69*(0), 211-232. Retrieved from `http://www.sciencedirect.com/science/article/pii/S1463500313000760` doi: http://dx.doi.org/10.1016/j.ocemod.2013.04.010

Robinson, C. L. K., Morrison, J., & Foreman, M. G. G. (2011). Oceanographic connectivity among marine protected areas on the north coast of British Columbia, Canada. *Canadian Journal of Fisheries and Aquartic Sciences*, *62*(6), 1350-1362. doi: 10.1139/f05-088

Satterfield, E. A., & Bishop, C. H. (2014, 08). Heteroscedastic ensemble postprocessing. *Monthly Weather Review*, *142*(9), 3484-3502. Retrieved from `https://doi.org/10.1175/MWR-D-13-00286.1` doi: 10.1175/MWR-D-13-00286.1

Stauffer, D. R., & Seaman, N. L. (1990). Use of four-dimensional data assimilation in a limited-area mesoscale model. part i: Experiments with synoptic-scale data. *Monthly Weather Review*, *118*(6), 1250-1277. Retrieved from `https://doi.org/10.1175/1520-0493(1990)118<1250:UOFDDA>2.0.CO;2` doi: 10.1175/1520-0493(1990)118⟨1250:UOFDDA⟩2.0.CO;2

Trémolet, Y. (2007). Incremental 4D–Var convergence study. *Tellus*, *59A*, 706–718.

Wei, J., Malanotte-Rizzoli, P., Eltahir, E. A. B., Xue, P., & Xu, D. (2014). Coupling of a regional atmospheric model (RegCM3)and a regional oceanic model (FV-COM) over the maritime continent. *Climate Dynamics*, *43*, 1575–1594. doi: 10.1007/s00382-013-1986-3

Weisberg, R. H., Barth, A., Alvera-Azcárate, A., & Zheng, L. (2009). A coordinated coastal ocean observing and modeling system for the West Florida Continental Shelf. *Harmful Algae*, *8*(4), 585 - 597. Retrieved from `http://www.sciencedirect.com/science/article/pii/S156898830800156X` (Understanding the causes and impacts of the Florida Red Tide and improving management and response) doi: https://doi.org/10.1016/j.hal.2008.11.003

Wolfram, P., Ringler, T., Maltrud, M., Jacobsen, D., & Petersen, M. (2015). Diagnosing isopycnal diffusivity in an eddying, idealized midlatitude ocean basin via Lagrangian, In Situ, Global, High-performance particle Tracking (LIGHT). *J. Phys. Oceanogr.*, *45*, 2114–2133.

Woodcock, F., & Engel, C. (2005, 02). Operational consensus forecasts. *Weather and Forecasting*, *20*(1), 101-111. Retrieved from `https://doi.org/10.1175/WAF-831.1` doi: 10.1175/WAF-831.1

Ye, F. (2017). *Developing efficient high-order transport schemes for cross-scale coupled estuary-ocean modeling scale coupled estuary-ocean modeling* (Unpublished doctoral dissertation). College of William and Mary.

Ye, X., Chu, P. Y., Anderson, E. J., Huang, C., Lang, G. A., & Xue, P. (2020). Improved thermal structure simulation and optimized sampling strategy for Lake Erie using a data assimilative model. *Journal of Great Lakes Research*, *46*(1), 144 - 158. Retrieved from `http://www.sciencedirect.com/science/article/pii/S0380133019302084` doi: https://doi.org/10.1016/j.jglr.2019.10.018

Zerfas, C. (2019). *Numerical methods and analysis for continuous data assimilation in fluid models in fluid models* (Unpublished doctoral dissertation). Clemson University.

Zerfas, C., Rebholz, L. G., Schneier, M., & Iliescu, T. (2019). Continuous data assimilation reduced order models of fluid flow. *arXiv: Numerical Analysis*.

Zhang, S., Wang, M., Ghan, S., Ding, A., Wang, H., Zhang, K., ... Fu, C. (2016). On the characteristics of aerosol indirect effect based on dynamic regimes in global climate models. *Atmospheric Chemsitry and Physics*, *16(5)*, 2765-2783. Retrieved from `https://doi.org/10.5194/acp-16-2765-2016` doi: 10.5194/acp-16-2765-2016

Zheng, L., & Weisberg, R. (2012). Modeling the west Florida coastal ocean by downscaling from the deep ocean, across the continental shelf and into the estuaries. *Ocean Modell.*, *48*, 10-29. doi: 10.1016/j.ocemod.2012.02.002