JAMES | Journal of Advances in Modeling Earth Systems®

RESEARCH ARTICLE
10.1029/2023MS003783

Special Collection:
Data assimilation for Earth system models

Vertical Localization for Strongly Coupled Data Assimilation: Experiments in a Global Coupled Atmosphere-Ocean Model

Zofia C. Stanley[1,2], Clara Draper[2], Sergey Frolov[2], Laura C. Slivinski[2], Wei Huang[1,2], and Henry R. Winterbottom[3]

[1]Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder, Boulder, CO, USA, [2]NOAA Physical Sciences Laboratory, Boulder, CO, USA, [3]Lynker Technologies/NOAA/EMC/EIB, Leesburg, VA, USA

Key Points:

- We investigate vertical localization for strongly coupled atmosphere-ocean data assimilation in a realistic global model
- Strong coupling can improve data assimilation effectiveness over weak coupling when large ensembles are used
- We present a method for optimal observation space localization, called EORL, and demonstrate its performance in offline experiments

Correspondence to:
Z. C. Stanley,
zofia.stanley@colorado.edu

Citation:
Stanley, Z. C., Draper, C., Frolov, S., Slivinski, L. C., Huang, W., & Winterbottom, H. R. (2024). Vertical localization for strongly coupled data assimilation: Experiments in a global coupled atmosphere-ocean model. *Journal of Advances in Modeling Earth Systems*, *16*, e2023MS003783. https://doi.org/10.1029/2023MS003783

Received 27 APR 2023
Accepted 25 JAN 2024

Author Contributions:
Conceptualization: Zofia C. Stanley, Clara Draper, Sergey Frolov, Laura C. Slivinski
Data curation: Wei Huang, Henry R. Winterbottom
Formal analysis: Zofia C. Stanley

Abstract Strongly coupled data assimilation allows observations of one Earth system component (e.g., the ocean) to directly update another component (e.g., the atmosphere). The majority of the information transfer in strongly coupled atmosphere-ocean systems is passed through vertical correlations between atmospheric boundary layer and ocean mixed layer fields. In this work we use correlations from a global, coupled model to study vertical observation-space localization techniques for strongly coupled data assimilation. We generate target correlations using a bootstrapping approach from a single 24 hr forecast from a realistic global, weakly coupled atmosphere-ocean cycling system with an 80-member ensemble, which is the ensemble size currently used by the NOAA operational global data assimilation system. We compare data assimilation methods with different localization schemes using single-update, offline experiments. We develop a new strategy for optimal observation space localization, called Empirical Optimal R-localization (EORL), to give an upper bound on the improvement we can expect with any localization scheme. We then evaluate Gaspari-Cohn localization, which is a commonly used parametric localization function and review its performance with respect to the optimal localization scheme. We investigate how the performance of these localization strategies changes with increasing ensemble sizes. Our results show that strongly coupled data assimilation has the potential to be an improvement over weakly coupled data assimilation when large ensembles are used. We also show that the Gaspari-Cohn localization function does not appear to be a particularly good choice for cross-fluid vertical localization.

Plain Language Summary Accurate Earth system forecasts rely on accurate estimates of the current state of the system. This initial state is estimated through a process called data assimilation, which combines the previous forecast with current observations. This process relies on accurate estimates of uncertainty in both the model and the observations. Inaccurate estimates of model and observation uncertainty can result in a degradation of future forecasts. A technique called localization is widely used to minimize the impact of distant observations due to unreliable uncertainty estimates over long distances. In this work we study localization in the context of a global, coupled atmosphere-ocean model. In particular, we look at strongly coupled data assimilation, where observations of the ocean are used to update the atmospheric model state and vice versa. This is in contrast to weakly coupled data assimilation, where observations of the ocean are only used to update the ocean model state. We investigate different types of localization and find that a commonly used localization function does not perform particularly well in strongly coupled model initialization.

1. Introduction

Coupled model development is an active area of research for operational weather prediction centers including the European Center for Medium Range Weather Forecasting and the National Oceanographic and Atmospheric Administration in the United States (de Rosnay et al., 2022; NOAA Science Advisory Board, 2021). In addition, recent outlooks on the future of Earth system reanalysis products have recognized the need for consistent atmosphere and ocean state estimates, which may be achieved through coupled models (Frolov et al., 2023). As coupled models are adopted in a growing number of applications, it is necessary to look closely at how we initialize these systems. In the case of numerical weather prediction, forecast models are initialized from real world observations of current conditions using a process called data assimilation (DA; Bengtsson et al., 1981; Lorenc, 1986). For a coupled dynamical model, such as an atmosphere-ocean model, there are multiple ways to assimilate observations. One option is to only assimilate observations of the atmosphere into the atmospheric

Funding acquisition: Clara Draper, Sergey Frolov, Laura C. Slivinski
Investigation: Zofia C. Stanley
Methodology: Zofia C. Stanley
Project administration: Clara Draper, Sergey Frolov, Laura C. Slivinski
Software: Zofia C. Stanley, Wei Huang, Henry R. Winterbottom
Supervision: Clara Draper, Sergey Frolov, Laura C. Slivinski
Visualization: Zofia C. Stanley
Writing – original draft: Zofia C. Stanley
Writing – review & editing: Zofia C. Stanley, Clara Draper, Sergey Frolov, Laura C. Slivinski, Henry R. Winterbottom

model and observations of the ocean into the ocean model. This is called weakly coupled data assimilation (WCDA; Penny & Hamill, 2017). In this case, all of the coupling between states comes through dynamical coupling of the forecast model. The DA step provides no additional coupling between the states. Alternatively, observations of both fluids can be used to update both model components directly. This is called strongly coupled data assimilation (SCDA; Frolov et al., 2016; Sluka et al., 2016; Smith et al., 2017).

Previous studies have not conclusively demonstrated a universal advantage for either SCDA or WCDA. For example, Lu et al. (2015) experimented with a very simple model with just a handful of degrees of freedom and found improvements for strongly coupled methods over weakly coupled methods. However, this improvement virtually disappeared when they introduced model error. Han et al. (2013) worked with a slightly more complex model, in the form of a Lorenz (1963) atmosphere coupled to a simple pycnocline ocean model. In biased model experiments they found that a prohibitively large ensemble size was needed to realize improvements for SCDA over WCDA. Goodliff and Penny (2022) used a quasigeostrophic model and variational DA scheme and found that WCDA and SCDA perform similarly in the atmosphere, while WCDA leads to the lowest error in the ocean. Tang et al. (2021) assimilated sea surface temperature observations into a global atmosphere-ocean model and found that WCDA and SCDA yield similar errors in the ocean. In the atmosphere, their results were state-variable and location dependent. For example, they found that SCDA yielded smaller errors for the temperature, wind velocity, and specific humidity in the Arctic region, while WCDA was generally preferable in the tropical region.

While the four studies above show mixed findings, there are physical reasons to believe that strong coupling may be beneficial. Namely, a strongly coupled scheme may be able to capitalize on the strong air-sea interfacial correlations in highly coupled regimes (Frolov et al., 2021) and may be able to make better use of satellite observations which are sensitive to both fluids. A number of studies argue unreservedly for the advantages of SCDA over WCDA. These include studies in a coupled Lorenz-Jin model (Liu et al., 2013), an idealized single-column atmosphere-ocean model (Smith et al., 2015, 2017, 2018), a linear inverse model (Hakim et al., 2022), a coupled quasi-geostrophic model (Penny et al., 2019), and an intermediate complexity coupled global circulation model (Sluka et al., 2016). Amidst the varied findings one thing is clear: the choice of coupled DA method depends on the model, model bias, observations, observation density, and underlying DA framework (e.g., ensemble or variational). Even within a single system the optimal method can vary spatially and by state variable (Tang et al., 2021; Yoshida & Kalnay, 2018). Given the value of coupled model initialization in subseasonal to seasonal forecasting and consistent reanalyses, it is important to investigate the potential benefits of SCDA and determine the conditions under which SCDA may outperform WCDA.

Successful applications of DA rely on accurate specifications of uncertainty in the initial conditions, as well as the uncertainty in the numerical forecast model and observations. In ensemble-based methods like EnKF and EnVar, the initial time and model uncertainty are specified using covariances within and between state variables. The cross-fluid correlations employed in strongly coupled initialization for atmosphere-ocean models are weak in an average sense, but can fluctuate to much larger values as a result of local coupling (Frolov et al., 2021; Smith et al., 2017). Ensemble forecasts, such as those used in NOAA's operational Global Ensemble Forecast System (GEFS; Zhou et al., 2022) are used to understand uncertainty in the forecast and can be leveraged to model the highly flow-dependent cross-fluid correlations necessary for SCDA. Due to computational limitations, ensemble sizes are limited (e.g., GEFS uses 80 ensemble members) and are typically many orders of magnitude smaller than the size of the state. This leads to significant sampling errors in the computed statistics. These sampling errors are known to degrade the quality of the state estimate and are typically mitigated through a process called localization (e.g., Gaspari & Cohn, 1999; Houtekamer & Mitchell, 2001; Morss & Emanuel, 2002). Localization is typically incorporated into ensemble-based DA schemes either through the background error covariance matrix (model space localization; Houtekamer & Mitchell, 2001), or the observation error matrix (observation space localization; Hunt et al., 2007). In this study we focus on observation space localization.

In atmospheric models, localization is often distance-based and predicated on the assumption that variables which are distant from each other will also have uncorrelated forecast errors. There is a long history of studies on localization in the atmosphere or ocean (e.g., Anderson & Lei, 2013; Bishop & Hodyss, 2007; Gaspari & Cohn, 1999; Houtekamer & Mitchell, 2001; Ménétrier et al., 2015; Necker et al., 2023). In contrast, localization methods for cross-fluid assimilation between the atmosphere and the ocean is much less established. Several studies have looked at this issue in reduced complexity models (Smith et al., 2018; Stanley et al., 2021). The majority of the information transferred across the fluid interface is captured in vertical correlations between the

atmospheric boundary layer and the ocean mixed layer (Frolov et al., 2016). Hence, understanding how to localize vertical correlations is the first step toward SCDA. To our knowledge, the study described in this manuscript is the first to investigate optimal vertical localization strategies for cross-fluid assimilation in a realistic atmosphere-ocean coupled system.

In this manuscript we use correlations from a global, coupled model to study vertical observation-space localization techniques for SCDA. We compute correlations from a single 24 hr forecast from a realistic global, weakly coupled atmosphere-ocean cycling system with an 80-member ensemble, which is the ensemble size currently used by the NOAA global data assimilation system. We use bootstrapping to increase the effective ensemble size and estimate target correlations. We then compare how well different localization schemes perform within single-update, offline experiments. We propose two benchmarks, which represent two extremes of performance: (a) no localization at all (this is generally understood to cause poor performance), and (b) optimal localization, which we calculate empirically (to give an upper bound on the improvement we can expect with any localization scheme). We then evaluate Gaspari-Cohn localization, which is a commonly used parametric localization function and review its performance with respect to these two benchmark measures. We investigate how the performance of these localization strategies changes with different ensemble sizes. Finally, we offer suggestions for important avenues of future research to determine an appropriate and practical localization scheme for global, coupled models.

2. Methods

2.1. Data

We evaluate vertical localization strategies using a single 24 hr ensemble forecast from a global, weakly coupled cycling system. This ensemble forecast comes from a NOAA Unified Forecast System (UFS; Unified Forecast System, 2023) research to operations project which couples the atmospheric model from the Finite Volume Cubed Sphere Global Forecast System (FV3 GFS; Harris et al., 2021), Modular Ocean Model version 6 (MOM6; Adcroft et al., 2019), and the Los Alamos sea ice model (CICE; Hunke et al., 2017). The forecast ensemble size is 80 members. The atmosphere and ocean components are run at a 1-degree horizontal resolution. The FV3 GFS and MOM6 models have different horizontal grid spacing, so to facilitate comparison, for the purposes of this study each forecast is interpolated to a uniform 1-degree grid. The forecast is valid at 03Z 6th December 2015. We use a 24 hr lead time for the forecast since it takes about 12 hr to equilibrate unbalanced atmospheric and oceanic initial conditions (Laloyaux et al., 2018).

The initial atmospheric ensemble is defined by the 80 member operational GFS ensemble. The ocean is initialized from the Next Generation Global Ocean Data Assimilation System (NG-GODAS; Kim et al., 2022) analysis and the respective ocean and ice ensemble members are computed via perturbations added by the JEDI-based Sea-ice Ocean and Coupled Analysis (JEDI-SOCA; JCSDA Project: SOCA, 2023) ensemble perturbation application, which generates perturbations by sampling from the square root of the 3DVAR covariance function. The weakly coupled ensemble, including the stochastic perturbations, is spun up for 10 days. We assume that the atmospheric states are near statistical equilibrium at this point, while the ocean ensemble is likely under-dispersive with spread primarily driven by the variability in the atmospheric forcing. However, we expect that any 1-degree ocean model will have low internal (eddy-driven) variability, leading to insufficient internal spread. Thus we expect that the variability in the ocean component is representative of variability in other realistic global coupled models at similar horizontal resolution.

Throughout this work we focus on atmosphere and ocean temperature fields. We assimilate synthetic observations of surface level temperature fields. For simplicity the surface level fields are taken to be the value of the temperature in the vertical level closest to the air-sea interface. That is, we take the lowest atmospheric level and call its temperature the "atmospheric surface temperature." Likewise, the temperature in the top level of the ocean model is taken to be the "sea surface temperature." Figure 1 shows the 24hr forecast ensemble mean and spread temperature fields from the ensemble that will be used to evaluate localization strategies.

2.2. Target and Sample Correlations

In order to understand the effects of localization, we establish "target" correlations (Figure 2) using a large (2,000 member) ensemble that was bootstrapped from the 80 member ensemble described in the previous section

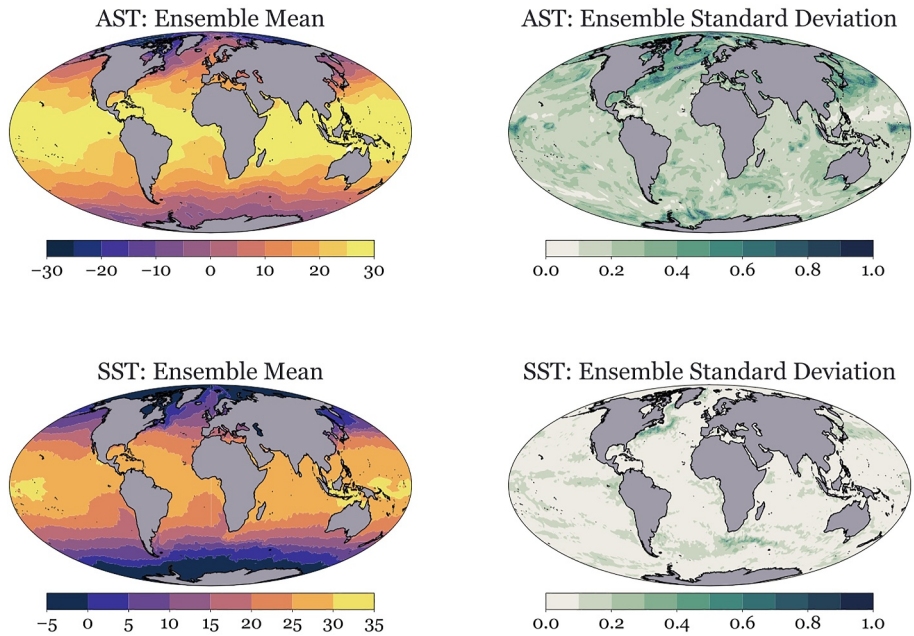

Figure 1. Four panels show ensemble mean (left; degrees C) and ensemble standard deviation (right; degrees C) for atmospheric surface temperature (top; AST) and sea surface temperature (bottom; SST) in the realistic 24 hr forecast used throughout this investigation. The surface level fields are taken to be the value of the temperature in the vertical level closest to air-sea interface. The forecast is valid for 6th December 2015 at 03Z and is produced using an 80 member ensemble.

(Tibshirani, 1994). We are unable to estimate the unknown true correlations. Instead, these bootstrapped correlations serve as a useful target for investigating the performance of different localization methods. The bootstrapping approach is similar to the approach used in Gasperoni et al. (2022) and serves to dampen spuriously large correlations. The target correlations can be viewed as estimates of correlations from ensembles that are much larger than what is currently used operationally. They form the basis of our evaluation of different localized sample correlations.

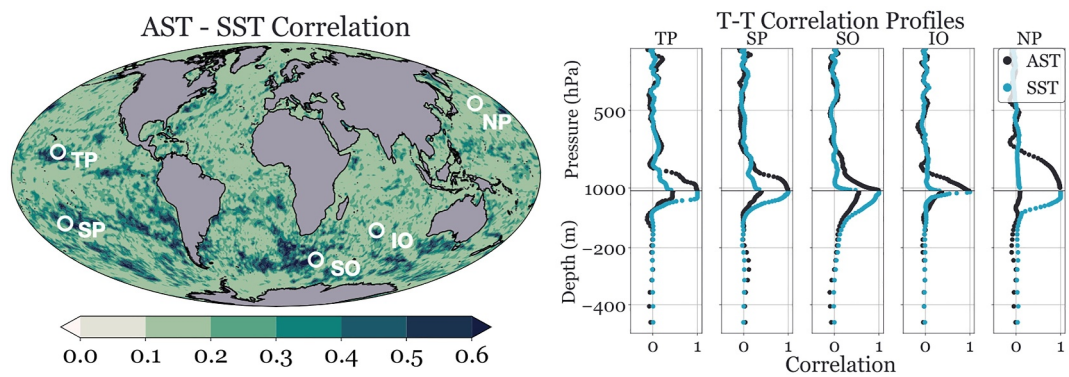

Figure 2. The panel on the left shows the correlation between atmospheric surface temperature (AST) and sea surface temperature (SST). The correlations are not strictly positive, but negative values are rare. Stronger correlations, greater than 0.3, appear to be highly flow-dependent. The right panel shows correlations between atmospheric surface temperature and the entire atmosphere-ocean temperature column (black). Correlations with sea surface temperature are shown in blue. These correlation columns are shown for five locations in the Tropical Pacific (TP), South Pacific (SP), Southern Ocean (SO), Indian Ocean (IO), and North Pacific (NP). Column locations are shown with black circles in the plot on the left. Three columns with large cross-correlations (in the upper 10%) are shown (TP, SP, SO). One location with moderate cross-correlation (close to the global mean) is chosen (IO), as well as one location with near-zero cross-correlation (NP). Across all column locations, there is a discontinuity in correlation across the fluid boundary, consistent with the plot on the right.

Our bootstrapping approach assumes local stationarity in the ensemble perturbations. Note that the ensemble mean is not locally stationary, but when we remove the mean, we see that the structure of the perturbations is similar in local neighborhoods. Thus, starting from the 80 member ensemble, we remove the ensemble mean at each grid cell i to create an ensemble of perturbations, $\boldsymbol{\eta}_b^i$, where the subscript b refers to the background state. At each horizontal grid cell we gather ensemble perturbations from 5×5 columns centered on the grid cell in question. This gives a collection of 2,000 (25×80) ensemble perturbations per grid point. From this collection of perturbations we compute the target covariance matrix $\mathbf{P}_b^*$ (Equation 1) and target correlation matrix $\mathbf{C}_b^*$ (Equation 2).

$$\mathbf{P}_b^* = \frac{1}{N-1}\sum_{i=1}^{N}\left(\boldsymbol{\eta}_b^i\right)\left(\boldsymbol{\eta}_b^i\right)^{\top} \tag{1}$$

$$\mathbf{C}_b^* = \left(\mathrm{diag}\left(\mathbf{P}_b^*\right)\right)^{-1/2}\mathbf{P}_b^*\left(\mathrm{diag}\left(\mathbf{P}_b^*\right)\right)^{-1/2} \tag{2}$$

To maintain a tractable scope, we focus exclusively on temperature-temperature correlations in the atmosphere and ocean. However, this experimental procedure is general and can be applied to any observing system and state variable pairs. Figure 2 (left) shows the correlation between background errors in sea surface temperature and background errors in atmospheric surface temperature. Figure 2 (right) shows the vertical correlations between background errors in surface temperatures and background errors in temperature throughout the entire atmosphere-ocean column for five selected locations: Tropical Pacific (TP), South Pacific (SP), Southern Ocean (SO), Indian Ocean (IO), and North Pacific (NP). A range of cross-interface correlations are shown: three locations have correlation strength in the upper 10% (TP, SP, SO), one location has a correlation strength that is close to the global average (IO), and one location has low correlation strength (NP).

With the target correlations described above, we now estimate sample correlations, to be used as a proxy for the correlations that would be used in practice. In our experiments we will test different approaches to localizing these sample correlations, and then compare the results to the target. To compute our sample correlations, we sample normally distributed ensembles of background errors $\{\boldsymbol{\eta}_{b,s}^i\}_{i=1}^{N_e}$ from the target covariance matrix $\mathbf{P}_b^*$, as follows:

$$\boldsymbol{\eta}_{b,s}^i = \left(\mathbf{P}_b^*\right)^{1/2}\boldsymbol{\varepsilon}, \text{ with } \boldsymbol{\varepsilon} \sim N(\mathbf{0},\mathbf{I}) \tag{3}$$

2.3. Vertical Localization

In this study we focus on R-matrix (or observation-space) localization, which inflates the observation error variance for distant observations and is commonly used in both atmosphere and ocean DA (Sluka et al., 2016). An alternative localization method which is commonly used in variational solvers is B-matrix (or model-space) localization, which applies a cutoff function to the background covariance matrix. We defer investigation of B-matrix localization to future studies and note that the requirement of maintaining positive definiteness of the localized B-matrix is not present with R-matrix localisation. R-matrix localization allows for a different localization strategy for each observation type, and hence is helpful for coupled localization where a wide number of different observations are used. In addition, R-matrix localization is used in many local DA schemes, such as the Local Ensemble Transform Kalman Filter (LETKF; Hunt et al., 2007). With R-matrix localization, the solution for each grid cell i is as follows:

$$\boldsymbol{x}^{\mathrm{inc}}(i) = \mathbf{S}_i\mathbf{P}_b\mathbf{H}_i^{\top}\left(\mathbf{H}_i\mathbf{P}_b\mathbf{H}_i^{\top} + \mathbf{L}_i \circ \mathbf{R}_i\right)^{-1}\left(\boldsymbol{y}_i - \mathbf{H}_i\boldsymbol{x}_b(i)\right) \tag{4}$$

where $\boldsymbol{x}^{\mathrm{inc}}(i)$ is the increment at the ith grid cell and $\mathbf{S}_i$ is the selection operator that selects the row of the state vector corresponding to the ith grid point. The local observation vector, $\boldsymbol{y}_i$, corresponds to a subset of observations in a region near grid cell i, and the local observation operator, $\mathbf{H}_i$, consists of the corresponding rows of the full observation operator $\mathbf{H}$. Likewise, $\mathbf{R}_i$ and $\mathbf{L}_i$ are the local observation error covariance and localization matrices corresponding to local observations $\boldsymbol{y}_i$. The $\cdot$ symbol represents the Schur, or element-wise, product. For the purposes of this investigation, observations are treated individually so that $\mathbf{L}_i = L_i$ and $\mathbf{R}_i = R_i$ are scalars and $\mathbf{L}_i \circ \mathbf{R}_i = L_i \times R_i$.

One common way to prescribe localization weights is through the use of distance-based functions, so that the observation error variance of distant observations is inflated and the background state is only minimally updated by the distant observation. For this investigation we use synthetic surface level observations (sea surface temperature and atmospheric surface temperature). The vertical distance (log hPa) from an atmospheric grid cell i to either of these observations is computed as follows:

$$d_i^A = \log(p_i) - \log(p_s) \tag{5}$$

where p_i is the pressure at grid cell i and p_s is atmospheric surface pressure. The distance (m) from an ocean grid cell i to either of the observations is simply the depth below the surface of the grid cell i. By considering only surface level observations, we never need to measure distances across the two fluids and thus do not need to contend with the different units used to measure distance in the atmosphere and ocean. In general to compute distance-based localization weights for observations that are not located at the surface, one could use normalized distances, as in Frolov et al. (2016).

Previous studies have constructed "optimal" localization methods to provide an upper bound on the performance of practical and tractable localization schemes (e.g., Anderson & Lei, 2013; Destouches et al., 2021; Flowerdew, 2015; Ménétrier et al., 2015; Necker et al., 2023). In particular, Necker et al. (2023) used an empirically derived localization method to provide guidance on vertical localization in an atmospheric model. They develop their Empirical Optimal Localization (EOL) for B-matrix localization. In this work we present a similar approach for R-matrix localization and call it Empirical Optimal R-localization (EORL). In this approach, we allow the localization weights that multiply the R-matrix to vary at each grid cell. We perform a separate scalar minimization at each grid cell to set the localization weight L_i in Equation 4. Specifically, we minimize the error in the Kalman gain with respect to a target Kalman gain, $\mathbf{K}^*$. As an example, consider a single horizontal location with an observation operator $\mathbf{H}$ which is 1 at the surface and 0 elsewhere. The Kalman gain at vertical level i is as follows:

$$\mathbf{K}_i = \mathbf{S}_i\mathbf{P}_b\mathbf{H}^\top\left(\mathbf{H}\mathbf{P}_b\mathbf{H}^\top + \mathbf{L}_i \circ \mathbf{R}\right)^{-1}. \tag{6}$$

Note that $\mathbf{H}\mathbf{P}_b\mathbf{H}^\top = \sigma^2_{\text{surface}}$ is a scalar, since the observation operator corresponds to a single surface-level observation. Further, $\mathbf{S}_i\mathbf{P}_b\mathbf{H}^\top = \sigma_{\text{surface},i}$ is the (scalar) covariance between the state at level i and the surface-level state. The observation error variance $\mathbf{R} = r^2$ is also a scalar since we have a single observation. Finally, the localization coefficient $\mathbf{L}_i = L_i$ is a scalar and is allowed to vary with vertical coordinate i. To illustrate that the Kalman gain $\mathbf{K}_i$ varies with different choices of localization coefficient L_i we abuse notation and write the Kalman gain as a function of localization coefficient $\mathbf{K}_i(L)$. To find the optimal localization coefficient, we define a target Kalman gain $\mathbf{K}_i^*$ and find the localization coefficient which solves the following minimization problem:

$$L_i = \min_L \|\mathbf{K}_i(L) - \mathbf{K}_i^*\| \tag{7}$$

The choice to focus on the error in the Kalman gain and the procedure for identifying the target $\mathbf{K}^*$ are discussed in the following section. To perform the minimization, we use a hybrid root-finding method known as Brent's algorithm (Press et al., 1992), as implemented in the `scipy.optimize` package in Python (Virtanen et al., 2020). EORL provides an upper bound on the improvement we can expect to achieve with an R-matrix localization scheme, but it relies on knowledge of the target correlations, which are unknown in a realistic setting.

In practice, localization schemes should be easily implementable. This often prompts the use of a parametric localization function. Hence, in addition to EORL, we investigate the performance of two optimal parametric localization strategies based on the Gaspari-Cohn function. This function was first introduced as a piecewise, compactly supported approximation to a Gaussian (Gaspari & Cohn, 1999). Its compact support means that localized covariance matrices have many 0 entries and this sparsity is computationally advantageous in DA systems. For these reasons, the Gaspari-Cohn localization function is widely used in DA studies and is currently used in operational NOAA DA systems. To provide an upper bound on the performance of the Gaspari-Cohn

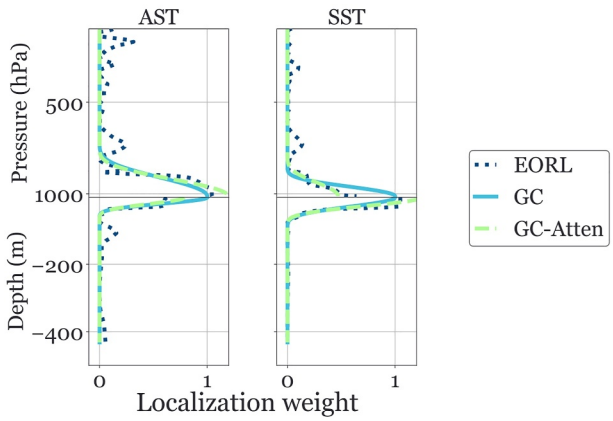

Figure 3. This figure shows the reciprocal of the functions used to inflate the observation error variances in R-matrix localization at a location in the Tropical Pacific (12.5 N, 150.5 E). Three localization methods are shown for each observation type with atmospheric surface temperature on the left and sea surface temperature on the right. With EORL (dark blue, dotted), localization weights are determined separately at each vertical level. With GC (light blue, solid), a separate length scale is estimated for each observation/fluid pair. GC-Atten (green, dashed) is also parametric, but allows for different function magnitudes for each observation/fluid pair in addition to different localization length scales.

localization functions, we estimate optimal parameters separately at each horizontal location. The first parametric localization method we consider simply uses the localization function f^{GC} from Gaspari and Cohn (GC; Gaspari & Cohn, 1999). That is, we set the localization weight at each vertical level i as follows:

$$L_i = 1/f^{GC}(d_i; r) \tag{8}$$

where d_i is the distance from the ith vertical level to the (surface-level) observation and r is the optimal localization length scale. We set the localization length scale r separately at each horizontal localization using the minimization procedure described above.

The Gaspari-Cohn function has a magnitude of one when distance $d_i = 0$. While this is appropriate for within-fluid assimilation, where correlations are close to one for neighboring vertical levels, it may not be appropriate for the small and fluctuating cross-fluid correlations that characterize SCDA (Figure 2). To understand this, note that linear filtering theory in Ménétrier et al. (2015) tells us that the optimal localization weight depends on the average (squared) correlation between two state variables. Further, the optimal localization weight decreases with decreasing correlation strength. Hence, we may expect that the optimal localization function for cross-fluid assimilation will have a magnitude smaller than one. Thus, with our second parametric localization method we allow the magnitude of the function to be an arbitrary positive number and set the localization weights as follows:

$$L_i = c/f^{GC}(d_i; r) \tag{9}$$

where d_i is the distance from the ith grid cell to the observation, c sets the magnitude of the function at distance $d_i = 0$, and r is the localization length scale. The parameters c and r are set separately at each horizontal localization using the direct-search simplex method (Nelder & Mead, 1965) which is capable of minimizing multivariate functions. The implementation we use is from the `scipy.optimize` package in Python (Virtanen et al., 2020). We call this approach Gaspari-Cohn with attenuation (GC-Atten). Figure 3 illustrates the localization weights resulting from these three schemes (EORL, GC, and GC-Atten) at a location in the Tropical Pacific (12.5 N, 150.5 E).

2.4. Metrics for Evaluating Localization

To evaluate the performance of the localization schemes, we consider different metrics for measuring the distance between target and localized sample correlations. Previous works on optimal localization strategies often focused on B-matrix localization and measured error in the localized B-matrix with respect to the target B-matrix (e.g., Anderes et al., 2013; Ménétrier et al., 2015; Necker et al., 2023). This error metric is not appropriate here because R-matrix localization does not regularize the B-matrix directly. Other possible metrics include measurements of error in the analysis mean, error in the mean increment, and error in the Kalman gain. This study uses the Kalman gain metric because (unlike errors in the analysis mean and mean increment) the Kalman gain metric is not a function of errors in the mean state (Equation 10) and most clearly measures the correlation structures which are the focus of this investigation. This is the strategy employed by Sakov and Bertino (2011) because it allows for the comparison of B-matrix and R-matrix localization strategies.

To compute the error in the Kalman gain, we first need a target Kalman gain. The target Kalman gain, $\mathbf{K}^*$, is computed from the target correlations discussed above as follows:

$$\mathbf{K}^* = \mathbf{P}_b^* \mathbf{H}^\top \left(\mathbf{H}\mathbf{P}_b^* \mathbf{H}^\top + \mathbf{R}^*\right)^{-1} \tag{10}$$

where $\mathbf{P}_b^*$ is the target background error covariance matrix and $\mathbf{H}$ is the observation operator. We select the observation error variance to be equal to the variance of the ensemble at the observed location, that is, $\mathbf{R}^* = \mathbf{H}\mathbf{P}_b^*\mathbf{H}^\top$. To facilitate comparisons among grid cells, we compute a relative error between the sample Kalman gain, K, and the target Kalman gain, $\mathbf{K}$*, as follows:

$$err_{\mathbf{K}} = \|\mathbf{K} - \mathbf{K}^*\|^2 / \|\mathbf{K}^*\|^2 \tag{11}$$

We choose to use the 2-norm. The relative error is bounded below by 0, which indicates a perfect update, $\mathbf{K} = \mathbf{K}^*$. When this error is equal to 1, we see that $\|\mathbf{K} - \mathbf{K}^*\| = \|\mathbf{K}^*\| = \|\mathbf{K}^* - \mathbf{0}\|$, so using $\mathbf{K}$ is no more accurate than simply persisting the background state with no update ($\mathbf{K} = \mathbf{0}$). Taking this relative error one step further, we compute the percent improvement over persistence, ϕ, as follows:

$$\phi = \left(1 - \|\mathbf{K} - \mathbf{K}^*\|^2 / \|\mathbf{K}^*\|^2\right) \times 100 \tag{12}$$

Throughout the rest of this manuscript we will discuss our results in terms of percent improvement over persistence. Percent improvement equal to zero indicates that our update did no better than a persistence, while values close to 100 show optimal improvement. Negative values indicate that the update degraded the analysis state and that it would have been better not to assimilate observations at all. Furthermore, in the case of cross-component assimilation we can think of improvement over persistence as improvement for SCDA over WCDA. While we do not have a WCDA system, since WCDA relies on the forecast step to provide the coupling, we use the terms SCDA and WCDA to differentiate between a single update with a SCDA versus a WCDA system. Consider the errors associated with the assimilation of only SST (and no atmospheric observations) into the atmospheric component of the model. In a WCDA system, SST would not be assimilated into the atmosphere. In other words, the atmospheric state would be persisted. By contrast, in a SCDA system, SST would be assimilated into the atmosphere. Thus, any improvements because of the assimilation of SST into the atmosphere can be thought of as improvement for SCDA over WCDA.

3. Experimental Design

To understand how the design of vertical localization impacts our ability to transfer innovations between fluids, we conduct a series of experiments. First, we fit three optimal localization functions to estimate best possible performance with a modest size ensemble (80 members). The non-parametric EORL localization provides an upper bound for possible performance. We also present two optimal GC formulations to provide an upper estimate of the performance with a popular parametric form of localization. We draw 1,000 ensembles, each containing 80 members, compute the sample correlations, and report the average ϕ (Equation 12) at each location.

To ensure that our results capture a wide variety of air-sea coupling regimes, we use all ($n = 24,801$) available coupled atmosphere-ocean grid points that are further than 200 km from the coastline. In all experiments we assimilate atmospheric surface temperature and sea surface temperature into both the atmosphere and the ocean separately. That is, we consider four pairs of observations and fluids and conduct a separate assimilation experiment for each observation/fluid pair, which we use to test the different localization strategies. Two of these pairs are classified as within fluid assimilation (atmospheric surface temperature assimilated into the atmosphere, sea surface temperature assimilated into the ocean) and the other two constitute cross-fluid assimilation. In this study we are especially interested in the implementation of SCDA, and hence are particularly interested in the cross-fluid pairs. Results from this experiment are shown in Figure 4 and discussed in the following section.

For a second experiment, to understand how ensemble size impacts the results, we increase the ensemble size by about a factor of three to 256 ensemble members. The setup is otherwise identical to the previous experiment. Results are shown in Figure 5 and discussed in the following section.

In a further series of experiments, we investigate how the number of ensemble members affects the asymptotic behavior of the optimal localization functions. We vary the number of the ensemble members exponentially (in powers of two) from 4 to 256. At each ensemble size we compare results with EORL, Gaspari-Cohn, and no localization. We compute the distribution of percent improvements as in the previous two experiments.

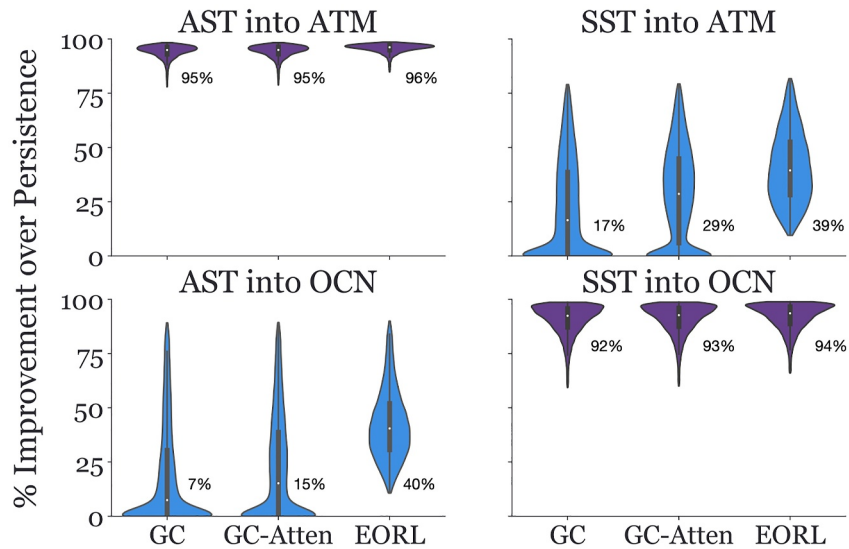

Figure 4. Violin plots show the distribution of percent improvement over persistence for three different "optimal" localization methods with an 80 member ensemble. For each of the $n = 24,801$ grid points we include the average of 1,000 trials. Thus the violin plots show spatial variability in performance. The median is marked on each violin plot with a small white dot. Numerical values of the median are also written to the right of each violin plot. The top and bottom 0.5% of data points are discarded as outliers and not shown. The violin plots are scaled so that, within each panel, each violin plot has the equal area. Results are separated by observation/fluid pair with atmospheric surface temperature observations on the left, sea surface temperature observations on the right, and atmospheric temperature states on top and ocean temperature states on bottom. The diagonal subplots (shown in purple) correspond to within-fluid assimilation, while the off-diagonal plots (shown in blue) correspond to cross-fluid assimilation. We also compute percent improvement over persistence when no localization is used. These values are not shown because they fall outside of the bounds of this figure. The median percent improvement over persistence with no localization is: (i) AST into ATM: 90%, (ii) SST into ATM: −295%, (iii) AST into OCN: −213%, (iv) SST into OCN: 61%.

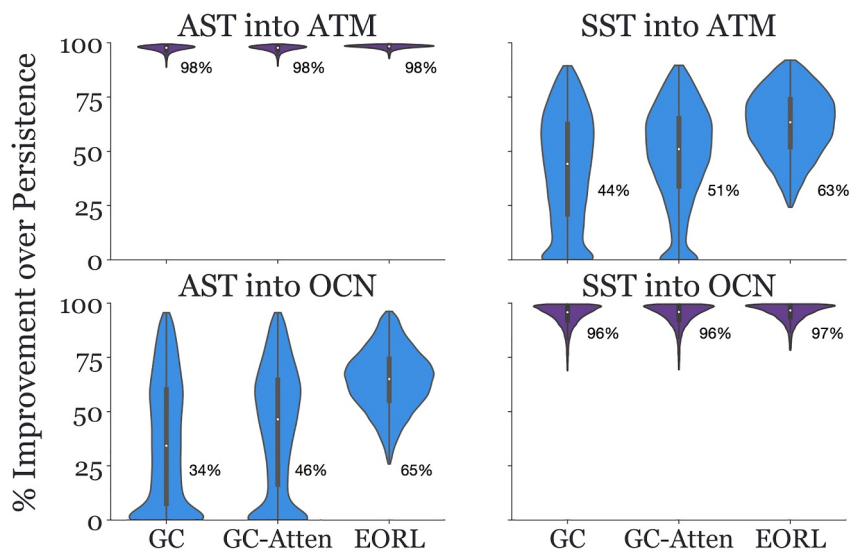

Figure 5. Same as Figure 4, but with 256 ensemble members. The median percent improvement over persistence with no localization (not shown) is: (i) AST into ATM: 96%, (ii) SST into ATM: 2%, (iii) AST into OCN: −23%, (iv) SST into OCN: 88%.

Finally, we recognize that in practical applications it might not be possible to find optimal EORL or GC parameters for each vertical column. To estimate the performance under such "practical" constraints, we examine performance of the localization with fixed values of the GC parameters. We first use a single localization length scale for the entire globe, which we set to be equal to the median of our estimated "optimal" length scales. We call this scheme OneRad. As in the previous cases, we allow the localization radius to vary with each fluid/observation pair. Previous studies have also proposed an adaptive approach to strong coupling, where observations of one fluid are allowed to update another fluid when modeled correlations exceed a prescribed threshold (Yoshida & Kalnay, 2018). Thus, we consider a second practical option, where we use Gaspari-Cohn localization with a single localization radius for the entire globe as in OneRad, but the localization radius is set to zero (and no assimilation is performed) when cross-fluid correlations are smaller than a prescribed threshold. We call this the Cutoff method and assimilate cross-fluid observations when the sample correlation between atmospheric surface temperature and sea surface temperature is at least 0.3. We use an ensemble size of 80 and compute ϕ as in the first experiment.

4. Results

Figure 4 shows the global distribution of improvements over persistence, ϕ, associated with the three localization schemes from the first experiment. In this figure we see that the improvement over persistence is largest for empirical optimal R-localization (EORL), followed by Gaspari-Cohn with attenuation (GC-Atten), followed by Gaspari-Cohn (GC). For example, we can look at the median percent improvement over persistence (shown with a white dot in the center of each violin plot), which is always larger with EORL than with either of the other localization methods. This is to be expected based on the number of degrees of freedom in each of these methods. EORL shows the upper bound on the improvement we can achieve with R-localization and has degrees of freedom equal to the number of vertical levels at each horizontal location. The GC-based methods impose a parametric form and hence have many fewer degrees of freedom, with just two degrees of freedom for GC-Atten, and one degree of freedom for GC. For within-component assimilation we see that all three localization methods perform quite similarly and all give close to 100% improvement, with even more improvements in the atmosphere (AST into ATM) when compared to the ocean (SST into OCN). This supports the decision to use a parametric localization function for within-fluid assimilation as there is not much to be gained by adding degrees of freedom.

For cross-component assimilation, the story is quite different. AST into OCN and SST into ATM (lower left and upper right subplots) look qualitatively similar and both show significant and consistent improvement over persistence with EORL. For example, with EORL almost none of the distribution is located at zero percent improvement over persistence, and indeed most of the mass is concentrated between about 30% and 60% improvement over persistence. This can also be interpreted as a significant improvement with SCDA over WCDA, when EORL is employed. It is possible that EORL is better at capturing the structure of the atmospheric boundary layer and marine mixed layer. Some of this improvement could be gained with a different parametric form, which is an interesting avenue for future research.

Both of the parametric localization methods we consider here have a large density of points with no improvement over persistence in cross-fluid assimilation, suggesting that there are grid points for which any assimilation of cross-fluid observations would degrade performance. Note that there are no negative values in Figure 4. We can explain this by considering the specific design of this experiment. When the Kalman gain $\mathbf{K}$ is identically zero we see in Equation 12 that $\phi = 0$ and there is no improvement over persistence. Intuitively this makes sense because $\mathbf{K} = \mathbf{0}$ means we are simply persisting the background state. One way to construct $\mathbf{K} = \mathbf{0}$ is by setting the localization radius to zero, since we update only the points with distance strictly less than the localization radius. Thus, it is always possible to construct a localization radius which gives $\phi = 0$. Assuming that the optimization schemes we use to maximize ϕ are able to explore the space of localization radii well, we should never see negative value of ϕ.

There is slight improvement in ϕ with GC-Atten over GC, despite the marginal addition in complexity. This is especially noticeable for the assimilation of sea surface temperature into the atmosphere. In general we see that, in this system, the assimilation of sea surface temperature into the atmosphere is more often beneficial compared to the assimilation of atmospheric surface temperature into the ocean. Although this is likely due to the way we set the observation error variance for each observation type, the result is encouraging, as the modern observing

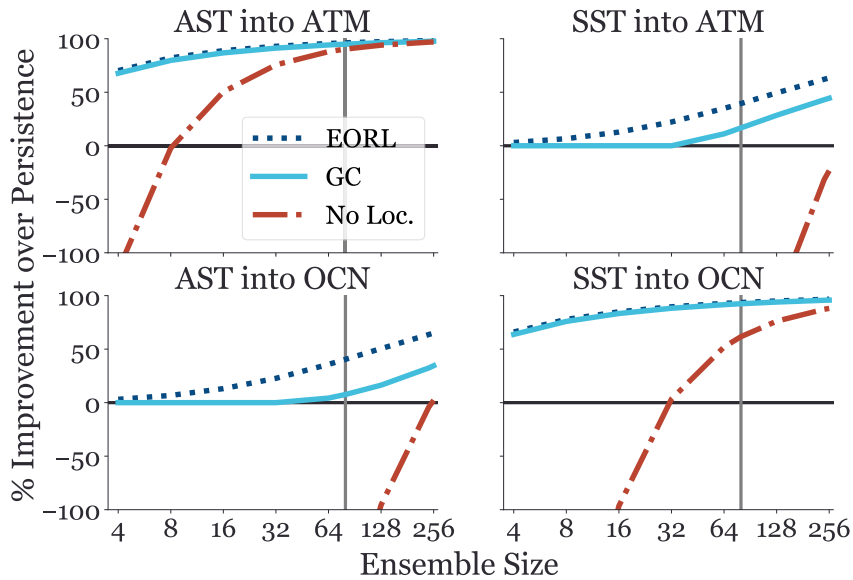

Figure 6. Median percent improvement over persistence is shown with different ensemble sizes (in powers of 2 from 4 to 256). Results are shown for EORL (dark blue, dotted), GC localization (light blue, solid), and no localization (red, dot-dashed). A solid gray vertical line marks 80 ensemble members, which is the size of current GEFS ensembles. Each observation/fluid pair is shown in a different subplot, similar to Figures 4 and 5.

system is saturated with sea surface temperature observations (order of ∼100 M in the 24 hr cycle) but is very sparse when it comes to observations of the air temperature in the marine boundary layer.

Figure 5 shows the effects of increasing the ensemble size approximately three-fold. Effects on within-fluid assimilation are small, suggesting that an ensemble size of 80 is sufficient for that purpose. The largest improvements in ϕ are seen in cross-fluid assimilation: the median ϕ with EORL increases from about 40% (Figure 4) to 60% (Figure 5). For the parametric methods, we also see a marked increase in the median ϕ, which increases almost three fold from about 15% when 80 members are used (Figure 4) to about 40% when 256 members are used (Figure 5). However, even with this large ensemble size, there are many locations where there is no improvement in ϕ with SCDA when compared to WCDA, as evidenced by the violin plots flaring out near zero percent improvement over persistence in the lower left and upper right subplots.

Figure 6 shows that as we systematically increase the ensemble size, the performance continues to improve for all assimilation and localization cases. In this experiment we choose not to include GC-Atten for clarity of exposition and because we see similar results for GC and GC-Atten in the results of the first two experiments. There is a stark divide between results from within-fluid and cross-fluid assimilation. While localization is certainly important for within-fluid assimilation, particularly with small ensemble sizes, we can see that it is absolutely necessary for cross-fluid assimilation. Indeed, without localization we see that assimilating observations is detrimental for all but the largest ensemble size (Ne = 256) considered here (red curve in Figure 6). Further, we see a large improvement with EORL when compared to GC in cross-fluid assimilation. EORL performs better than any operational localization scheme could hope to do because it is trained on target correlations which are unknown in realistic setups. We can see that special care will need to be taken when designing a vertical localization method for realistic applications.

Figure 6 shows that for our experiments assimilating surface temperatures to update model temperature in a single-update experiment, optimal localization with around 128 ensemble members is needed to achieve median percent improvement over persistence of at least 50% for cross-fluid assimilation. For within-ocean assimilation the same performance is achieved without any localization at all, and with half the number of ensemble members (64). For within-atmosphere assimilation this performance is achieved without any localization with just 16 ensemble members. We need 8 times more ensemble members for SCDA with perfect localization to realize the same level of performance as atmosphere-only assimilation with no localization in our single update experiment using correlations computed from a realistic global model. From these results, we can expect that SCDA will

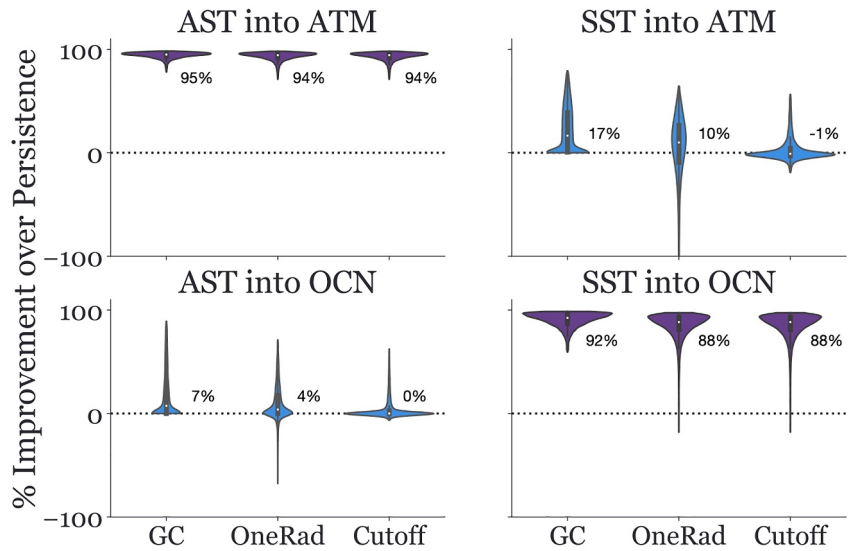

Figure 7. Violin plots show the distribution of percent improvement over persistence for two different "practical" localization methods, OneRad and Cutoff, as well as an optimally tuned parametric localization method, GC, with 80 ensemble members. The top and bottom 0.5% of data points are discarded as outliers and not shown. The violin plots are scaled so that, within each panel, each violin plot has equal area. OneRad uses the standard Gaspari-Cohn localization function with a single localization length scale for each observation/fluid pair. Localization length scales are chosen to be the median of the optimal length scales estimated for an 80 member ensemble. The length scales used here are: (1) AST into ATM, 0.81 log hPa; (2) SST into ATM, 0.10 log hPa; (3) AST into OCN, 30.5 m; (4) SST into OCN, 126.8 m. For within-fluid assimilation, Cutoff is identical to OneRad. For cross-fluid assimilation, Cutoff uses the same localization length scales and OneRad and only performs an update when the sample correlation between AST and SST is greater than 0.3. This cutoff is chosen to be approximately the 90th percentile of the target correlations. GC is as in Figure 4 and is provided here for reference. The dashed line at zero separates the locations where assimilation is helpful (positive values) and where it degrades performance compared to persisting the forecast (negative values).

require dramatic increases in ensemble size, well beyond the 80 members currently used in the NOAA operational global data assimilation system. Indeed, even with 256 ensemble members, we find greater errors in our experimental single-update strongly coupled system than what we would expect from an operational-sized atmosphere-only ensemble.

The optimal localization schemes investigated above perform better than any operational localization scheme could hope to do because they are trained on target correlations which are unknown in realistic setups. This motivates us to consider several practical localization schemes: OneRad where a single localization length scale is used globally, and Cutoff where cross-fluid updates are made only when sample correlations exceed a threshold of 0.3. As in Figures 4 and 5, the violin plots in Figure 7 show distributions of ϕ for different localization methods. Two practical localization methods (OneRad and Cutoff) are shown in comparison to an optimal localization scheme, GC. Results for both cross-fluid assimilation experiments show that using a single localization radius globally leads to performance that is an improvement over persistence in just over half of the locations (i.e., the median is just above zero), but can degrade the performance compared to persistence by over 100% in the worst cases. The cutoff method sharpens the distribution of ϕ so that nearly all of the locations show no improvement over persistence. The median percent improvement is slightly worse with the cutoff method when compared to using a single localization radius everywhere, but the cutoff method has a significantly better worst case performance. The decrease in median ϕ is likely due to sampling errors in the estimation of the ensemble correlation strength. Unlike the optimal localization methods which never degrade performance when compared to persistence, using either of these practical methods will lead to degradation of performance in about half of the locations.

5. Discussion

In this work we evaluated vertical localization strategies for SCDA with different ensemble sizes. We tested three different optimal localization strategies, two based on the Gaspari-Cohn localization function, and one which

gives the greatest possible error reduction. Our results show that SCDA has the potential to be an improvement over WCDA with correlations inferred from a realistic global model. However, localization will require some serious thought, as a simple parametric R-localization is unlikely to be sufficient to properly attenuate the small on average and highly flow-dependent cross-fluid correlations. In particular, Gaspari-Cohn parametric localization does not appear to be an especially good choice for cross-fluid vertical localization, when the localization radius is chosen globally. Realistic assimilation systems often allow localization radii to vary by latitude. In keeping with this, we looked at the differences in localization radii by latitude (not shown). However, we did not find meaningful differences between the optimal localization radii in different regions.

Our offline experiments make use of a single ensemble forecast from a realistic global coupled system, estimate target covariances using a neighborhood bootstrapping approach, and then sample ensembles from the target covariance model. This methodology allows us to investigate how to improve a DA system quickly and without the overhead needed when setting up a cycling system. In particular, our approach provides a clear routine for evaluating vertical localization in a coupled system. This framework is a valuable tool which can be used in the future to test different practical and implementable vertical localization strategies. Using this framework to inform localization strategies for other state variables, for example, atmospheric wind speed, is an exciting avenue for future research. While we assume a Gaussian distribution for the background errors, one could easily extend this experimental protocol for non-Gaussian background errors by employing a more sophisticated sampling technique when drawing ensemble members from the target background error distribution.

Although we focus on R-matrix localization, it is far from the only method to treat sampling errors in DA. Indeed, there is a rich history of studies on the difference between R-matrix localization, as we use here, and B-matrix localization, which localizes the background error covariance matrix directly (e.g., Lei & Whitaker, 2015; Nerger et al., 2012; Sakov & Bertino, 2011). With B-matrix localization one must take special care to ensure that the localized B-matrix is positive semidefinite (Gaspari & Cohn, 1999). This complication is avoided with R-matrix localization in the case of a diagonal R-matrix, as we have here. Any positive localization weight will ensure that R remains positive definite. A similar methodology to that presented here could also be used to look at B-matrix localization.

Using only one forecast ensemble limits our ability to understand how localization will impact strongly coupled assimilation at varying forecast lead times or in a cycling system. In particular, is possible that in a cycling system with balanced states we might see higher cross-fluid correlation values and less need for expanded ensemble sizes. This study also only included fields at a 1° resolution with this particular FV3-MOM6 coupled configuration. Further studies are needed to explore how localization may impact results with other models or at other forecast lead times and resolutions. For example, a cycling system could be used to test the benefit of SCDA in systems where one field is densely observed (e.g., sea surface temperature) and is used to constrain another well-correlated but poorly observed field like near surface atmospheric temperature. Finally, our chosen observing systems are not representative of realistic observing systems in that they do not address issues of representativity or observation density. Our results in a simplified system encourage us to further investigate the following questions in more complex configurations. For example,:

- How would these results change in a cycling experiment?
- How would these results change with B-matrix localization, or other types of localization methods?
- Do the results we found here generalize to other forecast lead times, state variables, resolutions, and models?
- How does observation density impact the performance of strongly coupled over weakly coupled data assimilation?

6. Conclusions

Here, we conducted experiments applying different localization strategies to a constructed case study, designed to represent a single analysis update to atmosphere and/or ocean temperature, from assimilation of surface level temperature observations. A central difficulty of strongly coupled atmosphere-ocean data assimilation is that the cross-fluid correlations are small on average and highly flow-dependent (Frolov et al., 2021). For example, in Figure 2 for our experiments, the atmosphere-ocean cross-component temperature correlations are near-zero over

most of the domain, with larger values occurring in localized regions associated with certain flow conditions. A successful strongly coupled data assimilation, in which states from one component are updated from observations of the other component, must then make use of the stronger correlations to make cross-component updates where appropriate, while avoiding making harmful updates in the bulk of the domain, where cross-component observations are not informative.

One consequence of the tendency for low cross-component correlations across much of the domain in coupled systems is that strongly coupled data assimilation is much more sensitive to sampling noise than within-fluid data assimilation. Hence, in our experiments, the strongly coupled data assimilation required much larger sample sizes, and were more sensitive to localization strategies than the within fluid updates. In particular, even with perfect localization, we required 128 ensemble members to see median percent improvement of at least 50% for cross-fluid updates over within-fluid updates. We saw additional improvements when we increased the ensemble size to 256, indicating that ensemble sizes much larger than the 80 members currently used in NOAA GEFS may be needed.

An additional complication revealed by this work, is that it is unlikely that simple parametric localization, as is standardly used for within-fluid data assimilation, is going to be sufficient for coupled data assimilation. For example, in Figure 7, applying the GC localization function using a single set of parameters (i.e., without optimizing it to each location) in the strongly coupled data assimilation updates resulted in degradations at up to 50% of the model locations with an 80 member ensemble. Instead, it is very likely that we will need to develop methods that are optimized according to the local coupled strength.

Consequently, in order to realize the potential improvements associated with strongly coupled data assimilation in an atmosphere-ocean model, we recommend focusing on developing better methods for addressing sampling noise. This could be achieved by increasing the ensemble size, with this work suggesting at least 250 may be necessary. Given the computational difficulties of increasing ensemble sizes through traditional approaches, it may be necessary to augment current ensembles using methods such as machine learning, or re-sampling methods. Additionally, computationally efficient optimal localization techniques are needed to enable applying localization schemes in Kalman filter methods that dynamically respond to local coupling strength. In all likelihood, advances on both of these fronts will be necessary before we can successfully apply strongly coupled data assimilation in full-scale ensemble systems.

Data Availability Statement

The code used for the experiments and analysis in this work is available here: https://doi.org/10.5281/zenodo.10966857. If this manuscript is accepted, all code and data will be transferred to Zenodo prior to publication.

Acknowledgments
The authors would like to thank Tobias Necker for many insightful conversations. This research was supported in part by the NOAA Physical Sciences Laboratory and NOAA cooperative agreements NA17OAR4320101 and NA22OAR4320151.

References

Adcroft, A., Anderson, W., Balaji, V., Blanton, C., Bushuk, M., Dufour, C. O., et al. (2019). The GFDL global ocean and sea ice model OM4.0: Model description and simulation features. *Journal of Advances in Modeling Earth Systems*, *11*(10), 3167–3211. https://doi.org/10.1029/2019MS001726

Anderes, E., Huser, R., Nychka, D., & Coram, M. (2013). Nonstationary positive definite tapering on the plane. *Journal of Computational & Graphical Statistics*, *22*(4), 848–865. https://doi.org/10.1080/10618600.2012.729982

Anderson, J., & Lei, L. (2013). Empirical localization of observation impact in ensemble Kalman filters. *Monthly Weather Review*, *141*(11), 4140–4153. https://doi.org/10.1175/MWR-D-12-00330.1

L. Bengtsson, M. Ghil, & E. Källén (Eds.) (1981). *Dynamic meteorology: Data assimilation methods* (Vol. 36). Springer. https://doi.org/10.1007/978-1-4612-5970-1

Bishop, C. H., & Hodyss, D. (2007). Flow-adaptive moderation of spurious ensemble correlations and its use in ensemble-based data assimilation. *Quarterly Journal of the Royal Meteorological Society*, *133*(629), 2029–2044. https://doi.org/10.1002/qj.169

de Rosnay, P., Browne, P., de Boisséson, E., Fairbairn, D., Hirahara, Y., Ochi, K., et al. (2022). Coupled data assimilation at ECMWF: Current status, challenges and future developments. *Quarterly Journal of the Royal Meteorological Society*, *148*(747), 2672–2702. https://doi.org/10.1002/qj.4330

Destouches, M., Montmerle, T., Michel, Y., & Ménétrier, B. (2021). Estimating optimal localization for sampled background-error covariances of hydrometeor variables. *Quarterly Journal of the Royal Meteorological Society*, *147*(734), 74–93. https://doi.org/10.1002/qj.3906

Flowerdew, J. (2015). Towards a theory of optimal localisation. *Tellus A: Dynamic Meteorology and Oceanography*, *67*(1), 25257. https://doi.org/10.3402/tellusa.v67.25257

Frolov, S., Bishop, C. H., Holt, T., Cummings, J., & Kuhl, D. (2016). Facilitating strongly coupled ocean–atmosphere data assimilation with an interface solver. *Monthly Weather Review*, *144*(1), 3–20. https://doi.org/10.1175/MWR-D-15-0041.1

Frolov, S., Reynolds, C. A., Alexander, M., Flatau, M., Barton, N. P., Hogan, P., & Rowley, C. (2021). Coupled ocean–atmosphere covariances in global ensemble simulations: Impact of an eddy-resolving ocean. *Monthly Weather Review*, *149*(5), 1193–1209. https://doi.org/10.1175/MWR-D-20-0352.1

Frolov, S., Rousseaux, C. S., Auligne, T., Dee, D., Gelaro, R., Heimbach, P., et al. (2023). Road map for the next decade of Earth system reanalysis in the United States. *Bulletin of the American Meteorological Society*, *104*(3), E706–E714. https://doi.org/10.1175/BAMS-D-23-0011.1

Gaspari, G., & Cohn, S. E. (1999). Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society*, *125*(554), 723–757. https://doi.org/10.1002/qj.49712555417

Gasperoni, N. A., Wang, X., & Wang, Y. (2022). Using a cost-effective approach to increase background ensemble member size within the GSI-based EnVar system for improved radar analyses and forecasts of convective systems. *Monthly Weather Review*, *150*(3), 667–689. https://doi.org/10.1175/MWR-D-21-0148.1

Goodliff, M., & Penny, S. G. (2022). Developing 4D-Var for strongly coupled data assimilation using a coupled atmosphere–ocean quasigeostrophic model. *Monthly Weather Review*, *150*(9), 2443–2458. https://doi.org/10.1175/MWR-D-21-0240.1

Hakim, G. J., Snyder, C., Penny, S. G., & Newman, M. (2022). Subseasonal forecast skill improvement from strongly coupled data assimilation with a linear inverse model. *Geophysical Research Letters*, *49*(11), e2022GL097996. https://doi.org/10.1029/2022GL097996

Han, G., Wu, X., Zhang, S., Liu, Z., & Li, W. (2013). Error covariance estimation for coupled data assimilation using a Lorenz atmosphere and a simple pycnocline ocean model. *Journal of Climate*, *26*(24), 10218–10231. https://doi.org/10.1175/JCLI-D-13-00236.1

Harris, L., Chen, X., Putman, W., Zhou, L., & Chen, J.-H. (2021). *A scientific description of the GFDL finite-volume cubed-sphere dynamical core*. (Technical Report). NOAA Geophysical Fluid Dynamics Laboratory (GFDL). Retrieved from https://library.noaa.gov/view/noaa/30725

Houtekamer, P. L., & Mitchell, H. L. (2001). A sequential ensemble Kalman filter for atmospheric data assimilation. *Monthly Weather Review*, *129*(1), 15–137. https://doi.org/10.1175/1520-0493(2001)129<0123:asekff>2.0.co;2

Hunke, E., Lipscomb, W., Jones, P., Turner, A., Jeffery, N., & Elliott, S. (2017). *CICE, the Los Alamos sea ice model*. (Technical Report No. CICE; 005315WKSTN00). Los Alamos National Laboratory (LANL). Retrieved from www.osti.gov/biblio/1364126

Hunt, B. R., Kostelich, E. J., & Szunyogh, I. (2007). Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica D: Nonlinear Phenomena*, *230*(1–2), 112–126. https://doi.org/10.1016/j.physd.2006.11.008

JCSDA Project: SOCA. (2023). JCSDA project: SOCA. Retrieved from www.jcsda.org/jcsda-project-soca

Kim, J., Teng, Y.-C., Paturi, S., Vernieres, G., Sluka, T., Zhu, J., et al. (2022). *The NOAA-NCEP 40 year reanalysis with the next generation global ocean data assimilation system (NG-GODAS): 1979 to 2019*. National Centers for Environmental Prediction (U.S.). https://doi.org/10.25923/26DS-Q363

Laloyaux, P., Frolov, S., Ménétrier, B., & Bonavita, M. (2018). Implicit and explicit cross-correlations in coupled data assimilation. *Quarterly Journal of the Royal Meteorological Society*, *144*(715), 1851–1863. https://doi.org/10.1002/qj.3373

Lei, L., & Whitaker, J. S. (2015). Model space localization is not always better than observation space localization for assimilation of satellite radiances. *Monthly Weather Review*, *143*(10), 3948–3955. https://doi.org/10.1175/MWR-D-14-00413.1

Liu, Z., Wu, S., Zhang, S., Liu, Y., & Rong, X. (2013). Ensemble data assimilation in a simple coupled climate model: The role of ocean-atmosphere interaction. *Advances in Atmospheric Sciences*, *30*(5), 1235–1248. https://doi.org/10.1007/s00376-013-2268-z

Lorenc, A. C. (1986). Analysis methods for numerical weather prediction. *Quarterly Journal of the Royal Meteorological Society*, *112*(474), 1177–1194. https://doi.org/10.1002/qj.49711247414

Lorenz, E. N. (1963). Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences*, *20*(2), 130–141. https://doi.org/10.1175/1520-0469(1963)020⟨0130:DNF⟩2.0.CO;2

Lu, F., Liu, Z., Zhang, S., & Liu, Y. (2015). Strongly coupled data assimilation using leading averaged coupled covariance (LACC). Part I: Simple model study. *Monthly Weather Review*, *143*(9), 3823–3837. https://doi.org/10.1175/MWR-D-14-00322.1

Ménétrier, B., Montmerle, T., Michel, Y., & Berre, L. (2015). Linear filtering of sample covariances for ensemble-based data assimilation. Part II: Application to a convective-scale NWP model. *Monthly Weather Review*, *143*(5), 1644–1664. https://doi.org/10.1175/MWR-D-14-00156.1

Morss, R. E., & Emanuel, K. A. (2002). Influence of added observations on analysis and forecast errors: Results from idealized systems. *Quarterly Journal of the Royal Meteorological Society*, *128*(579), 285–321. https://doi.org/10.1256/00359000260498897

Necker, T., Hinger, D., Griewank, P. J., Miyoshi, T., & Weissmann, M. (2023). Guidance on how to improve vertical covariance localization based on a 1000-member ensemble. *Nonlinear Processes in Geophysics*, *30*(1), 13–29. https://doi.org/10.5194/npg-30-13-2023

Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. *The Computer Journal*, *7*(4), 308–313. https://doi.org/10.1093/comjnl/7.4.308

Nerger, L., Janjić, T., Schröter, J., & Hiller, W. (2012). A regulated localization scheme for ensemble-based Kalman filters. *Quarterly Journal of the Royal Meteorological Society*, *138*(664), 802–812. https://doi.org/10.1002/qj.945

NOAA Science Advisory Board. (2021). A report on priorities for weather research. NOAA science advisory board report. Retrieved from https://sab.noaa.gov/wp-content/uploads/2021/12/PWR-Report_Final_12-9-21.pdf

Penny, S. G., Bach, E., Bhargava, K., Chang, C.-C., Da, C., Sun, L., & Yoshida, T. (2019). Strongly coupled data assimilation in multiscale media: Experiments using a quasi-geostrophic coupled model. *Journal of Advances in Modeling Earth Systems*, *11*(6), 1803–1829. https://doi.org/10.1029/2019MS001652

Penny, S. G., & Hamill, T. M. (2017). Coupled data assimilation for integrated Earth system analysis and prediction. *Bulletin of the American Meteorological Society*, *98*(7), ES169–ES172. https://doi.org/10.1175/bams-d-17-0036.1

Press, W., Teukolsky, S., Vetterling, W., & Flannery, B. (1992). *Numerical recipes in C*. Cambridge University Press.

Sakov, P., & Bertino, L. (2011). Relation between two common localisation methods for the EnKF. *Computational Geosciences*, *15*(2), 225–237. https://doi.org/10.1007/s10596-010-9202-6

Sluka, T. C., Penny, S. G., Kalnay, E., & Miyoshi, T. (2016). Assimilating atmospheric observations into the ocean using strongly coupled ensemble data assimilation. *Geophysical Research Letters*, *43*(2), 752–759. https://doi.org/10.1002/2015GL067238

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

Smith, P. J., Lawless, A. S., & Nichols, N. K. (2017). Estimating forecast error covariances for strongly coupled atmosphere–ocean 4D-Var data assimilation. *Monthly Weather Review*, *145*(10), 4011–4035. https://doi.org/10.1175/MWR-D-16-0284.1

Smith, P. J., Lawless, A. S., & Nichols, N. K. (2018). Treating sample covariances for use in strongly coupled atmosphere-ocean data assimilation. *Geophysical Research Letters*, *45*(1), 445–454. https://doi.org/10.1002/2017GL075534

Stanley, Z., Grooms, I., & Kleiber, W. (2021). Multivariate localization functions for strongly coupled data assimilation in the bivariate Lorenz 96 system. *Nonlinear Processes in Geophysics*, *28*(4), 565–583. https://doi.org/10.5194/npg-28-565-2021

Tang, Q., Mu, L., Goessling, H. F., Semmler, T., & Nerger, L. (2021). Strongly coupled data assimilation of ocean observations into an ocean-atmosphere model. *Geophysical Research Letters*, *48*(24), e2021GL094941. https://doi.org/10.1029/2021GL094941

Tibshirani, B. E. R. J. (1994). An introduction to the bootstrap. https://doi.org/10.1201/9780429246593

Unified Forecast System. (2023). Unified forecast system. Retrieved from https://ufscommunity.org/

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., et al. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, *17*(3), 261–272. https://doi.org/10.1038/s41592-019-0686-2

Yoshida, T., & Kalnay, E. (2018). Correlation-cutoff method for covariance localization in strongly coupled data assimilation. *Monthly Weather Review*, *146*(9), 2881–2889. https://doi.org/10.1175/MWR-D-17-0365.1

Zhou, X., Zhu, Y., Hou, D., Fu, B., Li, W., Guan, H., et al. (2022). The development of the NCEP global ensemble forecast system version 12. *Weather and Forecasting*, *37*(6), 1069–1084. https://doi.org/10.1175/WAF-D-21-0112.1