

# Lagrangian Data Assimilation of Surface Drifters in a Double-Gyre Ocean Model Using the Local Ensemble Transform Kalman Filter

LUYU SUN

*Department of Atmospheric and Oceanic Science, University of Maryland, College Park, College Park, Maryland*

STEPHEN G. PENNY

*Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder, and  
Physical Sciences Division, NOAA/Earth System Research Laboratory, Boulder, Colorado*

(Manuscript received 28 November 2018, in final form 27 July 2019)

## ABSTRACT

The assimilation of position data from Lagrangian observing platforms is underdeveloped in operational applications because of two main challenges: 1) nonlinear growth of model and observation error in the Lagrangian trajectories, and 2) the high dimensionality of realistic models. In this study, we propose a localized Lagrangian data assimilation (LaDA) method that is based on the local ensemble transform Kalman filter (LETKF). The algorithm is tested with an “identical twin” approach in observing system simulation experiments (OSSEs) using a simple double-gyre configuration of the Geophysical Fluid Dynamics Laboratory (GFDL) Modular Ocean Model. Results from the OSSEs show that with a proper choice of localization radius, the LaDA can outperform conventional assimilation of surface in situ temperature and salinity measurements. The improvements are seen not only in the surface state estimate, but also throughout the ocean column to 1000 m depth. The impacts of localization radius and model error in estimating accuracy of both fluid and drifter states are further investigated.

## 1. Introduction

While the Argo float program has been leveraged very successfully in operational ocean data assimilation applications, surface drifters have been underutilized. The trajectories of drifters transported by ocean currents provide information about the underlying dynamics and physical properties of the ocean. Poje et al. 2014 pointed out that Lagrangian experiments, aiming to track large numbers of flow-following instruments, are the most feasible means of simultaneously measuring the submesoscale of ocean surface structure. The Global Drifter Program (GDP) sponsored by the National Oceanic and Atmospheric Administration (NOAA) is a part in situ, part remote sensing program that provides the capacity to produce a rough estimate of near-surface currents by tracking a large number of surface drifters deployed at 15-m depth throughout the global ocean. Ocean data assimilation systems have typically used temperature and salinity measurements from surface drifters indirectly via SST and SSS retrieval products (Bitterman and Hansen

1993; Melnichenko et al. 2016), though they are sometimes assimilated directly (Dong et al. 2017). The position data of surface drifters are rarely used in any way, though there have been some pioneering investigations in reconstructing the mesoscale Eulerian velocity fields based on observations of the Lagrangian positions (Taillandier et al. 2006; Nilsson et al. 2012).

A traditional goal of data assimilation is to estimate the state of a dynamical system by combining observations with a numerical model approximation of that system. Lagrangian data assimilation (LaDA) focuses on the assimilation of the position information provided by Lagrangian instruments such as surface drifters. There are two main approaches to assimilate Lagrangian measurements. One approach applies a transformation from geophysical track locations to an estimated Eulerian velocity and assimilates this proxy measurement as a representation of the instantaneous flow velocity of the system

*Publisher's Note:* This article was revised on 20 November 2019 to correct a typographical error in the first part of Eq. (2.2) that occurred when originally published.

*Corresponding author:* Luyu Sun, lysun@umd.edu

DOI: 10.1175/MWR-D-18-0406.1

© 2019 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](https://www.ametsoc.org/PUBSReuseLicenses) ([www.ametsoc.org/PUBSReuseLicenses](https://www.ametsoc.org/PUBSReuseLicenses)).

(Molcard et al. 2003; Ozgokmen et al. 2003; Taillandier et al. 2006; Nilsson et al. 2012). Transforming position measurements to Eulerian velocity additionally requires determining a method to match the timing of the proxy observations with the modeled velocity, such as the degree of time averaging that should be applied. Assimilating such velocity observations has been attempted and implemented in forecasting systems such as the U.S. Navy's Coastal Ocean Model (NCOM) (Jacobs et al. 2014; Carrier et al. 2014, 2016) and the Mediterranean ocean Forecasting System (MFS) (Nilsson et al. 2012), both showing promising improvements in characterizing the ocean currents.

An alternative approach (Ide et al. 2002; Kuznetsov et al. 2003) extends the original fluid model state to include the drifter positions in an augmented state vector  $\mathbf{x} = (\mathbf{x}_F, \mathbf{x}_D)^T$  and evolves this extended dynamical system. Given a sufficiently short analysis cycle, this approach is able to update the model states with high accuracy by utilizing the quasi-linear error growth in the simulated drifter position and the correlations of errors between the ocean tracer and velocity fields. This Lagrangian approach may be more effective than the assimilation of transformed Eulerian velocity proxy observations in strongly nonlinear dynamic models (Kuznetsov et al. 2003; Vernieres et al. 2011). In this study, we investigate the Lagrangian approach.

Two issues that must be addressed in applying LaDA and transitioning to operational ocean data assimilation are: 1) nonlinear error growth in the modeled drifter trajectories, and 2) the ability to scale to high-dimensional systems. A thorough examination of issues regarding systematic model errors will be reserved for future study. Apte et al. (2008) and Apte and Jones (2013) showed that LaDA can be applied with the ensemble Kalman filter (EnKF) (Evensen 2004; Ide et al. 2002) with high-dimensional model states, though it fails to capture dynamics with the high-order nonlinearity because of the resulting non-Gaussian distribution of error from the nonlinear trajectory model. An alternative assimilation method that could be used in the presence of this nonlinearity is the particle filter (PF) (Salman et al. 2008; Santitissadeekorn et al. 2014). However, the canonical PF does not scale well to high dimensions (Snyder et al. 2008). While new types of PF have recently been proposed for high-dimensional geophysical systems (van Leeuwen 2003; Penny and Miyoshi 2016; Poterjoy 2016), none have yet been shown viable for operational applications. A hybrid method (Slivinski et al. 2015) has been proposed to address the issues due to nonlinearity and high dimensionality at the same time by updating the flow states using an EnKF while updating the drifter states using a PF. The algorithm of Slivinski et al. (2015) was examined using a framework of perfect twin experiments with the linear shallow-water equations. Results indicated that compared to the EnKF,

this hybrid approach produced improved estimates of the Bayesian posterior and better tracked the true state.

The augmented-state LaDA methods described above have so far only been applied to simplified models (e.g., the linear shallow-water equations), and so far there has been no attempt to transition the work for viable use in operational prediction. In this study, we propose a new augmented-state LaDA as an extension of an ocean data assimilation system (Penny et al. 2015) that is currently being transitioned to operations at NOAA's National Centers for Environmental Prediction. We further facilitate the transition to an operational environment by using the Geophysical Fluid Dynamics Laboratory (GFDL) Modular Ocean Model (MOM4p1) so that we can update not only the surface fluid velocity fields, but also tracer fields such as temperature and salinity spanning from the surface down to the thermocline and below.

An observing system simulation experiment (OSSE) is conducted to evaluate the performance of the proposed LaDA method. OSSEs are widely used in validating the impact of new observation (Atlas et al. 1985) and DA system in both atmosphere and ocean prior to deployment (Aksoy et al. 2012; Halliwell et al. 2014; Penny et al. 2015). An OSSE experiment basically consists of 1) a nature run of the numerical model representing the true atmosphere or ocean, 2) a DA system that includes DA solution method and a numerical forecast model, and 3) observations sampled from the nature run with realistic errors. An "identical twin" OSSE refers to the use of identical model configurations for both of the nature run and forecast model. As the work done by (Kuznetsov et al. 2003; Slivinski et al. 2015), we use this "identical twin" experiment as a preliminary verification tool to examine the performance of the proposed LaDA method. The perfect twin OSSE is a necessary validation step for any new DA method, which must ultimately be followed by experiments assimilating real observation data using an operational model configuration. Intermediate steps may include relaxing the perfect twin assumption by degrading the accuracy of the forecast model.

In section 2, we describe the general form of the augmented-state LaDA and review the original LETKF algorithm. The localized LaDA algorithm using LETKF with an augmented state vector is introduced at the end of this section. Section 3 details the design of the numerical experiments and definitions of different metrics to evaluate the performance. Section 4 presents results of applying LaDA to a simple configuration of MOM4p1, and demonstrates the capability of the localized LaDA algorithm to produce an accurate analysis of not only simulated trajectories of drifters, but also all prognostic ocean model variables such as temperature ( $T$ ), salinity ( $S$ ), and flow velocities ( $U$ ,  $V$ ). Section 5 provides discussion and future directions.

## 2. Review of augmented-state method and LETKF

### a. Augmented-state method

The augmented-state approach (Ide et al. 2002) extracts information carried by Lagrangian tracers by extending the model states as a combination of fluid and drifter states:

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_F \\ \mathbf{x}_D \end{bmatrix},$$

where  $\mathbf{x}_D$  contains data corresponding to the drifters simulated by the ocean model, such as: longitude, latitude, depth, as well as temperature and salinity when these are observed by the drifters. In this study, we mainly consider the case of drifter data consisting of position measurements of longitude and latitude, while assuming the drifters maintain a constant depth. Thus if there are  $N_D$  drifters, then  $\mathbf{x}_D$  is a  $2N_D$ -component vector and  $\mathbf{x}_F$  is of dimension  $N_F$ , the degrees of freedom of the fluid model.

The drifter advection equation is added to the original fluid dynamical system as

$$\begin{cases} \frac{d\mathbf{x}_F^f}{dt} = M_F(\mathbf{x}_F^f, t) \\ \frac{d\mathbf{x}_D^f}{dt} = M_D(\mathbf{x}_F^f, \mathbf{x}_D^f, t) \end{cases}.$$

As a clarifying example, suppose that only the position data carried by Lagrangian drifters are observed. Then the dimension of the observation space is  $2N_D$  and the observations can be represented in terms of the observation operator  $\mathbf{H}$  and true states  $\mathbf{x}^t$  as

$$\mathbf{y}^o = [\mathbf{0} \ \mathbf{I}] \begin{bmatrix} \mathbf{x}_F^t \\ \mathbf{x}_D^t \end{bmatrix} + \varepsilon = \mathbf{H}\mathbf{x}^t + \varepsilon,$$

where

$$\mathbf{H} = [\mathbf{0} \ \mathbf{I}] \quad \text{and} \quad \varepsilon \sim N(0, \mathbf{R}).$$

The matrix  $\mathbf{I}$  in the above formula is an identity matrix of dimension  $2N_D \times 2N_D$ . The quantity  $\varepsilon$  is a Gaussian random error, with the  $2N_D \times 2N_D$  observation error covariance matrix  $\mathbf{R}$ . The forecast time length must be chosen sufficiently small to maintain approximately Gaussian error statistics. The augmented-state LaDA provides an estimator of  $\mathbf{x}_{i+1} = (\mathbf{x}_F, \mathbf{x}_D)^T$  at time  $t_{i+1}$  given the observations  $\mathbf{y}_{i+1}^o$ . With the Gaussian assumption, the Kalman filter-based methods attempt to provide the best linear unbiased estimator by taking advantage of the estimated observation error covariance matrix  $\mathbf{R}$  and the estimated background error covariance matrix  $\mathbf{P}$ . Our background error covariance

matrix is composed of the combined fluid and drifter states defined as

$$\mathbf{P} = \begin{bmatrix} \mathbf{P}_{FF} & \mathbf{P}_{FD} \\ \mathbf{P}_{FD}^T & \mathbf{P}_{DD} \end{bmatrix},$$

where  $\mathbf{P}_{FF}$ ,  $\mathbf{P}_{FD}$ , and  $\mathbf{P}_{DD}$  denote the background error covariance matrices of the fluid state vector, the cross covariance between the fluid and drifters state vectors, and the drifters state vector, respectively. The detail definition of the components of these background error covariance matrices will be specified in the later section. The ensemble Kalman filter (EnKF) updates the prior error covariance matrix by computing the sample error covariance matrix from ensemble perturbations around the ensemble forecast mean.

### b. LETKF

The LETKF is an ensemble square root filter (EnSRF) proposed by Hunt et al. (2007) as an extension of works by Bishop et al. (2001), Hamill et al. (2001) and Houtekamer and Mitchell (1998), using the localization approach of Ott et al. (2004). There are generally two kinds of localization approach: in observation space (R-localization), and in model space (B-localization) (Greybush et al. 2011). The LETKF uses R-localization, which selects and weighs local observations in a prescribed region around each grid point while excluding observations outside this region. Salman et al. 2006 shows that a proper selection of the localization region is beneficial in using EnKF to assimilate the drifter positions within the shallow-water system. To preserve vertically consistent dynamics in each ocean column, no localization is applied in the vertical (Penny et al. 2015). As a consequence, surface observations impact the analysis of the entire water column. Because previous studies have found superior results using this approach to vertical localization when applying LETKF in the ocean (Penny et al. 2015; Sluka et al. 2016), we only consider experiments applying variations in the horizontal localization radius.

After the localized region is determined, we compute an analysis update to the center grid point at all depths. In general, the EnKFs assume Gaussian error statistics, which are estimated from the perturbations of the ensemble forecast around the ensemble forecast mean state. The analysis solution is thus confined to a maximum  $(K - 1)$ -dimensional linear space defined by the ensemble states, where  $K$  is the number of ensemble members. By applying the localization technique, we allow the global analysis to be formed from a larger dimensional space, though the localized solution is still formed within a linear space limited by the ensemble size. The localization makes it possible to approximate the solution of spatially extended high-dimensional nonlinear problems

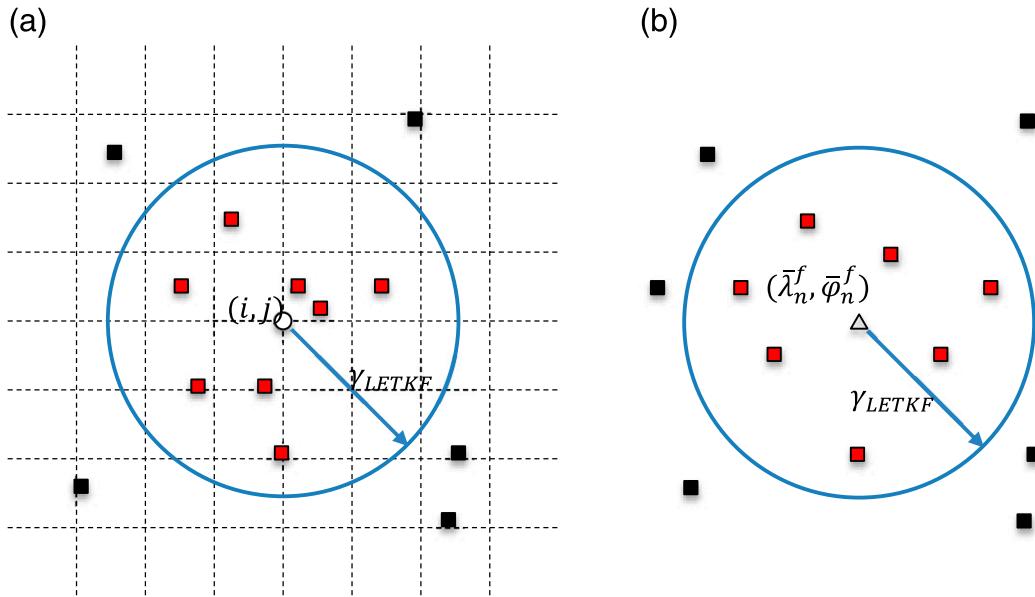


FIG. 1. (a) The localization region defined to update the fluid states  $\bar{\mathbf{x}}_{F[l]}^f$  at grid point  $(i, j)$ , the white circle at the center of the figure. The background dashed lines represent the mesh grids of the forecast model. (b) The localization region defined to update the drifter states  $\bar{\mathbf{x}}_{D[l]}^f$  associated to a forecasting drifter ID  $n$ . The center white triangle is located at the forecast ensemble mean position of drifter number  $n$ . In both figures,  $\gamma_{\text{LETKF}}$  denotes the radius of the localization region and the squares represents the observation drifter locations. Those observation drifters inside the circles (red squares) are marked by their IDs as the localized observation.

with better accuracy. At the analysis step, instead of minimizing the original cost function of EnKF:

$$J(\mathbf{x}) = (\mathbf{x} - \bar{\mathbf{x}}^f)^T (\mathbf{P}^f)^{-1} (\mathbf{x} - \bar{\mathbf{x}}^f) + (\mathbf{y}^o - \mathbf{H}\mathbf{x})^T \mathbf{R}^{-1} (\mathbf{y}^o - \mathbf{H}\mathbf{x}),$$

with respect to the model state vector  $\mathbf{x}$ , LETKF minimizes the equivalent cost function:

$$\tilde{J}(\mathbf{w}) = (k-1)\mathbf{w}^T \mathbf{w} + (\mathbf{y}^o - \bar{\mathbf{y}}^f - \mathbf{Y}^f \mathbf{w})^T \mathbf{R}^{-1} (\mathbf{y}^o - \bar{\mathbf{y}}^f - \mathbf{Y}^f \mathbf{w}), \quad (2.1)$$

with respect to  $\mathbf{w}$  within each localization region, where  $\mathbf{w}$  is defined as the ensemble weight vector using the formula  $\mathbf{x} = \bar{\mathbf{x}}^f + \mathbf{X}^f \mathbf{w}$ . The linear transform applied as part of the LETKF is an efficient implementing technique to update the ensemble deterministically by post-multiplying the forecast ensemble with a linear transform matrix (Bishop et al. 2001; Katzfuss et al. 2016). The LETKF algorithm is easily parallelized by calculating the analysis independently at each grid point, thus facilitates the technical transition to more realistic applications.

### c. LETKF for augmented-state LaDA

Combining sections 2a and 2b, we denote the LETKF method applied to solve the Lagrangian problem as

LETKF-LaDA. To simplify the description, we first consider the case when drifter locations are observed and the dimension of observation space is  $2N_D$ . A more general case considering the surface temperature and salinity measurement on drifters is elaborated at the end of this section.

- 1) Run the dynamical model to obtain the global ensemble forecast states  $\mathbf{x}_{[g]}^{f(k)} = (\mathbf{x}_{F[g]}^{f(k)}, \mathbf{x}_{D[g]}^{f(k)})^T$  ( $k = 1, 2, \dots, K$ ), then obtain the corresponding global ensemble mean  $\bar{\mathbf{x}}_{[g]}^f = (\bar{\mathbf{x}}_{F[g]}^f, \bar{\mathbf{x}}_{D[g]}^f)^T$  and the forecast error perturbation matrix  $\mathbf{X}_{[g]}^f$ , whose  $k$ th column is  $\mathbf{x}_{[g]}^{f(k)} - \bar{\mathbf{x}}_{[g]}^f$ . The subscript “[g]” indicates the global state vector.
- 2) Apply the specialized observation operator  $\mathbf{H}$  defined in section 2a to the augmented model states in order to form the ensemble  $\mathbf{y}_{[g]}^{f(k)}$  of the forecast observation vectors by  $\mathbf{y}_{[g]}^{f(k)} = \mathbf{H}\mathbf{x}_{[g]}^{f(k)}$ . Compute the corresponding mean  $\bar{\mathbf{y}}_{[g]}^f$ , and error perturbation matrix  $\mathbf{Y}_{[g]}^f$ , defined in observation space. In this case, we assume all observation data are carried by the drifters so that  $\mathbf{y}_{[g]}^{f(k)} = \mathbf{x}_{D[g]}^{f(k)}$ , the mean vector  $\bar{\mathbf{y}}_{[g]}^f = \bar{\mathbf{x}}_{D[g]}^f$  and  $2N_D \times K$  matrix  $\mathbf{Y}_{[g]}^f = \mathbf{X}_{D[g]}^f$ . For our experiments, the global observation is denoted as  $\mathbf{y}_{[g]}^o$  and the observation error covariance matrix  $\mathbf{R}_{[g]}$  is defined as a diagonal matrix with diagonal entries specified as the observation variance  $\sigma^2$ , which will be discussed in section 3.

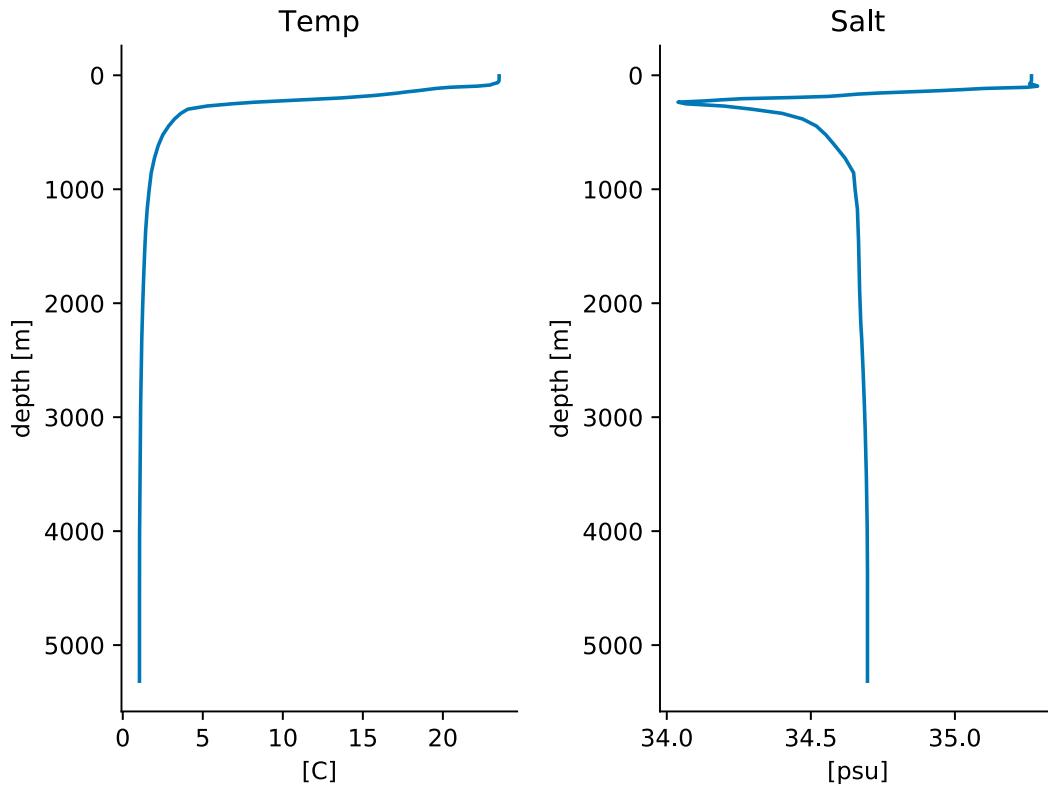


FIG. 2. Temperature and salinity initial conditions provided by SODA on 2 Jan 1981 at location 25°N, 175°E varying from 5 to 5316 m. The thermocline is located below 75 m and above 235 m.

3) Determine the local analysis of the LETKF-LaDA system, using the local arrays of the global terms:  $\bar{\mathbf{x}}_{[l]}^f$ ,  $\mathbf{X}_{[l]}^f$ ,  $\mathbf{Y}_{[l]}^f$ ,  $\mathbf{R}_{[l]}$ ,  $\mathbf{y}_{[l]}^o$  and  $\bar{\mathbf{y}}_{[l]}^f$ . The selection approach of the localization region depends on the type of the analysis variables to be updated. For the fluid variables, we apply the localization region as a cylinder centered at a horizontal grid point  $(i, j)$  with horizontal localization radius  $\gamma_{\text{LETKF}}$  (see Fig. 1a), which will be specified in section 3. For global model state variables  $\bar{\mathbf{x}}_{[g]}^f$  and  $\mathbf{X}_{[g]}^f$ , rows associated to the fluid variables at this grid point  $(i, j)$  from all the depth levels are

chosen to formulate their corresponding local variables  $\bar{\mathbf{x}}_{F[l]}^f$  and  $\mathbf{X}_{F[l]}^f$ . For example, suppose the prognostic flow states are gridded velocities  $u_F$  and  $v_F$ , then we have

$$\bar{\mathbf{x}}_{F[l]}^f = \begin{bmatrix} \bar{u}_F^f(i, j, 1) \\ \bar{v}_F^f(i, j, 1) \\ \vdots \\ \bar{u}_F^f(i, j, h) \\ \bar{v}_F^f(i, j, h) \end{bmatrix}_{2h \times 1}$$

$$\mathbf{X}_{F[l]}^f = \begin{bmatrix} u_F^{f(1)}(i, j, 1) - \bar{u}_F^f(i, j, 1) & \dots & u_F^{f(K)}(i, j, 1) - \bar{u}_F^f(i, j, 1) \\ v_F^{f(1)}(i, j, 1) - \bar{v}_F^f(i, j, 1) & \dots & v_F^{f(K)}(i, j, 1) - \bar{v}_F^f(i, j, 1) \\ \vdots & & \vdots \\ u_F^{f(1)}(i, j, h) - \bar{u}_F^f(i, j, h) & \dots & u_F^{f(K)}(i, j, h) - \bar{u}_F^f(i, j, h) \\ v_F^{f(1)}(i, j, h) - \bar{v}_F^f(i, j, h) & \dots & v_F^{f(K)}(i, j, h) - \bar{v}_F^f(i, j, h) \end{bmatrix}_{2h \times K}, \quad (2.2)$$

where  $h$  is the bottom level of the model. In Fig. 1a, all the observed drifters located in this localized region

are marked by their IDs and we define the number of localized observed drifters as  $N_{D[l]}$ . We then choose

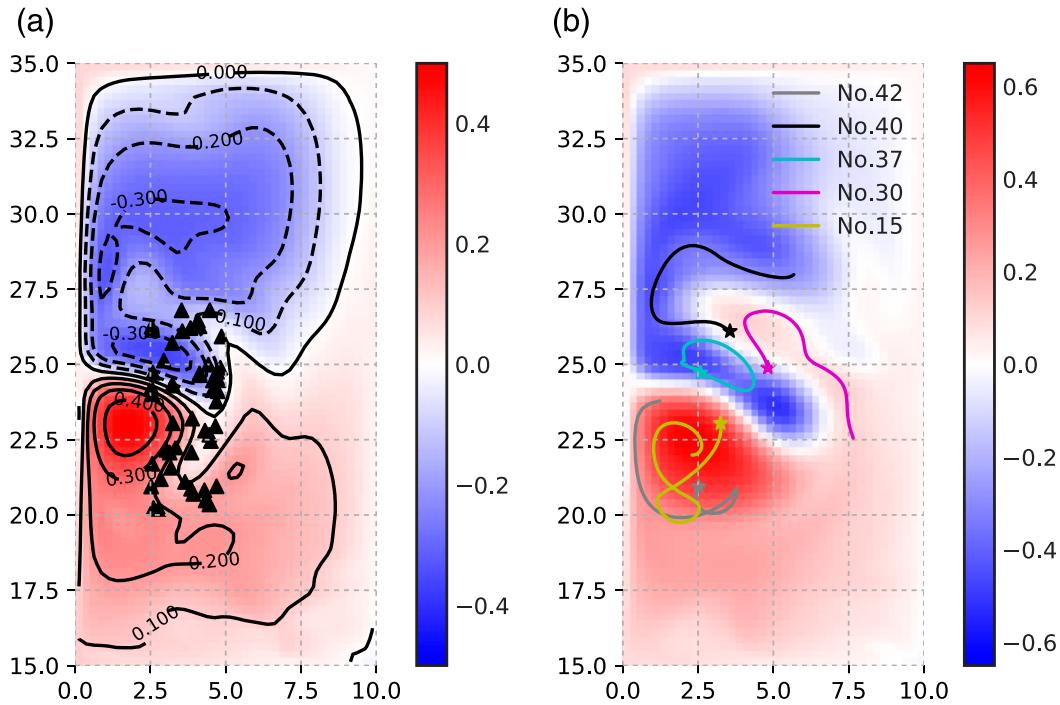


FIG. 3. (a) The sea surface height (SSH) of the nature run at the end of the 6-month spinup procedure (contour interval is 0.1 m); (b) sampled drifter trajectories in 91 days are shown with SSH (m) at the terminal time. Initial drifter locations are marked as triangles in (a) and asterisks in (b).

the rows of  $\mathbf{Y}_{[g]}^f, \mathbf{y}_{[g]}^o$  and  $\bar{\mathbf{y}}_{[g]}^f$  related to these marked drifter IDs to create  $2N_{D[l]} \times K$  matrix  $\mathbf{Y}_{[l]}^f$ , and  $2N_{D[l]} \times 1$  vectors  $\mathbf{y}_{[l]}^o$  and  $\bar{\mathbf{y}}_{[l]}^f$ . Similarly, rows and columns are chosen to formulate  $2N_{D[l]} \times 2N_{D[l]}$  diagonal matrix  $\mathbf{R}_{[l]}$ . Figure 1b illustrates the approach to define the localization region in order to update the local state variables of the simulated drifters. As shown in Fig. 1b, the selection of the localization region is associated with each drifter ID rather than each model grid point. For each simulated drifter ID  $n$ , its forecast ensemble mean location  $(\bar{\lambda}_n^f, \bar{\varphi}_n^f)$  is defined as the center of the corresponding localization region. Localized model state vector  $\bar{\mathbf{x}}_{D[l]}^f$  and error perturbation matrix  $\mathbf{X}_{D[l]}^f$  include all the entries associated to this ID  $n$ :

$$\bar{\mathbf{x}}_{D[l]}^f = \begin{bmatrix} \bar{\lambda}_{D,n}^f \\ \bar{\varphi}_{D,n}^f \end{bmatrix} \quad \text{and}$$

$$\mathbf{X}_{D[l]}^f = \begin{bmatrix} \bar{\lambda}_{D,n}^{f(1)} - \bar{\lambda}_{D,n}^f & \cdots & \bar{\lambda}_{D,n}^{f(K)} - \bar{\lambda}_{D,n}^f \\ \bar{\varphi}_{D,n}^{f(1)} - \bar{\varphi}_{D,n}^f & \cdots & \bar{\varphi}_{D,n}^{f(K)} - \bar{\varphi}_{D,n}^f \end{bmatrix}. \quad (2.3)$$

Similarly to the approach used to update the fluid state,  $\bar{\mathbf{y}}_{[l]}^f, \mathbf{Y}_{[l]}^f, \mathbf{y}_{[l]}^o$ , and  $\mathbf{R}_{[l]}$  can be defined accordingly based on the localization region. Among all the marked IDs, the observation associated to its own drifter ID  $n$  is included. To simplify the notation, the subscriptions

“ $F[l]$ ” in (2.2) and “ $D[l]$ ” in (2.3) of the model variables (i.e.,  $\bar{\mathbf{x}}^f$  and  $\mathbf{X}^f$ ) are all replaced by “[ $l$ ]” in the next step. 4) The remaining steps follow the original LETKF (Hunt et al. 2007), repeated here for completeness. Form the local analysis error covariance for the ensemble weight vector  $\mathbf{w}$  in (2.1) and the corresponding weight matrix as

$$\tilde{\mathbf{P}}_{[l]}^a = [(K - 1)\rho + (\mathbf{Y}_{[l]}^f)^T \mathbf{R}_{[l]}^{-1} \mathbf{Y}_{[l]}^f]^{-1} \quad \text{and}$$

$$\mathbf{W}_{[l]}^a = [(k - 1)\tilde{\mathbf{P}}_{[l]}^a]^{1/2},$$

where  $\rho$  is the covariance inflation factor. The ensemble weight vector used to compute the local mean analysis is

$$\bar{\mathbf{w}}_{[l]}^a = \tilde{\mathbf{P}}_{[l]}^a (\mathbf{Y}_{[l]}^f)^T \mathbf{R}_{[l]}^{-1} (\mathbf{y}_{[l]}^o - \bar{\mathbf{y}}_{[l]}^f), \quad (2.4)$$

where  $\bar{\mathbf{w}}^a$  is the minimizer of the cost function  $\tilde{J}$  in (2.1). This mean is added to each column of  $\mathbf{W}_{[l]}^a$  to form a  $K$ -by- $K$  weighting matrix with column vectors  $\{\mathbf{w}_{[l]}^{a(k)}\}$ . Each local analysis ensemble member is computed as

$$\mathbf{x}_{[l]}^{a(k)} = \bar{\mathbf{x}}_{[l]}^f + \mathbf{X}_{[l]}^f \mathbf{w}_{[l]}^{a(k)}.$$

The local analysis ensemble mean is determined as

$$\bar{\mathbf{x}}_{[l]}^a = \bar{\mathbf{x}}_{[l]}^f + \mathbf{X}_{[l]}^f \bar{\mathbf{w}}_{[l]}^a.$$

Specifically, the analysis mean associated to a flow state variable is determined by the corresponding mean weight  $\bar{\mathbf{w}}_{[l]}^a$ . The covariances between the given flow state variable and the drifter state variables are included in  $\mathbf{P}_{[l]}^a$  within the definition of  $\bar{\mathbf{w}}_{[l]}^a$  in (2.4).

- 5) Collect all the local analysis states at each grid point and each drifter ID to form the global analysis state vector:

$$\mathbf{x}_{[g]}^{a(k)} = \begin{bmatrix} \mathbf{x}_{F[g]}^{a(k)} \\ \mathbf{x}_{D[g]}^{a(k)} \end{bmatrix}.$$

This algorithm can be extended to a more general case in which additional measurements of the fluid variables are associated with each drifter. For example, if additional temperature and salinity measurements are made for each drifter ID  $n$ ,  $\bar{\mathbf{x}}_{D[l]}^f$  and  $\mathbf{X}_{D[l]}^f$  (2.3) in step 3 can be extended as

$$\bar{\mathbf{x}}_{D[l]}^f = \begin{bmatrix} \bar{\lambda}_{D,n}^f \\ \bar{\varphi}_{D,n}^f \\ \bar{T}_{D,n}^f \\ \bar{S}_{D,n}^f \end{bmatrix} \quad \text{and}$$

$$\mathbf{X}_{D[l]}^f = \begin{bmatrix} \lambda_{D,n}^{f(1)} - \bar{\lambda}_{D,n}^f & \dots & \lambda_{D,n}^{f(K)} - \bar{\lambda}_{D,n}^f \\ \varphi_{D,n}^{f(1)} - \bar{\varphi}_{D,n}^f & \dots & \varphi_{D,n}^{f(K)} - \bar{\varphi}_{D,n}^f \\ T_{D,n}^{f(1)} - \bar{T}_{D,n}^f & \dots & T_{D,n}^{f(K)} - \bar{T}_{D,n}^f \\ S_{D,n}^{f(1)} - \bar{S}_{D,n}^f & \dots & S_{D,n}^{f(K)} - \bar{S}_{D,n}^f \end{bmatrix}.$$

The remaining steps in the above algorithm are the same.

The above further extension of the system allows a direct differencing of the observed drifter measurements with the modeled drifter measurements (i.e.,  $\mathbf{y}^o - \mathbf{H}\mathbf{x}^f = \mathbf{y}^o - \mathbf{x}_D^f$ ). We assume the interpolation operator is automatically embedded in the drifter dynamic model  $M_D$  rather than explicitly relying on an observation operator  $\mathbf{H}$  to map the gridded temperature and salinity states to the drifter positions at the appropriate time.

### 3. Experimental setup

#### a. The numerical ocean model and spinup procedure

We use the “identical twin” approach in OSSEs to evaluate the impact of LETKF-LaDA. The nature run and the forecast model in all experiments use the GFDL subtropical double-gyre configuration of the B-grid

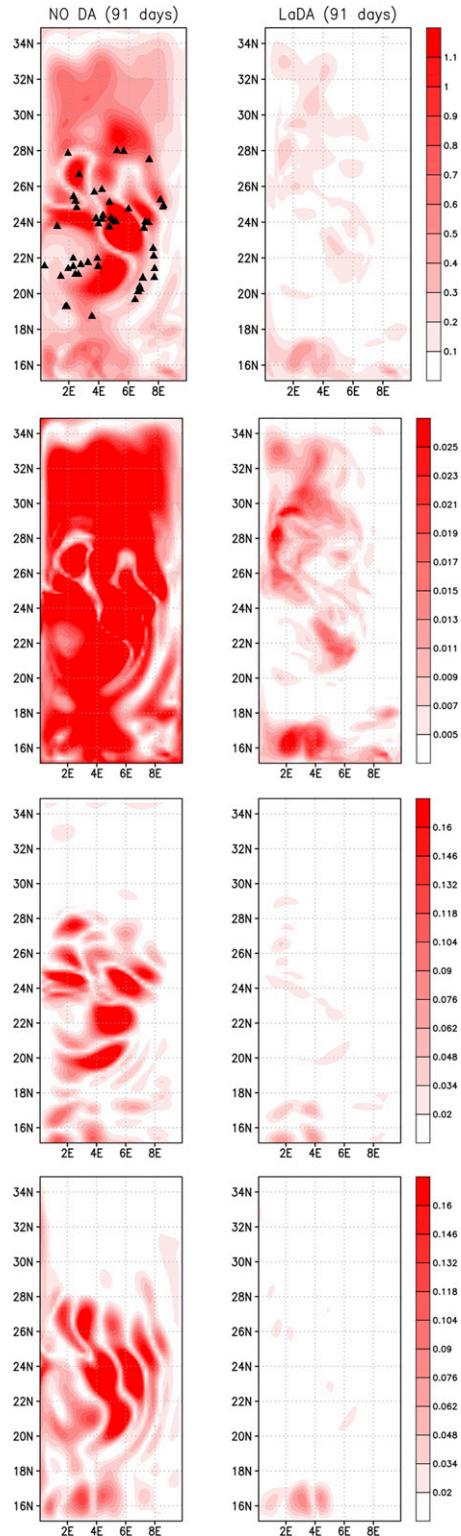


FIG. 4. Absolute error comparisons between the control run (i.e.,  $|\mathbf{x}^c - \mathbf{x}^f|$ ) and the LETKF-LaDA (i.e.,  $|\mathbf{x}^a - \mathbf{x}^f|$ ) in ocean flow fields temperature ( $T$ ), salinity ( $S$ ), and velocities ( $U$ ,  $V$ ) at end of 91 days. The errors are shown in the longitude–latitude plane and averaged from the top layer 5 m to the 235 m. The true drifter locations at the end of 91 days are marked by closed triangles in the first contour figure of the left column.

hydrostatic non-Boussinesq ocean model MOM4p1 (Griffies 2008). We work in a rectangular, closed basin on a beta plane with longitude ranging from  $0^\circ$  to  $10^\circ\text{E}$  and latitude ranging from  $15^\circ$  to  $35^\circ\text{N}$ . Ocean circulation is driven by the zonal wind stress defined as  $F_\lambda(\varphi) = 0.1 \sin[\pi(\varphi - 20^\circ\text{N})/10^\circ\text{N}] \text{ N m}^{-2}$ , where  $\varphi$  is the latitude. Fifty constant  $z$ -level coordinates are used extending down to 5500 m.

The nature run is generated using a  $1/4^\circ$  horizontal resolution with integration time step  $\Delta t = 1800 \text{ s}$ . We specify the initial conditions of the temperature and salinity fields by replicating a point profile of temperature and salinity state estimates from the Simple Ocean Data Assimilation (SODA), version 3 (Carton et al. 2018a,b), on 2 January 1981 (Fig. 2) and then initializing the model with this horizontally uniform stratification. The true drifter locations are simulated in the nature run, randomly initialized in the energetic region  $20^\circ\text{--}27^\circ\text{N}$ ,  $2.5^\circ\text{--}5^\circ\text{E}$ , as shown in Fig. 3a. The deployment depth, 15 m, is the same as the official fixed depth assigned for the NOAA GDP surface drifters.

We construct an ensemble of wind stress fields by adding a constant to the zonal component of the true wind field with the magnitude, which is randomly chosen by  $\theta \sim N(0, 0.1) \text{ N m}^{-2}$ . This addition of the constant is applied uniformly for all grid points and time steps [i.e.,  $F_\lambda^{(k)}(\varphi) = F_\lambda(\varphi) + \theta^{(k)}$  ( $k = 1, 2, \dots, K$ )]. To initialize the ensemble, we first spin up each ensemble ocean member for six months with the perturbed atmospheric conditions  $F_\lambda^{(k)}$ . All experiments use an ensemble size  $K = 40$ . We next populate 50 drifters ( $N_D = 50$ ) simulated at 15 m depth, initialized at the ‘true’ positions identically for each ensemble member. The ensemble members are then integrated for another 16 days to generate the initial ensemble spread of drifter positions. We take advantage of an embedded drifter module in MOM4p1 that provides output of drifter positions as well as temperature and salinity at the drifter locations. Within this module, the drifter positions are calculated by integrating the corresponding velocities obtained after applying a bilinear interpolation to the gridded ocean velocity fields at the corresponding depth. The temperature and salinity data are determined directly by bilinear interpolation of the gridded temperature and salinity fields to the corresponding drifter locations.

The LETKF-LaDA system uses a daily analysis cycle following Jacobs et al. (2014). The optimal choice of analysis cycle window depends on the error doubling time of the modeled drifter positions, and remains an open question. The observations are assimilated daily with positioning errors drawn from a normal distribution using a prescribed standard deviation of  $\sigma = 0.1^\circ$  in both longitude and latitude. The parameter  $\sigma$  is determined by averaging the daily error growth of all the 50 drifter locations

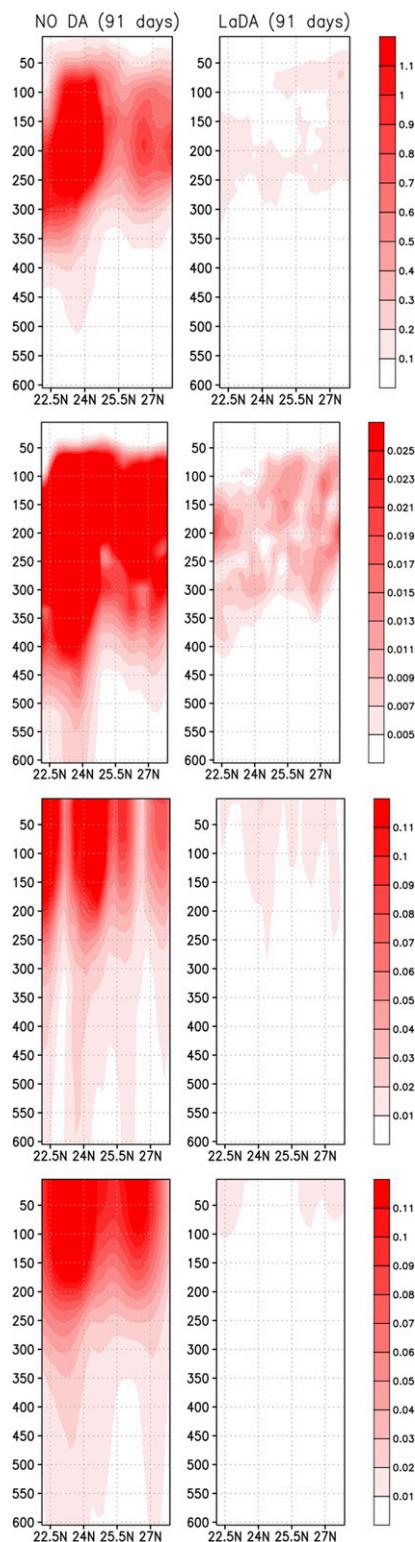


FIG. 5. Absolute error comparisons between the control run (i.e.,  $|\mathbf{x}^c - \mathbf{x}'|$ ) and the LETKF-LaDA (i.e.,  $|\mathbf{x}^d - \mathbf{x}'|$ ) in ocean flow fields temperature ( $T$ ), salinity ( $S$ ), and velocities ( $U$ ,  $V$ ). The errors are shown in the latitude–depth plane and averaged along longitude from  $0.625^\circ$  to  $7.125^\circ\text{E}$ . The  $x$  axis stands for the latitude interval  $22.125^\circ\text{--}27.875^\circ\text{N}$ , while the  $y$  axis represents the depth level from 5 to 600 m.

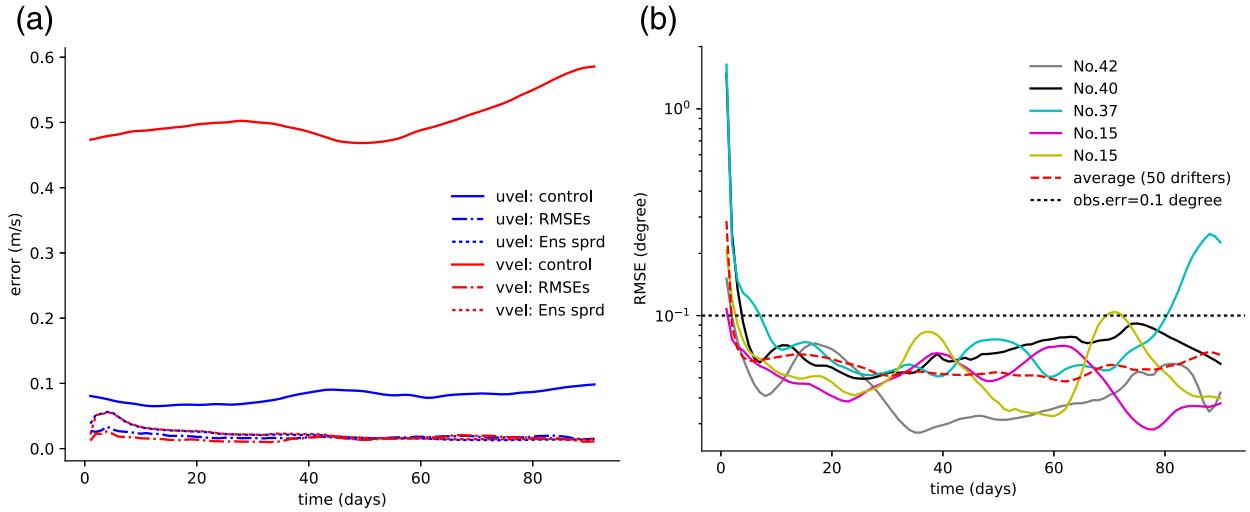


FIG. 6. (a) Time variation of analysis RMSE and ensemble spread in velocities at confluence region (23.625°–26.625°N, 0.125°–2.875°E) averaged through all the depth levels; (b) time variation of analysis RMSE (in degree) in drifter displacements.

deployed in 40-member flow ensemble generated by the first step of spinup process as previously described. We investigate the influence of the horizontal localization radius and model resolution on the performance of the LETKF-LaDA. We define the horizontal localization radius,  $\gamma_{\text{LETKF}}$ , as a multiple of the baroclinic Rossby radius of deformation (RRD) at the respective latitudes (Chelton et al. 1998). The choice of  $\gamma_{\text{LETKF}}$  in different experiments will be specified in the next section. None of the experiments use multiplicative inflation (i.e.,  $\rho = 1$ ). The control experiment is initialized using the ensemble mean ocean and drifter states at the end of the two-step spinup process.

*b. Error metrics*

The locations of drifters in this work are based on longitude–latitude coordinates on a spheroid surface instead of the Cartesian coordinates on a flat surface as used in previous augmented-state LaDA studies (Salman et al. 2006). In this section, we utilize the measurements of

geophysical distance to evaluate the difference between two locations of interest. The errors in drifter states are computed by taking the average distance between ensemble mean positions and the true positions of all the drifters:

$$\text{RMSE}_D = \frac{1}{N_D} \sum_{n=1}^{N_D} d[(\lambda_{D,n}^t, \varphi_{D,n}^t), (\bar{\lambda}_{D,n}, \bar{\varphi}_{D,n})],$$

where  $d[(\cdot, \cdot), (\cdot, \cdot)]$  is the function to compute the geophysical distance between two locations based on their latitude and longitude coordinates. Salman et al. (2008) defined a dimensionless fluid field norm in the horizontal direction that evaluates the error in terms of a percentage of the true fluid state. The error corresponding to the velocity field is formulated as a type of kinetic energy field. Because the double-gyre model used for our experiments has the additional complexity of including multiple vertical layers, we sum the horizontal errors from the top depth level  $h_o$  to level  $h$ :

$$|T| = \left[ \frac{\sum_{m=h_o}^h \sum_{i,j} [T_F(i,j,m) - T_F^t(i,j,m)]^2}{\sum_{m=h_o}^h \sum_{i,j} [T_F^t(i,j,m)]^2} \right]^{1/2}, \quad |S| = \left[ \frac{\sum_{m=h_o}^h \sum_{i,j} [S_F(i,j,m) - S_F^t(i,j,m)]^2}{\sum_{m=h_o}^h \sum_{i,j} [S_F^t(i,j,m)]^2} \right]^{1/2}$$

$$|\text{KE}| = \left[ \frac{\sum_{m=h_o}^h \sum_{i,j} [u_F(i,j,m) - u_F^t(i,j,m)]^2 + [v_F - v_F^t(i,j,m)]^2}{\sum_{m=h_o}^h \sum_{i,j} [u_F^t(i,j,m)]^2 + [v_F^t(i,j,m)]^2} \right]^{1/2}, \tag{3.1}$$

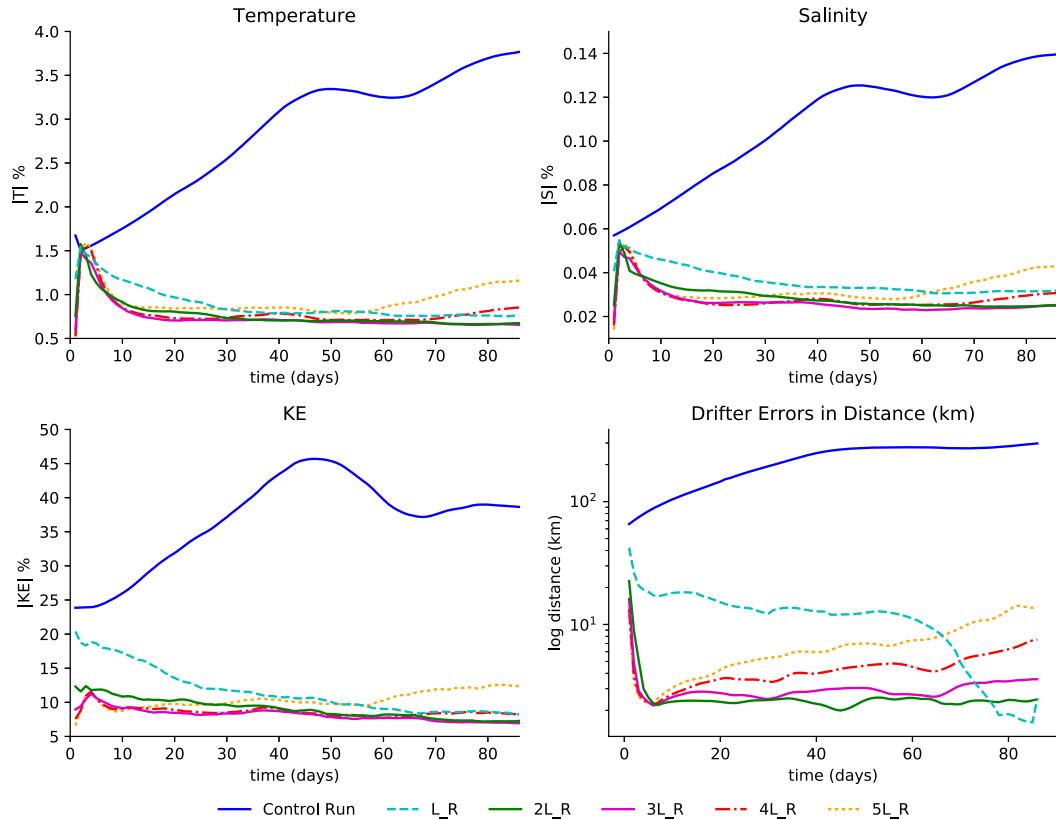


FIG. 7. Error norms of temperature, salinity, kinetic energy, and drifter states ( $RMSE_D$ ) in the given period  $[0, 91]$  using formula (3.1). The error norms of the ocean flow fields are aggregated from 5 to 1000 m depth (i.e.,  $h_o = 5$  m and  $h = 1000$  m). In each subplot, we show the experiment results of control run (blue solid line),  $5L_R$  (orange dotted line),  $4L_R$  (red dash-dotted line),  $3L_R$  (purple solid line),  $2L_R$  (green solid line), and  $L_R$  (cyan dashed line).

where  $T_F(i, j, m)$ ,  $S_F(i, j, m)$ ,  $u_F(i, j, m)$  and  $v_F(i, j, m)$  indicate the flow states at the grid point  $(i, j, m)$ , in order to verify the convergence of the proposed LETKF-LaDA approach. We use the  $RMSE_D$  defined above to quantify the norm of the drifter states.

Due to the relationship between geostrophic currents and sea surface height (SSH), we are also interested in the forecast error correlation between the drifter states and the surrounding flow fields. For a specific drifter with ID  $n$ , we denote its true position as  $(\lambda_{D,n}^t, \varphi_{D,n}^t)$  and the drifter state corresponding to this ID as  $\mathbf{x}_{D,n}^t$ . The fluid state at the horizontal grid point  $(i, j)$  with depth  $h$  is specified as the vector  $\mathbf{x}_F^f(i, j, h)$ . The error correlation between the drifter state and its closest vertical ocean field is denoted as

$$r_n(i, j, h) = \text{Corr}[\mathbf{e}_F^f(i, j, h), \mathbf{e}_{D,n}^f] \\ = \frac{E\{\mathbf{x}_F^f(i, j, h) - \mathbf{x}_F^t\}[\mathbf{x}_{D,n}^f - \mathbf{x}_{D,n}^t]^\top}}{\sqrt{E\{\mathbf{x}_F^f(i, j, h) - \mathbf{x}_F^t\}^2}} \sqrt{E\{[\mathbf{x}_{D,n}^f - \mathbf{x}_{D,n}^t]^2\}}, \quad (3.2)$$

where  $E(\bullet)$  denotes the mean. Similarly, the error correlation corresponding to SSH is

$$r_{SSH,n}(i, j, h) \\ = \text{Corr}[\mathbf{e}_F^f(i, j, h), \mathbf{e}_{SSH}^f] \\ = \frac{E\{\mathbf{x}_F^f(i, j, h) - \mathbf{x}_F^t\}(\mathbf{x}_{SSH}^f - \mathbf{x}_{SSH}^t)}}{\sqrt{E\{\mathbf{x}_F^f(i, j, h) - \mathbf{x}_F^t\}^2}} \sqrt{E\{(\mathbf{x}_{SSH}^f - \mathbf{x}_{SSH}^t)^2\}}, \quad (3.3)$$

where  $\mathbf{x}_{SSH}$  is selected at the closest gridpoint position to the  $n$ th drifter location.

#### 4. Results

We first set the localization radius of LETKF-LaDA as  $\gamma_{\text{LETKF}} = 3L_R$ , where  $L_R$  stands for the approximate RRD at the corresponding latitude, and implement the LETKF-LaDA for 91 days (about 3 months). Within this time period, some of the true drifters complete one circuit around the gyre (see Fig. 3b). Figure 4 depicts contours of absolute error averaged from the top layer

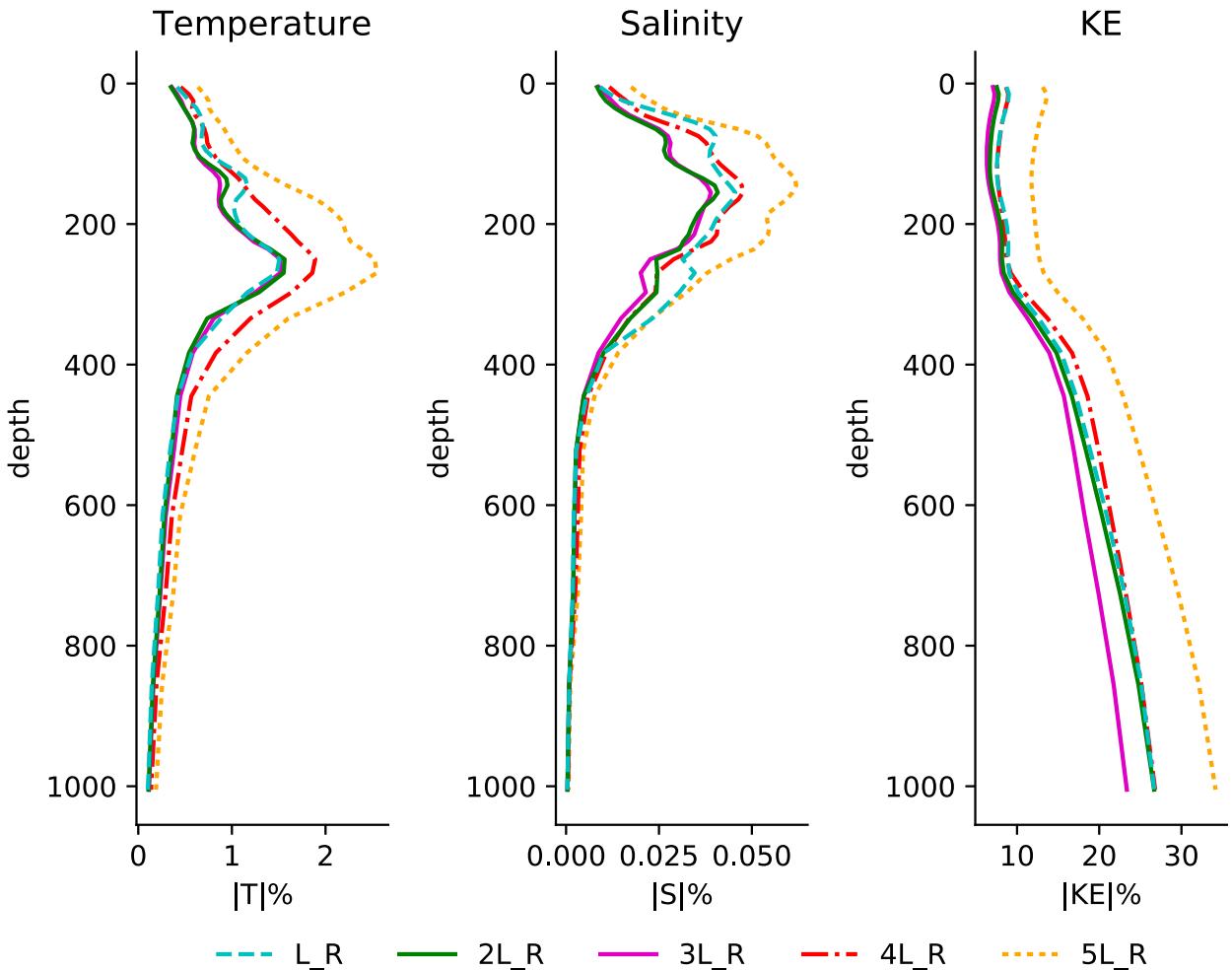


FIG. 8. Error in temperature, salinity, and kinetic energy at the end of 91st DA cycle along the vertical direction with experiments of localization radius  $5L_R$ ,  $4L_R$ ,  $3L_R$ ,  $2L_R$ , and  $L_R$ . These quantities are evaluated by formula (3.1) at each individual vertical level with  $h_o = h$ . The control run is not shown in these figures because its errors largely exceeded the scale of the given results in all fields.

5 m to the 235 m. Since the initial drifter positions are all close to the gyre, a large reduction in error occurs within the western-central region for all prognostic model variables. Figure 5 shows that the error is reduced not only in the surface fields but also at deeper levels. The relatively small error reduction in the northern and southern boundary regions is due to sparse observation coverage (Fig. 4). Figure 6a highlights the time variation of the analysis RMSEs and ensemble spreads in velocities at confluence region ( $23.625^\circ$ – $26.625^\circ$ N,  $0.125^\circ$ – $2.875^\circ$ E). The exponential decays in both of the RMSEs and ensemble spread of flow velocities at the confluence region are able to guarantee the ensemble drifters at this region distributed around the truth with small ensemble spreads and therefore provide accurate analysis states. Salman et al. (2008) shows that different drifter deployments can affect the convergence of the error in drifter displacements. In Fig. 6b, we display the time

variation of analysis RMSE for sampled drifters labeled in Fig. 3b. The RMSE of most of the drifters are below the prescribed observation error standard deviation (i.e.,  $0.1^\circ$ ), except for the drifter number 37 circuiting near the confluence. It is observed that this drifter has a sudden increase in the RMSE as it is about to complete one loop and then the error gradually decays in the end.

In general, the performance of the LETKF-LaDA depends on the localization radius, which affects both the extent of influence by observations and the reduction of nonlinear error in the prognostic model states. In the following discussions, we first determine an effective localization radius for the given system and then apply this radius for the remainder of the experiments.

a. Impacts of varying localization radius  $\gamma_{LETKF}$

Using the  $1/4^\circ$  horizontal resolution, we vary the localization radius:  $\gamma_{LETKF} = 5L_R, 4L_R, 3L_R, 2L_R$ , and  $L_R$ .

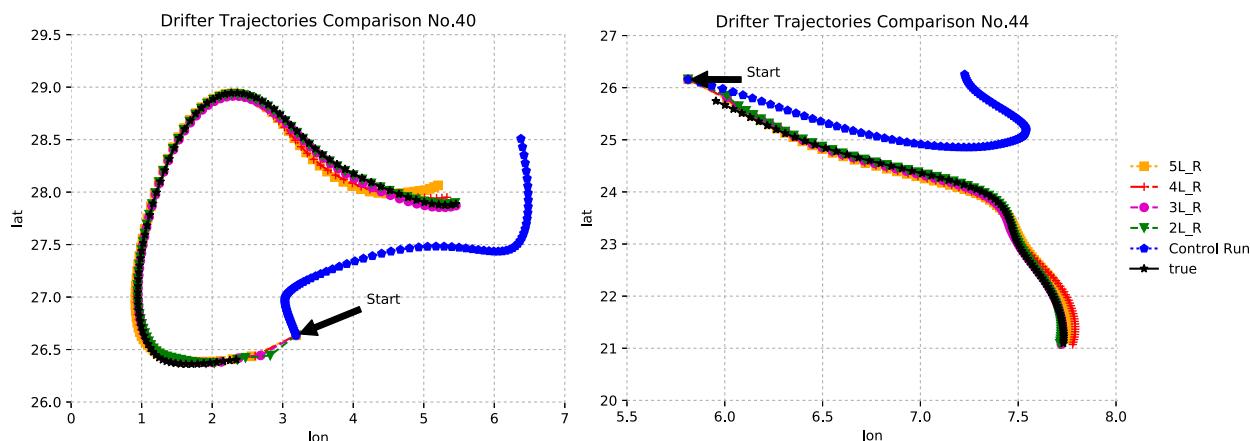


FIG. 9. Drifter trajectories of nature run (black) compared to the control run (blue) and LETKF-LaDA varying the localization radius:  $5L_R$ ,  $4L_R$ ,  $3L_R$ , and  $2L_R$  in 91 days. All cases use identical drifter starting positions. The green line ( $2L_R$ ) is the closest trajectory to the black line (obs) at the end for both of the drifters.

Each experiment applies the LETKF-LaDA over 91 analysis cycles. We compare the error variations of the cycled analysis states to the control run. Figure 7 shows the results of percentage error norm defined in Eq. (3.1). We compare the errors of the analysis mean with the control run. For the temperature, salinity, and kinetic energy fields, we sum the squared errors in percentage norm from 5 to 1000 m depth. Cases using  $3L_R$  and  $2L_R$  produce the lowest errors over the course of the experiments. The  $2L_R$  localization radius produces slightly lower errors than  $3L_R$  radius in controlling the mean drifter error distance below 3 km within the  $1/4^\circ$  meshgrid map. The experiment with  $5L_R$  radius generates the most accurate analysis in model prognostic fields at end of the first cycle, however it begins diverging after 60 cycles due to continuous error growth in drifter states beginning after 5 cycles. Using the smallest attempted localization radius  $L_R$ , the error decays relatively slowly in both the prognostic fields and estimated drifter positions in the first 60 cycles. The errors in prognostic fields reach a low value comparing with other radii at the end of this time period. With localization radius  $L_R$ , the LETKF-LaDA produces in a sudden reduction in error between 60 and 80 days, however this error reduction does not appear stable. The sudden “shock” in error after the first cycle of LETKF-LaDA with the various localization radii may be the consequence of an instability generated by a large change in the state variables.

To monitor the performance of the LETKF-LaDA in the vertical using varying localization radii, we consider the analysis states at the end of the 91st analysis cycle and compute the error in each horizontal layer from 5 to 1000 m depth (see Fig. 8). In the temperature field, all errors have a similar shape in the vertical direction with a maximum at approximately 270 m depth. Both of the  $3L_R$

and  $2L_R$  localization radii produce smaller errors at all depth levels than all the rest of the localization radii. The experiment with localization radius  $L_R$  behaves similarly to those with radii  $3L_R$  and  $2L_R$  below the thermocline, though a marginally larger error is produced in the top layer (5–200 m). For salinity, experiments using  $3L_R$  and  $2L_R$  radii have the lowest error above 270 m depth, while errors in all experiments decay to almost the same value at deeper levels. All LETKF-LaDA cases have difficulty in reducing error in kinetic energy below 300 m depth, though the case using a radius of  $3L_R$  produces the smallest error in all depth levels.

The LETKF-LaDA requires the growth in drifter position error to grow quasi-linearly in time for the duration of the analysis cycle. The degree of nonlinearity depends on the stability of the dynamics, being greatest in the gyre region, and for the purpose of LaDA depends upon the time between observations. To evaluate the performance of LETKF-LaDA in updating the state in the presence of varying degrees of nonlinearity, we examine trajectories for two different drifter IDs (Fig. 9). We select one drifter in the gyre (number 44) and another in a region with relatively slow approximately linear flow (number 40). For both the linear and nonlinear trajectories, the LETKF-LaDA is able to track the true drifter trajectories.

In summary, with the quantitative and qualitative comparison in all the states and drifter trajectories, we find that the LETKF-LaDA with localization radii  $3L_R$  and  $2L_R$  result in the lowest errors in the state estimate.

#### b. Comparing to conventional assimilation of in situ temperature and salinity

Drifters may be equipped not only with a GPS locator but also sensors providing in situ measurements

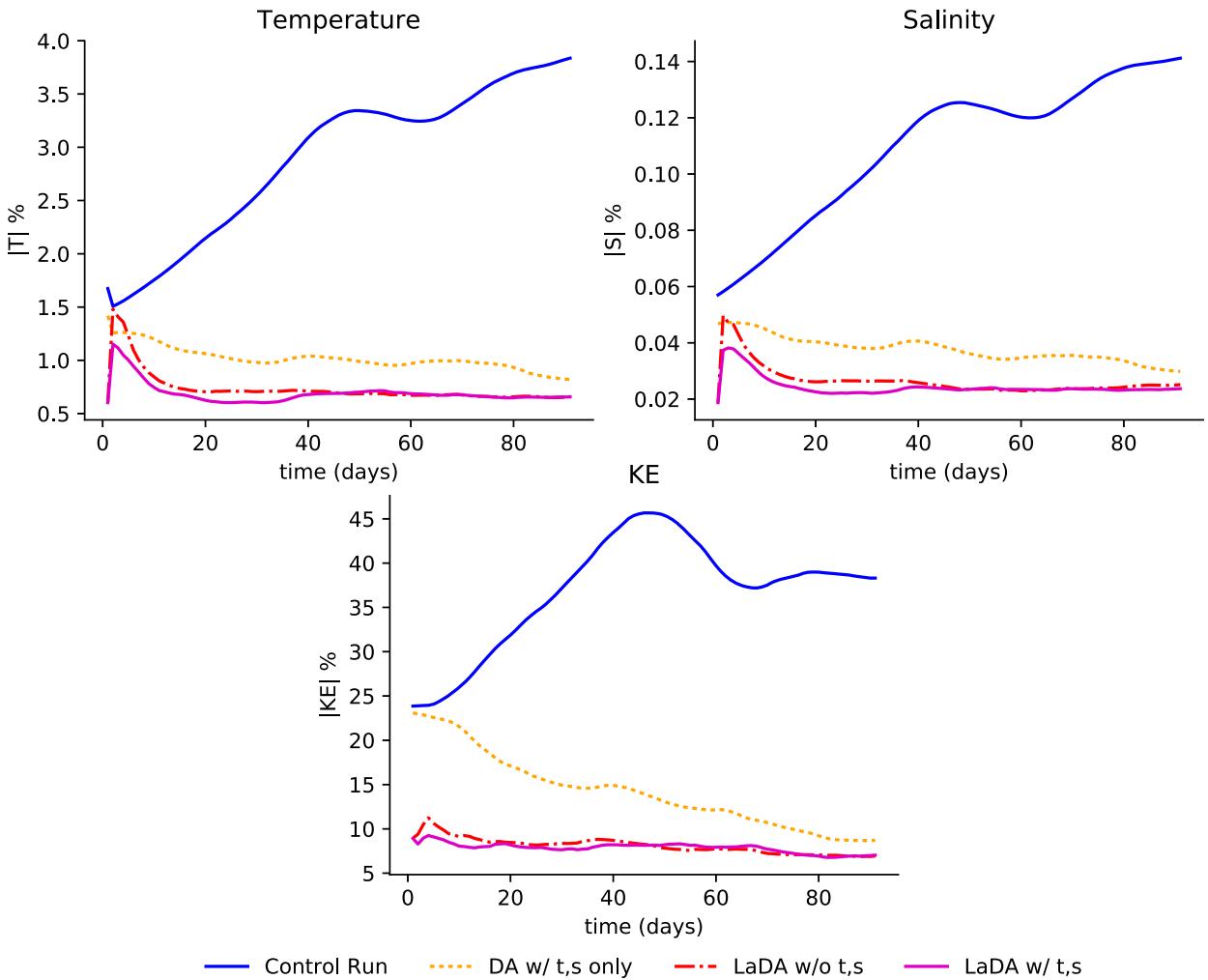


FIG. 10. Error norms of temperature, salinity, and kinetic energy in the given period  $[0, 91]$  using formula (3.1). The error norms of the ocean flow fields are aggregated from 5 to 1000 m depth (i.e., with  $h_o = 5$  m and  $h = 1000$  m). In each subplot, we show the experiment results of control run (blue solid line), assimilation of surface drifter measured  $T$  and  $S$  (orange dotted line), LETKF-LaDA assimilating only drifter positions (red dashdotted line), and LETKF-LaDA assimilating both drifter positions and surface drifter measured  $T$  and  $S$  (purple solid line).

(Lumpkin and Pazos 2007). In this section, we add temperature and salinity observations to the simulated Lagrangian drifters and compare the performance of three observing strategies: 1) using LETKF to assimilate only the in situ temperature and salinity observations at the surface (no position information), 2) using LETKF-LaDA to assimilate only drifter positions, and 3) using LETKF-LaDA to assimilate both drifter locations and in situ temperature and salinity observations with observation errors  $\sigma_T = 0.1^\circ\text{C}$  and  $\sigma_S = 0.1$  psu. All experiments use the localization radius  $\gamma_{\text{LETKF}} = 3L_R$ , as determined in the previous section.

Figure 10 summarizes the time variation of error in the estimated ocean state. The LETKF-LaDA

experiments provide more accurate estimates of the prognostic model state variables than that of the conventional T/S DA after passing through the “shock” period in all the flow fields. The addition of the in situ data to some extent reduces the “shock” in temperature and salinity error for LETKF-LaDA in the first few cycles. In the vertical comparison at the terminal time (see Fig. 11), the LETKF-LaDA outperforms assimilation of conventional observations at all depths using the metric given by formula (3.1). Assimilating the combination of in situ and Lagrangian position data further improves the accuracy of the LETKF-LaDA salinity estimates in all layers and temperature in most of the layers except for the levels between 100 and 200 m, though this results

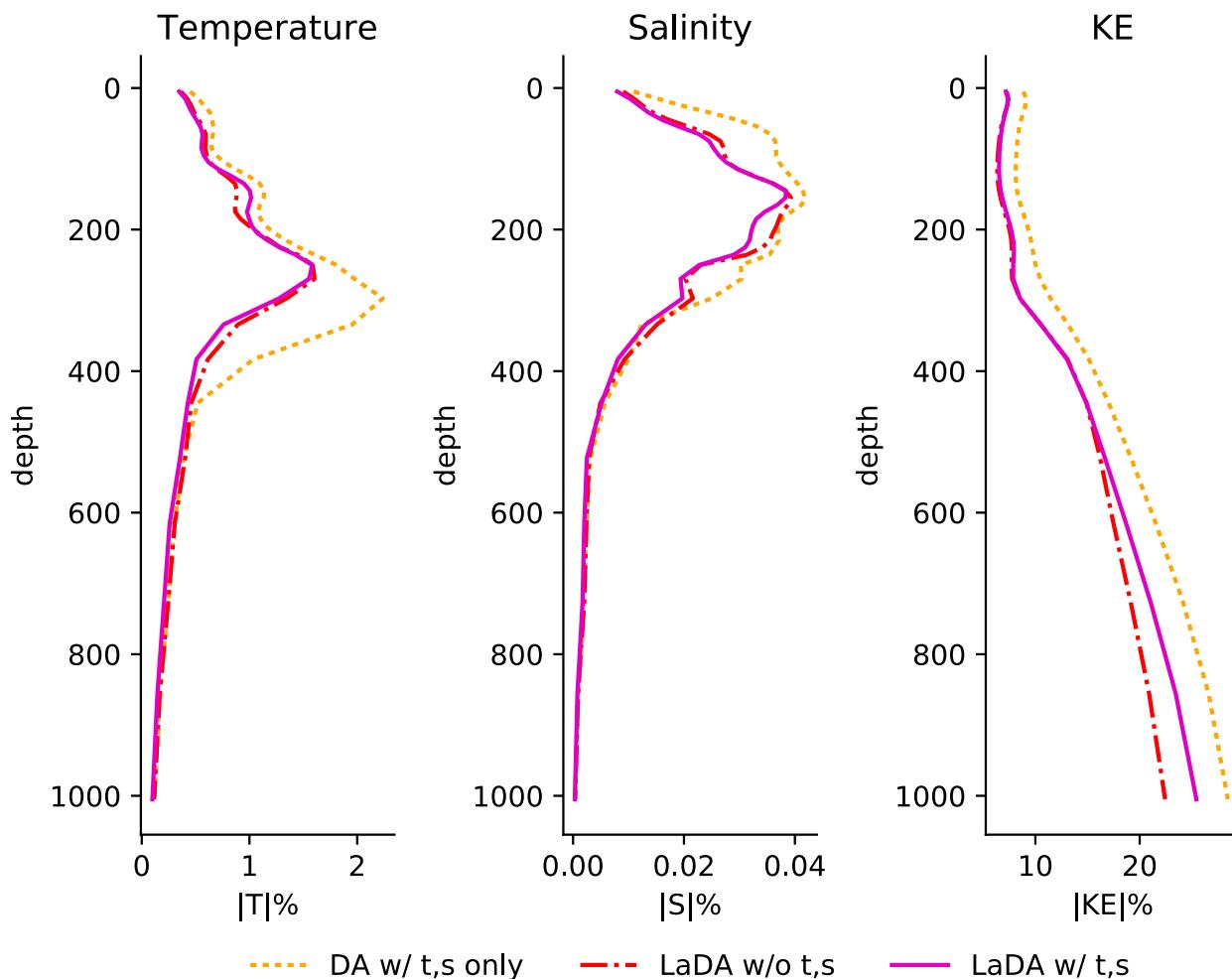


FIG. 11. Vertical profile of change in error for temperature, salinity and kinetic energy at the end of 91st DA cycle with experiments of traditional DA assimilating  $T$  and  $S$ , LETKF-LaDA with and without assimilation of  $T$  and  $S$ . These quantities are evaluated by formula (3.1) at each individual vertical level with  $h_o = h$ . The control run is not shown in these figures because its errors largely exceed the scale of the given results in all fields.

in a smaller improvement of the estimations in kinetic energy below 450 m.

### c. Model resolution

One factor that could have an impact on the convergence of the LETKF-LaDA is errors in drifter trajectories caused by the model grid resolution. The insufficient modeling flow caused by a coarse grid resolution can possibly result in a relatively large discrepancy between the forecast drifter and the truth. This is a particular concern for transitioning to the assimilation of real observational data. To examine sensitivity to resolution, model forecasts are obtained using configurations of  $1/3^\circ$  and  $1/2^\circ$  horizontal grid resolutions. The initial coarser ensemble fields are generated by applying bilinear interpolation to the previous initial ensemble fields defined in  $1/4^\circ$  gridding

system as in section 3a. Figure 12 shows a comparison of the SSH fields in one-day forecast generated by models of different resolutions before we start applying LETKF-LaDA. Forecast models of coarser grids fail to represent the scale of SSH at the center of the gyres and create larger errors at eastern edge of the gyre. LETKF-LaDA is performed with localization radius  $3L_R$ .

We observe in Fig. 13 that assimilating Lagrangian data is possible with a forecast model resolution that resolves the dynamics present in the observations (e.g., in this case with a perfect model). Experiments using both coarser resolutions fail to stabilize the error growth in ocean and drifter states due to the exponential growth in the drifter position errors.

The above results show that the accuracy of the forecasting model influences the performance of LETKF-LaDA. The relation between the convergence of

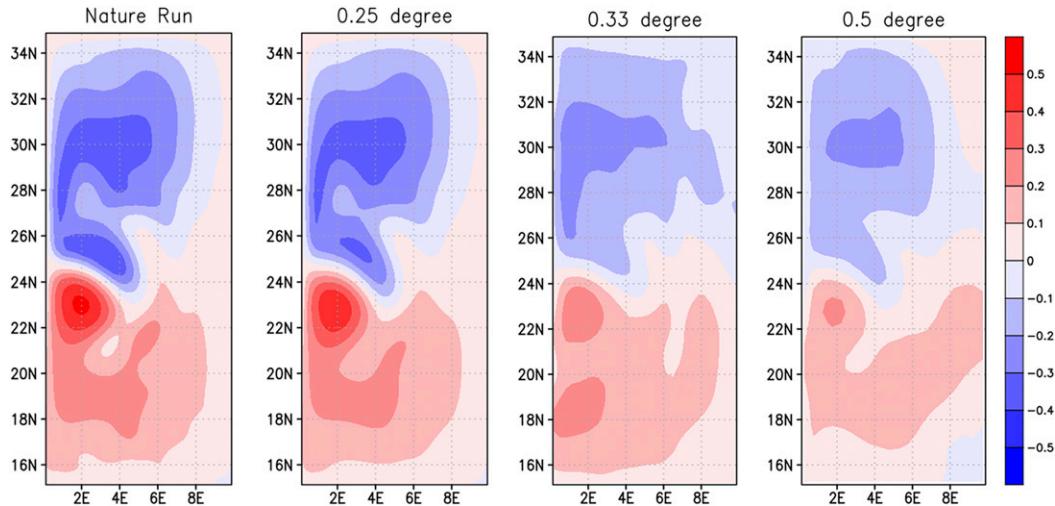


FIG. 12. SSH contour comparisons between the nature run and 1-day forecast ensemble means generated by forecast models of  $1/4^\circ$ ,  $1/3^\circ$ , and  $1/2^\circ$  resolutions before we start applying the LETKF-LaDA.

LETKF-LaDA and the prognostic flow scale of the forecasting model must be examined in the future with a high-resolution model.

#### d. Relation between the Lagrangian states and SSH

Due to the geostrophic relationship between the surface currents and pressure in the midlatitudes, and because SSH can be used as a proxy for integrated pressure in the upper ocean column, we expect some redundancy in the information provided by surface drifter position data and satellite altimeter measurements. For example, [Carrier et al. 2016](#) showed that assimilating the combination of the along-track SSH and the Eulerian flow velocity approximated by drifter locations can provide improved forecasts of SSH, compared to using either the observations of the along-track SSH or the Eulerian velocity alone. Similarly, we expect the LETKF-LaDA will be a viable method to improve forecasts of SSH by taking advantage of drifter position data.

We use the error correlation between the drifter state variables (e.g., longitude and latitude) and the ocean prognostic fields (e.g., temperature, salinity and velocity) to extract information from the Lagrangian position data that can be used to update the subsurface state. Because the general quasigeostrophic balance between the surface currents and dynamic ocean height fields indicates that a passive drifter will move following the contour of SSH, we infer that the correlation between the SSH forecast error and the Eulerian velocity forecast error is intertwined with the forecast errors corresponding to the drifter positions. To verify this, we spin up the ensemble as introduced in [section 3a](#) and compute the SSH ( $r_{SSH}$ ) and drifter forecast

error correlations ( $r_{lon}$  and  $r_{lat}$ ) using the formula (3.3) and (3.2) defined in section 3c before implementing the LETKF-LaDA.

We examine error correlations between the drifter positions and the 3D fluid and compare in the longitude–depth plane. For example, the  $r_{lon}$  and  $r_{lat}$  associated with drifter number 14 share similar patterns with the contour profile of  $r_{SSH}$  for all the flow fields ([Fig. 14](#)). Continuing the experiments for all the other drifters, we summarize the connection between  $r_{SSH}$  and  $r_{lon}$ ,  $r_{lat}$  in longitude–depth plane as following: for a specific drifter ID  $n$ ,

$$r_{SSH,n} \approx \alpha r_{lon,n} + \beta r_{lat,n}, \quad \alpha \propto -u_n \quad \text{and} \quad \beta \propto \text{sign}(u_n)v_n,$$

where  $(u_n, v_n)$  is the drifter velocity of drifter ID  $n$ . For those drifters located in unstable portions of the flow, a large ensemble size is needed to verify the above relation numerically. In the last row of [Fig. 14](#), the estimations of  $r_{SSH}$  are determined by this formula for drifter number 14, using the coefficients:  $\alpha = -u_n/(|u_n| + |v_n|)$  and  $\beta = \text{sign}(u_n) \times v_n/(|u_n| + |v_n|)$ . Comparing the first three rows with the last row in [Fig. 14](#), we observe that with this approximation formula,  $r_{lon}$  and  $r_{lat}$  partially recover the positivity and negativity of the correlations  $r_{SSH}$  in the given region, though the magnitude is different. The strongly linear relation between  $r_{lon}/r_{lat}$  and  $r_{SSH}$  suggests a similar relation in their corresponding error covariance matrices.

[Verrier et al. 2017](#) shows that the forecasts of SSH and ocean currents are continuously improved when transitioning from one altimeter to two altimeters with a relative error reduction of almost 30% under the

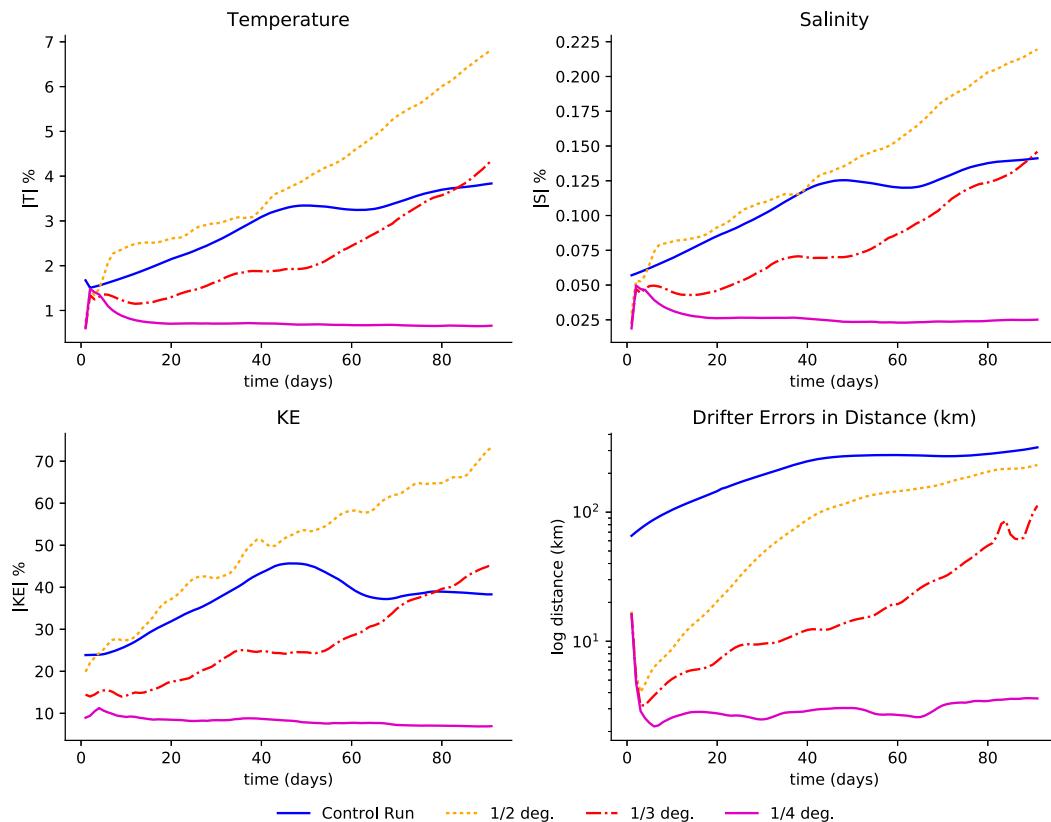


FIG. 13. Error norms of temperature, salinity, kinetic energy, and drifter states ( $RMSE_D$ ) in the given period [0, 91] using formula (3.1). The error norms of the ocean flow fields are aggregated from 5 to 1000 m depth (i.e.,  $h_o = 5$  m and  $h = 1000$  m). In each subplot, we show the experiment results of control run (blue solid line), LETKF-LaDA with forecast model of  $1/2^\circ$  (orange dotted line),  $1/3^\circ$  (red dashed-dotted line), and  $1/4^\circ$  (purple solid line) resolutions.

framework of OSSEs. The addition of the third altimeter further improves the results with additional relative error reduction of about 10%. Since increasing the number of altimeters improves forecasts of sea level and ocean currents, using LaDA of surface drifters, which have a similar error covariance/correlation relationship with the ocean prognostic variables as the sea surface height errors, should provide the potential for further improvement of the ocean surface currents, especially where there are gaps in the altimeter observations.

## 5. Discussion and outlook

We introduced a localized augmented-state Lagrangian data assimilation implemented with the local ensemble transform Kalman filter (LETKF-LaDA). Extending previous work with augmented-state LaDA methods, we applied this method to an ocean model with 3D dynamics and representation of temperature and salinity tracers. The LETKF-LaDA was found to provide more accurate analyses throughout the ocean

column from the surface to 1000 m depth by assimilating drifter position observations, compared to the conventional assimilation of in situ drifter measurements of temperature and salinity measurements. The reason why a growing misfit error for a drifter circuiting near the confluence is shown at the end of the first loop remains as an open question.

In the context of the LETKF-LaDA, we studied the impact of localization radius on stabilizing the error growth over time. We found that with radius  $5L_R$ , five times the baroclinic Rossby radius of deformation, the error decays fast in the first DA cycle but the nonlinear error grows unchecked over longer time. We also found that if the localization radius is too small (e.g.,  $L_R$ ) error often grows unconstrained in regions far away from the observations. The best performing localization radius in our experiments with the LETKF-LaDA assimilating 50 drifters deployed randomly around the gyre was between  $2L_R$  and  $3L_R$ .

Correlations between errors in the drifter locations and ocean states ( $r_{lon}$  and  $r_{lat}$ ) were also studied. For a

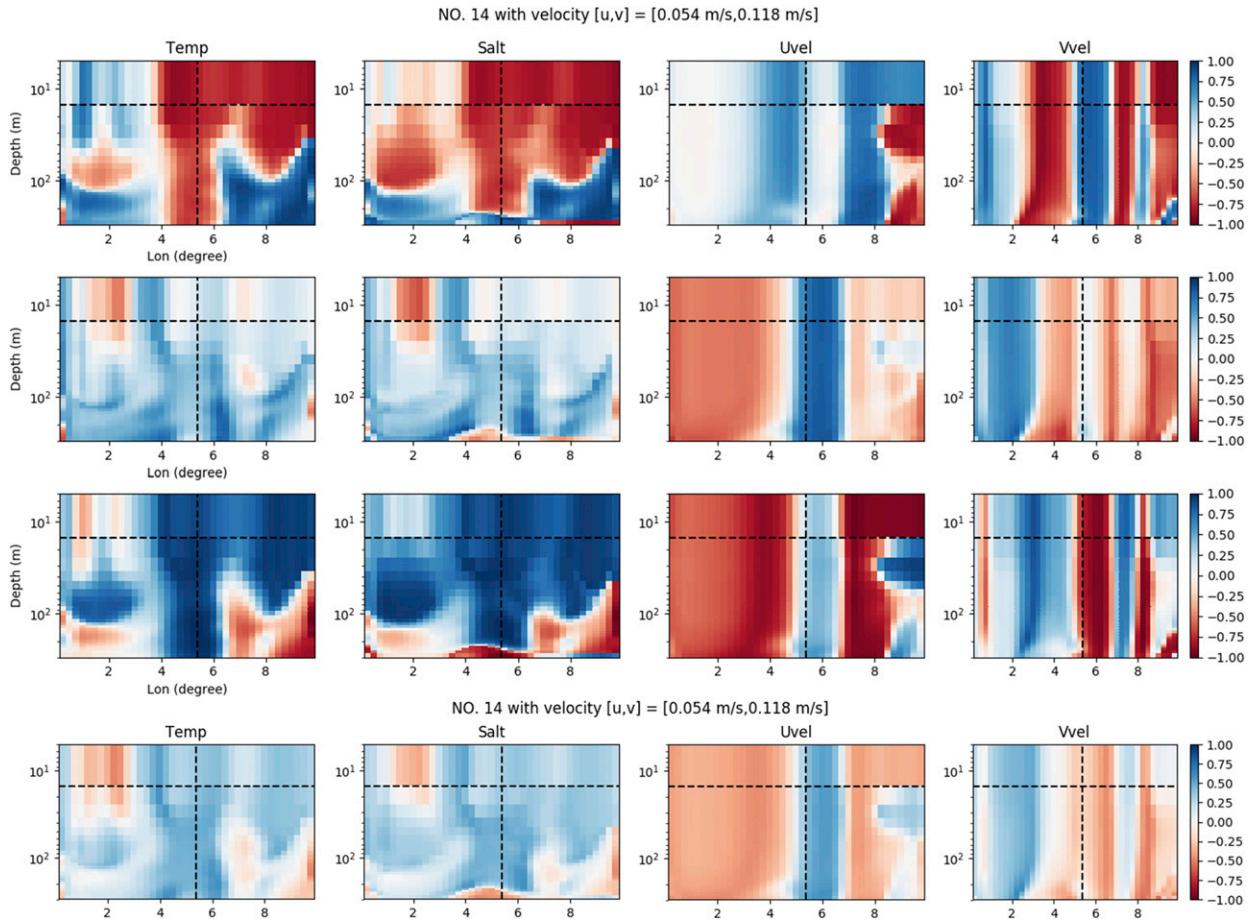


FIG. 14. Correlation comparisons among (first row)  $r_{lon}$ , (second row)  $r_{lat}$ , and (third row)  $r_{SSH}$  associated with drifter number 14 in the longitude–depth plane. (fourth row) The correlation estimation of  $r_{SSH}$  using the linear combination of  $r_{lon}$  and  $r_{lat}$  associated with drifter number 14 in the longitude–depth plane using coefficients:  $\alpha = -u_n/(|u_n| + |v_n|)$  and  $\beta = \text{sign}(u_n) \times v_n/(|u_n| + |v_n|)$ .

specific drifter,  $r_{lon}$  and  $r_{lat}$  had an approximately linear relation with the SSH error correlation ( $r_{SSH}$ ) at the drifter location. Since the error correlation is an influential factor in the performance of DA methods, future research will examine the use of LETKF-LaDA to simultaneously assimilate both drifter position and along-track SSH observations to improve the forecast accuracy of the SSH field.

The LETKF-LaDA was implemented using a parallel computing framework in order to accelerate the computation of the analysis in anticipation of scaling to an operational global ocean data assimilation system. An additional factor affecting the stability of the LaDA filter is the number of drifter observations. A software limitation in the MOM4p1 drifter module prevents a large number of drifters to be simulated. In future work we transition to using MOM6 with an updated drifters module.

Conducting an OSSE using the “identical twin” approach is an important first step for validation of any new DA method. However this type of experiment

ignores important concerns that are necessary to determine the viability of the LETKF-LaDA for operational use, such as the presence of systematic model errors. For example, in our experiments varying model resolution, results indicated that the LETKF-LaDA is sensitive to degraded model resolution. In future research, we will apply the LETKF-LaDA using a more realistic model of the latest GFDL MOM6, configured with various resolutions that span a range from eddy permitting to eddy resolving and forced with time-variant near-surface wind forcing, and transit the whole system into assimilating real historical drifter and hydrographic profile data.

*Acknowledgments.* LS and SGP acknowledge support from the National Oceanic and Atmospheric Administration (NOAA) National Environmental Satellite, Data, and Information Service (NESDIS) awarded via the National Aeronautics and Space Administration (NASA) Research Opportunities in Earth and Space Science (ROSES)

program (NA14NES4320003). SGP also acknowledges support from the NOAA Climate Program Office (CPO) (NA16OAR4310140), the NOAA Next Generation Global Prediction System (NGGPS) program (NA18NWS4680048 and NA16NWS4680011), and the Indian Institute of Tropical Meteorology (IITM) Monsoon Mission II (MM-II) (IITMMIIUNIVMARYLANDUSA2018INT2). Resources supporting this work were provided by the NASA High-End Computing (HEC) Program through the NASA Center for Climate Simulation (NCCS) at Goddard Space Flight Center. The SODA dataset of version 3.4.1 used in this research is publicly available ([http://www.atmos.umd.edu/~ocean/index\\_files/soda3.4.1\\_mn\\_download.htm](http://www.atmos.umd.edu/~ocean/index_files/soda3.4.1_mn_download.htm)).

## REFERENCES

- Aksoy, A., S. Lorsolo, T. Vukicevic, K. J. Sellwood, S. D. Aberson, and F. Zhang, 2012: The HWRP Hurricane Ensemble Data Assimilation System (HEDAS) for high-resolution data: The impact of airborne Doppler radar observations in an OSSE. *Mon. Wea. Rev.*, **140**, 1843–1862, <https://doi.org/10.1175/MWR-D-11-00212.1>.
- Apte, A., and C. K. R. T. Jones, 2013: The impact of nonlinearity in Lagrangian data assimilation. *Nonlinear Processes Geophys.*, **20**, 329–341, <https://doi.org/10.5194/npg-20-329-2013>.
- , —, and A. M. Stuart, 2008: A Bayesian approach to Lagrangian data assimilation. *Tellus*, **60A**, 336–347, <https://doi.org/10.1111/j.1600-0870.2007.00295.x>.
- Atlas, R., E. Kalnay, and M. Halem, 1985: Impact of satellite temperature sounding and wind data on numerical weather prediction. *Opt. Eng.*, **24**, 242341, <https://doi.org/10.1117/12.7973481>.
- Bishop, C. H., B. J. Etherton, and S. J. Majumdar, 2001: Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. *Mon. Wea. Rev.*, **129**, 420–436, [https://doi.org/10.1175/1520-0493\(2001\)129<0420:ASWTET>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<0420:ASWTET>2.0.CO;2).
- Bitterman, D. S., and D. V. Hansen, 1993: Evaluation of sea surface temperature measurements from drifting buoys. *J. Atmos. Oceanic Technol.*, **10**, 88–96, [https://doi.org/10.1175/1520-0426\(1993\)010<0088:EOSSTM>2.0.CO;2](https://doi.org/10.1175/1520-0426(1993)010<0088:EOSSTM>2.0.CO;2).
- Carrier, M. J., H. Ngodock, S. Smith, G. Jacobs, P. Muscarella, T. Ozgokmen, B. Haus, and B. Lipphardt, 2014: Impact of assimilating ocean velocity observations inferred from Lagrangian drifter data using the NCOM-4DVAR. *Mon. Wea. Rev.*, **142**, 1509–1524, <https://doi.org/10.1175/MWR-D-13-00236.1>.
- , H. E. Ngodock, P. Muscarella, and S. Smith, 2016: Impact of assimilating surface velocity observations on the model sea surface height using the NCOM-4DVAR. *Mon. Wea. Rev.*, **144**, 1051–1068, <https://doi.org/10.1175/MWR-D-14-00285.1>.
- Carton, J. A., G. A. Chepurin, and L. Chen, 2018a: SODA3: A new ocean climate reanalysis. *J. Climate*, **31**, 6967–6983, <https://doi.org/10.1175/JCLI-D-18-0149.1>.
- , —, —, and S. A. Grodzky, 2018b: Improved global net surface heat flux. *J. Geophys. Res. Oceans*, **123**, 3144–3163, <https://doi.org/10.1002/2017JC013137>.
- Chelton, D., R. A. deSzoeke, M. G. Schlax, K. El Naggar, and N. Siwertz, 1998: Geophysical variability of the first baroclinic Rossby radius of deformation. *J. Phys. Oceanogr.*, **28**, 433–460, [https://doi.org/10.1175/1520-0485\(1998\)028<0433:GVOTFB>2.0.CO;2](https://doi.org/10.1175/1520-0485(1998)028<0433:GVOTFB>2.0.CO;2).
- Dong, S., D. Volkov, G. Goni, R. Lumpkin, and G. R. Foltz, 2017: Near-surface salinity and temperature structure observed with dual-sensor drifters in the subtropical South Pacific. *J. Geophys. Res. Oceans*, **122**, 5952–5969, <https://doi.org/10.1002/2017JC012894>.
- Evensen, G., 2004: Sampling strategies and square root analysis schemes for the EnKF with correction. *Ocean Dyn.*, **54**, 539–560, <https://doi.org/10.1007/s10236-004-0099-2>.
- Greybush, S. J., E. Kalnay, T. Miyoshi, K. Ide, and B. Hunt, 2011: Balance and ensemble Kalman filter localization techniques. *Mon. Wea. Rev.*, **139**, 511–522, <https://doi.org/10.1175/2010MWR3328.1>.
- Griffies, S. M., 2008: Elements of MOM4p1. NOAA/Geophysical Fluid Dynamics Laboratory, 32 pp., [https://data1.gfdl.noaa.gov/~arl/pubrel/r/mom4p1/src/mom4p1/doc/mom4p1\\_synopsis.pdf](https://data1.gfdl.noaa.gov/~arl/pubrel/r/mom4p1/src/mom4p1/doc/mom4p1_synopsis.pdf).
- Halliwell, G. R., A. Srinivasan, V. Kourafalou, H. Yang, D. Willey, M. Le Hénaff, and R. Atlas, 2014: Rigorous evaluation of a fraternal twin ocean OSSE system for the open Gulf of Mexico. *J. Atmos. Oceanic Technol.*, **31**, 105–130, <https://doi.org/10.1175/JTECH-D-13-00011.1>.
- Hamill, T. M., J. S. Whitaker, and C. Snyder, 2001: Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Mon. Wea. Rev.*, **129**, 2776–2790, [https://doi.org/10.1175/1520-0493\(2001\)129<2776:DDFOBE>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<2776:DDFOBE>2.0.CO;2).
- Houtekamer, P. L., and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.*, **126**, 796–796, [https://doi.org/10.1175/1520-0493\(1998\)126<0796:DAUAEK>2.0.CO;2](https://doi.org/10.1175/1520-0493(1998)126<0796:DAUAEK>2.0.CO;2).
- Hunt, B. R., E. J. Kostelich, and I. Szunyogh, 2007: Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica D*, **230**, 112–126, <https://doi.org/10.1016/j.physd.2006.11.008>.
- Ide, K., L. Kuznetsov, and C. K. R. T. Jones, 2002: Lagrangian data assimilation for point vortex system. *J. Turbulence*, **3**, N53, <https://doi.org/10.1088/1468-5248/3/1/053>.
- Jacobs, G., and Coauthors, 2014: Data assimilation considerations for improved ocean predictability during the Gulf of Mexico Grand Lagrangian Deployment (GLAD). *Ocean Modell.*, **83**, 98–117, <https://doi.org/10.1016/j.ocemod.2014.09.003>.
- Katzfuss, M., J. R. Stroud, and C. K. Wikle, 2016: Understanding the ensemble Kalman filter. *Amer. Stat.*, **70**, 350–357, <https://doi.org/10.1080/00031305.2016.1141709>.
- Kuznetsov, L., K. Ide, and C. K. R. T. Jones, 2003: A method for assimilation of Lagrangian data. *Mon. Wea. Rev.*, **131**, 2247, [https://doi.org/10.1175/1520-0493\(2003\)131<2247:AMFAOL>2.0.CO;2](https://doi.org/10.1175/1520-0493(2003)131<2247:AMFAOL>2.0.CO;2).
- Lumpkin, R., and M. Pazos, 2007: Measuring surface currents with Surface Velocity Program drifters: The instrument, its data and some recent results. *Lagrangian Analysis and Prediction of Coastal and Ocean Dynamics*, A. Griffa et al., Eds., Cambridge University Press, 39–67.
- Melnichenko, O., P. Hacker, N. Maximenko, G. Lagerloef, and J. Potemra, 2016: Optimum interpolation analysis of Aquarius sea surface salinity. *J. Geophys. Res. Oceans*, **121**, 602–616, <https://doi.org/10.1002/2015JC011343>.
- Molcard, A., L. I. Piterbarg, A. Griffa, T. M. Özgökmen, and A. J. Mariano, 2003: Assimilation of drifter positions for the reconstruction of the Eulerian circulation field. *J. Geophys. Res.*, **108**, 3056, <https://doi.org/10.1029/2001JC001240>.
- Nilsson, J. A. U., D. Pettenuzzo, S. Dobricic, N. Pinardi, and P.-M. Poulain, 2012: Variational assimilation of Lagrangian trajectories

- in the Mediterranean Ocean forecasting system. *Ocean Sci.*, **8**, 249–259, <https://doi.org/10.5194/os-8-249-2012>.
- Ott, E., and Coauthors, 2004: A local ensemble Kalman filter for atmospheric data assimilation. *Tellus*, **56A**, 415–428, <https://doi.org/10.1111/j.1600-0870.2004.00076.x>.
- Ozgekmen, T. M., A. Molcard, T. M. Chin, L. I. Piterbarg, and A. Griffa, 2003: Assimilation of drifter positions in primitive equation models of midlatitude ocean circulation. *J. Geophys. Res.*, **108**, 3238, <https://doi.org/10.1029/2002JC001719>.
- Penny, S. G., and T. Miyoshi, 2016: A local particle filter for high-dimensional geophysical systems. *Nonlinear Processes Geophys.*, **23**, 391–405, <https://doi.org/10.5194/npg-23-391-2016>.
- , D. W. Behringer, J. A. Carton, and E. Kalnay, 2015: A Hybrid Global Ocean Data Assimilation System at NCEP. *Mon. Wea. Rev.*, **143**, 4660–4677, <https://doi.org/10.1175/MWR-D-14-00376.1>.
- Poje, A. C., and Coauthors, 2014: Submesoscale dispersion in the vicinity of the Deepwater Horizon spill. *Proc. Natl. Acad. Sci. USA*, **111**, 12 693–12 698, <https://doi.org/10.1073/pnas.1402452111>.
- Poterjoy, J., 2016: A localized particle filter for high-dimensional nonlinear systems. *Mon. Wea. Rev.*, **144**, 59–76, <https://doi.org/10.1175/MWR-D-15-0163.1>.
- Salman, H., L. Kuznetsov, C. K. R. T. Jones, and K. Ide, 2006: A method for assimilating Lagrangian data into a shallow-water-equation ocean model. *Mon. Wea. Rev.*, **134**, 1081–1101, <https://doi.org/10.1175/MWR3104.1>.
- , K. Ide, and C. K. R. T. Jones, 2008: Using flow geometry for drifter deployment in Lagrangian data assimilation. *Tellus*, **60A**, 321–335, <https://doi.org/10.1111/j.1600-0870.2007.00292.x>.
- Santitissadeekorn, N., E. T. Spiller, C. Jones, R. Rutarindwa, L. Liu, and K. Ide, 2014: Observing system simulation experiments of cross-layer Lagrangian data assimilation. *Dyn. Atmos. Oceans*, **66**, 77–93, <https://doi.org/10.1016/j.dynatmoce.2014.02.002>.
- Slivinski, L., E. Spiller, A. Apte, and B. Sandstede, 2015: A hybrid particle–ensemble Kalman filter for Lagrangian data assimilation. *Mon. Wea. Rev.*, **143**, 195–211, <https://doi.org/10.1175/MWR-D-14-00051.1>.
- Sluka, T. C., S. G. Penny, E. Kalnay, and T. Miyoshi, 2016: Assimilating atmospheric observations into the ocean using strongly coupled ensemble data assimilation. *Geophys. Res. Lett.*, **43**, 752–759, <https://doi.org/10.1002/2015GL067238>.
- Snyder, C., T. Bengtsson, P. Bickel, and J. Anderson, 2008: Obstacles to high-dimensional particle filtering. *Mon. Wea. Rev.*, **136**, 4629–4640, <https://doi.org/10.1175/2008MWR2529.1>.
- Taillandier, V., A. Griffa, and A. Molcard, 2006: A variational approach for the reconstruction of regional scale Eulerian velocity fields from Lagrangian data. *Ocean Modell.*, **13**, 1–24, <https://doi.org/10.1016/j.ocemod.2005.09.002>.
- van Leeuwen, P. J., 2003: A variance-minimizing filter for large-scale applications. *Mon. Wea. Rev.*, **131**, 2071–2084, [https://doi.org/10.1175/1520-0493\(2003\)131<2071:AVFFLA>2.0.CO;2](https://doi.org/10.1175/1520-0493(2003)131<2071:AVFFLA>2.0.CO;2).
- Vernieres, G., C. K. R. T. Jones, and K. Ide, 2011: Capturing eddy shedding in the Gulf of Mexico from Lagrangian observations. *Physica D*, **240**, 166–179, <https://doi.org/10.1016/j.physd.2010.06.008>.
- Verrier, S., P.-Y. Le Traon, and E. Remy, 2017: Assessing the impact of multiple altimeter missions and Argo in a global eddy-permitting data assimilation system. *Ocean Sci.*, **13**, 1077–1092, <https://doi.org/10.5194/os-13-1077-2017>.