

Supporting Information for “LEMDA: A Lagrangian-Eulerian Multiscale Data Assimilation Framework”

Quanling Deng¹, Nan Chen², Samuel N. Stechmann^{2,3}, and Jiuhua Hu²

¹School of Computing, Australian National University, Canberra, ACT 2601, Australia

²Department of Mathematics, University of Wisconsin–Madison, Madison, WI 53706, USA

³Department of Atmospheric and Oceanic Sciences, University of Wisconsin–Madison, Madison, WI 53706, USA

1. Discretization Scheme of the Eulerian Equations

In general, the equations (5) and (6) in the main text are of the following form of conservation law

$$\frac{\partial w}{\partial t} + \nabla \cdot (\mathbf{v}w) = f(\mathbf{x}, t) \quad (1)$$

with an initial solution w_0 .

This section describes the spatiotemporal discretization scheme implemented in Eulerian DA. Let (h_x, h_y) be the spatial grid spacing and dt be the time step. The spatial and temporal derivatives are approximated using finite differences. In the DA implementation, the first-order upwind finite difference scheme, finite volume method, and center-finite difference scheme for the discretization of the term $\nabla \cdot (\mathbf{v}w)$ are tested. They all yield

Corresponding author: Nan Chen, Department of Mathematics, University of Wisconsin–Madison, Madison, WI 53706, USA (chennan@math.wisc.edu)

similar results (based on the RMSE and pattern correlation skill scores). As has been pointed out in past studies (Grote & Majda, 2006; Majda & Grote, 2007), it is possible for an unstable discretization (such as the standard centered scheme) to be stabilized through the assimilation of observational data. Below, the centered finite difference discretization of $\nabla \cdot (\mathbf{v}w)$ is utilized for its symmetry and simplicity and good results.

The temporal discretization can be performed using an explicit time-stepping scheme, such as the forward Euler method:

$$\frac{\partial w}{\partial t} \approx \frac{w_{j,i}^{n+1} - w_{j,i}^n}{dt}, \quad (2)$$

where $w_{j,i}^n$ is the value of w at spatial grid point (x_i, y_j) and time step n . The spatial discretization is using a central difference scheme for the flux term:

$$\nabla \cdot (\mathbf{v}w) \approx \frac{(\mathbf{v}w)_{j,i+1}^n - (\mathbf{v}w)_{j,i-1}^n}{2h_x} + \frac{(\mathbf{v}w)_{j+1,i}^n - (\mathbf{v}w)_{j-1,i}^n}{2h_y}, \quad (3)$$

where (j, i) represents the indices in y and x dimensions following the Matlab matrix representation.

Substituting the spatial and temporal discretizations into (1) yields the following scheme:

$$\frac{w_{j,i}^{n+1} - w_{j,i}^n}{dt} + \nabla \cdot (\mathbf{v}w) \approx \frac{w_{j,i}^{n+1} - w_{j,i}^n}{dt} + \frac{(\mathbf{v}w)_{j,i+1}^n - (\mathbf{v}w)_{j,i-1}^n}{2h_x} + \frac{(\mathbf{v}w)_{j+1,i}^n - (\mathbf{v}w)_{j-1,i}^n}{2h_y} = f_{j,i}^n. \quad (4)$$

Lastly, the boundary terms are updated using periodic boundary conditions.

2. Estimating the Random Noise Coefficient in the Processes of Eulerian Observations

The processes of the statistical moment equations associated with the Eulerian observations, e.g., the momentum equation (6) in the main text, are a set of partial differential

equations. However, the actual Eulerian observations are the averaged values of these quantities at coarse mesh grids. Therefore, the observational processes in Eqn. (11a) in the main text are the spatial discrete form of Eqn. (6) in the main text and other statistical moment equations. At least two sources of errors need to be considered in carrying out Eulerian DA. One is the discretization error, and another is the error due to using a finite number of particles in computing the observed statistics. These potential errors are characterized by the random noise in Eqn. (11a) in the main text. Since the noise describes the uncertainty and influences the DA solution, it is essential to provide an effective estimation of the random noise coefficient $\Sigma_{\mathbf{Z}}$, which is written as $\Sigma_{\mathbf{Z}} = \Sigma_{\mathbf{Z},dis} + \Sigma_{\mathbf{Z},spl}$, representing the uncertainty associated with the discretization and the finite number of samples, respectively. For simplicity, assume both $\Sigma_{\mathbf{Z},dis}$ and $\Sigma_{\mathbf{Z},spl}$ are diagonal matrices, implying the independence of the noise at different grid cells.

Consider the spatial discretized equation of Eqn. (6) in the main text evaluated at a specific location $\mathbf{x}^*$:

$$\left. \frac{d\langle \rho \mathbf{v} \rangle}{dt} \right|_{\mathbf{x}^*} = G(\rho \mathbf{v}), \quad (5a)$$

$$\left. \frac{d\langle \widetilde{\rho \mathbf{v}} \rangle}{dt} \right|_{\mathbf{x}^*} = \widetilde{G}(\widetilde{\rho \mathbf{v}}), \quad (5b)$$

where a fine discretization is used in (5a) while the spatial discretization of (5b) is based on a coarse mesh as in LEMDA. The solution in (5a) can be regarded as the truth when studying the discretization error in (5b). Taking the difference between these two equations yields

$$\frac{d\zeta_{\mathbf{x}^*}}{dt} := \left. \frac{d(\langle \rho \mathbf{v} \rangle - \langle \widetilde{\rho \mathbf{v}} \rangle)}{dt} \right|_{\mathbf{x}^*} = G(\rho \mathbf{v}) - \widetilde{G}(\widetilde{\rho \mathbf{v}}). \quad (6)$$

Note that since $\langle \rho \mathbf{v} \rangle$ is computed using a much finer solver, a spatial average over the multiple nearby locations within the same grid cell as $\langle \widetilde{\rho \mathbf{v}} \rangle$ needs to be taken in computing (6). The equation (6) is run forward by only one time step from t to $t + \Delta t$ with the initial condition $\zeta_{\mathbf{x}^*}(t) = 0$. Therefore, the error is solely due to the difference between the functions G and $\widetilde{G}$ that accounts for the discretization error. By computing the one-step solution of $\zeta_{\mathbf{x}^*}((j+1)\Delta t)$ with the initial condition of $\langle \rho \mathbf{v} \rangle(j\Delta t)$ being taken at different time instants for $j = 1, 2, \dots, J$, the resulting J points of $\zeta_{\mathbf{x}^*}$ are then used to estimate one diagonal component of the noise coefficient $\Sigma_{\mathbf{Z},dis}$, which is the standard deviation of these $\zeta_{\mathbf{x}^*}$.

Next, the diagonal component of $\Sigma_{\mathbf{Z},spl}$ is estimated in light of the standard Monte Carlo error analysis. Denote by n the number of particles average over time in one specific grid cell. This gives the corresponding diagonal component of $\Sigma_{\mathbf{Z},spl} = \sigma_x / \sqrt{n}$.

A practical simplification of $\Sigma_{\mathbf{Z},dis}$ and $\Sigma_{\mathbf{Z},spl}$ is to assume both of them are constant diagonal matrices, which is appropriate if the underlying flow field is statistically homogeneous. Recall that L is the total number of particles in the domain. Denote by $N = N_x \times N_y$ the number of the discrete grid cells, where N_x and N_y are the number of grids in the x and the y directions, respectively. Denote by ζ the set that contains all the NJ points from the $\zeta_{\mathbf{x}^*}$ at different $\mathbf{x}^*$. Similarly, the average number of particles in each grid cell is L/N . The constant diagonal entries of $\Sigma_{\mathbf{Z},dis}$ and $\Sigma_{\mathbf{Z},spl}$ are thus given by $\text{std}(\zeta)$ and $\sigma_x N/L$, respectively. These will be the estimated noise coefficients $\Sigma_{\mathbf{Z}}$ used in the numerical experiments. In some cases, especially in the presence of model error, noise inflation can be further added based on the calculated values from the above procedure.

3. Interpolating Missing Values at Eulerian Grid Points

It is not always guaranteed that each grid cell contains at least one Lagrangian particle throughout time, especially when the total number of tracers is limited or the underlying flow field is strongly compressible. To fill in the missing data, a standard thin-plate spline interpolation method is applied (Duchon, 1977).

A numerical simulation is shown in Figure S1 as a proof-of-concept validation. In this simulation, an incompressible flow field with $K_{\max} = 1$ is adopted such that there are a total of 9 Fourier modes in the random flow field. It is represented using the red arrows in the figure. The contour plot in Panel (a) shows the spatial distribution of the zonal momentum (e.g., the first component of $\langle \rho \mathbf{v} \rangle$ in Eqn. (6) in the main text) with three different spatial discretizations: $N_x = N_y = 5, 10$ and 25 . First, the Eulerian observations are computed based on $L = 5000$ particles. Panels (a)–(c) include the spatial distribution of the zonal momentum at a specific time $t = 2$ with different mesh grids. Notably, with a spatial resolution of $N_x = N_y = 25$, there are, on average, only 8 particles in each grid cell. Despite having a small number of particles, the zonal momentum has a smooth transition in space. Panel (d) shows the spatial distribution of the zonal momentum with $N_x = N_y = 25$ but using only $L = 500$ Lagrangian particles. As expected, several grid cells contain no particles. Then, the spatial interpolation is applied, and the result is shown in Panel (e). The spatial distribution after the interpolation is similar to the case using $L = 5000$ particles. Their difference, as is shown in Panel (f), is negligible.

References

Duchon, J. (1977). Splines minimizing rotation-invariant semi-norms in Sobolev spaces.

In *Constructive theory of functions of several variables: Proceedings of a conference*

held at oberwolfach april 25–may 1, 1976 (pp. 85–100).

Grote, M. J., & Majda, A. J. (2006). Stable time filtering of strongly unstable spatially extended systems. *Proceedings of the National Academy of Sciences*, *103*(20), 7548–7553.

Majda, A. J., & Grote, M. J. (2007). Explicit off-line criteria for stable accurate time filtering of strongly unstable spatially extended systems. *Proceedings of the National Academy of Sciences*, *104*(4), 1124–1129.

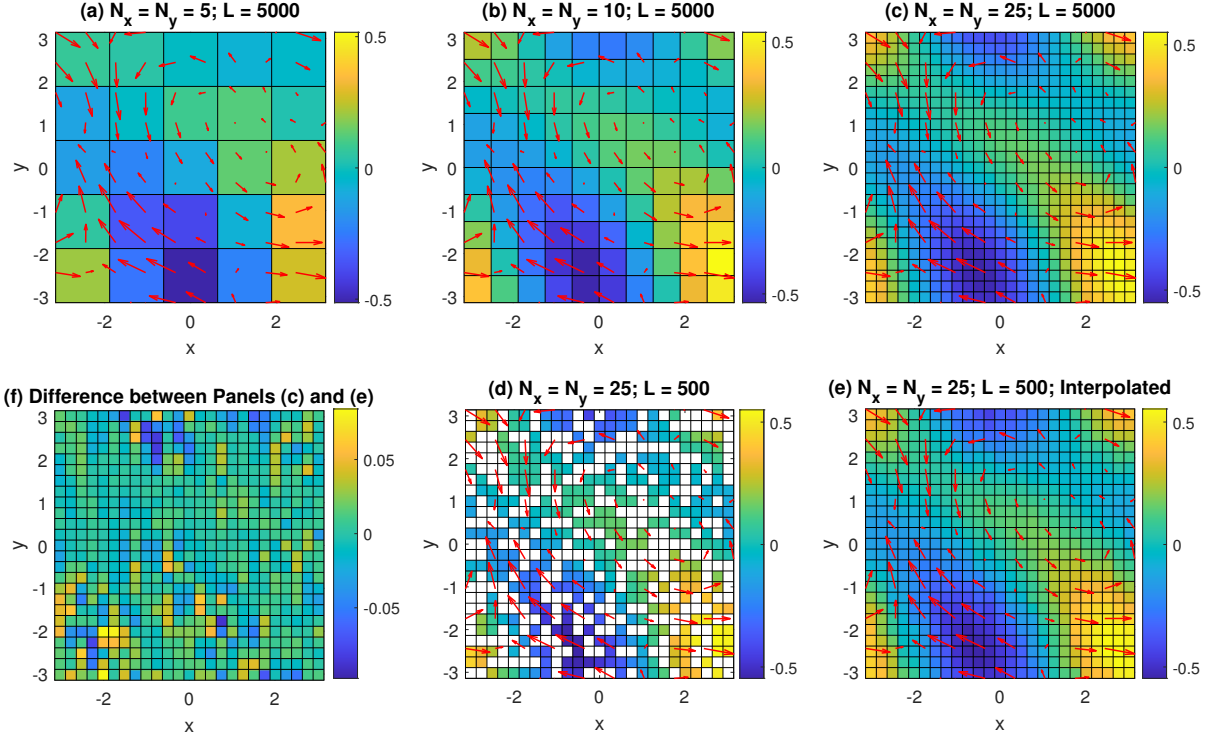


Figure S1. A proof-of-concept validation. Panels (a)–(c) include the spatial distribution of the zonal momentum at a specific time $t = 2$ with different mesh grids using $L = 5000$ particles. Panel (d) shows the spatial distribution of the zonal momentum with $N_x = N_y = 25$ but using only $L = 500$ Lagrangian particles. Panel (e) shows the result after applying the spatial interpolation to Panel (d). Panel (f) shows the difference between Panel (d) and Panel (e).