Check for updates

JAMES | Journal of Advances in Modeling Earth Systems

RESEARCH ARTICLE
10.1029/2025MS005702

Machine Learning-Based Representation of Model Error and Observation Error With Estimated Cross-Correlation in Data Assimilation

Yingjie Chen[1,2], Yuefei Zeng[1,2], Kefeng Zhu[3,4], Yuxuan Feng[5], and Tijana Janjić[6]

[1]State Key Laboratory of Climate System Prediction and Risk Management/Key Laboratory of Meteorological Disaster, Ministry of Education/Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disasters, Nanjing University of Information Science and Technology, Nanjing, China, [2]School of Atmospheric Sciences, Nanjing University of Information Science and Technology, Nanjing, China, [3]Nanjing Innovation Institute for Atmospheric Sciences, Chinese Academy of Meteorological Sciences-Jiangsu Meteorological Service, Nanjing, China, [4]Jiangsu Key Laboratory of Severe Storm Disaster Risk/Key Laboratory of Transportation Meteorology of CMA, Nanjing, China, [5]Jiangsu Meteorological Observatory, Jiangsu Meteorological Bureau, Nanjing, China, [6]Mathematical Institute for Machine Learning and Data Science, Catholic University of Eichstätt-Ingolstadt (KU), Ingolstadt, Germany

Key Points:

- New machine learning-based schemes for model and observation error representation are developed
- The machine learning-based model error scheme successfully recovers smaller-scale structures and achieves the lowest error
- Including forecast-observation error cross-correlation yields a further RMSE reduction

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

Y. Zeng,
yuefei.zeng@nuist.edu.cn

Citation:

Chen, Y., Zeng, Y., Zhu, K., Feng, Y., & Janjić, T. (2026). Machine learning-based representation of model error and observation error with estimated cross-correlation in data assimilation. *Journal of Advances in Modeling Earth Systems*, *18*, e2025MS005702. https://doi.org/10.1029/2025MS005702

Received 7 FEB 2026
Accepted 22 OCT 2026

Abstract The current study presents machine learning (ML)-based schemes that independently represent model error and observation error during data assimilation. The Model Error ML scheme (MLME) uses a generative adversarial network (GAN) to produce flow-dependent model corrections. The Observation Error ML scheme (MLOE) also uses the GAN to adaptively estimate the observation error covariance matrix. Furthermore, a generalized formulation (CorrFOE) is derived to incorporate the often-neglected cross-correlations between forecast and observation errors. The proposed schemes are rigorously evaluated via twin experiments with the Lorenz III model. Results show that the ML-based schemes reduce the root-mean-square error (RMSE) of analyses relative to the control and, for model error, relative to the additive-noise baseline. Additionally, the MLME scheme is particularly effective at recovering small-scale structures. Crucially, the synergistic combination of both ML schemes with the explicit inclusion of cross-correlation between forecast and observation errors produces the most accurate analyses. The added computational costs for MLME, MLOE and CorrFOE are moderate, confirming their operational potential.

Plain Language Summary Errors in weather prediction come from two main sources: the model itself and the observations used to update it. Using a simplified atmospheric model, we show that our machine learning approaches can estimate these errors more accurately than traditional techniques. Specifically, our method not only reduces overall forecast errors but also helps recover small-scale weather features that are often not well represented. Additionally, we found that considering how model errors and observation errors relate to each other further improves forecast accuracy. Together, these advances offer a more complete way to handle uncertainties in weather forecasting, paving the way for more reliable predictions.

1. Introduction

Data Assimilation (DA) is a process that derives analyses by comparing model backgrounds and observations, based on knowledge or assumptions about background error and observation error characteristics. Model error is defined as the discrepancy between simulated and true atmospheric states. This error arises from various sources, including limited model resolution and imperfect parameterizations of subgrid-scale physical processes. Model error is an important part of background error and its representation impacts the performance of DA algorithms. For ensemble-based approaches, techniques such as (adaptive) multiplicative inflation (Anderson, 2001) and relaxation methods (Whitaker & Hamill, 2012; Zhang et al., 2004) have been often used to account for model error. However, those methods do not correct systematic model error directly but merely dampen its impact on filter overconfidence. In convection-scale DA more principled additive noise schemes are the preferred approaches. For instance, Snyder and Zhang (2003) applied Gaussian noise to horizontal wind and potential temperature fields in storm regions, leading to reduced spurious convection. Dowell et al. (2004) introduced smoothed elliptical perturbations during assimilation cycles, achieving faster updrafts and larger spreads. Dowell and Wicker (2009) restrict perturbations within regions of considerable radar reflectivities, and Sobash and Wicker (2015) extended this by using threshold-dependent perturbations to minimize temperature/humidity

© 2026 The Author(s). Journal of Advances in Modeling Earth Systems published by Wiley Periodicals LLC on behalf of American Geophysical Union. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

19422466, 2026, 11, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025MS005702 by Katholische Universitat Eichstatt-Ingolstadt, Wiley Online Library on [24/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

biases. Zeng et al. (2018, 2019) combined the Large-scale Additive Noise (LAN) based on random draws from the background error covariance of the global model with the Small-scale Additive Noise (SAN) based on random draws of samples for model truncation error, resulting in improved short-term precipitation forecasts particularly under weak forcing conditions. However, the random-noise-based schemes omit the consistency to model states and thus lack of flow-dependency (Zeng et al., 2020).

In recent years, Machine Learning (ML) techniques proved to be powerful tools to account for model error. Brajard et al. (2021) used the Convolutional Neural Network (CNN) to learn model truncation error due to lower resolutions and incorporated it into DA. Legler and Janjic (2022) and Peng et al. (2024) employed Bayesian Neural Network (BNN) and CNN, respectively, to online learn model error arising from parameter uncertainties during DA. Lu (2025) introduced a UNet for learning flow dependent background error covariances, which can yield improved DA analysis for small ensembles. While those studies are based on simplified models, the first attempt with the operational model was carried out by Bonavita and Laloyaux (2020), in which an Artificial Neural Network (ANN)-model is trained on data of analysis increments of the ECMWF and corrects systematic model bias. More recently, Gregory et al. (2024, 2026) implemented an online sea ice bias correction in a global ice-ocean model using the CNN trained on data assimilation increments, improving initial conditions and subsequent forecasts. Chapman and Berner (2025) introduced a CNN approach that predicts real-time nudging increments from reanalysis data to correct climate model biases online, significantly improving state variable accuracy and enhancing the simulation of the Madden-Julian Oscillation. Furthermore, the generative ML has also permeated into DA. For instance, Manshausen et al. (2025) performed the so-called score-based DA to assimilate sparse station data into a ML (surrogate) model to generate km-scale precipitation and surface wind maps over the central US, showing physical consistency and skill improvements over a state-of-the-art analysis product, representing a promising advancement in efficient, deep generative DA for operational weather models.

Observation error is composed of instrument error and representation error, while the latter consists of scale-mismatch error (arising from unresolved model scales and processes), error in data pre-processing or quality control, and observation-operator error due to the approximations required when formulating the mapping from state to observation space (Janjic et al., 2018). To estimate observation errors, a number of modeling centers use the diagnostic method proposed by Desroziers et al. (2005) (hereafter, D05), which is based on observation-minus-background and observation-minus-analysis statistics. However, D05 is based on specific assumptions (e.g., background and observation errors are uncorrelated and precisely provided) and its accuracy considerably depend on settings of data assimilation systems (e.g., inflation schemes and ensemble size for the ensemble-based DA Walsworth et al. (2023)). There are other approaches to estimate or account for observation error statistics, for example, the three-cornered hat method based on the comparison of observations to two independent references (Anthes & Rieckh, 2018) and a non-parametric method that estimates the observation error based on ensemble innovations (Hu et al., 2024). Moreover, ML has also demonstrated its great potential here, for example, Cheng and Qiu (2022) proposed a Long Short-Term Memory (LSTM)-based approach to estimate observation error characteristics online.

Among other error components, representation error due to unresolved scales and processes (hereafter denoted by "UE") has received particular attention. UE is an important source of error for conventional and radar observations Waller et al. (2021) and Zeng et al. (2021). Furthermore, due to existence of UE, both model and forecast error are correlated with observation error (Janjic et al., 2018). Janjic and Cohn (2006) introduced a mathematical formalism to account for UE during DA. Recently, Ohishi et al. (2025) reformulated DA algorithm to include a presumed and simplistic cross-correlation between forecast error and observation error and investigated the performance of the algorithm with a perfect model (i.e., no model error). Nevertheless, studies on DA algorithms accounting for the cross-correlation still remain highly insufficient.

In the current study, for model error representation, a GAN-based scheme is introduced, which generates flow-dependent corrections with small-scale structure recovery, going beyond random-noise-based methods in Zeng et al. (2019). For the observation error representation, the paradigm of Cheng and Qiu (2022) is adapted to a GAN architecture to better exploit the structured spatial dependencies of geophysical fields. Furthermore, the most important innovation of this work is to refine the DA algorithm to incorporate the cross-correlation between forecast and observation errors in a general form. The paper is organized as follows. In Section 2, after a short introduction to the Lorenz III model, the detailed descriptions of ML-based schemes and derivation of the new

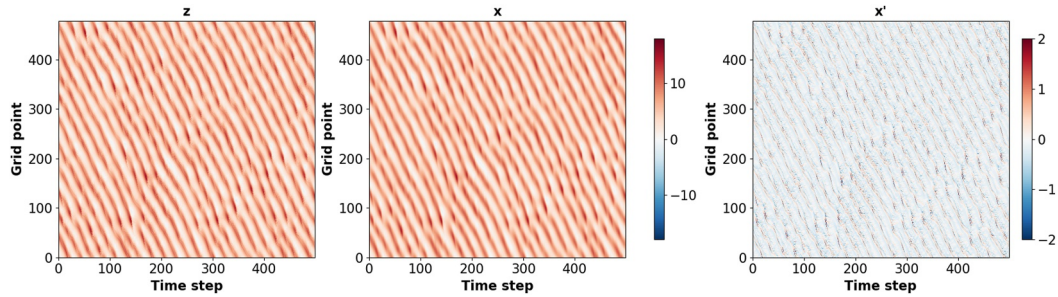


Figure 1. Hovmöller diagrams of **z** (left), **x** (middle) and **x′** (right).

formulation to account for cross-correlation are provided. Section 3 describes experimental setup and results, followed by Section 4 for conclusion and outlook.

2. Methodology

In this study, the hybrid-scale Lorenz model III (Lorenz, 2005) is used. Its hybrid-scale model variable z can be divided into the large-scale x and the small-scale x'. x is considered here as the model state. For investigation of model error, z is considered as the truth and the difference between truth z and large-scale x is the model error x'. For investigation of observation error, z is treated as observations and x' represents the error due to unresolved scales, namely UE. First, It is assumed that neither instrument error, forward operator error, nor pre-processing error is introduced, therefore, observation error only consists of UE. It is noted that model error and observation error are identical and forecast error and observation error are correlated due to the existence of UE. Furthermore, instrument error is introduced in addition to UE. In the following, Generative Adversarial Network (GAN) is used to construct ML-models to predict x' and its covariance. Furthermore, a refined formulation of the Ensemble Kalman Transform Filter (ETKF, Bishop et al., 2001) to include the cross-correlation.

2.1. Lorenz Model III

The Lorenz model III simulates two-scale interaction and nonlinear dynamics of mid-latitude atmospheric motions:

$$\frac{dz_n}{dt} = [x,x]_{K,n} + b^2[x',x']_{1,n} + c[x',x]_{1,n} - x_n - bx'_n + F, \tag{1}$$

where $[x,x]_{K,n}$, $[x',x']_{1,n}$ and $[x',x]_{1,n}$ are the "modified" summations defined in Lorenz (2005). The index $n = 1,\ldots,N$, where N is the total number of grid points. x and x' represent the large- and small-scale components of a geophysical flow, respectively; c is the coupling parameter of two scales, F is the model forcing and b is the small-scale forcing parameter. K determines the characteristic spatial scale of the model's large-scale waves, and I determines the width of the smoothing window used to separate scales. It holds

$$x_n = \sum_{i=-I}^{I} (\alpha - \beta|i|) z_{n+i}, \tag{2}$$

$$x'_n = z_n - x_n, \tag{3}$$

where $\alpha = \left(3I^2 + 3\right)/\left(2I^3 + 4I\right)$, $\beta = \left(2I^2 + 1\right)/\left(I^4 + 2I^2\right)$ and I is a smoothing parameter.

In the current study, the model is run on a 1-dimensional 480-gridpoint space with model integration time step of 0.001. Parameters are set as follows: $F = 12.0$, $b = 10.0$, $c = 2.5$, $K = 8$ and $I = 12$. The state vectors of variables x, x' and z are denoted by **x**, **x′** and **z**. Figure 1 illustrates the temporal evolution of **x**, **x′** and **z** at the first grid point. It can be seen that **x** has large amplitude and slow variability, **x′** has small amplitude and fast variability and **z** is the sum of both.

19422466, 2026, 11, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025MS005702 by Katholische Universitat Eichstatt-Ingolstadt, Wiley Online Library on [24/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

19422466, 2026, 11, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025MS005702 by Katholische Universitat Eichstatt-Ingolstadt, Wiley Online Library on [24/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

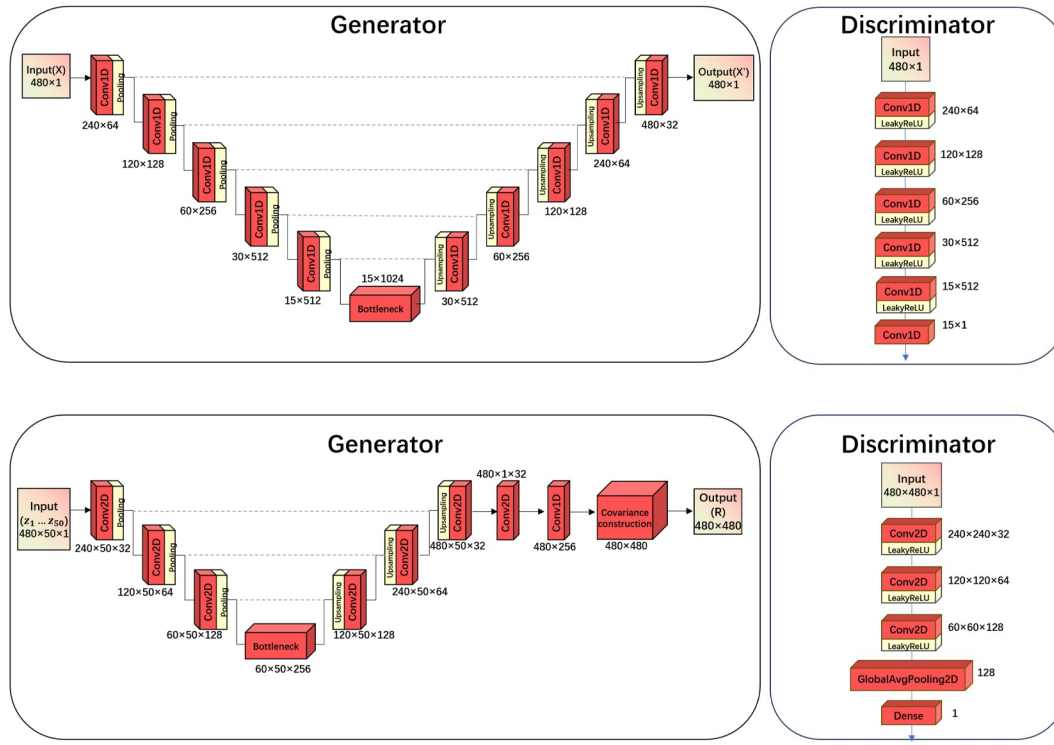


Figure 2. Neural network architecture for the generator and the discriminator for model error (upper panel) and for observation error (lower panel).

2.2. Machine Learning

The ML-based model for model error representation (MLME) is based on a GAN framework (Goodfellow et al., 2014). The GAN consists of a generator and a discriminator (see Figure 2). The generator learns a mapping from a input to its corresponding error representation, while the discriminator evaluates the realism of generated outputs by comparing them against reference data. In the current work, the generator follows a one-dimensional U-Net–like encoder–decoder structure (Isola et al., 2017), designed to capture spatial patterns of model errors. The encoder applies successive 1D convolutional layers with batch normalization and LeakyReLU activations, combined with average pooling to progressively compress spatial resolution while expanding the feature channels. At the bottleneck, a convolutional block with wider kernels is employed to enhance multi-scale feature extraction. The decoder restores the original resolution through upsampling, with skip connections linking encoder and decoder layers to preserve fine-scale details. A final convolutional layer outputs the predicted model error field. The discriminator adopts a PatchGAN design (Isola et al., 2017), consisting of stacked 1D convolutional layers and LeakyReLU activations, progressively downsampling the input model error field. Unlike global classifiers, PatchGAN assigns realism scores to overlapping local regions, thereby enforcing both local consistency and global realism in the generated error fields. This adversarial training strategy enhances the model's ability to capture subtle statistical features and produces error estimates that are more consistent with the underlying dynamics. A composite loss function that combines adversarial learning with reconstruction and structural constraints is defined as

$$L = L_{\mathrm{GAN}} + \lambda_1 L_1 + \lambda_{\mathrm{grad}} L_{\mathrm{grad}} + \lambda_{\mathrm{lap}} L_{\mathrm{lap}}, \tag{4}$$

$$L_1 = \frac{1}{N_s}\sum_{i=1}^{N_s} \left\| \hat{\mathbf{e}}^{(i)} - \mathbf{e}^{(i)} \right\|_{l1}, L_{\mathrm{grad}} = \frac{1}{N_s}\sum_{i=1}^{N_s} \left\| \nabla\hat{\mathbf{e}}^{(i)} - \nabla\mathbf{e}^{(i)} \right\|_{l1}, L_{\mathrm{lap}} = \frac{1}{N_s}\sum_{i=1}^{N_s} \left\| \Delta\hat{\mathbf{e}}^{(i)} - \Delta\mathbf{e}^{(i)} \right\|_2^2, \tag{5}$$

where N_s is the number of samples, $\hat{\mathbf{e}}^{(i)}$ denotes the predicted error field produced by the generator for sample i, and $\mathbf{e}^{(i)}$ denotes the corresponding target error field. In this formulation, L_{GAN} is the least-squares adversarial loss

19422466, 2026, 11, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025MS005702 by Katholische Universitat Eichstatt-Ingolstadt, Wiley Online Library on [24/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

that drives the generator to fool the discriminator, L_1 aims to reduce the mean absolute prediction error, and L_{grad} and L_{lap} are structural constraints. L_{grad} matches the first-order differences between prediction and target to preserve edges and fronts and reduces phase errors without over-smoothing fronts. L_{lap} matches curvature via the discrete Laplacian and this high-frequency–weighted matching suppresses spurious oscillations while recovering genuine small-scale structure present in the target. The loss function combines adversarial loss with strong reconstruction terms, which heavily penalizes the generator for collapsing to a single mode, as it must accurately reconstruct a wide variety of target error fields from the training set. Through sensitivity analysis of the loss weights (see Supporting Information S1), the configuration $(\lambda_1, \lambda_{\text{grad}}, \lambda_{\text{lap}}) = (50, 5, 5)$ was found to be optimal. This balance ensures that the model achieves both global accuracy and the retention of fine-scale spatial features.

The ML-based model for observation error representation (MLOE) shares a similar design as the MLME. Following Cheng and Qiu (2022), its input is a time-series window of observations and the output is the predicted observation error covariance $\mathbf{R}$. The loss function is defined as

$$L = L_{\text{GAN}} + \lambda \, \|\hat{\mathbf{R}} - \mathbf{R}\|_F^2, \tag{6}$$

where $\|\cdot\|$ is the Frobenius norm and $\hat{\mathbf{R}}$ is the target. For each window, the observation-error samples are used to compute an observation-error covariance matrix,

$$\mathbf{R}^{(i)} = \frac{1}{L-1} \sum_{t=1}^{L} \left(\boldsymbol{\epsilon}_{o,t}^{(i)} - \bar{\boldsymbol{\epsilon}}_o^{(i)}\right)\left(\boldsymbol{\epsilon}_{o,t}^{(i)} - \bar{\boldsymbol{\epsilon}}_o^{(i)}\right)^{\mathrm{T}}, \tag{7}$$

where L is the number of samples. Each $\mathbf{R}^{(i)}$ is then converted to a correlation matrix:

$$\rho_{mn}^{(i)} = \frac{R_{mn}^{(i)}}{\sqrt{R_{mm}^{(i)} R_{nn}^{(i)}}}. \tag{8}$$

To prepare the training data for the MLME and the MLOE, the model is run 500,000 time steps. The first 80% of data are used for training, the next 10% for validation and the last 10% for testing. For the MLME, the input is the large-scale state $\mathbf{x}$ and the output is small-scale state $\mathbf{x}'$, while for the MLOE the input is a sequence $\left(\mathbf{z}^1, \mathbf{z}^2, \ldots, \mathbf{z}^{50}\right)$. The training data for the MLOE are segmented into non-overlapping time series of 50 consecutive steps, resulting in 10,000 samples.

Figure 3 gives an example that the MLME use the large-scale state $\mathbf{x}$ as the input to generate its corresponding model error $\mathbf{x}'$. The RMSE of the predicted background state (the sum of $\mathbf{x}$ and $\mathbf{x}'$) is 0.188, much smaller than that of the background state, which is 0.393. Furthermore, Figure 4 compares the observation error correlation predicted by the MLOE to the target. It can be seen that the MLOE produces similar basic pattern of the correlations. By the experimental design, observation error is equivalent to model error, therefore, wave-like structure in Figure 4 mirrors small-scale model dynamics.

In this study, we use a GAN as one generative-model implementation to examine whether conditional, flow-dependent error representations can be learned and incorporated into the proposed data-assimilation framework. The GAN formulation provides a practical setting for this investigation because the generator produces an error estimate through a single forward pass after training and can be trained using paired input-target samples. To promote stable training and preserve the structure of the generated error fields, we employ batch normalization and a composite loss comprising reconstruction, gradient-consistency, and Laplacian-consistency terms. These terms encourage agreement between the generated and reference fields while retaining spatial structures at multiple scales.

2.3. Data Assimilation

2.3.1. Augmented Ensemble Transform Kalman Filter

This study employs the Augmented Ensemble Transform Kalman Filter (AETKF; Sommer and Janjić (2018); Farchi and Bocquet (2019)), which extends the ETKF via modulation-wise localization. The AETKF is an

19422466, 2026, 11, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025MS005702 by Katholische Universitat Eichstatt-Ingolstadt, Wiley Online Library on [24/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

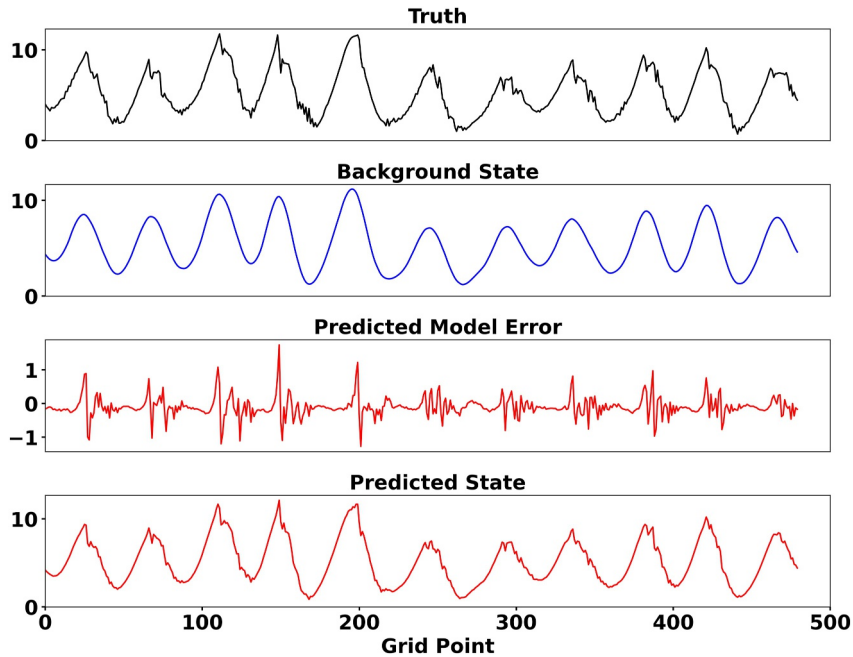


Figure 3. An example: truth (first panel), background state (second panel), model error generated by the MLME (third panel) and the sum of background state and model error (fourth panel). The RMSE of the background state is 0.393 and the RMSE of the predicted background state is 0.188.

operationally viable approach, having been implemented at centers such as Environment and Climate Change Canada (Buehner, 2005) for ensemble-informed background error covariances. Its formulation is detailed below.

Let the background error covariance $\mathbf{P}_b = \rho \circ \left(\mathbf{X}_b \mathbf{X}_b^T\right)$, where $\rho \in \mathbb{R}^{N_x \times N_x}$ is localization matrix and $\mathbf{X}_b = \frac{1}{\sqrt{N_e - 1}}\left[\mathbf{x}_{b(1)} - \bar{\mathbf{x}}_b, \ldots, \mathbf{x}_{b(N_e)} - \bar{\mathbf{x}}_b\right] \in \mathbb{R}^{N_x \times N_e}$ is background ensemble perturbation, $\bar{\mathbf{x}}_b \in \mathbb{R}^{N_x}$ is the background ensemble mean, N_x is the model dimension and N_e is the ensemble size. Suppose $\rho \approx \mathbf{L}\mathbf{L}^T$ and $\mathbf{L} \in \mathbb{R}^{N_x \times N_m}$ contain the N_m leading modes of ρ, which can be obtained by the Singular Value Decomposition of the localization matrix ρ. Then,

$$\rho \circ \left(\mathbf{X}_b \mathbf{X}_b^T\right) \approx \mathbf{L}\mathbf{L}^T \circ \left(\mathbf{X}_b \mathbf{X}_b^T\right) = \widetilde{\mathbf{X}}_b \widetilde{\mathbf{X}}_b^T, \tag{9}$$

where

$$\widetilde{\mathbf{X}}_b = \left[\mathrm{diag}\left(\mathbf{X}^b(:,1)\right)\mathbf{L}, \ \ldots, \ \mathrm{diag}\left(\mathbf{X}^b(:, N_e)\right)\mathbf{L}\right] \tag{10}$$

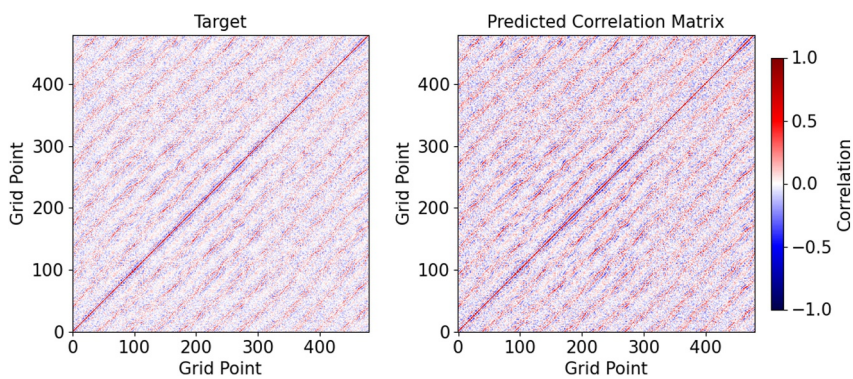


Figure 4. Target observation error correlation, estimated from 50 steps (left panel) and prediction by the MLOE (right panel).

19422466, 2026, 11, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025MS005702 by Katholische Universitat Eichstatt-Ingolstadt, Wiley Online Library on [24/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

is the modulation product of size $N_x \times \tilde{N}_e$ with $\tilde{N}_e = N_e \cdot N_m$ and it holds,

$$\widetilde{\mathbf{X}}_b \widetilde{\mathbf{X}}_b^T \approx \mathbf{P}_b, \tag{11}$$

$$\widetilde{\mathbf{X}}_b \mathbf{1} = 0, \tag{12}$$

where $\mathbf{1}$ is an $\tilde{N}_e \times 1$ vector of ones.

The transform matrix $\widetilde{\mathbf{T}} \in \mathbb{R}^{\tilde{N}_e \times \tilde{N}_e}$ and the analysis ensemble perturbation $\widetilde{\mathbf{X}}_a$ are given by

$$\widetilde{\mathbf{T}} = \left(\mathbf{I} + \widetilde{\mathbf{Y}}_b^T \mathbf{R}^{-1} \widetilde{\mathbf{Y}}_b\right)^{-\frac{1}{2}}, \tag{13}$$

$$\widetilde{\mathbf{X}}_a = \widetilde{\mathbf{X}}_b \widetilde{\mathbf{T}}, \tag{14}$$

where I is the identity matrix of size $\tilde{N}_e \times \tilde{N}_e, \widetilde{\mathbf{Y}}_b = \mathbf{H}\widetilde{\mathbf{X}}_b$ and $\mathbf{H}$ is the linearized version of the observation operator $\mathcal{H}$. Therefore, the update equations for analysis ensemble mean $\bar{\mathbf{x}}_a$ and ensemble members $\mathbf{X}_a$ are given by

$$\bar{\mathbf{x}}_a = \bar{\mathbf{x}}_b + \mathbf{K}(\mathbf{y} - \mathcal{H}(\bar{\mathbf{x}}_b)), \quad \text{where} \quad \mathbf{K} = \widetilde{\mathbf{X}}_b \left(\mathbf{I} + \widetilde{\mathbf{Y}}_b^T \mathbf{R}^{-1} \widetilde{\mathbf{Y}}_b\right)^{-1} \widetilde{\mathbf{Y}}_b^T \mathbf{R}^{-1}, \tag{15}$$

$$\mathbf{X}_a = \bar{\mathbf{x}}_a + \lambda\left(\mathcal{S}\left(\widetilde{\mathbf{X}}_a\right)\right), \tag{16}$$

where $\mathcal{S}$ randomly selects N_e from $\widetilde{\mathbf{X}}_a$ and λ is the inflation coefficient.

2.3.2. Cross-Correlation Between Forecast Error and Observation Error

To incorporate the cross-correlation between forecast error $\boldsymbol{\epsilon}_b$ and observation error $\boldsymbol{\epsilon}_o$ into the ETKF, Ohishi et al. (2025) introduced a reformulation assuming linear relation between forecast error and observation error. However, this assumption does not always hold in practice. The objective of the current work is to overcome this limitation by deriving a more general and flexible formulation for integration of cross-correlation without a prescribed error model. Although the derivation is conducted here within the AETKF framework, the approach is applicable to other ETKF-type algorithms.

As known, without allowing for correlation between $\boldsymbol{\epsilon}_b$ and $\boldsymbol{\epsilon}_o$, it holds

$$\mathbf{P}_a = (\mathbf{I} - \mathbf{KH})\,\mathbf{P}_b(\mathbf{I} - \mathbf{KH})^T + \mathbf{KRK}^T. \tag{17}$$

Now let $\mathbf{C} = \mathrm{Cov}(\boldsymbol{\epsilon}_b, \boldsymbol{\epsilon}_o)$ be the covariance between $\boldsymbol{\epsilon}_b$ and $\boldsymbol{\epsilon}_o$. At first, it holds (Ohishi et al., 2025):

$$\mathbf{P}_a = (\mathbf{I} - \mathbf{KH})\,\mathbf{P}_b + \mathbf{KC}^T, \tag{18}$$

$$\mathbf{K} = \left(\mathbf{P}_b\mathbf{H}^T - \mathbf{C}\right)\left(\mathbf{HP}_b\mathbf{H}^T + \mathbf{R} - \mathbf{HC} - \mathbf{C}^T\mathbf{H}^T\right)^{-1}, \tag{19}$$

Note that Equation 18 can also be written in symmetric form

$$\mathbf{P}_a = (\mathbf{I} - \mathbf{KH})\,\mathbf{P}_b(\mathbf{I} - \mathbf{KH})^T + \mathbf{KRK}^T + (\mathbf{I} - \mathbf{KH})\,\mathbf{CK}^T + \mathbf{KC}^T(\mathbf{I} - \mathbf{KH})^T. \tag{20}$$

In ensemble Kalman filter methods, the background error covariance is represented by the low-rank matrix $\widetilde{\mathbf{X}}_b\widetilde{\mathbf{X}}_b^T$, whose column space has dimension at most $\tilde{N}_e$ (typically much smaller than the full state dimension N_x). Consequently, any covariance that interacts with the background, including the forecast–observation cross-covariance $\mathbf{C}$, can only be reliably estimated and represented within this same subspace. Thus, $\mathbf{C}$ is projected onto the range of $\widetilde{\mathbf{X}}_b$ via

$$\mathbf{C} = \widetilde{\mathbf{X}}_b\left(\widetilde{\mathbf{X}}_b^{\dagger}\mathbf{C}\right) = \widetilde{\mathbf{X}}_b\mathbf{D}^T, \tag{21}$$

where $\widetilde{\mathbf{X}}_b^{\dagger}$ denotes the pseudo-inverse of $\widetilde{\mathbf{X}}_b$. It is acknowledged that the cross-covariance matrix $\mathbf{C}$ does not necessarily lies in the column space of $\widetilde{\mathbf{X}}_b$ and therefore Equation 21 should be understood as an approximation rather than an exact algebraic identity. This projection is a practical necessity for the derivation of the transform matrix $\widetilde{\mathbf{T}}$ and for ensuring computational tractability.

Replacing $\mathbf{C}$ in Equation 18 by Equation 21 leads to

$$\mathbf{K} = \left(\widetilde{\mathbf{X}}_b\widetilde{\mathbf{X}}_b^T\mathbf{H}^T - \widetilde{\mathbf{X}}_b\mathbf{D}^T\right)\left(\mathbf{H}\widetilde{\mathbf{X}}_b\widetilde{\mathbf{X}}_b^T\mathbf{H}^T + \mathbf{R} - \mathbf{H}\widetilde{\mathbf{X}}_b\mathbf{D}^T - \mathbf{D}\widetilde{\mathbf{X}}_b^T\mathbf{H}^T\right)^{-1} \tag{22}$$

$$= \left(\widetilde{\mathbf{X}}_b\,\widetilde{\mathbf{Y}}_b^T - \widetilde{\mathbf{X}}_b\mathbf{D}^T\right)\left(\widetilde{\mathbf{Y}}_b\,\widetilde{\mathbf{Y}}_b^T + \mathbf{R} - \widetilde{\mathbf{Y}}_b\mathbf{D}^T - \mathbf{D}\widetilde{\mathbf{Y}}_b^T\right)^{-1} \tag{23}$$

Since $\mathbf{P}_a = (\mathbf{I} - \mathbf{KH})\mathbf{P}_b + \mathbf{KC}^T = (\mathbf{I} - \mathbf{KH})\widetilde{\mathbf{X}}_b\widetilde{\mathbf{X}}_b^T + \mathbf{KD}\widetilde{\mathbf{X}}_b^T = \widetilde{\mathbf{X}}_b\widetilde{\mathbf{X}}_b^T - \mathbf{K}\left(\widetilde{\mathbf{Y}}_b - \mathbf{D}\right)\widetilde{\mathbf{X}}_b^T$ and $\mathbf{P}_a = \widetilde{\mathbf{X}}_b\widetilde{\mathbf{T}}\widetilde{\mathbf{T}}^T\widetilde{\mathbf{X}}_b^T$, it holds

$$(\mathbf{I} - \mathbf{KH})\widetilde{\mathbf{X}}_b\widetilde{\mathbf{X}}_b^T + \mathbf{KD}\widetilde{\mathbf{X}}_b^T = \widetilde{\mathbf{X}}_b\widetilde{\mathbf{T}}\widetilde{\mathbf{T}}^T\widetilde{\mathbf{X}}_b^T. \tag{24}$$

By right-multiplying both sides of Equation 24 with $\left(\widetilde{\mathbf{X}}_b^{\dagger}\right)^T$, it results in

$$\widetilde{\mathbf{X}}_b\widetilde{\mathbf{T}}\widetilde{\mathbf{T}}^T = \widetilde{\mathbf{X}}_b - \mathbf{K}\left(\widetilde{\mathbf{Y}}_b - \mathbf{D}\right) \tag{25}$$

$$= \widetilde{\mathbf{X}}_b - \left(\widetilde{\mathbf{X}}_b\,\widetilde{\mathbf{Y}}_b^T - \widetilde{\mathbf{X}}_b\mathbf{D}^T\right)\left(\widetilde{\mathbf{Y}}_b\,\widetilde{\mathbf{Y}}_b^T + \mathbf{R} - \widetilde{\mathbf{Y}}_b\mathbf{D}^T - \mathbf{D}\widetilde{\mathbf{Y}}_b^T\right)^{-1}\left(\widetilde{\mathbf{Y}}_b - \mathbf{D}\right), \tag{26}$$

consequently,

$$\widetilde{\mathbf{T}}\widetilde{\mathbf{T}}^T = \mathbf{I} - \left(\widetilde{\mathbf{Y}}_b^T - \mathbf{D}^T\right)\left(\widetilde{\mathbf{Y}}_b\,\widetilde{\mathbf{Y}}_b^T + \mathbf{R} - \widetilde{\mathbf{Y}}_b\mathbf{D}^T - \mathbf{D}\widetilde{\mathbf{Y}}_b^T\right)^{-1}\left(\widetilde{\mathbf{Y}}_b - \mathbf{D}\right), \tag{27}$$

and

$$\widetilde{\mathbf{T}} = \left(\mathbf{I} - \left(\widetilde{\mathbf{Y}}_b^T - \mathbf{D}^T\right)\left(\widetilde{\mathbf{Y}}_b\,\widetilde{\mathbf{Y}}_b^T + \mathbf{R} - \widetilde{\mathbf{Y}}_b\mathbf{D}^T - \mathbf{D}\widetilde{\mathbf{Y}}_b^T\right)^{-1}\left(\widetilde{\mathbf{Y}}_b - \mathbf{D}\right)\right)^{\frac{1}{2}}, \tag{28}$$

Note that in case of no cross-correlation (i.e., $\mathbf{C} = 0$, hence $\mathbf{D} = 0$), Equation 28 becomes

$$\widetilde{\mathbf{T}} = \left(\mathbf{I} - \widetilde{\mathbf{Y}}_b^T\left(\widetilde{\mathbf{Y}}_b\,\widetilde{\mathbf{Y}}_b^T + \mathbf{R}\right)^{-1}\widetilde{\mathbf{Y}}_b\right)^{\frac{1}{2}} = \left(\mathbf{I} + \widetilde{\mathbf{Y}}_b^T\mathbf{R}^{-1}\widetilde{\mathbf{Y}}_b\right)^{-\frac{1}{2}}, \tag{29}$$

which is in line with Equation 13.

The refined formulation incorporating $\mathbf{C}$ is named hereafter "CorrFOE."

3. Experimental Setup and Results

3.1. Experimental Setup

In the following, a series of twin experiments are performed (see Table 1). To spin-up, the model Equation 1 is run at first for 3,000 time steps. During DA cycles, first, the truth denoted by $\mathbf{z}_t$ is considered as observations (i.e., the

19422466, 2026, 11, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025MS005702 by Katholische Universitat Eichstatt-Ingolstadt, Wiley Online Library on [24/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

19422466, 2026, 11, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025MS005702 by Katholische Universitat Eichstatt-Ingolstadt, Wiley Online Library on [24/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 1
Experimental Setups

EXP	SAN	MLME	MLOE	CorrFOE
E_Control	Off	Off	Off	Off
E_Control_0.08	Off	Off	Off, IR = 0.08	Off
E_Control_0.16	Off	Off	Off, IR = 0.16	Off
E_Control_0.32	Off	Off	Off, IR = 0.32	Off
E_Control_50%obs	Off	Off	Off, OD = 50%	Off
E_Control_25%obs	Off	Off	Off, OD = 25%	Off
E_SAN	On	Off	Off	Off
E_MLME	Off	On	Off	Off
E_MLOE	Off	Off	51-st	Off
E_MLOE_0.08	Off	Off	51-st, IR = 0.08	Off
E_MLOE_0.16	Off	Off	51-st, IR = 0.16	Off
E_MLOE_0.32	Off	Off	51-st, IR = 0.16	Off
E_MLOE_50%obs	Off	Off	51-st, OD = 50%	Off
E_MLOE_25%obs	Off	Off	51-st, OD = 25%	Off
E_MLMOE	Off	On	51-st	Off
E_CorrFOE_T	Off	On	51-st	101-st, T
E_CorrFOE_A	Off	On	51-st	101-st, A

Note. Four schemes are available. "Off" and "on" mean that the scheme is switched off and on, respectively. The ordinal number means from which step the scheme is switched on. Note that due to the randomness of the model error sampling in E_SAN, five runs are conducted for it. "IR = number" indicates that the instrumental error is taken into account with its variance equal to the given number. "OD = number%" indicates that the given percentage of the 480 grid points is observed at each assimilation step; experiments without an OD label use 100% observations. "T" ("A") means using truth (analysis) for calculation of $\mathbf{C}$.

truth is not perturbed by random noise to create observations, and, unless otherwise specified, every point is observed) and the x-component of the hybrid-scale state $\mathbf{z}$ is treated as the background state denoted by $\mathbf{x}_b$. In all DA experiments, the assimilation window is 30 time steps, 200 assimilation steps are conducted (i.e., $N_k = 200$), the Gaspari-Cohn function (Gaspari & Cohn, 1999) is applied to the localization with a radius of 10 grid points (after tuning) and an adaptive multiplicative inflation scheme (Anderson, 2009) is employed to calculate λ. Unless otherwise specified, the ensemble size is 40.

The model error representation is in the manner of additive noise, that is, for each background ensemble member

$$\mathbf{x}_{b(i)} \leftarrow \mathbf{x}_{b(i)} + \gamma \boldsymbol{\eta}_{(i)}, \tag{30}$$

where γ is a tunable coefficient and $\boldsymbol{\eta}_{(i)}$ is a sample for model error. Two additive noise schemes are implemented: As (Zeng et al., 2019, 2020), the SAN and the MLME. For each ensemble member $\mathbf{x}_{b(i)}$, the SAN randomly samples $\boldsymbol{\eta}_{(i)}$ from the model error data set (the same data set as used for training of the MLME), while the MLME predicts the appropriate perturbation $\boldsymbol{\eta}_{(i)}$ based on the input $\mathbf{x}_{b(i)}$. γ is set to be 1 in this study, since model error stems solely from unresolved small scales, consequently, inflating $\boldsymbol{\eta}$ is unwarranted. To compare performance of different schemes for model error representation, three experiments E_Control, E_SAN and E_MLME are run, in which the assigned observation error ($\mathbf{R} = 0.8 \cdot \mathbf{I}$) is a priori estimate of the D05 (not shown). It is noted that due to the randomness of the model error sampling in E_SAN, five runs are conducted for it. In addition, to test the sensitivity to the ensemble size, experiments with 20, 40, 60, 80, and 100 are conducted.

To isolate observation-error representation, E_MLOE retains the model-error treatment of E_Control rather than adding the explicit perturbations used in E_MLME. After the first 50 assimilation steps, the MLOE GAN replaces the prescribed diagonal observation-error covariance with a full covariance estimated from the preceding 50 observation batches. The sensitivity to sparsity of observations is examined through two additional sparsity levels. In E_Control_50%obs and E_MLOE_50%obs, every second grid point is observed; likewise, in

19422466, 2026, 11, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025MS005702 by Katholische Universitat Eichsatt-Ingolstadt, Wiley Online Library on [24/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

E_Control_25%obs and E_MLOE_25%obs, every fourth grid point is observed. In E_Control_0.08 and E_MLOE_0.08, the truth is then perturbed by Gaussian noise with a variance of 0.08 to generate observations; likewise, E_Control_0.16 and E_MLOE_0.16 use a noise variance of 0.16, and E_Control_0.32 and E_MLOE_0.32 use a noise variance of 0.32. Thus, the total observation error in these cases includes both representation error and uncorrelated instrumental error. For the control runs under instrumental error, $\mathbf{R}$ is prescribed as a fixed diagonal matrix with values estimated offline via the D05 (specifically, $\mathbf{R} = 0.82 \cdot \mathbf{I}$ for the 0.08 case and $\mathbf{R} = 0.85 \cdot \mathbf{I}$ for the 0.16 case). In contrast, E_MLOE_0.08 and E_MLOE_0.16 estimated $\mathbf{R}$ adaptively online, following the same procedure as E_MLOE.

Finally, E_MLMOE combines the MLME model-error perturbations with the MLOE-estimated observation-error covariance. Building on this combined setup, E_CorrFOE_T and E_CorrFOE_A additionally account for forecast–observation error cross-correlation estimated from truth and analyses, respectively. In E_CorrFOE_T, $\mathbf{C} = \frac{\sum_{k=1}^{100}\left(\bar{\mathbf{x}}_b^k - \mathbf{z}_t^k\right)\left(\mathbf{x}_t^k - \mathbf{z}_t^k\right)^T}{99}$ is estimated, where $\bar{\mathbf{x}}_b^k$, $\mathbf{z}_t^k$ and $\mathbf{x}_t^k$ are the ensemble mean of the background state, the truth and its x-component at the k-th assimilation step, respectively, that is, E_CorrFOE_T follows the E_Control setup for the first 100 steps, after which the CorrFOE is activated to determine $\mathbf{C}$ based on $\bar{\mathbf{x}}_b^k$, $\mathbf{z}_t^k$ and $\mathbf{x}_t^k$ from the preceding 100 steps. It is noted that a localization with radius of three grid points is applied to $\mathbf{C}$ due to the limited size of samples, otherwise, DA becomes unstable (not shown). In realistic applications, the true state $\mathbf{z}_t^k$ is unavailable. A higher-resolution model (Zeng et al., 2019) or analyses can be used as a proxy for truth. Therefore, an additional experiment E_CorrFOE_A employs $\bar{\mathbf{x}}_a^k$ to replace $\mathbf{z}_t^k$ for calculating $\mathbf{C}$.

3.2. Results

For verification, the analyses of different experiments are compared in terms of

$$\mathrm{RMSE} = \frac{1}{N_k}\sum_{k=1}^{N_k}\sqrt{\frac{1}{N_x}\sum_{j=1}^{N_x}\left(\mathbf{z}_t^{j,k} - \bar{\mathbf{x}}_a^{j,k}\right)^2} \tag{31}$$

$$\mathrm{Spread} = \frac{1}{N_k}\sum_{k=1}^{N_k}\sqrt{\frac{1}{N_x}\sum_{j=1}^{N_x}\left[\frac{1}{N_e - 1}\sum_{i=1}^{N_e}\left(\mathbf{x}_{a(i)}^{j,k} - \bar{\mathbf{x}}_a^{j,k}\right)^2\right]}, \tag{32}$$

$$\mathrm{SSR} = \frac{\mathrm{Spread}}{\mathrm{RMSE}}, \tag{33}$$

where N_k denotes the total number of assimilation cycles. An SSR value closer to unity indicates better ensemble calibration; values below and above unity indicate under-dispersion and over-dispersion, respectively. In addition, the energy spectrum is used to evaluate the capability in scale reconstruction.

Figure 5 compares the performance of E_Control, E_SAN, and E_MLME in representing model errors. Due to the randomness in model-error sampling inherent to E_SAN, five independent experiments were conducted, with the average results shown here. It is noted that the RMSE, spread, and SSR of E_SAN are averaged across five runs, whereas the spectrum in the main figure is from one individual, unaveraged E_SAN realization; a Figures S1–S3 in Supporting Information S1 compares this individual realization with the mean spectrum across the five runs, and both show weaker small-scale representation relative to E_MLME, indicating that this difference is not caused by averaging across runs. All experiments demonstrate strong reliability, with SSR values approaching unity. In terms of accuracy, E_MLME achieves the smallest RMSE (=0.322), outperforming both E_SAN (=0.442) and E_Control (=0.481). The result for E_SAN, which also surpasses E_Control, aligns with the finding reported in Zeng et al. (2019) that the SAN method effectively enhances analysis accuracy. Energy spectrum analysis shows that while all experiments match the reference at larger scales, E_Control exhibits substantial energy deficits at smaller scales. In contrast, E_SAN and E_MLME remain much closer to the reference spectrum, showing only slight underestimation. This is more evident in Figure 6. For the large-scale component, all experiments produce patterns broadly similar to the truth, indicating that the large-scale dynamics are well captured regardless of the error representation method in this case. However, for the small-scale component—where capturing nuanced, high-wavenumber structures is more challenging—E_MLME stands

19422466, 2026, 11, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025MS005702 by Katholische Universitat Eichstatt-Ingolstadt, Wiley Online Library on [24/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

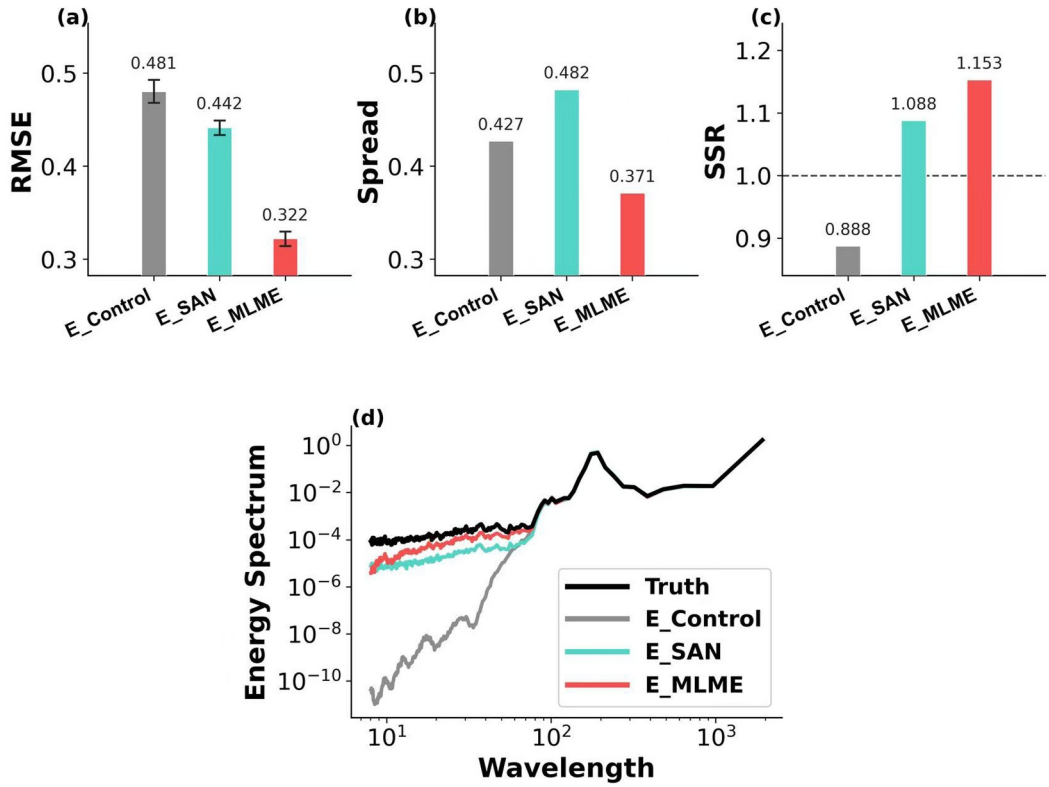


Figure 5. The RMSE, spread, and spread-skill ratio (SSR; upper panels) and energy spectra of truth and analyses (lower panel) for E_Control, E_SAN, and E_MLME. 95% confidence intervals for the RMSE are denoted. Results are averaged over the last 100 assimilation steps.

out as the closest to the truth, successfully reproducing fine-scale features with notable fidelity. E_SAN ranks next, showing reasonable but less accurate small-scale patterns, while E_Control deviates most significantly from the reference, underscoring its limitations in resolving small-scale variability. Moreover, Figure 7 presents the RMSE for all experiments across different ensemble sizes. The RMSE generally decreases as ensemble size increases, with E_MLME performing substantially better than E_SAN, followed by E_Control.

For observation-error representation, Figure 8 compares E_Control and E_MLOE under 100%, 50%, and 25% observation densities. Under 100% observations, E_MLOE reduced the mean RMSE from 0.481 to 0.430 (10.5%), but also reduced the SSR from 0.888 to 0.712, indicating greater under-dispersion. Reducing the observation density increased the RMSE for both methods. Nevertheless, E_MLOE retained lower RMSEs than the corresponding control experiments: 0.472 versus 0.536 at 50% observations and 0.536 versus 0.608 at 25% observations, corresponding to reductions of approximately 12% in both cases. The associated SSRs were 0.779 and 0.839 for E_MLOE, compared with 0.934 and 0.973 for the respective control experiments. Thus, within these sparse-observation experiments, adaptive estimation of $\mathbf{R}$ was associated with lower mean analysis error, while the lower SSR values indicate that the ensemble remained under-dispersive. The spectra are broadly similar at the resolved large scales; therefore, the present comparison does not support a strong claim of improved small-scale reconstruction. Figure 9 separately examines instrumental-error sensitivity under 100% observations. When instrumental error was introduced, the RMSE increased with the added noise magnitude, whereas E_MLOE_0.08, E_MLOE_0.16, and E_MLOE_0.32 retained lower RMSEs than their corresponding control experiments.

Finally, both model error and observation error representation as well as cross correlation between forecast error and observation error are accounted for within assimilation cycles. As shown in Figure 10, the RMSE of E_MLMOE (=0.260) is approximately 45.7% smaller than that of E_Control (=0.481). The RMSE of E_Corr-FOE_T (=0.243) is even smaller than that of E_MLMOE. It is also observed that E_CorrFOE_T produces a much larger spread (0.435, with SSR $\approx$ 1.79). This may be because the additional term $(\mathbf{I}-\mathbf{KH})\mathbf{CK}^T+$

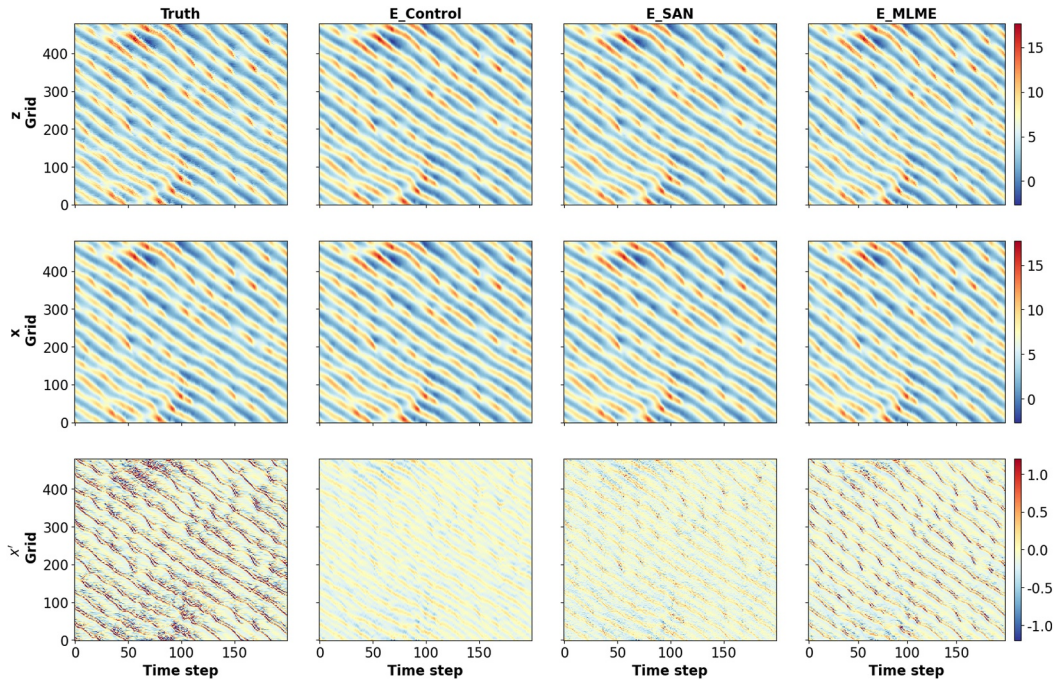


Figure 6. Hovmöller diagrams of truth and the analyses from the E_Control, E_SAN and E_MLME, presenting **z** and its decomposition into components **x** and **x′**.

$\mathbf{KC}^T(\mathbf{I}-\mathbf{KH})^T$ in Equation 29 turns to contribute to variance. Moreover, when the cross-covariance **C** is estimated using the true state, the resulting correlations are strong and idealized because the truth is perfect (free of representation and observation errors) and highly correlated with the background, which leads to a considerable injection of variance, artificially inflating the ensemble spread (the SSR considerably larger than unity). In $\mathrm{E}_{\mathrm{CorrFOE_A}}$, the analysis state (which already contains errors) is used instead. This may provide an implicit damping or smoothing of the cross-covariance, reducing the magnitude of the injected variance and bringing the SSR to 1.204. This indicates that using analyses to estimate the cross-correlation is feasible choice in practice. Furthermore, all three methods—E_MLMOE, E_CorrFOE_T, and E_CorrFOE_A also exhibit spectra that are closer to the truth at smaller scales. This indicates that combining the MLME and the MLOE leads to complementary and synergistic improvements in both overall accuracy and spectral fidelity.

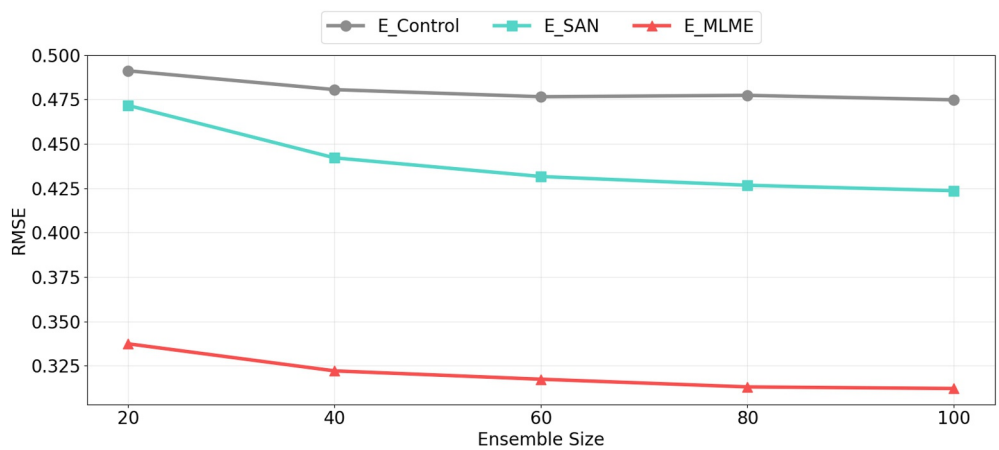


Figure 7. The RMSE of E_Control, E_SAN and E_MLME for different ensemble sizes 20, 40, 60, 80, 100. Results are averaged over the last 100 assimilation steps.

19422466, 2026, 11, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025MS005702 by Katholische Universitat Eichstatt-Ingolstadt, Wiley Online Library on [24/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

19422466, 2026, 11, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025MS005702 by Katholische Universitat Eichstatt-Ingolstadt, Wiley Online Library on [24/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

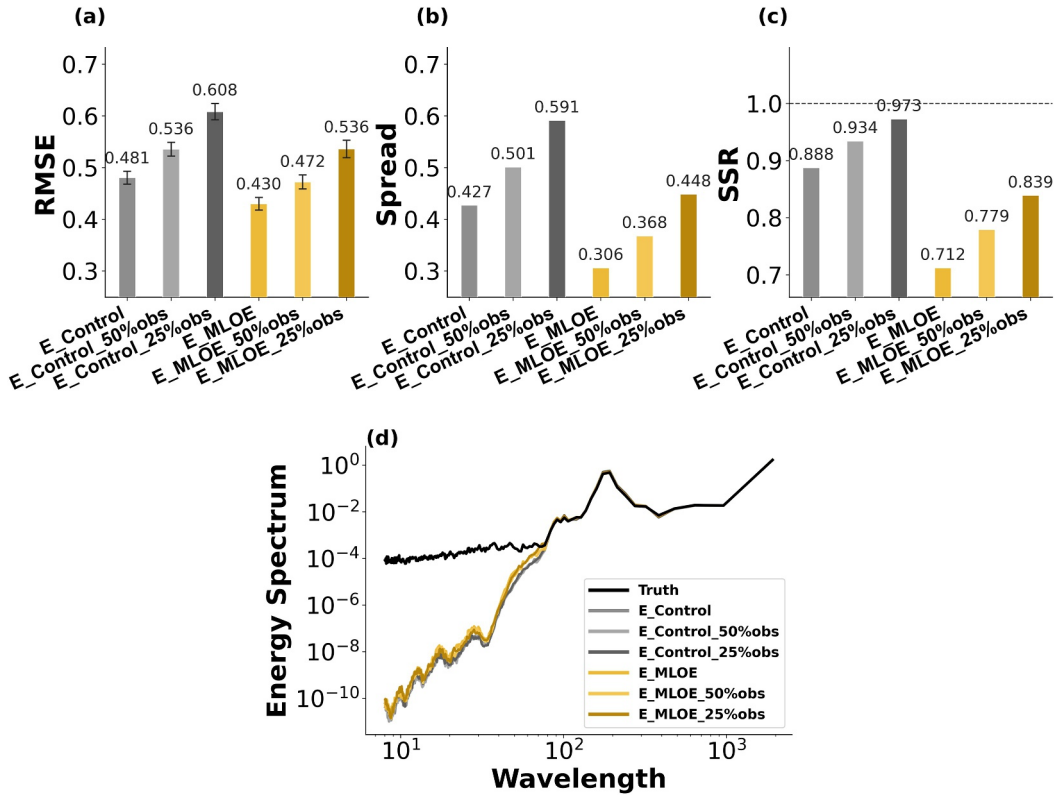


Figure 8. The RMSE, spread, and spread-skill ratio (SSR; upper panels) and energy spectra of truth and analyses (lower panel) for E_Control and E_MLOE under 100%, 50%, and 25% observation densities. 95% confidence intervals for the RMSE are denoted. Results are averaged over the last 100 assimilation steps.

Table 2 lists the computational times for different experiments. While all advanced methods introduce some overhead, the increase in computational times remain modest—under 7% in total runtime. E_SAN adds only 0.4%, whereas E_MLME and E_MLOE increase the computational time by 4.1% and 3.2%, respectively. The most expensive configurations, E_MLMOE and E_CorrFOE_T, require about 6.6%–6.7% more time than E_Control. These results confirm that the proposed techniques are computationally feasible, with only a mild penalty relative to the control.

4. Conclusion and Discussion

This study has introduced a novel, integrated ML framework to enhance error treatment in ensemble DA, with a focused investigation on the independent representation of model error and observation error, and the explicit inclusion of their often-neglected cross-correlation. Through a rigorous series of twin experiments with the controlled, two-scale Lorenz III model, we have demonstrated that ML-based error representation schemes outperform traditional approaches. The GAN-based Model Error scheme (MLME) not only achieves the lowest RMSE but also successfully recovers small-scale structures, providing physically consistent, flow-dependent corrections. The MLOE generates a stable observation error covariance matrix across assimilation cycles, leading to improved analysis accuracy and slight benefits in small-scale reconstruction. Most crucially, we derived a generalized formulation (CorrFOE) to incorporate the cross-correlation between forecast and observation errors within the AETKF framework without restrictive linearity assumptions. The synergistic combination of both ML schemes with the explicit inclusion of this cross-correlation yields the most accurate analyses, achieving the lowest RMSE and best spectral fidelity. The added computational overhead for these advanced methods is modest (under 7%), confirming their operational potential within efficient DA cycles.

The promising results from this idealized, low-dimensional model establish a critical proof-of-concept and provide a validated methodological foundation for several essential future directions. The primary and most immediate challenge is the transition to high-dimensional, operational systems. While the Lorenz III model (480

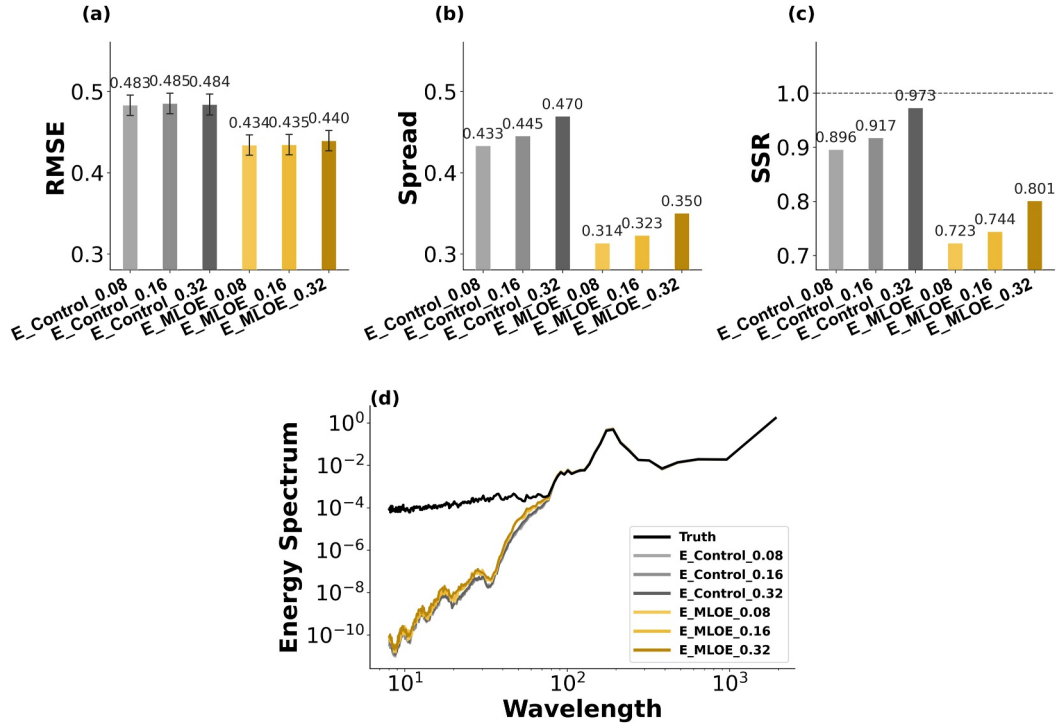


Figure 9. The same as Figure 5 but for the comparison of E_Control_0.08, E_Control_0.16 and E_Control_0.32, E_ML_OE_0.08, E_ML_OE_0.16 and E_ML_OE_0.32 from the 51-st assimilation step.

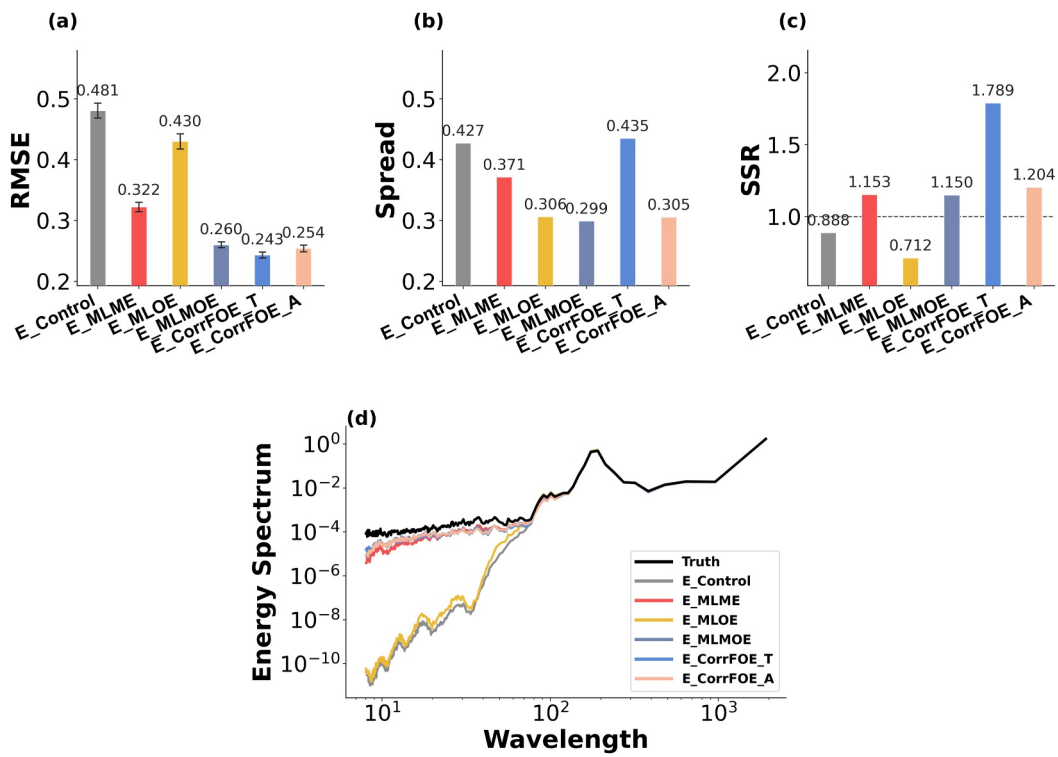


Figure 10. The same as Figure 5 but for the comparison of E_Control, E_MLME, E_MLOE, E_MLMOE, E_CorrFOE_T and E_CorrFOE_A.

19422466, 2026, 11, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025MS005702 by Katholische Universitat Eichstatt-Ingolstadt, Wiley Online Library on [24/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

19422466, 2026, 11, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025MS005702 by Katholische Universitat Eichstatt-Ingolstadt, Wiley Online Library on [24/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 2
Average and Total Computational Time of Different Experiments As Well As Their Relative Increases to E_Control

Name	Avg time [s]	Total time [s]	Relative increase [%]
E_Control	7.89	1578.33	0.0%
E_SAN	7.92	1584.17	0.4%
E_MLME	8.23	1646.07	4.1%
E_MLOE	8.14	1628.67	3.2%
E_MLMOE	8.41	1682.33	6.6%
E_CorrFOE_T	8.42	1684.22	6.7%

grid points) is invaluable for algorithmic development and validation, operational numerical weather prediction models operate on $\mathcal{O}(10^9)$ grid points with vastly more complex error structures. Our framework is designed with this scalability in mind: the AETKF operates in the ensemble subspace and employs modulation-based localization, a technique already proven in operational centers like Environment Canada. Furthermore, the convolutional neural network (CNN) architectures of the MLME and MLOE exploit local spatial correlations and parameter sharing, meaning their computational cost does not scale quadratically with resolution. The logical next step involves applying this integrated framework to intermediate-complexity models, such as quasi-geostrophic models or limited-area convection-permitting models with 10^4 to 10^6 grid points. This will test the scalability of training data generation, the stability of online ML inference, and the practical estimation of cross-correlations using proxies like high-resolution model runs or analysis fields, as preliminarily demonstrated by our experiment. In addition, the latent space data assimilation can be regarded as a viable and practical pathway to address the curse of dimensionality (Cheng et al., 2023; Melinc & Zaplotnik, 2024). Instead of operating directly on the full high-dimensional state, the system's dynamics and error statistics can be projected onto a low-dimensional latent space, learned for instance by an autoencoder or a physics-informed dimensionality reduction technique. Within this latent space, the proposed MLME, MLOE, and CorrFOE schemes can be efficiently implemented, as the latent representation preserves the essential structural and statistical information while significantly reducing computational cost. For practical generalization of the new schemes, their first rigorous application and testing will be focused on a specific, well-defined context: convective-scale events during the summer season using a convective-allowing model such as Zeng et al. (2019, 2020) This confines the problem to a regime where model errors related to moist convection and boundary layer processes are dominant. The ML models will be trained to capture a wide variety of summer convective scenarios. Subsequent work will explore the generalization and adaptability of the approach across different weather regimes and seasons.

Regarding the ML architecture, while GANs were selected for this study due to their computational efficiency, single-step inference, and established success in learning geophysical patterns, the rapid advancement of generative models cannot be ignored. Diffusion models have recently gained prominence for their superior sample quality and training stability. A valuable future exploration would be to adapt our framework to use a diffusion model as the core generator. This would involve addressing the associated challenges, such as the higher computational cost of iterative denoising during inference and ensuring spatiotemporal consistency in the generated error fields. A comparative study between GAN and diffusion-based implementations would provide crucial insights into the trade-offs between sample fidelity, computational cost, and operational practicality for this specific application.

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Availability Statement

The source code of data assimilation system and machine learning as well as the experimental scripts and configuration files are publicly available from Chen and Zeng (2025).

Acknowledgments
This research was primarily supported by the Foundation of Beiji Ge (BJG202303). Thanks are also given to the National Natural Science Foundation of China (Grant 42275159) and the National Key Research and Development Program of China (Grant 2023YFF0804703). The work of Yuxuan Feng is supported by Jiangsu Meteorological Bureau (KQ202305). We thank the reviewers for their constructive comments, which helped improve the manuscript.

References

Anderson, J. L. (2001). An ensemble adjustment Kalman filter for data assimilation. *Monthly Weather Review*, *129*(12), 2884–2903. https://doi.org/10.1175/1520-0493(2001)129<2884:aeakff>2.0.co;2

Anderson, J. L. (2009). Spatially and temporally varing adaptive covariance inflation for ensemble filters. *Tellus*, *61A*, 72–83.

Anthes, R., & Rieckh, T. (2018). Estimating observation and model error variances using multiple data sets. *Atmospheric Measurement Techniques*, *11*(7), 4239–4260. https://doi.org/10.5194/amt-11-4239-2018

Bishop, C. H., Etherton, B. J., & Majumdar, S. J. (2001). Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. *Monthly Weather Review*, *129*(3), 420–436. https://doi.org/10.1175/1520-0493(2001)129<0420:aswtet>2.0.co;2

Bonavita, M., & Laloyaux, P. (2020). Machine learning for model error inference and correction. *Journal of Advances in Modeling Earth Systems*, *12*, e2020MS002232. https://doi.org/10.1029/2020ms002232

Brajard, J., Carrassi, A., Bocquet, M., & Bertino, L. (2021). Combining data assimilation and machine learning to infer unresolved scale parametrization. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *379*(2194), 20200086. https://doi.org/10.1098/rsta.2020.0086

Buehner, M. (2005). Ensemble-derived stationary and flow-dependent background-error covariances: Evaluation in a quasi-operational NWP setting. *Quarterly Journal of the Royal Meteorological Society*, *131*(607), 1013–1043. https://doi.org/10.1256/qj.04.15

Chapman, W., & Berner, J. (2025). Improving climate bias and variability via CNN-based state-dependent model-error corrections. *Geophysical Research Letters*, *52*(6), e2024GL114106. https://doi.org/10.1029/2024gl114106

Chen, Y., & Zeng, Y. (2025). Machine learning-based representation of model error and observation error with estimated cross-correlation in data assimilation [Software]. https://doi.org/10.5281/zenodo.17035869

Cheng, S., Chen, J., Anastasiou, C., Angeli, P., Matar, O., Guo, Y., et al. (2023). Generalised latent assimilation in heterogeneous reduced spaces with machine learning surrogate models. *Journal of Scientific Computing*, *94*(1), 11. https://doi.org/10.1007/s10915-022-02059-4

Cheng, S., & Qiu, M. (2022). Observation error covariance specification in dynamical systems for data assimilation using recurrent neural networks. *Neural Computing & Applications*, *34*(16), 13149–13167. https://doi.org/10.1007/s00521-021-06739-4

Desroziers, G., Berre, L., Chapnik, B., & Poli, P. (2005). Diagnosis of observation, background and analysis-error statistics in observation space. *Quarterly Journal of the Royal Meteorological Society*, *131*(613), 3385–3396. https://doi.org/10.1256/qj.05.108

Dowell, D. C., & Wicker, L. J. (2009). Additive noise for storm-scale ensemble data assimilation. *Monthly Weather Review*, *26*(5), 911–927. https://doi.org/10.1175/2008jtecha1156.1

Dowell, D. C., Zhang, F., Wicker, L. J., Snyder, C., & Crook, N. A. (2004). Wind and temperature retrievals in the 17 May 1981 Arcadia, Oklahoma, supercell: Ensemble Kalman filter experiments. *Monthly Weather Review*, *132*(8), 1982–2005. https://doi.org/10.1175/1520-0493(2004)132<1982:watrit>2.0.co;2

Farchi, A., & Bocquet, M. (2019). On the efficiency of covariance localisation of the ensemble Kalman filter using augmented ensembles. *Frontiers in Applied Mathematics and Statistics*, *5*, 3. https://doi.org/10.3389/fams.2019.00003

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

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al. (2014). Generative adversarial nets. *Advances in Neural Information Processing Systems*, *27*.

Gregory, W., Bushuk, M., Zhang, Y., Adcroft, A., & Zanna, L. (2024). Machine learning for online sea ice bias correction within global ice-ocean simulations. *Geophysical Research Letters*, *51*(3), e2023GL106776. https://doi.org/10.1029/2023gl106776

Gregory, W., Bushuk, M., Zhang, Y., Adcroft, A., Zanna, L., McHugh, C., & Jia, L. (2026). Advancing global sea ice prediction capabilities using a fully coupled climate model with integrated machine learning. *Science Advances*, *12*(1), eady8957. https://doi.org/10.1126/sciadv.ady8957

Hu, C., van Leeuwen, P., & Geer, A. (2024). A non-parametric way to estimate observation errors based on ensemble innovations. *Quarterly Journal of the Royal Meteorological Society*, *150*(761), 2296–2315. https://doi.org/10.1002/qj.4710

Isola, P., Zhu, J.-Y., Zhou, T., & Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1125–1134).

Janjic, T., Bormann, N., Bocquet, M., Carton, J. A., Cohn, S. E., Dance, S. L., et al. (2018). On the representation error in data assimilation. *Quarterly Journal of the Royal Meteorological Society*, *144*(713), 1257–1278. https://doi.org/10.1002/qj.3130

Janjic, T., & Cohn, S. E. (2006). Treatment of observation error due to unresolved scales in atmospheric data assimilation. *Monthly Weather Review*, *134*(10), 2900–2915. https://doi.org/10.1175/mwr3229.1

Legler, S., & Janjic, T. (2022). Combining data assimilation and machine learning to estimate parameters of a convective-scale model. *Quarterly Journal of the Royal Meteorological Society*, *148*(743), 860–874. https://doi.org/10.1002/qj.4235

Lorenz, E. (2005). Designing chaotic models. *Journal of the Atmospheric Sciences*, *62*(5), 1574–1587. https://doi.org/10.1175/jas3430.1

Lu, F. (2025). U-net Kalman filter (UNETKF): An example of machine learning-assisted data assimilation. *Journal of Advances in Modeling Earth Systems*, *17*(4), e2023MS003979. https://doi.org/10.1029/2023ms003979

Manshausen, P., Cohen, Y., Harrington, P., Pathak, J., Pritchard, M., Garg, P., et al. (2025). Generative data assimilation of sparse weather station observations at kilometer scales. *Journal of Advances in Modeling Earth Systems*, *17*(10), e2024MS004505. https://doi.org/10.1029/2024ms004505

Melinc, B., & Zaplotnik, Z. (2024). 3d-var data assimilation using a variational autoencoder. *Quarterly Journal of the Royal Meteorological Society*, *150*(761), 2273–2295. https://doi.org/10.1002/qj.4708

Ohishi, S., Kobayashi, Y., & Miyoshi, T. (2025). Including cross correlation between forecast and observation errors in an ensemble Kalman filter. *Monthly Weather Review*, *153*(6), 1035–1043. https://doi.org/10.1175/mwr-d-25-0016.1

Peng, Z., Lei, L., & Tan, Z. (2024). A hybrid deep learning and data assimilation method for model error estimation. *Science China Earth Sciences*, *67*(12), 3655–3670. https://doi.org/10.1007/s11430-024-1395-7

Snyder, C., & Zhang, F. (2003). Assimilation of simulated Doppler radar observations with an ensemble Kalman filter. *Monthly Weather Review*, *131*(8), 1663–1677. https://doi.org/10.1175/2555.1

Sobash, R., & Wicker, L. (2015). On the impact of additive noise in storm-scale EnKF experiments. *Monthly Weather Review*, *143*(8), 3067–3086. https://doi.org/10.1175/mwr-d-14-00323.1

Sommer, M., & Janjić, T. (2018). A flexible additive inflation scheme for treating model error in ensemble Kalman filters. *Quarterly Journal of the Royal Meteorological Society*, *144*(716), 2026–2037. https://doi.org/10.1002/qj.3254

Waller, J., Dance, S. L., & Lean, H. (2021). Evaluating errors due to unresolved scales in convection permitting numerical weather prediction. *Quarterly Journal of the Royal Meteorological Society*, *147*(738), 2657–2669. https://doi.org/10.1002/qj.4043

19422466, 2026, 11, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025MS005702 by Katholische Universitat Eichstatt-Ingolstadt, Wiley Online Library on [24/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Walsworth, A., Poterjoy, J., & Satterfield, E. (2023). Challenges for inline observation error estimation in the presence of misspecified background uncertainty. *Monthly Weather Review*, *151*(9), 2295–2306. https://doi.org/10.1175/mwr-d-22-0298.1

Whitaker, J. S., & Hamill, T. M. (2012). Evaluating methods to account for system errors in ensemble data assimilation. *Monthly Weather Review*, *140*(9), 3078–3089. https://doi.org/10.1175/mwr-d-11-00276.1

Zeng, Y., Janjić, T., Sommer, M., de Lozar, A., Blahak, U., & Seifert, A. (2019). Representation of model error in convective-scale data assimilation: Additive noise based on model truncation error. *Journal of Advances in Modeling Earth Systems*, *11*(3), 752–770. https://doi.org/10.1029/2018ms001546

Zeng, Y., Janjić, T., de Lozar, A., Blahak, U., Reich, H., Keil, C., & Seifert, A. (2018). Representation of model error in convective-scale data assimilation: Additive noise, relaxation methods and combinations. *Journal of Advances in Modeling Earth Systems*, *10*(11), 2889–2911. https://doi.org/10.1029/2018ms001375

Zeng, Y., Janjić, T., de Lozar, A., Rasp, S., Blahak, U., Seifert, A., & Craig, G. C. (2020). Comparison of methods accounting for subgrid-scale model error in convective-scale data assimilation. *Monthly Weather Review*, *148*(6), 2457–2477. https://doi.org/10.1175/mwr-d-19-0064.1

Zeng, Y., Janjic, T., Feng, Y., Blahak, U., de Lozar, A., Bauernschubert, E., et al. (2021). Interpreting estimated observation error statistics of weather radar measurements using the ICON-LAM-KENDA system. *Atmospheric Measurement Techniques*, *14*(8), 5735–5756. https://doi.org/10.5194/amt-14-5735-2021

Zhang, F., Snyder, C., & Sun, J. (2004). Impacts of initial estimate and observation availability on convective-scale data assimilation with an ensemble Kalman filter. *Monthly Weather Review*, *132*(5), 1238–1253. https://doi.org/10.1175/1520-0493(2004)132<1238:ioieao>2.0.co;2

19422466, 2026, 11, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025MS005702 by Katholische Universitat Eichstatt-Ingolstadt, Wiley Online Library on [24/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License