



# 1    **A Unified Framework for Deep Learning-Based**

## 2    **Ensemble-Free Data Assimilation and Forecast Error**

### 3    **Correction**

4    Xinyue Li<sup>1</sup>, Youmin Tang<sup>1\*</sup>, Zheqi Shen<sup>1</sup>, Yi Li<sup>1</sup>, Zikang He<sup>2</sup>

5    <sup>1</sup>College of Oceanography, Hohai University, Nanjing, 210098, China

6    <sup>2</sup>State Key Laboratory of Satellite Ocean Environment Dynamics, Second Institute of Oceanography, Ministry  
7    of Natural Resources, Hangzhou, 310012, China

8    \*Correspondence to: Youmin Tang ([tangy@hhu.edu.cn](mailto:tangy@hhu.edu.cn))

9    **Abstract.** Filter-based data assimilation, particularly the ensemble Kalman filter (EnKF), has been widely used  
10    for forecasting across a broad range of scales. However, maintaining and updating multiple ensemble members  
11    incurs substantial computational costs, and accumulated forecast errors progressively degrade forecast skill. To  
12    address these challenges, we use the Lorenz-63 system as a testbed and develop a unified deep learning  
13    framework that integrates ensemble-free data assimilation emulation — based on a data-assimilation multilayer  
14    perceptron (DAMLP) — with forecast-error correction.

15    The experiments show that DAMLP substantially improves computational efficiency while producing analysis  
16    states with accuracy comparable to, or better than, that of EnKF. The periodic correction strategy effectively  
17    extends the useful forecast range, and its combination with a more accurate initial state yields a clear synergistic  
18    improvement, further slowing the growth of forecast errors. In addition, local Lyapunov exponents (LEs) are  
19    introduced to stratify the results according to local dynamical instability. These results demonstrate the potential  
20    of combining learned assimilation emulation with observation-free forecast-error correction for efficient  
21    forecasting of chaotic dynamical systems.

### 22    **3    Introduction**

23    The skill of numerical forecasting is strongly constrained by errors in the initial conditions. These errors, arising  
24    from factors such as sparse observations, observational noise, and imperfect assimilation analyses, can be rapidly  
25    amplified with model integration, thereby severely limiting the useful forecast range. Data assimilation (DA) is  
26    an important means of reducing initial-condition errors (Carrassi et al., 2018) by combining model forecasts with



27 observations to obtain an improved estimate of the system state. Among traditional DA methods, variational  
28 approaches formulate the analysis as an optimization problem and have played a central role in numerical weather  
29 prediction (Le Dimet and Talagrand, 1986; Courtier et al., 1994). The ensemble Kalman filter (EnKF), in contrast,  
30 uses a finite ensemble to approximate the background-error covariance and thereby represents flow-dependent  
31 uncertainty (Evensen, 1994; Houtekamer and Mitchell, 1998). Modern operational DA systems increasingly  
32 employ variational, ensemble-based, and hybrid ensemble–variational frameworks, reflecting the need to combine  
33 uncertainty estimation with efficient state estimation (Wang and Lei, 2014; Houtekamer and Zhang, 2016;  
34 Inverarity et al., 2023; Yokota et al., 2024).

35 Despite their established theoretical foundations and operational use, conventional DA methods still face  
36 important challenges in strong nonlinearity, under sparse observations, and constrained computational resources.  
37 For the EnKF, the method approximates the background-error covariance with a finite ensemble, so assimilation  
38 performance depends on ensemble size. An ensemble that is too small can suffer from sampling errors and  
39 inadequate representation of the background-error covariance, whereas increasing the ensemble size substantially  
40 raises the computational cost of ensemble propagation and analysis updates. Moreover, under sparse observations  
41 or strong nonlinearity, the finite-ensemble approximation and the linear analysis update may not fully capture  
42 complex error relationships. Reducing the computational cost of online assimilation while maintaining analysis  
43 accuracy and forecast stability therefore remains an important challenge.

44 In recent years, deep learning (DL) has been increasingly introduced into data assimilation and dynamical-system  
45 forecasting because of its strong capability for nonlinear mapping and high-dimensional feature representation  
46 (LeCun et al., 2015; Pathak et al., 2018). Applications have expanded from stand-alone neural-network forecasting  
47 to topics including analysis-operator learning, latent-space assimilation, background-error covariance  
48 representation, and uncertainty quantification (Bocquet et al., 2020; Brajard et al., 2020; Geer, 2021; Cheng et al.,  
49 2023). Among these, a representative direction is assimilation emulation, in which neural networks are trained to  
50 approximate the analysis operator of conventional data assimilation. The motivation for developing assimilation  
51 emulation is operational real-time forecasting: although training requires ensemble-based assimilation to generate  
52 target analyses, the trained network performs online inference on a single state without maintaining or propagating  
53 an ensemble, thereby substantially reducing computational cost.



54 Early studies used neural networks as auxiliary components within variational assimilation or incomplete  
55 dynamical systems (Tang and Hsieh, 2001), rather than learning the analysis operator itself. More recent work  
56 has moved toward end-to-end learning of the analysis mapping, which is the essence of assimilation emulation.  
57 For example, Arcucci et al. (2021) used deep learning to infer corrections between model forecasts and assimilated  
58 states, illustrating how learned mappings can support state estimation under incomplete observational information.  
59 Bocquet et al. (2024) more directly learned the sequential analysis step in a spatiotemporally chaotic system ---  
60 Lorenz-96, showing that a learned analysis operator can approach the performance of a tuned EnKF while  
61 propagating only a single forecast state during online filtering. These studies demonstrate that machine learning  
62 can replace computationally expensive components of the conventional assimilation cycle rather than merely  
63 serving as a forecast surrogate.

64 A complementary direction is forecast correction, which aims to mitigate the accumulation of forecast errors  
65 during model integration without continuously assimilating observations. In recursive forecasting, even small  
66 initial errors can continuously grow as the model evolves. Periodically correcting the model state with a neural  
67 network and using the corrected state to initialize subsequent integration therefore provides a potential means of  
68 slowing forecast-error growth. Previous studies have used analysis increments generated during data assimilation  
69 to train machine-learning models for forecast state correction (Farchi et al., 2021; He et al., 2025). However, these  
70 approaches generally make use of observational information available within the assimilation process. In contrast,  
71 the present study considers a forecast-only correction strategy that operates without observational input during  
72 inference and periodically adjusts the forecast state to mitigate the cumulative amplification of initial-condition  
73 errors.

74 Despite these advances, several questions remain regarding the reliability of assimilation emulation across  
75 different observation and dynamical regimes, as well as its integration with forecast correction. First, previous  
76 studies have focused primarily on the feasibility and implementation of assimilation emulation, while the detailed  
77 influences of observation frequency, training-data coverage, and local dynamical regime remain insufficiently  
78 understood. In particular, how learned analysis mappings behave when observational constraints are weakened  
79 and local instability varies requires systematic characterization. Second, assimilation emulation and forecast  
80 correction have largely been investigated as separate tasks, and their combined effect on recursive forecasting



81 remains less explored. In chaotic systems, the quality of the analysis state determines the initial forecast error,  
82 whereas subsequent model integration governs its accumulation and amplification. Evaluating these two processes  
83 within a unified framework is therefore important for understanding whether improvements in the initial state and  
84 forecast correction provide complementary benefits. Third, forecast-error growth is strongly state dependent in  
85 chaotic systems and is modulated by local dynamical instability and attractor geometry. Global error metrics alone  
86 may therefore obscure when and where learning-based methods are most effective, necessitating a dynamical-  
87 regime-based evaluation to characterize their strengths and limitations.

88 To address these questions, we use the classical Lorenz-63 chaotic system as a testbed to develop and evaluate  
89 two deep-learning approaches. The first is a data-assimilation multilayer perceptron (DAMLP). Its training data  
90 consist of EnKF-generated background–analysis ensemble-member pairs and corresponding observations, and  
91 the network learns a mapping from background states and observations to analysis states. During online  
92 application, DAMLP takes only a single background state and the current observation as input and directly outputs  
93 an analysis state through the trained mapping network. The second is a forecast-correction network that learns the  
94 mapping from the background states to the analysis states without using observations. During recursive  
95 forecasting, this network is applied periodically to correct model states that serve as initial conditions for the next  
96 step of model integration, thereby mitigating the accumulation of forecast errors. We further introduce the local  
97 Lyapunov exponent (LE) to examine how local dynamical instability affects the performance of the two  
98 approaches.

99 The main contributions of this study are threefold. First, we develop and evaluate an MLP-based analysis mapping  
100 that emulates the EnKF update and examine its computational efficiency and assimilation accuracy under different  
101 observation conditions. Second, we develop a periodic forecast-correction strategy that operates without  
102 observational input during inference and evaluate its ability to mitigate forecast-error growth and extend the useful  
103 forecast range. Third, we integrate assimilation emulation and forecast correction into a unified assimilation–  
104 forecast framework and use LE-based diagnostics to characterize their performance across different local  
105 dynamical regimes.

106 The remainder of this paper is organized as follows: Sect. 2 describes the Lorenz-63 model, data generation,  
107 evaluation metrics, and the experimental designs. Section 3 presents the results of assimilation emulation, forecast



108 correction, and their combined application. Section 4 summarizes the main findings and discusses the limitations  
 109 and directions for future work.

## 110 **4 Data and Methods**

### 111 **4.2 Lorenz-63**

112 The Lorenz-63 model is a three-variable nonlinear dynamical system introduced by Edward N. Lorenz (1963) to  
 113 study atmospheric convection. It is governed by a system of ordinary differential equations:

$$\begin{aligned}
 \frac{dx}{dt} &= \sigma(y - x) \\
 \frac{dy}{dt} &= \rho x - y - xz \\
 \frac{dz}{dt} &= xy - \beta z
 \end{aligned}
 \tag{1}$$

115 where  $x$ ,  $y$ , and  $z$  denote the state variables associated with the amplitudes of three Fourier modes describing  
 116 fluid motion and temperature variations. The parameters  $\sigma$ ,  $\rho$ , and  $\beta$  are control parameters related to the Prandtl  
 117 number, the normalized Rayleigh number, and the geometric properties of the convective system, respectively.  
 118 The three equations represent key physical processes including buoyancy forcing, dissipation, and nonlinear  
 119 advection.

120 The Lorenz-63 system is well known for its strong sensitivity to initial conditions and provides a canonical  
 121 example of deterministic chaotic systems. Owing to its simple structure and rich nonlinear dynamics, it has been  
 122 widely used as a testbed for nonlinear dynamics, chaos theory, and data assimilation. In this study, the Lorenz-63  
 123 system is used to examine the performance of the proposed deep-learning-based assimilation and forecast-  
 124 correction methods under controlled experimental conditions.

### 125 **4.3 EnKF**

126 The Ensemble Kalman Filter (EnKF) is used as the baseline data assimilation method in this study. Originally  
 127 introduced by Evensen (1994), the EnKF is a sequential data assimilation approach that represents forecast  
 128 uncertainty using a finite ensemble. At each observation time, the ensemble forecast is updated using the available  
 129 observations according to the Kalman analysis equation.



130 Let the model state variable be  $x$ , the observation be  $y$ , the ensemble mean be  $x_k^f$ , and the background-error  
131 covariance matrix be  $P^f$ . The Kalman analysis update is given by:

$$132 \quad x_i^a = x_i^f + K(y + \varepsilon_i - Hx_i^f) \quad (2)$$

133 where the Kalman gain matrix  $K$  is defined as:

$$134 \quad K = P^f H^T (H P^f H^T + R)^{-1} \quad (3)$$

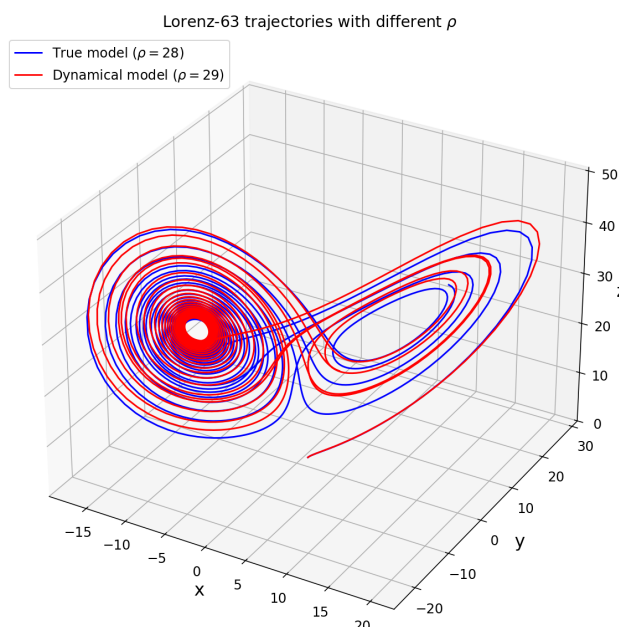
135 Here,  $H$  denotes the observation operator,  $R$  is the observation-error covariance matrix, and  $\varepsilon_i$  represents the  
136 perturbation added to the observations to maintain an appropriate ensemble spread (Burgers et al., 1998). The  
137 forecast-error covariance  $P_k^f$  is estimated directly from the finite ensemble:

$$138 \quad P_k^f \approx P_k^f = \frac{1}{N-1} \sum_{i=1}^N (x_{k,i}^f - \overline{x_k^f})(x_{k,i}^f - \overline{x_k^f})^T \quad (4)$$

139 Compared with the extended Kalman filter (EKF), the EnKF avoids the need for a tangent-linear approximation  
140 of the dynamical system, thereby reducing computational complexity. In addition, the EnKF provides a natural  
141 framework for representing uncertainty in both the model state and the observations, and has been widely used in  
142 numerical weather prediction, climate modeling, and other complex dynamical-system applications (Houtekamer  
143 and Zhang, 2016).

#### 144 **4.4 Data Generation**

145 To construct a high-quality training dataset for the subsequent neural-network models, a twin-experiment  
146 framework was designed within the Lorenz-63 system. Two dynamical models were used: a truth model and a  
147 forecast model, with a slight difference in their parameter settings. The truth model adopted the standard Lorenz-  
148 63 parameters  $\sigma = 10.0$ ,  $\beta = 8/3$ , and  $\rho = 28$ , with an integration time step of  $\Delta t = 0.01$ . In the forecast  
149 model,  $\rho$  was set to 29, while all other parameters remained unchanged. This parameter mismatch introduces a  
150 systematic model error and causes the forecast trajectory to gradually diverge from the truth trajectory, thereby  
151 representing the accumulation of forecast error associated with model-state evolution. Figure 1 compares the  
152 attractors generated by the truth and forecast models in phase space.



153

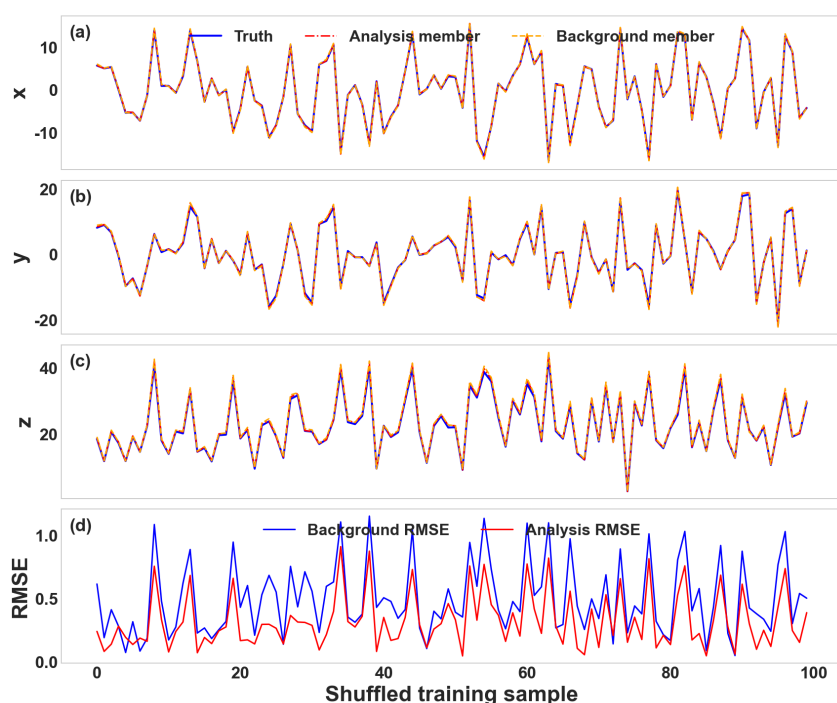
154 **Figure 1: Attractors of the truth model (blue) and forecast model (red) in phase space.**

155 A dimensionless model time unit was defined such that 1 MTU corresponds to 100 integration steps, or a physical  
 156 time of 1.0. The truth model was first freely integrated from the initial state  $[1, 1, 1]$  for 2000 MTU to eliminate  
 157 the initial transient, after which the integration was continued to generate the truth trajectory used in the  
 158 subsequent experiments. Observations were generated by adding zero-mean Gaussian noise with a standard  
 159 deviation of 0.15 to the truth, with an observation interval of 0.1 MTU, corresponding to every 10 integration  
 160 steps. During data assimilation, the ensemble size was set to 20. The initial ensemble is generated by adding a  
 161 background-error perturbation to the truth state to define an ensemble center, around which zero-mean random  
 162 perturbations are applied to construct the ensemble members. The forecast model is then integrated forward from  
 163 the initial ensemble, and a covariance-inflation factor of 1.4 is applied to the background ensemble to mitigate  
 164 covariance underestimation associated with the finite ensemble size.

165 The forecast–analysis cycle is repeated sequentially to generate a standardized assimilation dataset containing the  
 166 background ensemble members, the corresponding analysis ensemble members, observations, and truth state. At  
 167 each observation time, all ensemble members are retained as individual training samples. In total, 75, 000 samples  
 168 are generated. Of these, 60,000 samples are used for network training, while the remaining 15,000 samples are



169 reserved for validation and hyperparameter tuning. The test samples are taken from the subsequent portion of the  
 170 assimilation sequence and are not included in either the training or validation sets.



171  
 172 **Figure 2: Representative shuffled training samples used in the experiments. (a)–(c) Values of the truth state (blue),**  
 173 **background ensemble member (orange), and analysis ensemble member (red) for x, y, and z, respectively; (d) overall**  
 174 **RMSE of the background and analysis members relative to the truth for each sample. The samples are shown in**  
 175 **randomly shuffled order rather than chronological order.**

176 Figure 2 illustrates a representative subset of the shuffled training dataset generated from the EnKF assimilation  
 177 experiments. Each training sample is constructed from an individual background ensemble member and its  
 178 corresponding analysis ensemble member at a given observation time, together with the associated observation.  
 179 Panels (a)–(c) show the x, y and z components of these shuffled samples, while panel (d) shows the corresponding  
 180 sample-wise RMSE. Although the samples are no longer arranged in temporal order, the analysis members  
 181 generally exhibit smaller errors than the background members. This systematic difference provides the  
 182 supervised-learning signal used by the neural networks to learn the mapping from background states to analysis  
 183 states, with or without observational input depending on the application.



184 **4.5 Evaluation Metrics**

185 **2.4.1 Root Mean Square Error (RMSE)**

186 Root Mean Square Error (RMSE) is used to evaluate the overall magnitude of the deviation between predicted  
187 and true values, and is defined as:

188 
$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - y_i)^2} \quad (5)$$

189 where  $x_i$  denotes the model prediction,  $y_i$  denotes the corresponding truth, and  $N$  is the number of samples. A  
190 smaller RMSE indicates closer agreement between the prediction and the truth. RMSE is used as a primary metric  
191 for evaluating both assimilation and forecast accuracy.

192 **2.4.2 Correlation Coefficient (Corr)**

193 The correlation coefficient is used to evaluate the temporal consistency between the predicted and truth sequences  
194 and is defined as:

195 
$$Corr = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2 (y_i - \bar{y})^2}} \quad (6)$$

196 where  $\bar{x}$  and  $\bar{y}$  denote the means of the predicted and truth sequences, respectively, and  $N$  is the number of samples.  
197 A correlation coefficient closer to 1 indicates better agreement in temporal variability between the prediction and  
198 the truth. This metric is used to assess how well the predicted trajectory captures the temporal evolution of the  
199 dynamical system.

200 **2.4.3 Relative Skill Score (Skill)**

201 To quantify the performance of a method relative to a reference method, a relative Skill score based on RMSE is  
202 introduced:

203 
$$skill = \frac{RMSE_{ref} - RMSE_{method}}{RMSE_{ref}} \quad (7)$$



204 where  $\text{RMSE}_{ref}$  denotes the RMSE of the reference method and  $\text{RMSE}_{method}$  denotes that of the method being  
205 evaluated. A value of  $\text{Skill} > 0$  indicates a reduction in RMSE relative to the reference method,  $\text{Skill} = 0$   
206 indicates equal RMSE, and  $\text{Skill} < 0$  indicates a larger RMSE than the reference method.

#### 207 **2.4.4 Standard Deviation and Statistical Uncertainty**

208 The standard deviation across independent experiments is used to characterize the variability of the reported  
209 results. At a given forecast or assimilation time  $t_k$ , let  $\{e^{(1)}(t_k), e^{(2)}(t_k), \dots, e^{(M)}(t_k)\}$  denote the RMSE values  
210 obtained from  $M$  independent experiments. The mean and standard deviation are defined as:

$$211 \quad \bar{e}(t_k) = \frac{1}{M} \sum_{m=1}^M e^{(m)}(t_k), \sigma_e(t_k) = \sqrt{\frac{1}{M-1} \sum_{m=1}^M (e^{(m)}(t_k) - \bar{e}(t_k))^2} \quad (8)$$

212 The shaded regions surrounding the curves represent the mean  $\pm$  one standard deviation, illustrating the variability  
213 associated with different initial-ensemble realizations and other sources of experimental randomness.

#### 214 **2.4.5 Paired t-test**

215 A paired t-test is used to assess whether the difference in error between two methods under identical experimental  
216 conditions is statistically significant. Let  $e_i^{(1)}$  and  $e_i^{(2)}$  denote the errors of the two methods for the  $i$ -th paired  
217 sample, where  $i = 1, \dots, N$  and  $N$  is the number of paired samples, such as paired experimental runs or rolling-  
218 start samples considered in this study. The paired difference is defined as:

$$219 \quad d_i = e_i^{(1)} - e_i^{(2)} \quad (9)$$

220 Assuming that the paired differences  $d_i$  are approximately normally distributed, their sample mean  $\bar{d}$  and sample  
221 standard deviation  $s_d$  are calculated, and the t-statistic is given by:

$$222 \quad t = \frac{\bar{d}}{s_d/\sqrt{N}} \quad (10)$$

223 Under the null hypothesis  $H_0: \bar{d} = 0$ , which assumes no systematic difference in error between the two methods,  
224 a  $p$ -value below the significance level ( $p < 0.05$ ) indicates a statistically significant difference. With the  
225 definition in Eq. (9), if  $p < 0.05$  and  $\bar{d} > 0$ , Method 2 has a significantly lower error than Method 1.



#### 226    **2.4.6 Local Lyapunov Exponent (LE)**

227    The local Lyapunov exponent (LE) is introduced to characterize variations in local dynamical instability along  
228    the Lorenz-63 trajectory. It provides a finite-time measure of the growth of small perturbations and is used here  
229    as a diagnostic of local instability.

230    In this study, LE is estimated using a sliding-window approach. At a given time  $t_i$  along the reference trajectory,  
231    a small perturbation is introduced to the state  $\mathbf{x}(t_i)$ , yielding the perturbed state  $\mathbf{x}'(t_i)$ . The initial separation  
232    between the two states is defined as:

$$233 \qquad d_0 = \|\mathbf{x}(t_i) - \mathbf{x}'(t_i)\| \qquad (11)$$

234    The reference and perturbed states are then integrated forward over the same time interval  $\Delta t$ , giving the  
235    separation:

$$236 \qquad d'_i = \|\mathbf{x}(t_i + \Delta t) - \mathbf{x}'(t_i + \Delta t)\| \qquad (12)$$

237    Within a given sliding window, the local Lyapunov exponent is approximated as:

$$238 \qquad LE = \frac{1}{M\Delta t} \sum_{i=1}^M \ln \frac{d'_i}{d_0} \qquad (13)$$

239    where  $M$  denotes the number of samples included in the calculation window. In the present experiments, the  
240    initial perturbation magnitude is set to  $10^{-4}$ , with a new random unit direction generated at each starting time.  
241    The reference and perturbed trajectories are integrated forward for 1 MTU using the same dynamical model, with  
242    the perturbation renormalized after each integration step. The first 0.5 MTU is excluded as a transient, and the LE  
243    is evaluated from the perturbation-growth rates over the remaining 0.5 MTU. A larger positive LE indicates  
244    stronger local instability, whereas a smaller or negative value indicates relatively stable local dynamics.

245    LE is used here as a diagnostic of dynamical instability rather than as a direct measure of forecast skill. In the  
246    assimilation-emulation experiments, 5-MTU sliding-window mean LE values are used to identify representative  
247    high- and low-instability periods. In the forecast-correction experiments, each correction event is paired with the



248 LE of the truth state at the same valid time. Samples from all correction times and forecast initializations are then  
 249 pooled and grouped according to LE quantiles to examine how correction effectiveness varies with local instability.

## 250 **4.6 Experimental Setup**

251 This section describes the neural-network design, training strategy, and implementation of the two learning-based  
 252 approaches. Both models are trained using the EnKF assimilation dataset described in Sect. 2.3. Assimilation  
 253 emulation learns an observation-conditioned mapping from background ensemble members  $x_{i,j}^f$  to the  
 254 corresponding analysis ensemble members  $x_{i,j}^a$ , whereas forecast correction learns the corresponding  
 255 background-to-analysis mapping without observational input. Although ensemble-based data are used for training,  
 256 the trained models do not require explicit ensemble propagation during online inference. A multilayer perceptron  
 257 (MLP) is adopted as the basic network architecture because of its ability to approximate nonlinear mappings and  
 258 its relatively simple training procedure (Hornik et al., 1989).

### 259 **2.5.1 Experiment 1: Assimilation Emulation**

#### 260 **Neural Network Architecture and Input-Output Design**

261 DAMLP is trained to emulate the EnKF analysis update through an end-to-end mapping from a background  
 262 ensemble member and the corresponding observation to the corresponding analysis ensemble member:

$$263 \quad x_{i,j}^a = F_{\theta}(x_{i,j}^f, y_i^o) \quad (14)$$

264 where  $x_{i,j}^f \in \mathbb{R}^d$  denotes the  $j$ -th background ensemble member at observation time  $i$ ,  $y_i^o \in \mathbb{R}^d$  is the observation  
 265 vector, and  $x_{i,j}^a \in \mathbb{R}^d$  denotes the target analysis ensemble member. The network input is a six-dimensional vector  
 266 formed by concatenating  $x_{i,j}^f$  and  $y_i^o$ , while the output is a three-dimensional estimate of  $x_{i,j}^a$ .

267 DAMLP is trained by minimizing the mean squared error (MSE) between the predicted and target analysis states.  
 268 Given  $N$  training samples, let  $z_k$  denote the input of the  $k$ -th sample and  $x_k^a$  the corresponding target. The loss  
 269 function is defined as:

$$270 \quad L(\theta) = \frac{1}{N} \sum_{k=1}^N \| F_{\theta}(z_k) - x_k^a \|^2 \quad (15)$$



271 DAMLP consists of three fully connected hidden layers with widths [16, 32, 16] and ReLU activation functions.  
272 Batch normalization and dropout with a rate of 0.2 are introduced to improve generalization. The network also  
273 includes a linear residual connection from the input to the output. Training is performed using the Adam optimizer  
274 with an initial learning rate of  $1 \times 10^{-3}$ , together with a cosine-annealing learning-rate schedule. The model is  
275 trained for 120 epochs with a batch size of 32, and early stopping is used to select the best-performing model.

## 276 **Coupling with the Assimilation System**

277 DAMLP and EnKF are evaluated in parallel assimilation experiments for the Lorenz-63 system using the same  
278 initial conditions, model parameters, and observation sequences. DAMLP is trained only at  $\Delta t_m = 0.1$  MTU and  
279 is directly applied to larger observation intervals ( $\Delta t_m = 0.2 - 0.6$  MTU) without retraining to evaluate its  
280 generalization to sparser observations. The two experiments differ only in the analysis-update procedure.

281 EnKF scheme: At each observation time, all ensemble members are updated using the standard EnKF algorithm.

282 DAMLP scheme: At each observation time, the current background state  $x_i^f$  is concatenated with the observation  
283 vector  $y_i^o$  to form a six-dimensional input. A forward pass through DAMLP directly produces the analysis state  
284  $\hat{x}_i^a$ . This analysis state is then used as the initial condition for subsequent single-trajectory integration, and the  
285 ensemble members are no longer maintained or updated.

286 In essence, this approach replaces the linear EnKF analysis update with a nonlinear mapping learned by the neural  
287 network:

$$288 \quad x_{i,j}^a = x_{i,j}^f + K(y_i^o - Hx_{i,j}^f) \rightarrow \hat{x}_i^a = f_\theta([x_i^f, y_i^o]) \quad (16)$$

289 Thus, online DAMLP assimilation propagates only a single state rather than a full ensemble, reducing the  
290 computational cost associated with ensemble propagation and updating.

## 291 **2.5.2 Experiment 2: Forecast Correction Scheme**

### 292 **Neural Network Architecture and Input-Output Design**



293 The forecast-correction network is trained using the same background-to-analysis learning framework as DAMLP,  
 294 but without observational input. Specifically, it learns an end-to-end mapping from a background ensemble  
 295 member to the corresponding analysis ensemble member:

$$296 \quad x_{i,j}^a = \mathcal{F}_\theta(x_{i,j}^f) \quad (17)$$

297 where both the inputs and outputs are three-dimensional. The loss function follows the same MSE formulation as  
 298 Eq. (15), with the input reduced to the background state  $z_k = (x_{i,j}^f)$ .

299 The network uses the same three-layer MLP with a linear residual connection as in Experiment 1. The training  
 300 data are normalized using RobustScaler, and optimization is performed using AdamW with gradient clipping. The  
 301 model is trained for 120 epochs with a batch size of 32, with dynamic learning-rate adjustment and early stopping.

### 302 **Forecast Correction Scheme**

303 The forecast-correction experiment compares an uncorrected dynamical-model forecast with a neural-network-  
 304 corrected forecast (NN-Corrected). In the NN-Corrected experiment, the neural network is applied every 10  
 305 integration steps to correct the current model state. The corrected state is then used to initialize the subsequent  
 306 model integration. In contrast, the uncorrected forecast evolves solely according to the original dynamical  
 307 equations.

308 The forecast experiments use a rolling-start strategy. A total of 1000 initial states are selected sequentially from  
 309 the analysis sequence obtained from the assimilation experiments, with a spacing of 0.1 MTU between  
 310 consecutive initial states. An 8-MTU free forecast is performed from each initial state.

## 311 **5 Results**

312 This section provides a systematic evaluation of the two neural network approaches discussed earlier—  
 313 assimilation emulation using DAMLP and Forecast Correction. It is important to note that all data used in the  
 314 experiments in this section come from a test set that is independent of both the training and validation sets (see  
 315 Sect. 2.3); the networks were never exposed to these test samples during the training phase. Therefore, the results  
 316 below accurately reflect the models' generalization performance rather than a memorized fit to the training data.



## 317 5.2 Assimilation Emulation

318 This section evaluates the performance of DAMLP in the assimilation-emulation task, with particular emphasis  
 319 on computational efficiency, assimilation accuracy under different observation conditions, the influence of local  
 320 dynamical instability, and the impact of the resulting analysis states on subsequent forecast performance. DAMLP  
 321 is trained at  $\Delta t_m = 0.1$  MTU. To reduce the influence of random sampling in the initial ensemble on the statistical  
 322 results, 10 independent assimilation experiments are conducted for each observation setting. For each observation  
 323 interval, 400 analysis times were retained per run, and the assimilation-window length was adjusted accordingly.  
 324 All experiments use the same truth trajectory and observation sequence, while different random seeds are used to  
 325 generate the initial background ensembles. EnKF and DAMLP are compared under identical initial-ensemble and  
 326 observation conditions. For each experiment, the instantaneous RMSE over the assimilation window and the  
 327 corresponding time-mean RMSE are calculated. The final results are reported as the mean and standard deviation  
 328 across the 10 independent experiments. Using EnKF as the baseline, the skill of DAMLP is defined as:

$$329 \quad \text{skill} = \frac{\overline{RMSE}_{\text{EnKF}} - \overline{RMSE}_{\text{DAMLP}}}{\overline{RMSE}_{\text{EnKF}}} \quad (18)$$

330 where  $\text{Skill} > 0$  indicates that DAMLP has a lower RMSE than EnKF, whereas  $\text{Skill} < 0$  indicates that DAMLP  
 331 has a higher RMSE than EnKF.

### 332 3.1.1 Computational Efficiency Analysis

333 **Table 1: Total online assimilation times and corresponding RMSEs of EnKF and DAMLP under different ensemble**  
 334 **sizes**

| Ensemble size  | 20                  | 50                  | 100                 | 200                 |
|----------------|---------------------|---------------------|---------------------|---------------------|
| EnKF time (s)  | 1.3972              | 3.4069              | 6.6744              | 13.2718             |
| DAMLP time (s) | 0.0907              | 0.0907              | 0.0907              | 0.0907              |
| EnKF RMSE      | $0.4025 \pm 0.0024$ | $0.3929 \pm 0.0013$ | $0.3897 \pm 0.0007$ | $0.3887 \pm 0.0010$ |
| DAMLP RMSE     | $0.4075 \pm 0.0005$ | $0.4075 \pm 0.0005$ | $0.4075 \pm 0.0005$ | $0.4075 \pm 0.0005$ |

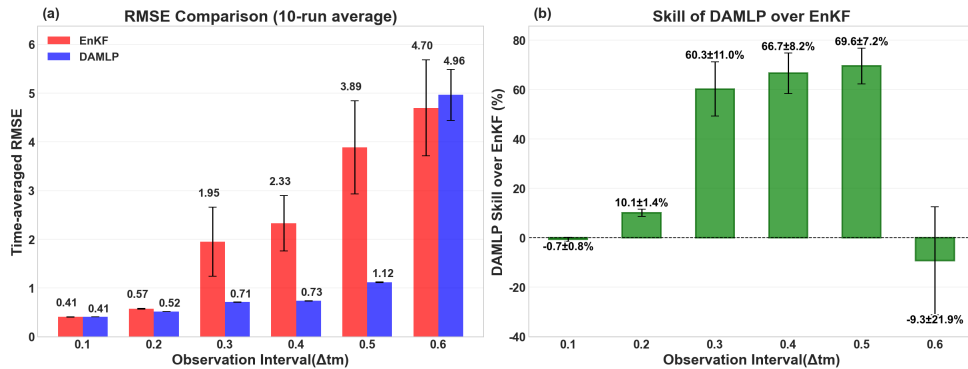
335 To evaluate the overall online computational efficiency, EnKF and DAMLP were compared using EnKF ensemble  
 336 sizes of 20, 50, 100, and 200. For each setting, 10 assimilation experiments were conducted over a 40-MTU



337 window with an observation interval of 0.1 MTU. The reported computational time includes both model  
 338 integration and analysis updates over the complete assimilation cycle.

339 The computational efficiency and assimilation accuracy of EnKF and DAMLP were summarized in Table 1.  
 340 The computational cost of EnKF increases markedly with ensemble size, from 1.3972 s for 20 members to 13.2718  
 341 s for 200 members, while the DAMLP runtime remains nearly constant at approximately 0.09 s. Meanwhile, the  
 342 RMSE of DAMLP remains approximately 0.408 and is comparable to that of EnKF across the tested ensemble  
 343 sizes, although EnKF achieves slightly lower mean RMSE as the ensemble size increases. DAMLP also shows  
 344 consistently low variability across experiments. These results indicate that the ensemble-free online operation of  
 345 DAMLP substantially reduces computational cost while maintaining comparable assimilation accuracy, with the  
 346 efficiency advantage becoming increasingly pronounced as the EnKF ensemble size increases.

### 347 3.1.2 Assimilation Accuracy under Different Observation Intervals



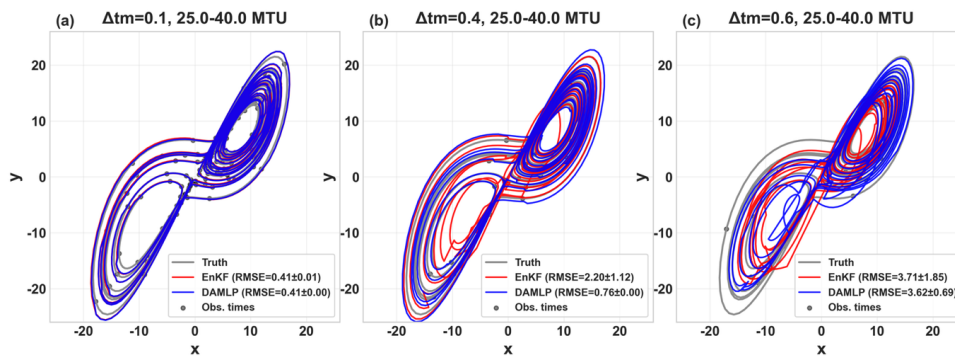
348 **Figure 3: Assimilation performance at different observation intervals. (a) Time-mean RMSE of EnKF and DAMLP**  
 349 **for  $\Delta t_m = 0.1$ – $0.6$  MTU, shown as the mean and standard deviation across 10 independent experiments; (b) Skill of**  
 350 **DAMLP relative to EnKF, expressed as a percentage, with error bars denoting the standard deviation across**  
 351 **experiments.**  
 352

353 Figure 3 compares the assimilation performance of DAMLP and EnKF at different observation intervals across  
 354 10 independent experiments. As shown in Fig. 3a, when  $\Delta t_m \leq 0.2$  MTU, the two methods exhibit similar error  
 355 levels, indicating that DAMLP closely reproduces the statistical performance of EnKF. As the observation interval  
 356 increases to  $\Delta t_m = 0.3$ – $0.5$  MTU, the mean RMSE of DAMLP becomes lower than that of EnKF. In particular,  
 357 DAMLP achieves Skill values of approximately 60.3%–69.6% over this range (Fig. 3b). When the observation  
 358 interval is further increased to 0.6 MTU, however, the DAMLP error rises sharply and the Skill becomes negative.



359 These results indicate that DAMLP and EnKF respond differently to increasing observation intervals. Under  
 360 frequent observations ( $\Delta t_m \leq 0.2$  MTU), the system state is constrained frequently by observational information  
 361 and the background error remains relatively small, resulting in similar performance between DAMLP and EnKF.  
 362 At moderately sparse observation intervals ( $\Delta t_m = 0.3\text{--}0.5$  MTU), background errors undergo stronger nonlinear  
 363 growth between successive observations. Under these conditions, the EnKF analysis update may become less  
 364 effective in representing the increasingly nonlinear error structure, whereas DAMLP directly learns a nonlinear  
 365 mapping from the background state and observations to the analysis state without explicitly relying on covariance  
 366 propagation or a linear Kalman-gain update. This enables DAMLP to maintain relatively stable assimilation  
 367 accuracy under moderately sparse observations.

368 When the observation interval is increased further and approaches the mean timescale of transitions between the  
 369 two wings of the Lorenz-63 attractor (approximately  $0.6\text{--}0.7$  MTU; Miller et al., 1994), the trajectory may switch  
 370 between attractor wings between two consecutive observations. In this regime, the deterministic mapping used by  
 371 DAMLP does not retain the probabilistic information represented by an ensemble. Consequently, a single  
 372 background state together with the current observation may be insufficient to distinguish the subsequent trajectory  
 373 evolution, leading to larger analysis errors and a rapid degradation in performance.



374  
 375 **Figure 4: Phase-space distributions of the EnKF and DAMLP analysis trajectories at different observation intervals.**  
 376 **Panels (a)–(c) correspond to  $\Delta t_m = 0.1, 0.4$ , and  $0.6$  MTU, respectively. Red lines denote the EnKF analysis**  
 377 **trajectories, blue lines denote the DAMLP analysis trajectories, and gray dots indicate observation times.**

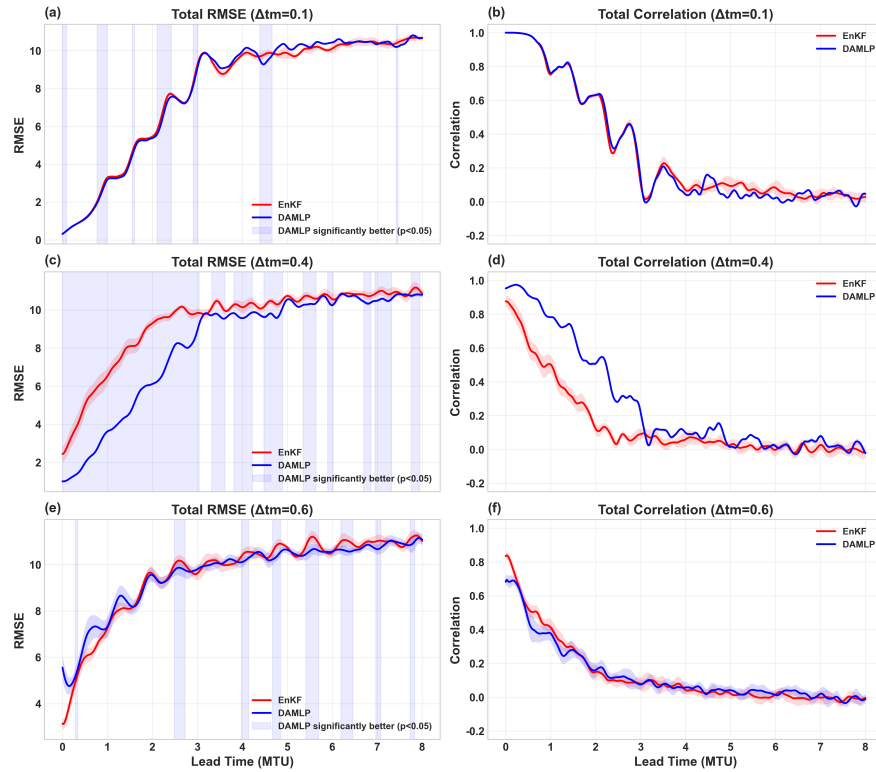
378 Figure 4 shows the analysis trajectories of EnKF and DAMLP in phase space for different observation intervals.  
 379 Because the trajectories are densely clustered over the full 40-MTU assimilation window, a 15 MTU segment



380 from 25 to 40 MTU is shown for clarity. When the observation interval is small ( $\Delta t_m = 0.1$  MTU), the analysis  
381 trajectories of both EnKF and DAMLP closely follow the truth trajectory, and their RMSE values are similar,  
382 indicating that DAMLP can effectively reproduce the analysis performance of EnKF.

383 When the observation interval increases to  $\Delta t_m = 0.4$  MTU, the EnKF trajectory exhibits pronounced deviations  
384 during transitions between the two attractor wings and near the outer parts of the attractor, whereas the DAMLP  
385 trajectory remains more closely distributed around the truth attractor. Its RMSE is also lower than that of EnKF,  
386 indicating more stable assimilation performance under moderately sparse observations. The error distribution is  
387 closely related to the attractor geometry: larger errors tend to occur along the outer branches of the two wings,  
388 while the inner regions of the wings and the transition passages between them are associated with relatively  
389 smaller errors. This contrast is particularly evident for EnKF, whereas DAMLP reduces the error growth along  
390 the outer branches to some extent.

391 When the observation interval is further increased to  $\Delta t_m = 0.6$  MTU, both methods show clear trajectory  
392 deviations. This suggests that, when observations are too sparse, a deterministic neural-network mapping may be  
393 insufficient to represent the uncertainty associated with transitions between attractor wings, thereby reducing the  
394 assimilation advantage of DAMLP. These phase-space trajectories are consistent with the results described above:  
395 DAMLP shows its clearest advantage under moderately sparse observation conditions, whereas under either  
396 frequent or extremely sparse observations, its relative performance compared with EnKF is less consistent.



397

398 **Figure 5: Comparison of rolling-forecast performance initialized from DAMLP and EnKF analysis states. (a, c, e)**  
 399 **RMSE and (b, d, f) correlation as functions of lead time for observation intervals of (a, b)  $\Delta t_m = 0.1$ , (c, d)  $\Delta t_m =$**   
 400  **$0.4$ , and (e, f)  $\Delta t_m = 0.6$  MTU. Shading denotes lead times at which DAMLP is significantly better than EnKF ( $p <$**   
 401  **$0.05$ ).**

402 To further evaluate the usefulness of the DAMLP analysis states for subsequent forecasting, we conducted rolling-  
 403 forecasts initialized from DAMLP and EnKF analyses. For each observation interval, 400 analysis states from  
 404 each of the 10 independent assimilation runs were used as initial conditions for 8-MTU free forecasts with the  
 405 dynamical model.

406 Figure 5 compares the forecast performance obtained from the initialization schemes. When  $\Delta t_m = 0.1$  MTU,  
 407 the RMSE and correlation curves for the two schemes nearly overlap, and statistically significant differences  
 408 occur only sporadically. This indicates that, under frequent observations, DAMLP closely reproduces the EnKF  
 409 analysis but provides little additional forecast benefit. When  $\Delta t_m = 0.4$  MTU, forecasts initialized from DAMLP  
 410 analyses maintain lower RMSE and higher correlation for approximately the first 3 MTU, with a relatively

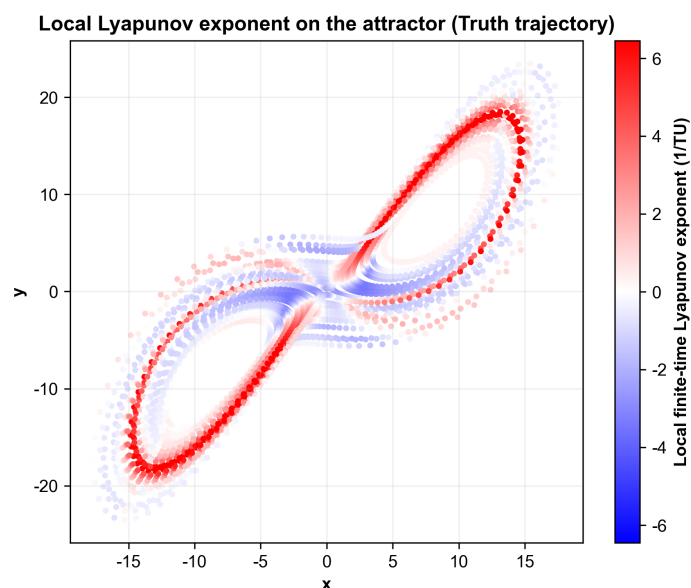


411 continuous region of statistical significance. This suggests that, under moderately sparse observations, DAMLP  
412 produces higher-quality initial states that effectively delay forecast-error growth. When  $\Delta t_m = 0.6$  MTU, the  
413 DAMLP-initialized forecast starts with a larger RMSE and lower correlation than the EnKF-initialized forecast,  
414 reflecting the degraded quality of the DAMLP analysis under extremely sparse observations. The errors of both  
415 schemes subsequently increase rapidly and gradually converge, with only intermittent lead times at which  
416 DAMLP shows lower RMSE.

417 Overall, the forecast benefit of DAMLP depends clearly on the observation interval: its advantage is most  
418 pronounced under moderately sparse observations, whereas the additional gain relative to EnKF is limited when  
419 observations are either frequent or extremely sparse.

### 420 **3.1.3 Modulation of DAMLP Performance by Local Dynamical Instability**

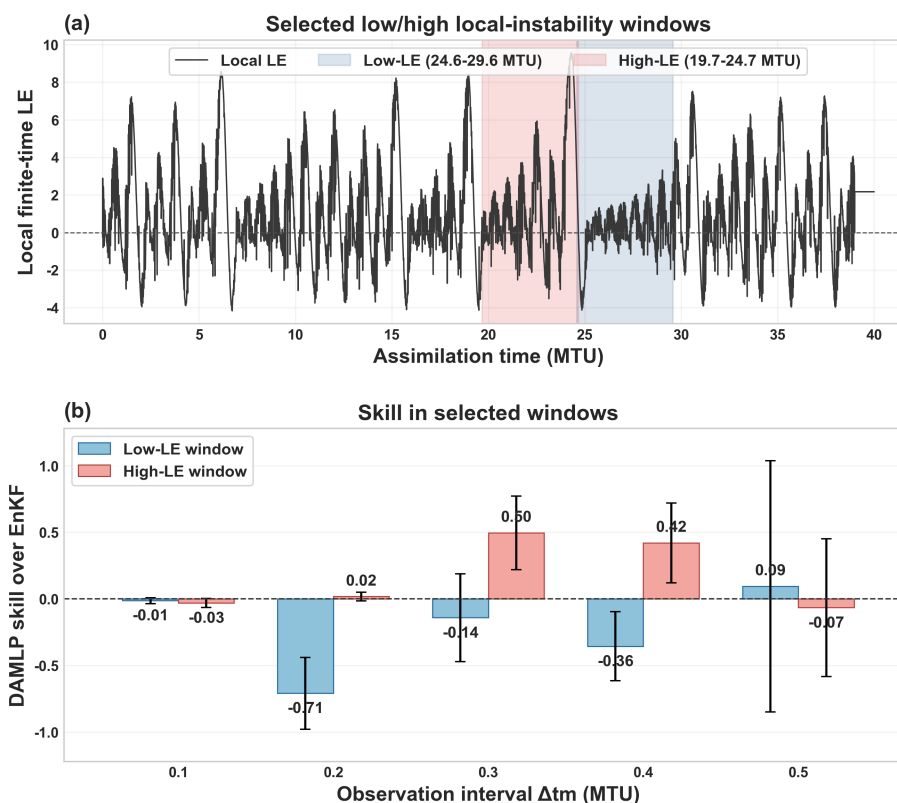
421 The preceding experiments show that the performance of DAMLP exhibits a clear three-regime dependence on  
422 the observation interval, suggesting that, in addition to observation frequency, the local dynamical state of the  
423 system may also influence its performance. Because the strength of local instability in the Lorenz-63 system is  
424 not uniform throughout phase space, error-growth rates can differ substantially along different parts of the  
425 trajectory. It is therefore necessary to examine the performance differences between DAMLP and EnKF from the  
426 perspective of local dynamical stability.



427

428 **Figure 6: Phase-space distribution of the local Lyapunov exponent (LE) for the Lorenz-63 system. Points represent**  
 429 **the truth trajectory in phase space, and colors indicate the corresponding LE values (units: MTU<sup>-1</sup>). Positive values**  
 430 **(red) indicate stronger local dynamical instability, whereas negative values (blue) indicate relatively stable dynamical**  
 431 **evolution.**

432 Figure 6 shows the local finite-time Lyapunov exponent (LE) along the attractor. The LE exhibits a clear spatial  
 433 structure: the outer branches of the two wings have large positive values, indicating strong local instability and  
 434 rapid error growth, whereas the inner regions and parts of the transition zones between the wings are characterized  
 435 by smaller or negative LE values. Notably, the regions where the EnKF trajectory in Fig. 4 deviates most—the  
 436 outer branches and the transition zones—are closely linked to this LE structure. The largest deviations occur on  
 437 the outer branches, i.e., the high-LE regions where errors grow fastest; these errors are then carried into the  
 438 transition zones and appear as trajectory displacements. This correspondence suggests that the relative advantage  
 439 of DAMLP under moderately sparse observations may be associated with dynamically sensitive, high-LE regimes,  
 440 motivating a more direct comparison between high- and low-LE periods.



441

442 **Figure 7: Selection of high- and low-instability periods based on local LE and comparison of DAMLP assimilation**  
 443 **Skill. (a) Temporal evolution of the local finite-time LE over the assimilation window; blue shading indicates the low-**  
 444 **LE period selected using a 5-MTU sliding window, and red shading indicates the high-LE period. (b) Mean Skill of**  
 445 **DAMLP relative to EnKF during the high- and low-LE periods at different observation intervals; error bars denote**  
 446 **the standard deviation across 10 independent experiments.**

447 To avoid subjective selection, the high- and low-instability periods were identified objectively as the 5-MTU  
 448 windows with the maximum and minimum window-mean LE, respectively. Using this criterion, Fig. 7a identifies  
 449 a high-LE period from 19.7 to 24.7 MTU and a low-LE period from 24.6 to 29.6 MTU. Figure 7b compares the  
 450 Skill of DAMLP relative to EnKF within these two periods for different observation intervals. Under frequent  
 451 observations ( $\Delta t_m = 0.1$  MTU), the Skill values are close to zero in both regimes. As the observation interval  
 452 increases to 0.2–0.4 MTU, a clear separation emerges: DAMLP shows positive Skill during the high-LE period,  
 453 while its Skill is negative during the low-LE period. At  $\Delta t_m = 0.5$  MTU, the difference between the two regimes  
 454 becomes small and the uncertainty increases substantially.



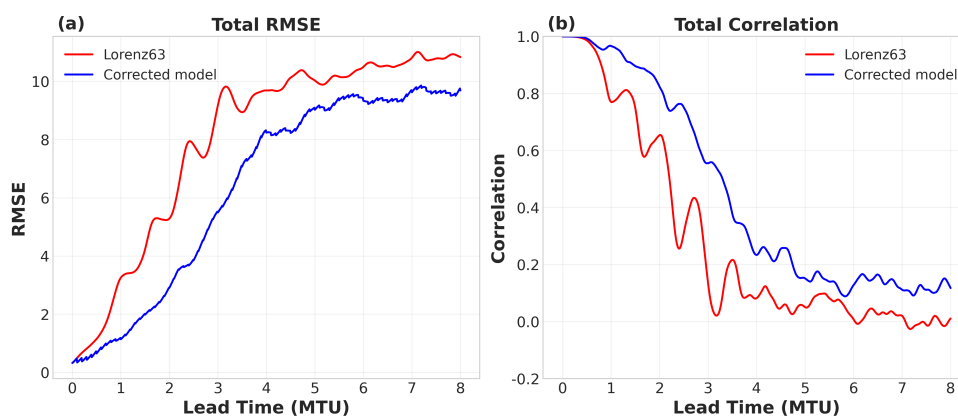
455 These results indicate that the relative performance of DAMLP is modulated by local dynamical instability, but  
 456 the relationship is not monotonic. During low-LE periods, error growth is relatively weak and the EnKF analysis  
 457 remains effective, leaving limited room for DAMLP to improve upon it. During high-LE periods, the background-  
 458 error structure is more strongly affected by nonlinear error growth, and the nonlinear mapping learned by DAMLP  
 459 can provide greater benefit. However, under very sparse observations, other factors such as accumulated  
 460 background error and reduced observational constraint become increasingly important.

461 Overall, the advantage of DAMLP is most evident under moderately sparse observations and strong local  
 462 instability. This is consistent with the phase-space results in Fig. 4 and the LE distribution in Fig. 6: the largest  
 463 EnKF deviations occur on the outer branches where LE is high, and DAMLP provides its clearest benefit precisely  
 464 in those dynamically sensitive regions. It also agrees with Fig. 3, where the overall Skill of DAMLP is positive  
 465 for  $\Delta t_m = 0.3$ – $0.5$  MTU and becomes negative at  $0.6$  MTU.

### 466 5.3 Forecast Correction

#### 467 5.3.1 Performance Evaluation of the Forecast-Correction Scheme

468 To evaluate the effectiveness of forecast-error correction in extending the useful forecast range, a fixed  
 469 observation interval of  $\Delta t_m = 0.1$  MTU was used. Based on 1000 rolling-start experiments, the forecast  
 470 performance of the standard Lorenz-63 model was compared with that of the periodic correction scheme over 800  
 471 integration steps.

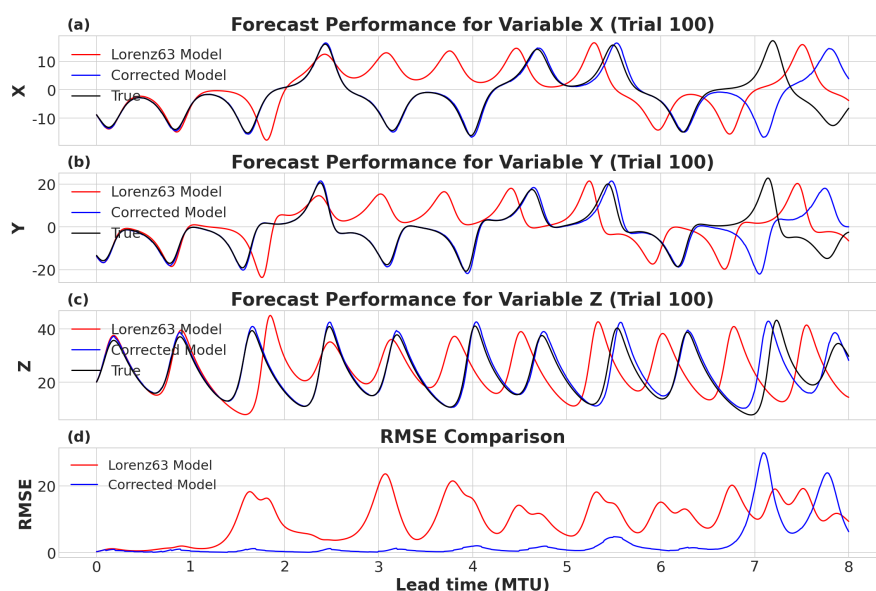




473 **Figure 8: Statistical evaluation of the forecast-correction scheme. (a) RMSE of the two schemes as a function of**  
 474 **forecast lead time; (b) correlation coefficient of the two schemes as a function of forecast lead time. The red solid line**  
 475 **denotes the standard Lorenz-63 forecast, and the blue solid line denotes the forecast with the correction model.**

476 Figure 8 shows the evolution of the overall RMSE (Fig. 8a) and correlation coefficient (Fig. 8b) with forecast  
 477 lead time for the standard Lorenz-63 model and the corrected model. As shown in Fig. 8a, the RMSE of the  
 478 standard model increases rapidly with lead time, exhibiting the characteristic error growth of a chaotic system,  
 479 and approaches saturation by 8 MTU, with a mean RMSE close to 11. By contrast, the corrected model applies  
 480 an error correction every 10 integration steps, which effectively slows the accumulation of forecast error and  
 481 produces a more gradual increase in RMSE. This result indicates that periodic correction can suppress forecast-  
 482 error growth and delay the loss of forecast skill.

483 The evolution of the correlation coefficient (Fig. 8b) further supports this result. As lead time increases, the  
 484 correlation of the standard model decreases rapidly and approaches zero at approximately 3 MTU, after which it  
 485 becomes negative and remains low at longer lead times. In contrast, the correlation of the corrected model  
 486 decreases more slowly, remaining above 0.5 until approximately 3.5 MTU. This indicates that the correction  
 487 scheme can maintain phase consistency between the forecast and truth trajectories over a longer forecast period.



488



489 **Figure 9: Forecast trajectories and overall RMSE for a randomly selected initialization. Panels (a)–(c) show the**  
490 **forecast trajectories of  $x$ ,  $y$ , and  $z$  as functions of forecast lead time, respectively; panel (d) shows the overall RMSE**  
491 **as a function of forecast lead time. Red lines denote the standard Lorenz-63 forecast, blue lines denote the neural-**  
492 **network-corrected forecast, and black lines denote the truth trajectory.**

493 To further illustrate the effect of the correction scheme during forecasting, Fig. 9 shows the forecast trajectories  
494 of the three state variables for a forecast initialized at a randomly selected time. For the  $x$  variable (Fig. 9a), the  
495 standard model (red line) begins to develop a phase error relative to the truth trajectory (black line) after  
496 approximately 1.5 MTU, followed by pronounced amplitude divergence after about 2.5 MTU. By contrast, the  
497 corrected forecast (blue line) maintains good agreement with the truth in both phase and amplitude until  
498 approximately 6 MTU. Its corresponding RMSE also remains markedly lower during the early and intermediate  
499 forecast period. Similar behavior is observed for the  $y$  (Fig. 9b) and  $z$  (Fig. 9c) variables. These results indicate  
500 that small corrections applied every 0.1 MTU can periodically bring the forecast state back toward the truth  
501 trajectory, thereby effectively suppressing forecast-error growth.

### 502 5.3.2 Modulation of Correction Performance by Local Dynamical Instability

503 The preceding experiments show that periodic correction can improve overall forecast performance. Because error  
504 growth in chaotic systems is closely related to the local dynamical state, the effectiveness of the correction may  
505 also vary with the degree of local instability. Figure 10 presents the Skill of the correction scheme relative to the  
506 standard Lorenz-63 model, stratified according to LE. Skill is defined as:

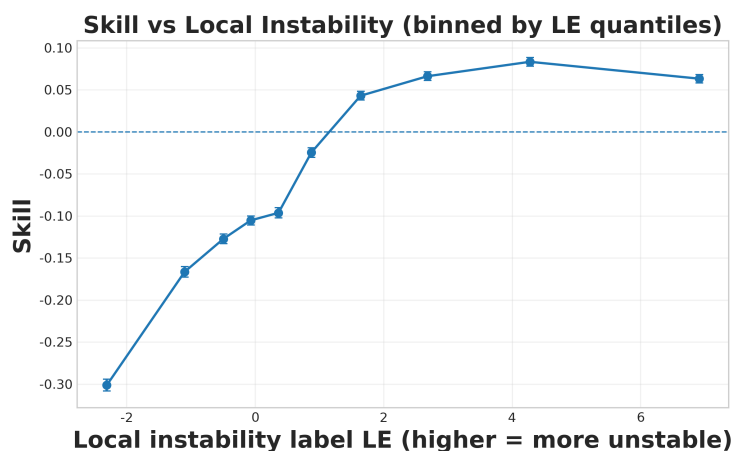
$$507 \quad Skill(\tau, i) = \frac{RMSE_{L63}(\tau, i) - RMSE_{Corr}(\tau, i)}{RMSE_{L63}(\tau, i)} \quad (19)$$

508 where  $\tau$  denotes the forecast lead time and  $i$  denotes the  $i$ -th rolling-start sample. A value of  $Skill > 0$  indicates  
509 that the corrected forecast has a lower RMSE than the standard Lorenz-63 forecast, whereas  $Skill < 0$  indicates  
510 that the correction increases the forecast error.

511 Corrections are applied every 0.1 MTU over the 8-MTU forecast window, giving 80 correction times for each  
512 rolling forecast. At each correction time, Skill is calculated from the errors immediately before and after correction,  
513 both evaluated against the truth at the same valid time, and is paired with the finite-time LE of that truth state.



514 Samples from all lead times (0.1–8.0 MTU) and all rolling-start forecasts are pooled and divided into 10 LE-  
 515 quantile bins. The mean Skill is then calculated for each bin.



516  
 517 **Figure 10: Relative Skill of the forecast-correction scheme as a function of local dynamical instability (LE).** Samples  
 518 from all correction times and rolling-start forecasts are pooled and divided into 10 LE-quantile bins. Points denote  
 519 the mean LE and mean Skill within each bin, and error bars indicate the standard error of the mean Skill.

520 The results show that, in low-instability regimes ( $LE < 1.5$ ), the Skill values remain mostly negative, indicating  
 521 limited benefit from periodic correction when the forecast is already relatively stable. As LE increases, Skill rises  
 522 steadily and becomes positive, reaching its maximum at relatively high LE ( $LE = 2-4$ ) before decreasing slightly  
 523 in the most unstable regime ( $LE > 4$ ). Overall, the correction effect is clearly state dependent, with the greatest  
 524 benefit occurring under moderately to strongly unstable conditions and a slight reduction at the highest LE values.

#### 525 **5.4 Combined Optimization Scheme**

526 The preceding two sections examined, respectively, the improvement in analysis-state quality achieved by  
 527 DAMLP and the effectiveness of periodic state correction in suppressing forecast-error growth. In this section,  
 528 the two components are combined to evaluate forecast performance when DAMLP-derived initial conditions are  
 529 used together with the correction model. Four forecast pathways are compared: EnKF analysis state + original  
 530 model, EnKF analysis state + corrected model, DAMLP analysis state + original model, and DAMLP analysis  
 531 state + corrected model.

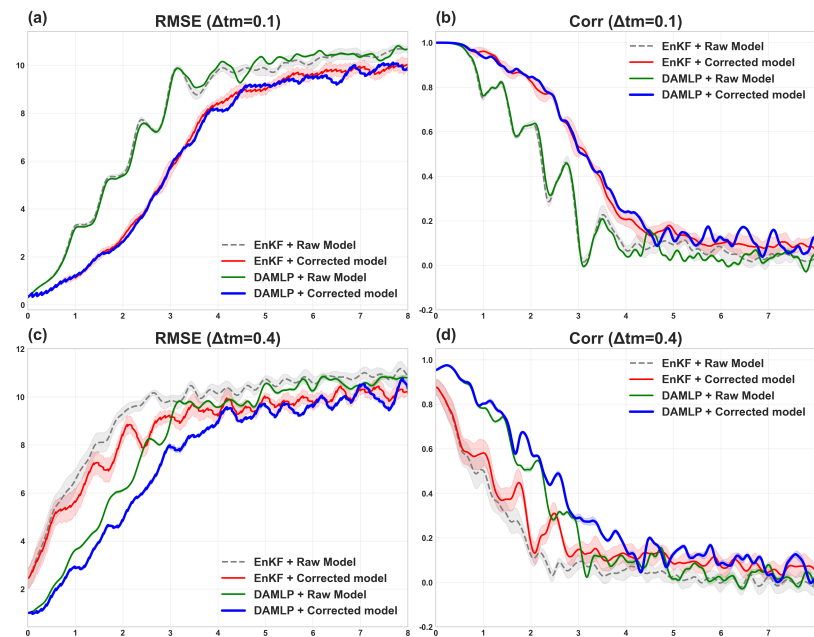


**Figure 11: Forecast performance of the four pathways in the unified framework under two observation scenarios: (a–b) frequent observations ( $\Delta t_m = 0.1$  MTU); (c–d) sparse observations ( $\Delta t_m = 0.4$  MTU).**

Figure 11 shows the evolution of RMSE and correlation for the four forecast pathways under two representative observation scenarios: frequent observations ( $\Delta t_m = 0.1$  MTU) and sparse observations ( $\Delta t_m = 0.4$  MTU).

Under frequent observations (Figs. 11a–b), the analysis states produced by DAMLP and EnKF have similar accuracy, and the initial forecast errors of the two initialization schemes are therefore nearly identical. For the uncorrected forecasts (gray and green dashed lines), RMSE increases rapidly with lead time and approaches saturation at approximately 11.0 after about 5 MTU. By contrast, the corrected forecasts (red and blue solid lines) exhibit a markedly slower increase in RMSE and a more gradual decrease in correlation, while the two corrected curves almost completely overlap. This indicates that, under frequent observations, the primary advantage of DAMLP lies in computational efficiency rather than in providing additional forecast-accuracy gains.

Under sparse observations (Figs. 11c–d), the four forecast pathways become clearly separated. The DAMLP-initialized forecasts have lower initial RMSE than the EnKF-initialized forecasts, indicating that DAMLP produces more accurate analysis states under sparse-observation conditions. However, without forecast correction, the advantage of DAMLP + original model persists for only a limited period, and its error gradually approaches



548 that of EnKF + original model after approximately 3 MTU. After the correction model is introduced, forecast-  
549 error growth is substantially reduced. Among the four pathways, DAMLP + corrected model maintains the lowest  
550 RMSE and relatively high correlation throughout the forecast window. The corresponding blue shaded region is  
551 also the narrowest, indicating lower sensitivity to initial-ensemble sampling and greater statistical stability.

552 These results show that forecast errors are influenced by both the accuracy of the initial state and the accumulation  
553 of error during model integration. The more accurate initial states produced by DAMLP reduce the initial forecast  
554 error, while the correction model suppresses its subsequent amplification through periodic state adjustment. Their  
555 combined effect therefore helps extend the useful forecast range and provides experimental support for integrating  
556 deep-learning-based assimilation emulation with forecast correction.

## 557 **6 Conclusions and Discussions**

558 This study develops a deep-learning framework that combines assimilation emulation with forecast-error  
559 correction in the Lorenz-63 system. DAMLP replaces the online ensemble-based analysis procedure with a  
560 learned state–observation mapping, substantially reducing computational cost while maintaining assimilation  
561 accuracy comparable to EnKF and providing additional benefits under moderately sparse observations. The  
562 forecast correction model is periodically applied during model integration without observational input and slows  
563 the growth of forecast errors. Their combination further shows that improving the initial state and reducing error  
564 accumulation during integration can provide complementary benefits for forecasting chaotic systems.

565 The assimilation-emulation experiments show that DAMLP offers two main advantages: substantially reduced  
566 online computational cost and improved performance under moderately sparse observations. In terms of  
567 computational efficiency, the total online assimilation time of EnKF increases markedly with ensemble size, from  
568 1.3972 s for 20 members to 13.2718 s for 200 members, whereas the runtime of DAMLP remains nearly constant  
569 at approximately 0.09 s. This contrast reflects the different online computational structures of the two methods:  
570 EnKF requires the propagation and update of all ensemble members, while DAMLP propagates only a single  
571 forecast state and performs one network inference at each observation time. In terms of assimilation accuracy,  
572 DAMLP closely reproduces the performance of EnKF when observations are frequent. As the observation interval  
573 increases to moderately sparse conditions, DAMLP achieves lower RMSE than EnKF, indicating that the learned



574 nonlinear mapping can provide additional benefit when observational constraints are weakened. However, this  
575 advantage decreases when observations become very sparse, suggesting that DAMLP performance depends not  
576 only on observation frequency but also on the accumulated background error and local dynamical regime.

577 The LE-based analysis further shows that the relative performance of DAMLP depends on the local dynamical  
578 regime. The distribution of LE over the Lorenz-63 attractor is highly nonuniform, indicating substantial variation  
579 in local instability along the trajectory. Under frequent observations, DAMLP and EnKF perform similarly in both  
580 high- and low-LE periods, indicating that strong observational constraints largely suppress the influence of local  
581 instability on their performance difference. As the observation interval increases to moderately sparse conditions,  
582 DAMLP exhibits positive Skill during the high-LE period but negative Skill during the low-LE period, suggesting  
583 that the nonlinear mapping learned by DAMLP provides the greatest relative benefit when observational  
584 constraints are weakened and local error growth is sufficiently strong. However, this relationship is not monotonic:  
585 under very sparse observations, the difference between the two regimes decreases and the variability increases,  
586 indicating that accumulated background error and reduced observational constraint begin to outweigh the effect  
587 of local instability alone. Therefore, LE modulates rather than solely determines the relative performance of  
588 DAMLP, which is also influenced by observation frequency and accumulated background error.

589 The forecast correction experiments show that periodic state correction slows forecast-error growth, producing  
590 lower RMSE and a slower decline in correlation than the uncorrected model. The LE-stratified results further  
591 reveal a clear state dependence of the correction effect. In low-instability regimes, the Skill of the correction can  
592 be negative, indicating that unnecessary correction may slightly degrade an already stable forecast. As LE  
593 increases, the correction becomes increasingly beneficial, with the largest improvement occurring under  
594 moderately to strongly unstable conditions. The benefit decreases slightly in the most unstable regime, showing  
595 that the correction effect does not increase monotonically with LE. These results indicate that periodic correction  
596 is most effective when forecast errors are actively growing but have not yet become too large to be reliably  
597 corrected.

598 The combined use of DAMLP and the forecast-correction model further demonstrates that improvements in  
599 initial-state quality and corrections applied during the forecast have complementary effects in recursive  
600 forecasting. In the sparse-observation case examined here, DAMLP provides a more accurate initial state than



601 EnKF, but this initial advantage gradually diminishes during uncorrected model integration. Periodic correction  
602 reduces the subsequent growth of forecast error and allows the benefit of the improved initial state to persist for  
603 a longer lead time. As a result, the DAMLP-initialized corrected forecast maintains lower RMSE, higher  
604 correlation, and smaller across-experiment variability over much of the forecast window. These results indicate  
605 that improving the analysis state and controlling error growth during integration address two different stages of  
606 forecast-error development and can therefore be effectively combined in a unified assimilation–forecast  
607 framework, while retaining the computational advantage of ensemble-free online assimilation. This framework  
608 thus offers a promising pathway toward efficient assimilation–forecast systems for chaotic dynamical systems.

609 Several limitations should be noted when interpreting these results. First, the present experiments are conducted  
610 only with the low-dimensional Lorenz-63 system under controlled observation and dynamical configurations.  
611 Although learning-based assimilation emulation has already been explored in higher-dimensional systems, the  
612 present framework still needs to be evaluated under more complex dynamical and observational configurations.  
613 Second, although DAMLP eliminates ensemble propagation during online assimilation, its deterministic single-  
614 state inference does not explicitly represent flow-dependent uncertainty, which may become increasingly  
615 important under very sparse observations or strong nonlinear error growth. Third, the forecast-correction model  
616 is currently applied at a fixed interval, whereas the LE-based results indicate that the benefit of correction varies  
617 substantially with the local dynamical regime. Future work will therefore focus on extending the framework to  
618 more complex systems, developing uncertainty-aware assimilation emulation, and exploring adaptive correction  
619 strategies based on local instability. In addition, incorporating physical or dynamical constraints into network  
620 training may further improve physical consistency, robustness, and generalization across different regimes.

#### 621 **Code and data availability**

622 The code and data that support the findings of this study are available upon reasonable request.

#### 623 **Author contributions**

624 XL developed the experimental code, conducted the numerical experiments, analyzed the results, prepared the  
625 figures, and wrote the original manuscript. YT supervised the research and reviewed and revised the manuscript.



626 ZS and YL provided scientific guidance and reviewed the manuscript. ZH assisted in identifying and resolving  
 627 issues encountered during the experiments.

### 628 **Competing interests**

629 The authors declare that they have no conflict of interest.

### 630 **Financial support**

631 This research was supported by the National Natural Science Foundation of China (grant nos. 42130409 and  
 632 U24A20606).

### 633 **References**

- 634 Arcucci, R., Zhu, J., Hu, S., and Guo, Y.-K.: Deep data assimilation: Integrating deep learning with data  
 635 assimilation, *Appl. Sci.*, 11, 1114, <https://doi.org/10.3390/app11031114>, 2021.
- 636 Bocquet, M., Brajard, J., Carrassi, A., and Bertino, L.: Bayesian inference of chaotic dynamics by merging data  
 637 assimilation, machine learning and expectation-maximization, *Found. Data Sci.*, 2, 55–80,  
 638 <https://doi.org/10.3934/fods.2020004>, 2020.
- 639 Bocquet, M., Farchi, A., Finn, T. S., Durand, C., Cheng, S., Chen, Y., Pasmans, I., and Carrassi, A.: Accurate  
 640 deep learning-based filtering for chaotic dynamics by identifying instabilities without an ensemble, *Chaos*, 34,  
 641 091104, <https://doi.org/10.1063/5.0230837>, 2024.
- 642 Brajard, J., Carrassi, A., Bocquet, M., and Bertino, L.: Combining data assimilation and machine learning to  
 643 emulate a dynamical model from sparse and noisy observations: A case study with the Lorenz 96 model, *J. Comput.*  
 644 *Sci.*, 44, 101171, <https://doi.org/10.1016/j.jocs.2020.101171>, 2020.
- 645 Burgers, G., van Leeuwen, P. J., and Evensen, G.: Analysis scheme in the ensemble Kalman filter, *Mon. Weather*  
 646 *Rev.*, 126, 1719–1724, [https://doi.org/10.1175/1520-0493\(1998\)126<1719:ASITEK>2.0.CO;2](https://doi.org/10.1175/1520-0493(1998)126<1719:ASITEK>2.0.CO;2), 1998.
- 647 Carrassi, A., Bocquet, M., Bertino, L., and Evensen, G.: Data assimilation in the geosciences: An overview of  
 648 methods, issues, and perspectives, *WIREs Clim. Change*, 9, e535, <https://doi.org/10.1002/wcc.535>, 2018.
- 649 Cheng, S., Quilodr n-Casas, C., Ouala, S., Farchi, A., Liu, C., Tandeo, P., Fablet, R., Lucor, D., Iooss, B., Brajard,  
 650 J., Xiao, D., Janjic, T., Ding, W., Guo, Y., Carrassi, A., Bocquet, M., and Arcucci, R.: Machine learning with data



- 651 assimilation and uncertainty quantification for dynamical systems: A review, *IEEE/CAA J. Autom. Sinica*, 10,  
652 1361–1387, <https://doi.org/10.1109/JAS.2023.123537>, 2023.
- 653 Courtier, P., Thépaut, J.-N., and Hollingsworth, A.: A strategy for operational implementation of 4D-Var, using  
654 an incremental approach, *Q. J. Roy. Meteor. Soc.*, 120, 1367–1387, <https://doi.org/10.1002/qj.49712051912>, 1994.
- 655 Evensen, G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods  
656 to forecast error statistics, *J. Geophys. Res.*, 99, 10143–10162, <https://doi.org/10.1029/94JC00572>, 1994.
- 657 Farchi, A., Laloyaux, P., Bonavita, M., and Bocquet, M.: Using machine learning to correct model error in data  
658 assimilation and forecast applications, *Q. J. Roy. Meteor. Soc.*, 147, 3067–3084, <https://doi.org/10.1002/qj.4116>,  
659 2021.
- 660 Geer, A. J.: Learning earth system models from observations: Machine learning or data assimilation?, *Philos. T.*  
661 *Roy. Soc. A*, 379, 20200089, <https://doi.org/10.1098/rsta.2020.0089>, 2021.
- 662 He, Z., Brajard, J., Wang, Y., Wang, X., and Shen, Z.: Improving dynamical climate predictions with machine  
663 learning: Insights from a twin experiment framework, *Nonlin. Processes Geophys.*, 32, 397–409,  
664 <https://doi.org/10.5194/npg-32-397-2025>, 2025.
- 665 Houtekamer, P. L. and Mitchell, H. L.: Data assimilation using an ensemble Kalman filter technique, *Mon.*  
666 *Weather Rev.*, 126, 796–811, [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), 1998.
- 667 Houtekamer, P. L. and Zhang, F.: Review of the ensemble Kalman filter for atmospheric data assimilation, *Mon.*  
668 *Weather Rev.*, 144, 4489–4532, <https://doi.org/10.1175/MWR-D-15-0440.1>, 2016.
- 669 Hornik, K., Stinchcombe, M., and White, H.: Multilayer feedforward networks are universal approximators,  
670 *Neural Networks*, 2, 359–366, [https://doi.org/10.1016/0893-6080\(89\)90020-8](https://doi.org/10.1016/0893-6080(89)90020-8), 1989.
- 671 Inverarity, G. W., Tennant, W. J., Anton, L., Bowler, N. E., Clayton, A. M., Jardak, M., Lorenc, A. C., Rawlins,  
672 F., Thompson, S. A., Thurlow, M. S., Walters, D. N., and Wlasak, M. A.: Met Office MOGREPS-G initialisation  
673 using an ensemble of hybrid four-dimensional ensemble variational (En-4DEnVar) data assimilations, *Q. J. Roy.*  
674 *Meteor. Soc.*, 149, 1138–1164, <https://doi.org/10.1002/qj.4431>, 2023.
- 675 Le Dimet, F.-X. and Talagrand, O.: Variational algorithms for analysis and assimilation of meteorological  
676 observations: Theoretical aspects, *Tellus A*, 38A, 97–110, <https://doi.org/10.3402/tellusa.v38i2.11706>, 1986.
- 677 LeCun, Y., Bengio, Y., and Hinton, G.: Deep learning, *Nature*, 521, 436–444, <https://doi.org/10.1038/nature14539>,  
678 2015.



- 679 Lorenz, E. N.: Deterministic nonperiodic flow, *J. Atmos. Sci.*, 20, 130–141, [https://doi.org/10.1175/1520-](https://doi.org/10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2)  
680 0469(1963)020<0130:DNF>2.0.CO;2, 1963.
- 681 Miller, R. N., Ghil, M., and Gauthiez, F.: Advanced data assimilation in strongly nonlinear dynamical systems, *J.*  
682 *Atmos. Sci.*, 51, 1037–1056, [https://doi.org/10.1175/1520-0469\(1994\)051<1037:ADAISN>2.0.CO;2](https://doi.org/10.1175/1520-0469(1994)051<1037:ADAISN>2.0.CO;2), 1994.
- 683 Pathak, J., Hunt, B., Girvan, M., Lu, Z., and Ott, E.: Model-free prediction of large spatiotemporally chaotic  
684 systems from data: A reservoir computing approach, *Phys. Rev. Lett.*, 120, 024102,  
685 <https://doi.org/10.1103/PhysRevLett.120.024102>, 2018.
- 686 Tang, Y. and Hsieh, W. W.: Coupling neural networks to incomplete dynamical systems via variational data  
687 assimilation, *Mon. Weather Rev.*, 129, 818–834, [https://doi.org/10.1175/1520-](https://doi.org/10.1175/1520-0493(2001)129<0818:CNNTID>2.0.CO;2)  
688 0493(2001)129<0818:CNNTID>2.0.CO;2, 2001.
- 689 Wang, X. and Lei, T.: GSI-based four-dimensional ensemble–variational (4DEnsVar) data assimilation:  
690 Formulation and single-resolution experiments with real data for NCEP Global Forecast System, *Mon. Weather*  
691 *Rev.*, 142, 3303–3325, <https://doi.org/10.1175/MWR-D-13-00303.1>, 2014.
- 692 Yokota, S., Banno, T., Oigawa, M., Akimoto, G., Kawano, K., and Ikuta, Y.: JMA operational hourly hybrid  
693 3DVar with singular vector-based mesoscale ensemble prediction system, *J. Meteorol. Soc. Jpn.*, 102, 129–150,  
694 <https://doi.org/10.2151/jmsj.2024-006>, 2024.