# Geophysical Research Letters®

**Key Points:**

- The FuXi-En4Dvar employ automatic differentiation to compute gradients eliminating the need for tangent linear models and adjoint models
- Using the rapid ensemble generation capabilities of ML-based weather forecasting model to construct the background error covariance matrix
- The FuXi-En4DVar demonstrates flow-dependent characteristics, constraining analysis increments that adhere to physical balance relationships

**Supporting Information:**

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

**Correspondence to:**

W. Han and H. Li,
hanwei@cma.gov.cn;
lihao_lh@fudan.edu.cn

**Author Contributions:**

**Conceptualization:** Yonghui Li, Wei Han, Hao Li, Wansuo Duan, Xiaohui Zhong, Jincheng Wang, Yongzhu Liu
**Data curation:** Yonghui Li
**Formal analysis:** Yonghui Li, Lei Chen, Xiaohui Zhong, Jincheng Wang, Yongzhu Liu, Xiuyu Sun
**Funding acquisition:** Wei Han
**Investigation:** Yonghui Li, Wei Han, Hao Li, Lei Chen, Xiaohui Zhong, Xiuyu Sun
**Methodology:** Yonghui Li, Wei Han, Hao Li, Wansuo Duan, Lei Chen,

# FuXi-En4DVar: An Assimilation System Based on Machine Learning Weather Forecasting Model Ensuring Physical Constraints

**Yonghui Li[1,2], Wei Han[3] , Hao Li[4,5], Wansuo Duan[1,2] , Lei Chen[5], Xiaohui Zhong[4] , Jincheng Wang[3], Yongzhu Liu[3], and Xiuyu Sun[5]**

[1]State Key Laboratory of Numerical Modeling for Atmospheric Sciences and Geophysical Fluid Dynamics, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China, [2]University of Chinese Academy of Sciences, Beijing, China, [3]CMA Earth System Modeling and Prediction Centre(CEMC), China Meteorological Administration, Beijing, China, [4]Artificial Intelligence Innovation and Incubation Institute, Fudan University, Shanghai, China, [5]Shanghai Academy of Artificial Intelligence for Science (SAIS), Shanghai, China

**Abstract** Recent machine learning (ML)-based weather forecasting models have improved the accuracy and efficiency of forecasts while minimizing computational resources, yet still depend on traditional data assimilation (DA) systems to generate analysis fields. Four dimensional variational data assimilation (4DVar) enhances model states, relying on the prediction model to propagate observation to the initial field. Consequently, the initial fields from traditional DA are not optimal for ML-based models, necessitating a customized DA system. This paper introduces an ensemble 4DVar system integrated with the FuXi model (FuXi-En4DVar), which can independently generate accurate analysis fields. It utilizes automatic differentiation to compute gradients, and demonstrates the equivalence of these gradients with those derived from adjoint models. Experimental results indicate that this system preserves the physical balance of the analysis field and exhibits flow-dependent characteristics. These features enhance the propagation and assimilation of observation into the initial analysis field, thereby improving the accuracy of the analysis fields.

**Plain Language Summary** Machine learning (ML)-based weather forecasting models have made significant progress, offering fast and accurate weather predictions. However, a critical limitation of these models is their dependence on externally provided initial fields, which they are unable to generate independently. This study addresses this limitation by developing a data assimilation (DA) system with FuXi, a state-of-the-art ML-based weather forecasting model, enabling it to generate these initial fields. Experimental results confirm the rationality and effectiveness of this system.

## 1. Introduction

Accurate weather forecasts are essential across various sectors of society. Traditional numerical weather prediction (NWP) relies on solving differential equations based on physical laws using high-performance computing clusters (Kalnay, 2003), which require substantial computational resources. However, recent advancements in ML-based weather forecasting models offer promising alternatives with enhanced computational efficiency, several orders of magnitude faster than traditional NWP models. These ML models leverage vast amounts of data to learn complex weather patterns and produce accurate forecast. Several ML-based weather forecasting models have been developed worldwide, including FourCastNet (Pathak et al., 2022), SwinVRNN (Hu et al., 2023), Pangu-weather (Bi et al., 2023), FengWu (K. Chen et al., 2023), GraphCast (Lam et al., 2023), and FuXi (L. Chen et al., 2023). These models are data-driven, and their prediction accuracy can match or even surpass traditional NWP models such as the Integrated Forecasting System (IFS) from the European Center for Medium-Range Weather Forecasts (ECMWF). For instance, FuXi has demonstrated forecast performance comparable to the ECMWF ensemble mean in 15 day forecasts. This enhanced efficiency and forecasting accuracy facilitate faster decision-making and improves the timeliness of weather-related warnings.

Despite significant achievements in ML-based weather forecasting models, they remain dependent on DA system of traditional NWP, which provides initial conditions for forecasts. This reliance prevents ML models from performing forecasts independently, in contrast to NWP systems that incorporate both DA and prediction components. DA integrates recent observations from various sources, including satellites, weather stations,

Xiaohui Zhong, Jincheng Wang, Yongzhu Liu, Xiuyu Sun
**Project administration:** Wei Han
**Resources:** Yonghui Li, Hao Li, Lei Chen, Xiaohui Zhong, Xiuyu Sun
**Software:** Yonghui Li, Hao Li
**Supervision:** Yonghui Li, Wei Han, Hao Li, Wansuo Duan
**Validation:** Yonghui Li, Wei Han, Wansuo Duan
**Visualization:** Yonghui Li
**Writing – original draft:** Yonghui Li, Xiaohui Zhong
**Writing – review & editing:** Yonghui Li, Wei Han, Wansuo Duan, Xiaohui Zhong

aircraft soundings, and ships to refining the background field to produce the analysis field. Subsequently, the prediction model forecasts the evolution of weather states. These two components are mutually dependent: the prediction model provides the background field for DA, which then optimizes a cost function to refine the initial field. The cost function essentially measures the distance of the solution to the background and observations. This process yields a more accurate analysis field for the prediction model. Within the DA components in the operational NWP systems, the cost function includes terms predicted by the model, highlighting a strong coupling between the prediction model and DA. A primary limitation of ML-based weather forecasting models is their lack of an exclusive DA component. Therefore, it is important to develop a dedicated DA system tailored for ML-based weather forecasting models.

Efforts to integrate ML-based weather forecasting models with DA have led to two primary methodologies. The first employs ECMWF Reanalysis v5 (ERA5) data, which is the fifth generation ECMWF atmospheric reanalysis of the global climate, as ground truth, and uses both ERA5 and satellite data to train supervised DA models such as FengWu-Adas (K. Chen et al., 2024) and FuXi-DA (Xu et al., 2024). These models aim to replace the entire traditional DA process with a ML model, providing initial fields for corresponding ML-based weather forecasting models. Despite their innovation, these systems have significant limitations. For instance, FengWu-Adas is restricted to assimilating conventional observations, whereas FuXi-DA is limited to assimilate satellite observations, specifically from the Advanced Geostationary Radiation Imager of FY-4B. Notably, both models overlook the intrinsic physical balance in their network architectures.

The second approach focuses on modifying the cost function of DA by replacing the traditional physics-based prediction model with a ML-based forecasting model, solving the DA optimization problem to obtain the analysis field. An example of this is FengWu-4DVar (Xiao et al., 2023), which constructs its framework based on a simplified FengWu model with a reduced spatial resolution of 1.4° and assimilates ERA5 data. Consequently, the analysis field generated cannot be utilized directly by the original FengWu model. Additionally, FengWu-4DVar uses an idealized background error covariance matrix (**B**), accounting only for the self-correlation of variables and ignoring any inter-variable correlations, which compromises the ability to impose reasonable physical constraints on the analysis increments. A common and significant issue with these two approaches is the lack of physical constraints in the DA process. An effective DA system must integrate observational data into the model state and ensure its propagation throughout all variables while maintaining physical balance constraints. For instance, assimilating the geopotential height at a specific location should trigger adjustments in the surrounding wind field, reflecting to these physical relationships.

This work presents an ensemble four dimension variational data assimilation (En4DVar) system based on the FuXi model, which employs Perlin noise (Perlin, 1985) to generate ensembles. The FuXi-En4DVar system, a hybrid model, merges the state-of-the-art ML-based FuXi model with the globally recognized En4DVar assimilation algorithm proposed by Liu et al. (C. Liu et al., 2008). The En4DVar system has been applied for WRF model (C. Liu et al., 2009) and compared with other DA method and summarized (Bannister, 2017). This integration facilitates the continuous refinement of the FuXi model's predictions through observational data, thereby reducing dependency on traditional DA and NWP systems. Ultimately, this approach may improve DA processes, potentially enabling the independent generation of analysis fields and forecasts in the future.

## 2. FuXi Ensemble Four Dimension Variational Data Assimilation

### 2.1. FuXi

The FuXi Model (L. Chen et al., 2023), developed by the Artificial Intelligence Innovation and Incubation Institute at Fudan University, implements a cascaded model architecture to provide a 15 day global forecast with a temporal resolution of 6 hr and a spatial resolution of 0.25°. It consists of 70 weather variables which include geopotential, relative humidity, temperature, $u$ component of wind, $v$ component of wind, 6–hourly total precipitation, and mean sea-level pressure (see Table 1 of the paper (L. Chen et al., 2023) for details), utilizing the input data with dimensions of $2*70*721*1{,}440$, where 2, 70, 721 and 1,440 represent the two preceding time steps (t-1 and t), the total number of input variables, the grid points of latitude and longitude, respectively. This model is developed utilizing a 39 year ECMWF ERA5 reanalysis dataset. Forecast performance is evaluated using latitude-weighted root mean square error (RMSE) and anomaly correlation coefficient (ACC). Remarkably, the results show that its 15 day forecast are comparable to those of the ECMWF ensemble mean, making it the first ML model to achieve such a level of accuracy (L. Chen et al., 2023).

**Table 1**
*Experiment Setup* [a]

| Name | Variable | Start of assimilation window | Point/Area of assimilation and corresponding assimilation time |
|---|---|---|---|
| EXP_1 | Z500 | 0000 UTC 30 July 2023 | 21°N,129°E at 00 [b] |
| EXP_2 | R500 | 0000 UTC 30 July 2023 | 52°N,86°E at 00, 52°N,86°E at 06, 17°N,135°E at 00, 17°N,135°E at 06, respectively |
| EXP_3 | T500, U500, V500 and R500 | 0000 UTC 30 July 2023 | 20°N–60°N, 80°E–140°E, spaced at 1°intervals, at 00 and 06 |
| EXP_4 | T, U, V and RH at 250 , 300, 500, 850 and 1,000 hPa, respectively | 0000 UTC 30 July 2023 | 20°N–60°N, 80°E–140°E, spaced at 1°intervals, at 00 and 06 |

[a]Z500, T500, U500, V500 and R500 represent geopotential, temperature, u-component of wind, v-component of wind and relative humidity at 500 hPa pressure level, respectively. [b]00 and 06 represent the start and end times of the assimilation window, respectively.

## 2.2. FuXi En4DVar

Over the past few decades, many operational centers have implemented the 4DVar method to develop global NWP systems (Gauthier et al., 2007; Gauthier & Thepaut, 2001; Rabier et al., 2000; Rawlins et al., 2007; Zhang et al., 2019). The 4DVar method incorporate dynamic and physical constraints into its cost function but relies on a static **B** matrix, which does not change over time, posing challenges in adapting to the temporally variable atmospheric states (Buehner et al., 2010). En4DVar addresses this issue by combining the core components of Ensemble Kalman Filter (EnKF) (Evensen, 2003) with traditional 4DVar (Le Dimet & Talagrand, 1986; Lewis & Derber, 1985). En4DVar enhances adaptability of the **B** matrix by incorporating flow-dependent characteristics and physical constraints, facilitating observation assimilation across multiple time steps similar to 4DVar. Additionally, En4DVar utilizes ensembles to refine the **B**, capturing the correlations between variables and solving the problems of FengWu-4DVar to a certain extent. This capability is particularly effective in modeling complex, nonlinear processes or in scenarios where traditional geophysical balances are unsuitable. Due to these enhancements and the rapid ensemble generation capabilities of ML-based weather forecasting models, En4DVar has been selected as the DA system for integration with the FuXi model.

In the 4DVar method, the analysis $\mathbf{x}_a$ is obtained by minimizing a cost function $J$. This function quantifies the discrepancy between the modeled trajectory $H_i(M_i(\mathbf{x}))$ and corresponding observations $\mathbf{y}_i$ at a sequential time points $t_i$, $i = 1, \ldots, n$, and the discrepancy between solution $\mathbf{x}$ and background $\mathbf{x_b}$. Given certain assumptions (Bouttier & Courtier, 2002), the cost function $J$ of 4DVar is formulated as follows:

$$J(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}_b) + \frac{1}{2}\sum_{i=0}^{n}\left(\mathbf{y}_i - H_i(M_i(\mathbf{x}))\right)^T \mathbf{R}_i^{-1}\left(\mathbf{y}_i - H_i(M_i(\mathbf{x}))\right) \qquad (1)$$

where $\mathbf{x}$ denotes the model state at the beginning of the assimilation window. $\mathbf{x}_b$ is the background or prior estimate of $\mathbf{x}$. The matrices **B** and **R** correspond to the background and observational error covariance matrices, respectively. $H_i$ represents the observational operator, facilitating the nonlinear transformation from model space to observational space, while $M_i$ is the nonlinear propagator operator, integrating model states from $t_0$ to $t_i$.

In the En4DVar framework, given $K$ ensemble members, the background error is estimated as follows:

$$\mathbf{X}_b = \frac{1}{\sqrt[2]{K-1}}(\mathbf{x}_{b1} - \overline{\mathbf{x}_b}, \ldots, \mathbf{x}_{bi} - \overline{\mathbf{x}_b}, \ldots, \mathbf{x}_{bK} - \overline{\mathbf{x}_b}) \qquad (2)$$

where $\mathbf{x}_{bi}$ represents the $i$-th member of $\mathbf{x}_b$, and $\overline{\mathbf{x}_b}$ is the mean of the $K$ ensemble members. The $\mathbf{x}_{bi}, \overline{\mathbf{x}_b} \in \mathbb{R}^N$ and $\mathbf{X}_b \in \mathbb{R}^{N*K}$ where $N$ denotes the dimension of the model space. Then the **B** matrix is approximated by:

$$\mathbf{B} \approx \mathbf{X}_b\mathbf{X}_b^T \qquad (3)$$

By transformation, we obtain an alternative expression for the state vector $\mathbf{x}$:

$$x = x_b + X_b w \tag{4}$$

where $\mathbf{w}$ is the control variable and $\mathbf{w} \in \mathbb{R}^K$. Then cost function in control variable space becomes

$$J(\mathbf{w}) = \frac{1}{2}\mathbf{w}^{\mathrm{T}}\mathbf{w} + \frac{1}{2}\sum_{i=0}^{n}(\mathbf{d}_i - \mathbf{H}_i(\mathbf{M}_i(\mathbf{X}_b\mathbf{w})))^{\mathrm{T}}\mathbf{R}_i^{-1}(\mathbf{d}_i - \mathbf{H}_i(\mathbf{M}_i(\mathbf{X}_b\mathbf{w}))) \tag{5}$$

where $\mathbf{d}_i = \mathbf{y}_i - H_i(M_i(\mathbf{x}_b))$, are the innovations at different time points within the assimilation window. $\mathbf{H}_t = \frac{\partial H_t(\mathbf{x})}{\partial \mathbf{x}}$ and $\mathbf{M}_t = \frac{\partial M_t(\mathbf{x})}{\partial \mathbf{x}}$ are respectively the tangent linear modes of the nonlinear observation $H_t$ and propagation operators $M_t$. The gradient of the cost function with respect to the control variables $\mathbf{w}$ is

$$\nabla J(\mathbf{w}) = \mathbf{w} - \sum_{i=0}^{n}\mathbf{X}_b^{\mathrm{T}}\mathbf{M}_i^{\mathrm{T}}\mathbf{H}_i^{\mathrm{T}}\mathbf{R}_i^{-1}(\mathbf{d}_i - \mathbf{H}_i(\mathbf{M}_i(\mathbf{X}_b\mathbf{w}))) \tag{6}$$

The superscript T denotes transpose of a matrix, with $\mathbf{M}_t^{\mathrm{T}}$ and $\mathbf{H}_t^{\mathrm{T}}$ representing the adjoint models corresponding to the propagation and observation operators, respectively.

This study utilizes 200 ensemble members ($K = 200$) generated by first randomly creating 200 sets of Perlin noise, then adding these noise fields to the initial conditions (ERA5) 6 hr before the start of the assimilation window, and finally using FuXi inference one step to the start of the assimilation window. It is noteworthy that the **B** matrix, derived from these members, is singular and exhibits a rank significantly smaller than the dimension of the model space, potentially introducing spurious correlations, where variables that are geographically distant exhibit a high degree of correlation. To address the rank deficiency in the **B** matrix, localization techniques are employed (Gaspari & Cohn, 1999; Hamill et al., 2001). For a detailed composition of the **B** matrix, refer to Text S1 in Supporting Information S1.

In FuXi-En4DVar, the $M_t$ from Equation 1 is replaced with the FuXi model, under the assumption that it conforms to the established hypotheses (Bouttier & Courtier, 2002), thereby ensuring consistency of its cost function with that of Equation 5. Due to the FuXi model's specific design, including a 6 hr time step, the following cost function is employed to seamlessly integrate it within the En4DVar framework:

$$J(\mathbf{w}) = \frac{1}{2}\mathbf{w}^{\mathrm{T}}\mathbf{w} + \frac{1}{2}(\mathbf{y}_0 - H_0(\mathbf{x}_b + \mathbf{X}_b\mathbf{w}))^{\mathrm{T}}\mathbf{R}_0^{-1}(\mathbf{y}_0 - H_0(\mathbf{x}_b + \mathbf{X}_b\mathbf{w}))$$
$$+ \frac{1}{2}(\mathbf{y}_6 - H_6(M_6(\mathbf{x}_b + \mathbf{X}_b\mathbf{w})))^{\mathrm{T}}\mathbf{R}_6^{-1}(\mathbf{y}_6 - H_6(M_6(\mathbf{x}_b + \mathbf{X}_b\mathbf{w}))) \tag{7}$$

where $M_6$ represents the FuXi short model with a temporal resolution of 6 hr. Here we employ a 6 hr assimilation window, incorporating observations only from the current and 6 hr forward. For clarity, in the following sections, we will refer to the terms in the right-hand side of Equation 7 as follows: the first term will be denoted as $J_b$, the second as $J_{o_0}$, and the third as $J_{o_6}$.

Unlike traditional NWP, ML models can bypass need for an adjoint model by directly computing gradients via automatic differentiation (Paszke et al., 2017). In essence, both ML and DA are optimization problems. Specifically, ML models utilize automatic differentiation to calculate gradients of neural network parameters, as shown in Figure 1a, aiming to optimize the model for fixed inputs (training data) using an optimization algorithm. DA is solved in a similar way, as shown in Figure 1b, the parameters remain constant while gradients of the cost function relative to the inputs are computed, leading to an optimal solution through optimization. This study employs the L-BFGS algorithm (D. C. Liu & Nocedal, 1989) for optimization purposes. To construct FuXi-En4DVar system, the 'model' in Figure 1b is strictly constructed based on the model described in Equation 7 and illustrated in Figure 1c, with parameters including $\mathbf{X}_b, \mathbf{y}_0, \mathbf{y}_6, \mathbf{x}_b$ and the control variable $\mathbf{w}$ as the input. If we ignore differences in symbolic representation, two methodologies are equivalent. The FuXi-En4DVar model facilitates accurate calculations of derivatives with respect to the inputs by adhering to the strict application of the chain rule.
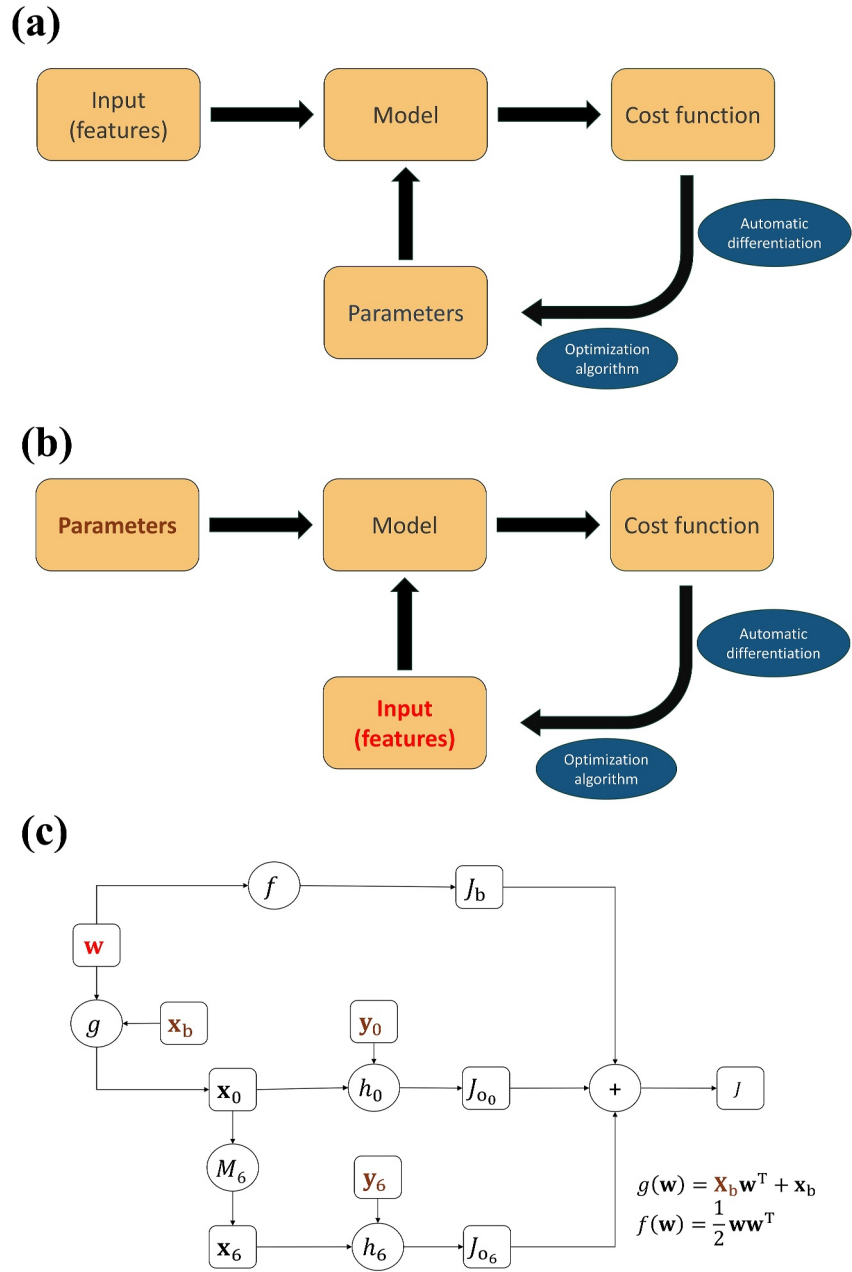
**Figure 1.** The general process of machine learning (a). Figure (b) illustrates the methodology used to solve the FuXi-En4DVar optimization problem, while Figure (c) represents the corresponding model. In Figure (c), **w** aligns with the 'Input' of Figure (b) (red color) and the parameters $\mathbf{X}_b, \mathbf{y}_0, \mathbf{y}_6, \mathbf{x}_b$ are depicted in brown color.

$$\nabla J(\mathbf{w}) = \frac{\partial J}{\partial J_b} \frac{\partial J_b}{\partial \mathbf{w}} + \frac{\partial J}{\partial J_{o_0}} \frac{\partial J_{o_0}}{\partial \mathbf{x}_0} \frac{\partial \mathbf{x}_0}{\partial \mathbf{w}} + \frac{\partial J}{\partial J_{o_6}} \frac{\partial J_{o_6}}{\partial \mathbf{x}_6} \frac{\partial \mathbf{x}_6}{\partial \mathbf{x}_0} \frac{\partial \mathbf{x}_0}{\partial \mathbf{w}} \tag{8}$$

$$\begin{aligned} = \ & \mathbf{w} + \mathbf{X}_b^T \mathbf{H}_0^T \mathbf{R}_0^{-1} \big( H_0(\mathbf{x}_b + \mathbf{X}_b \mathbf{w}) - \mathbf{y}_0 \big) \\ & + \mathbf{X}_b^T \mathbf{M}_6^T \mathbf{H}_6^T \mathbf{R}_6^{-1} \big( H_6(M_6(\mathbf{x}_b + \mathbf{X}_b \mathbf{w})) - \mathbf{y}_6 \big) \end{aligned} \tag{9}$$

The above analysis demonstrates that the gradients computed through automatic differentiation (Equation 8) are identical to those obtained using the adjoint model (Equation 9).

### 2.3. Experiment Setup

To validate where a DA system adheres to physical constraints and evaluate its rationality, four sets of experiments are designed, as detailed in table 1. The assimilation window for these experiments is from 0000 UTC on 30 July 2023 to 0600 UTC on 30 July 2023. The background field is derived by inferring the FuXi model for 12 hr, starting with the initial conditions from ERA5 data (Hersbach et al., 2020) at 1200 UTC on 29 July 2023.

- EXP_1: A single Z500 observation at 21°N,129°E at the initial time (0000 UTC, 30 July 2023) is introduced to evaluate the system's ability to satisfy physical constraint satisfaction and ensure proper localization.
- EXP_2: A single R500 observation is introduced ahead of the trough and behind the ridge (52°N, 86°E) at the beginning (000 UTC 30 July 2023) and ending (0600 UTC 30 July 2023) of the assimilation window, respectively. A similar test is conducted at a location near typhoon (17°N, 135°E) to verify the system's capability in propagating observational information reasonably and exhibiting flow-dependent characteristics. The reason for choosing water vapor as a tracer (Bosilovich & Schubert, 2002), as it aids in the estimation of water vapor sources through its transport, thus enhancing the demonstration of the system's propagation capabilities and flow-dependent characteristics.
- EXP_3: Regional observations (20°N–60°N, 80°E−140°E) of multiple variables within the same layer (T500, U500, V500 and R500) are assimilated simultaneously within the assimilation window, using an ideal observational grid arranged at 1°intervals.
- EXP_4: This experiment simulates real observational scenarios (T, U, V and RH at 250 , 300, 500, 850 and 1000 hPa) by assimilating data from multiple atmospheric layers and variables within the specified region (EXP_4), thereby verifying the system's effectiveness through comprehensive multi-layer assimilation.

### 2.4. Data

This study employed a basic interpolation method as the observation operator. For simplicity, the observation error covariance matrix **R** was assumed to be diagonal, with constant error variances. Specifically, the standard deviations for geopotential, relative humidity, temperature, u-component of wind and v-component of wind are set at constant values of 25 $m^2s^{-2}$, 2 %, 0.1 $K$, 0.5 $ms^{-1}$, and 0.5 $ms^{-1}$, respectively.

The observations used in conjunction with the FuXi model consist of ERA5 reanalysis data and Gaussian noise. This addition mimics real-world observational uncertainties by introducing Gaussian noise is with a mean of 0 and a variance equal to 0.001 times the standard deviations of observation errors. These observations are recorded every 6 hr, aligning with the temporal resolution of the FuXi model.

## 3. Results

### 3.1. Physical Balance Constraints of FuXi-En4DVar

For mid-latitude synoptic-scale disturbances, the Coriolis force and the pressure gradient force are approximately balanced (Holton & Hakim, 2012). An effective assimilation system should align its analysis increments with this geostrophic balance. To verify this, a single-point observation of geopotential at the 500 hPa pressure level (Z500) was introduced initially (EXP_1).

The results show that the observation minus background (OMB) for Z500 at this location is greater than 0 (see Table S1 in Supporting Information S1 for detail). Following the assimilation, the observation minus analysis (OMA) also remained positive, with the absolute values of OMA being smaller than the OMB values. Moreover, the analysis increment is positive (Figures 2a and 2b), consistent with the OMB, suggesting a reasonable adjustment (Desroziers et al., 2005). The localization prevented significant spurious correlations both vertically and horizontally in the experiments (Figures 2a and 2b).

Consistent with the geostrophic balance principle, near the analysis increment (positive geopotential anomaly), there is a noticeable anticyclonic response in the horizontal wind direction (Figure 2a). Vertically, an approximate anticyclonic structure is evident in the v-wind component (Figure 2b). These findings underscore that the FuXi-En4DVar assimilation system effectively adjusts the analysis increment within the constraints of physical balance.
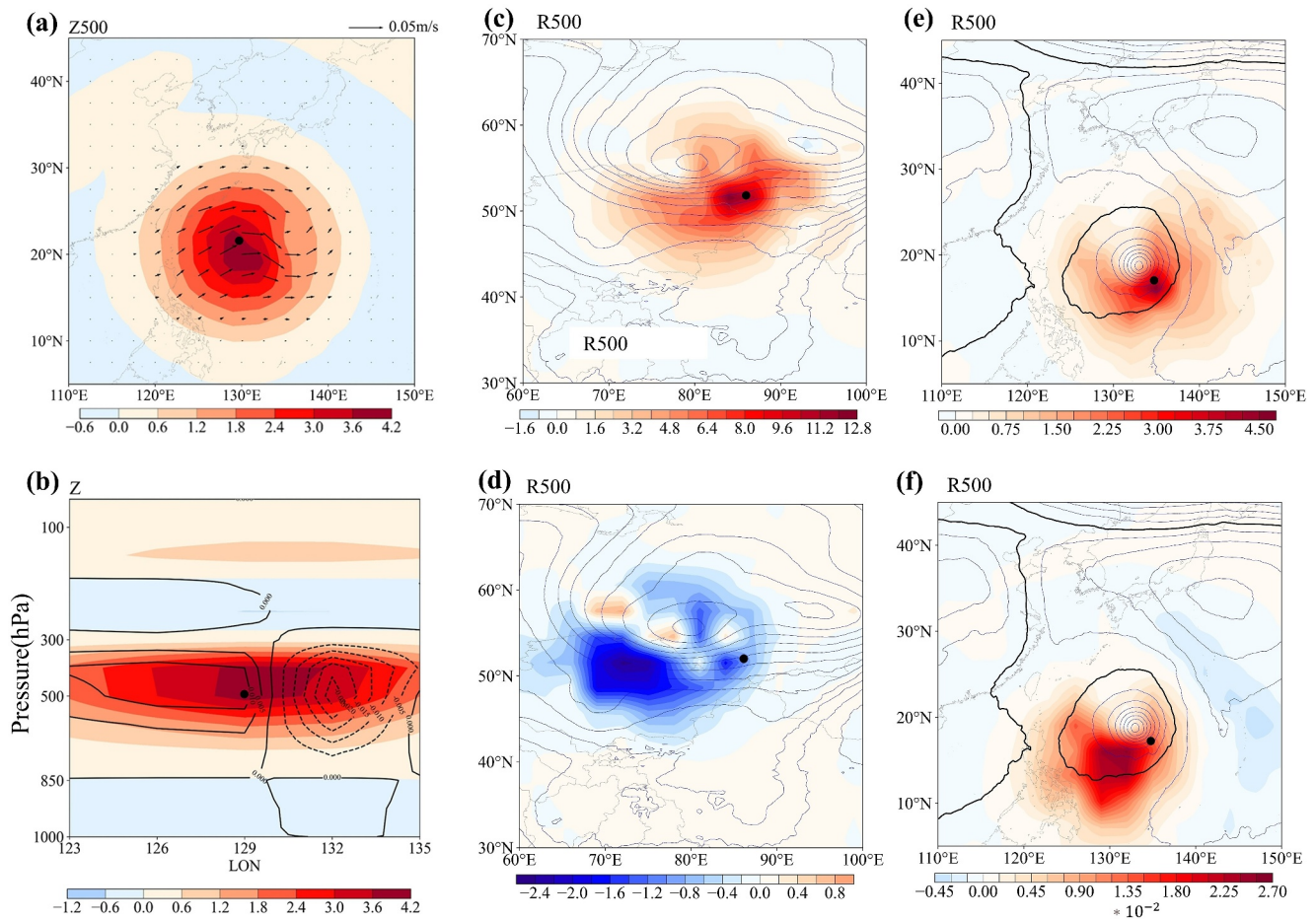
**Figure 2.** Analysis Increments. In Figure (a), the shaded area represents the Z500 analysis increment, with arrows indicating the increments of the wind field at 500 hPa pressure level. In Figure (b), the shaded area represents the vertical increment of the meridional geopotential, with black contour lines indicating the increments of the v-wind. Figures (c) and (e) represent the R500 increments at the observation time of 0000 UTC 30 July 2023, while Figures (d) and (f) represent the R500 increments at the observation time of 0600 UTC 30 July 2023. The observation point for Figures (a) and (b) is at 21°N, 129°E (black dot), for Figures (c) and (d) is at 52°N, 86°E (black dot), while for Figures (e) and (f), it is at 17°N, 135°E (black dot). The contour line represent the 500 hPa geopotential height of background field, with the thick solid line corresponding to the 5,880 line for Figures (b,c,d,e).

### 3.2. Propagation of Observational Information Over Time

Single-point observation experiments elucidate how data collected at various times contributes to the initial states via the assimilation system. EXP_2 demonstrates the system's capacity for such integration.

The first observation point, located ahead of the trough and behind the ridge (for details on OMB and OMA for EXP_2, see Table S1 in Supporting Information S1). Notably, when observations are exclusively added at the initial time, the analysis increment concentrates near this point (Figure 2c). In contrast, observation placed solely at the end of the assimilation window results in higher analysis increment values southwest of the observation point (Figure 2d), consistent with the influence of southwesterly winds of background fields.

The second observation point, located in the southeast quadrant of the typhoon, exhibits similar patterns. Initial observation results in analysis increments near this point (Figure 2e), whereas observations at the end of the window yields increments predominantly located upstream of the observation point (Figure 2f), reflecting a cyclonic structure consistent with the background field. These observations confirms the assimilation system's capability in accurately propagating and integrating data into the initial field according to the laws of temporal evolution, thereby adjusting the analysis field to represent flow-dependent characteristics effectively. The evolution of the vertical profile of the increments over time is illustrated Figure S1 in Supporting Information S1.
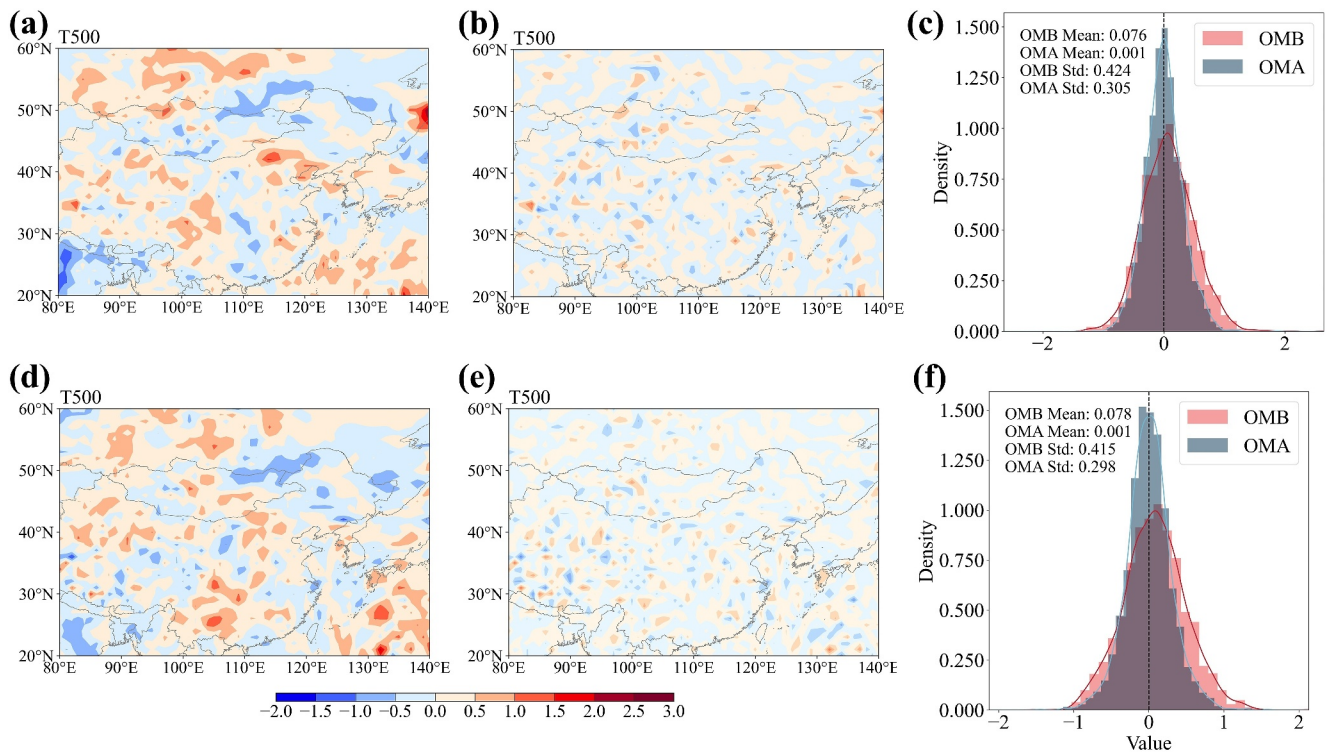
**Figure 3.** OMB and OMA of T500. Figure (a) represents OMB $\left(\mathbf{y}_0 - \boldsymbol{H}_0(\mathbf{x}_b)\right)$, while Figure (b) represents OMA $\left(\mathbf{y}_0 - \boldsymbol{H}_0(\mathbf{x}_a)\right)$, with Figure (c) displaying the PDF distribution of Figure (a) and (b). Similarly, Figure (d) depicts OMB_6 $\left(\mathbf{y}_6 - \boldsymbol{H}_6(\boldsymbol{M}_6(\mathbf{x}_b))\right)$, Figure (e) depicts OMA_6 $\left(\mathbf{y}_6 - \boldsymbol{H}_6(\boldsymbol{M}_6(\mathbf{x}_a))\right)$, and Figure (f) displays the PDF distribution of Figures (d) and (e).

### 3.3. Effectiveness of the FuXi-En4DVar

In EXP_3, we validate the effectiveness of the system from multiple perspectives. First, within FuXi-En4DVar framework, the convergence of the assimilation optimization problem is examined. The objective function $J$ decreases with an increase in the number of optimization iterations, as shown in Figure S2a in Supporting Information S1. This demonstrates the validity of gradients calculated via automatic differentiation, and confirms their convergence. Moreover, the increases in $J_b$ (Equation 7) and the decreases in $J_o$ where $J_o = J_{o0} + J_{o6}$ are also reasonable (Figure S2b in Supporting Information S1), underscoring a balanced adjustment between the background field and observations.

Subsequently, we verify whether the observational information is effectively assimilated by the system within the observation space. For instance, the distribution of OMB for T500, depicted at the beginning of the assimilation window in Figures 3a and 3ure 3d, shows negligible bias. After assimilating the T500, the OMA distributions for T500, as shown in Figures 3b and 3e, demonstrate that the absolute values of OMA for the majority of T500 observation points are smaller than OMB, closer to 0. For example, in regions from 20°N to 30°N and 80°E to 90°E, OMB exhibits a significant negative bias (Figure 3a). After assimilating the T500 (Figure 3b), the corresponding positions of OMA no longer exhibit a significant negative bias. Furthermore, the probability distribution function (PDF) of OMA (Figures 3c and 3f) more closely approximates a normal distribution with a mean of 0 and a reduced standard deviation. Similar trends were also observed for R500, U500 and V500 variables, though these are not shown here. Additionally, analysis increments in the model space were explored beyond the observation space, as seen in Figures 4a and 4b. The analysis increments occurred mainly within the observation area, delineated by gray dashed lines, without any apparent spurious correlations. And the distribution of increments is similar to that of OMB (Figure 3a), indicating a reasonable adjustment for background field and suggesting alignment with the observational data.

Finally, the proximity of the analysis field to the target data (ERA5), was evaluated using the root mean squared error (RMSE) of the analysis field relative to the background field (Figure 4e). RMSE calculations are focused on
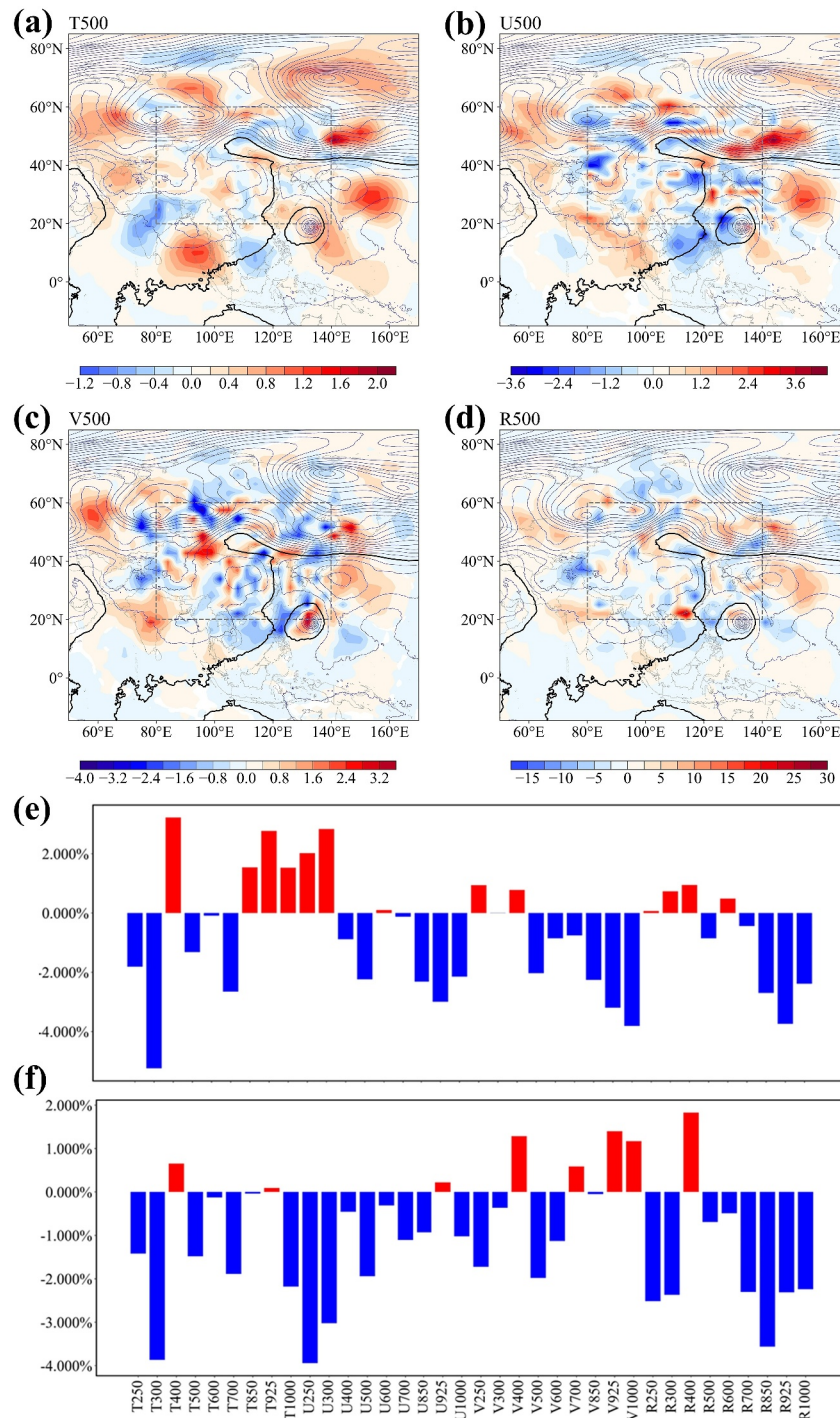
**Figure 4.** Analysis increments for T500 (a), U500 (b), V500 (c) and R500 (d). The blue thin lines represent contours of geopotential height at 500 hPa of background field, while the black solid line represents the contour at 5,880 m. The gray dashed line delineates the observation area. Figures (e), (f) show the percentage improvement in RMSE of the analysis field in the observation area compared to the RMSE of the background field $\left( Improvment = \frac{RMSE_{x_a} - RMSE_{x_b}}{RMSE_{x_b}} \right)$. Figure (e) corresponds to EXP_3, and Figure (f) corresponds to EXP_4.

the 0.25°grids within the observation area to highlight observational impact, revealing decreased RMSE of T500, U500, V500 and R500 and most other variables also showed reduced RMSE, demonstrating the system's ability to effectively propagate observational information in a reasonable manner. RMSE calculations excluded variables above 200 hPa, surface variables, and geopotential, as ensemble members constructed using Perlin noise may not

effectively capture correlations between upper-level and mid-to-lower-level. Additionally, ERA5 reanalysis surface variables exhibit relatively large uncertainty (Haiden et al., 2023). Additionally, in a simulation closer to real observation scenarios, temperature, wind, and relative humidity fields at 250 hPa, 300 hPa, 500 hPa, 850 hPa and 1,000 hPa (EXP_4) are assimilated, similar to the dropsonde observation. The results, as shown in Figure 4f, showing improvements across nearly all variables.

## 4. Conclusion and Discussion

This paper presents the development of a FuXi-En4DVar system, which incorporates the FuXi model's global weather forecasting capabilities with En4DVar assimilation techniques. This integration potentially allows ML-based weather forecasting models to operate independently from traditional NWP assimilation systems.

Unlike traditional NWP DA systems that rely on complex adjoint models, the FuXi model leverages automatic differentiation to automatically compute gradients like ML. This approach simplifies obtaining the optimal solution for the En4DVar optimization problem, as demonstrated by the equivalence of gradients obtained through automatic differentiation and adjoint models. FuXi-En4DVar incorporates observational data into the model state, adhering to constraints imposed by time-evolutionary laws and physical properties. Specially, single-point experiments using observations from ERA5 demonstrate that the FuXi-En4DVar system effectively constrains analysis increments through physical balance relationships, and propagates observational information to the initial field, displaying flow-dependent characteristics. Regional assimilation experiments using observations from ERA5 have further validated the system's effectiveness.

The results confirm the rationality and effectiveness of the FuXi-En4DVar system we developed. While the methodology for constructing such a system remains consistent across different ML-based weather forecasting models, there are areas for enhancement. For instance, ensemble members generated with Perlin noise inadequately capture variable correlations, such as spurious correlations, necessitating a large number of members and complex localization strategies for constructing the **B** matrix. Moreover, the use of ERA5 data, rather than actual observations, and a simplistic observational operator limit the system's ability to incorporate non-conventional data sources like satellite and radar.

Future work will focus on optimizing ensemble member selection to improve the **B** matrix and integrating unconventional observations using advanced machine learning techniques. For instance, neural networks could be employed to emulate the radiative transfer model (Liang et al., 2022), modeling complex observational operators which are crucial components of satellite DA. This would allow for a more complex observation error covariance matrix and could lead to the replacement of traditional components in DA systems with more sophisticated ML-driven alternatives. Through these efforts, ML-based weather forecasting models, encompassing both prediction and assimilation modules, may eventually replace traditional NWP systems.

## Data Availability Statement

The ERA5 reanalysis data is available at https://www.ecmwf.int/en/forecasts/dataset/ecmwf-reanalysis-v5 (Hersbach et al., 2020).

## References

Bannister, R. N. (2017). A review of operational methods of variational and ensemble-variational data assimilation. *Quarterly Journal of the Royal Meteorological Society*, *143*(703), 607–633. https://doi.org/10.1002/qj.2982

Bi, K., Xie, L., Zhang, H., Chen, X., Gu, X., & Tian, Q. (2023). Accurate medium-range global weather forecasting with 3d neural networks. *Nature*, *619*(7970), 533–538. https://doi.org/10.1038/s41586-023-06185-3

Bosilovich, M. G., & Schubert, S. D. (2002). Water vapor tracers as diagnostics of the regional hydrologic cycle. *Journal of Hydrometeorology*, *3*(2), 149–165. https://doi.org/10.1175/1525-7541(2002)003<0149:wvtado>2.0.co;2

Bouttier, F., & Courtier, P. (2002). *Data assimilation concepts and methods*. ECMWF.

Buehner, M., Houtekamer, P., Charette, C., Mitchell, H. L., & He, B. (2010). Intercomparison of variational data assimilation and the ensemble Kalman filter for global deterministic nwp. part i: Description and single-observation experiments. *Monthly Weather Review*, *138*(5), 1550–1566. https://doi.org/10.1175/2009mwr3157.1

Chen, K., Bai, L., Ling, F., Ye, P., Chen, T., Luo, J.-J., et al. (2024). Towards an end-to-end artificial intelligence driven global weather forecasting system. Retrieved from https://arxiv.org/abs/2312.12462

Chen, K., Han, T., Gong, J., Bai, L., Ling, F., Luo, J.-J., et al. (2023a). Fengwu: Pushing the skillful global medium-range weather forecast beyond 10 days lead. Retrieved from https://arxiv.org/abs/2304.02948

Chen, L., Zhong, X., Zhang, F., Cheng, Y., Xu, Y., Qi, Y., & Li, H. (2023b). Fuxi: A cascade machine learning forecasting system for 15-day global weather forecast. *npj Climate and Atmospheric Science*, *6*(1), 190. https://doi.org/10.1038/s41612-023-00512-1

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: A journal of the atmospheric sciences, applied meteorology and physical oceanography*, *131*(613), 3385–3396. https://doi.org/10.1256/qj.05.108

Evensen, G. (2003). The ensemble kalman filter: Theoretical formulation and practical implementation. *Ocean Dynamics*, *53*(4), 343–367. https://doi.org/10.1007/s10236-003-0036-9

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

Gauthier, P., Tanguay, M., Laroche, S., Pellerin, S., & Morneau, J. (2007). Extension of 3dvar to 4dvar: Implementation of 4dvar at the meteorological service of Canada. *Monthly Weather Review*, *135*(6), 2339–2354. https://doi.org/10.1175/mwr3394.1

Gauthier, P., & Thepaut, J.-N. (2001). Impact of the digital filter as a weak constraint in the preoperational 4dvar assimilation system of météo-France. *Monthly Weather Review*, *129*(8), 2089–2102. https://doi.org/10.1175/1520-0493(2001)129<2089:iotdfa>2.0.co;2

Haiden, T., Janousek, M., Vitart, F., Ben-Bouallegue, Z., & Prates, F. (2023). Evaluation of ecmwf forecasts, including the 2023 upgrade (No. 911). *ECMWF*. Retrieved from https://doi.org/10.21957/d47ba5263c

Hamill, T. M., Whitaker, J. S., & Snyder, C. (2001). Distance-dependent filtering of background error covariance estimates in an ensemble kalman filter. *Monthly Weather Review*, *129*(11), 2776–2790. https://doi.org/10.1175/1520-0493(2001)129<2776:ddfobe>2.0.co;2

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020). The era5 global reanalysis [dataset]. *Wiley Online Library*, *146*(730), 1999–2049. Retrieved from https://doi.org/10.1002/qj.3803

Holton, J. R., & Hakim, G. J. (2012). *An introduction to dynamic meteorology*. Academic press.

Hu, Y., Chen, L., Wang, Z., & Li, H. (2023). Swinvrnn: A data-driven ensemble forecasting model via learned distribution perturbation. *Journal of Advances in Modeling Earth Systems*, *15*(2), e2022MS003211. https://doi.org/10.1029/2022ms003211

Kalnay, E. (2003). *Atmospheric modeling, data assimilation and predictability*. Cambridge university press.

Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Alet, F., et al. (2023). Learning skillful medium-range global weather forecasting. *Science*, *382*(6677), 1416–1421. https://doi.org/10.1126/science.adi2336

Le Dimet, F.-X., & Talagrand, O. (1986). Variational algorithms for analysis and assimilation of meteorological observations: Theoretical aspects. *Tellus A: Dynamic Meteorology and Oceanography*, *38*(2), 97–110. https://doi.org/10.1111/j.1600-0870.1986.tb00459.x

Lewis, J. M., & Derber, J. C. (1985). The use of adjoint equations to solve a variational adjustment problem with advective constraints. *Tellus*, *37*(4), 309–322. https://doi.org/10.1111/j.1600-0870.1985.tb00430.x

Liang, X., Garrett, K., Liu, Q., Maddy, E. S., Ide, K., & Boukabara, S. (2022). A deep-learning-based microwave radiative transfer emulator for data assimilation and remote sensing. *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *15*, 8819–8833. https://doi.org/10.1109/jstars.2022.3210491

Liu, C., Xiao, Q., & Wang, B. (2008). An ensemble-based four-dimensional variational data assimilation scheme. part i: Technical formulation and preliminary test. *Monthly Weather Review*, *136*(9), 3363–3373. https://doi.org/10.1175/2008mwr2312.1

Liu, C., Xiao, Q., & Wang, B. (2009). An ensemble-based four-dimensional variational data assimilation scheme. part ii: Observing system simulation experiments with advanced research wrf (arw). *Monthly Weather Review*, *137*(5), 1687–1704. https://doi.org/10.1175/2008mwr2699.1

Liu, D. C., & Nocedal, J. (1989). On the limited memory bfgs method for large scale optimization. *Mathematical Programming*, *45*(1), 503–528. https://doi.org/10.1007/bf01589116

Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., et al. (2017). Automatic differentiation in pytorch.

Pathak, J., Subramanian, S., Harrington, P., Raja, S., Chattopadhyay, A., Mardani, M., et al. (2022). *Fourcastnet: A global data-driven high-resolution weather model using adaptive fourier neural operators*. arXiv preprint arXiv:2202.11214.

Perlin, K. (1985). An image synthesizer. *ACM Siggraph Computer Graphics*, *19*(3), 287–296. https://doi.org/10.1145/325165.325247

Rabier, F., Järvinen, H., Klinker, E., Mahfouf, J.-F., & Simmons, A. (2000). The ecmwf operational implementation of four-dimensional variational assimilation. i: Experimental results with simplified physics. *Quarterly Journal of the Royal Meteorological Society*, *126*(564), 1143–1170. https://doi.org/10.1002/qj.49712656415

Rawlins, F., Ballard, S., Bovis, K., Clayton, A., Li, D., Inverarity, G., et al. (2007). The met office global four-dimensional variational data assimilation scheme. *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography*, *133*(623), 347–362. https://doi.org/10.1002/qj.32

Xiao, Y., Bai, L., Xue, W., Chen, K., Han, T., & Ouyang, W. (2023). Fengwu-4dvar: Coupling the data-driven weather forecasting model with 4d variational assimilation. *arXiv preprint arXiv:2312.12455*.

Xu, X., Sun, X., Han, W., Zhong, X., Chen, L., & Li, H. (2024). Fuxi-da: A generalized deep learning data assimilation framework for assimilating satellite observations. Retrieved from https://arxiv.org/abs/2404.08522

Zhang, L., Liu, Y., Liu, Y., Gong, J., Lu, H., Jin, Z., et al. (2019). The operational global four-dimensional variational data assimilation system at the China meteorological administration. *Quarterly Journal of the Royal Meteorological Society*, *145*(722), 1882–1896. https://doi.org/10.1002/qj.3533