A Hybrid 3DVar-EnKF Data Assimilation Approach Based on Multilayer Perceptron

1st Lilan Huang
College of Meteorology and Oceanography, College of Computer
National University of Defense Technology,
Changsha, China
huanglilan18@nudt.edu.cn

2nd Hongze Leng
College of Meteorology and Oceanography
National University of Defense Technology,
Changsha, China
hzleng@nudt.edu.cn

3rd Junqiang Song
College of Meteorology and Oceanography
National University of Defense Technology,
Changsha, China
junqiang@nudt.edu.cn

4th Juan Zhao
College of Meteorology and Oceanography
National University of Defense Technology,
Changsha, China
zhaojuan@nudt.edu.cn

5th Rui Chen
College of Computer
National University of Defense Technology,
Changsha, China
chenrui16@nudt.edu.cn

6th Dongzi Wang
College of Computer
National University of Defense Technology,
Changsha, China
wangdongzi18@nudt.edu.cn

Abstract—The quality and accuracy of Numerical Weather Prediction (NWP) is based on its initial conditions (ICs), boundary conditions and forecast models. Data assimilation (DA) is a crucial procedure to optimally estimate the actual atmospheric state (known as the analysis field) as ICs for NWP by integrating available information, including the observation and the background field. Instead of only focusing on the speed-up for DA in virtue of the customized neural networks, this paper exploratively introduces the spatial-temporal peculiarities to construct a new hybrid data assimilation approach based on multilayer perceptron (MLP); and, its effectiveness and validity are verified in two classical nonlinear dynamic models. The results of experiments demonstrate that the Cache-MLP generally produces similar or smaller root mean square errors (RMSE) with much less time consuming, compared to the conventional 3D-Var and EnKF DA methods, and noticeably, the Cache-MLP has a more robust representation of turning points in the trajectories of the state variables. The final Backtracked-MLP learns a propriate weight matrix to couple previous two traditional DA methods and increases the accuracy by 10.32% in the Lorenz-63 system while 14.03% in the Lorenz-96 system, in comparison with the empirical hybrid DA method. To some extent, this method could be a reference to further researches to optimize the quality of the analysis field, in the meantime, saving significant computing time and resources by deep learning.

Index Terms—Data Assimilation, MLP, deep learning, three-dimensional variational data assimilation (3D-Var), ensemble Kalman filter (EnKF)

I. Introduction

Numerical Weather Prediction (NWP) was defined as an initial/boundary value problem by Vihelm Bjerknes(1904)[1]. The future state of the atmosphere is determined by its detailed initial state and boundary conditions, together with forecast models. Data assimilation (DA) could be described as a process to integrate all the available sources——the observation and the previous forecast(a.k.a. the background field or first guess), to obtain the "analysis field", which is hoped to be the best estimate of the probability density for the actual atmospheric state. Namely, the pursuit of DA is to produce the most accurate initial conditions (ICs) for NWP.

Based on Bayes' Theorem[2], two mainstream methods have flourished in the past three decades. One is variational DA (Var), another is ensemble DA, which is based upon the ensemble Kalman filter (EnKF). Basically, to deal with uncertainty from diverse information, these methods are guided by a same strategy, using fresh observations to amend the first guess at the specified grid-point[3], namely, combining the background fields and observations by minimizing a cost function that is a summation of two aspects: one part that penalizes the distance to the background mean and another penalizes the distance to the observations. However, the specific implementation varies from each other. First of all, Var assimilates straightway a wide range of observations by giving a static background-error covariance matrix ($\boldsymbol{B}$), which is prescribed by once in accordance with climatological data. Hence, the static $\boldsymbol{B}$ is incapable to represent properly the true spatio-temporal error statistics. In terms of time, Var is divided into 3D-Var[2] and 4D-Var[4], the former utilizes single moment data, while the latter has data from a time window of a given length to estimate an initial state. Although, in 4D-Var $\boldsymbol{B}$ evolves implicitly with the flow-dependence and could gain a

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

much better initial state than 3D-Var, it needs expensive calculations and relies on linearized and adjoint versions of the numerical model, which are hard to establish. Secondly, EnKF[5] offers ensemble-derived $\boldsymbol{B}$, which is dramatic and changes as the system evolves, outperforms $\boldsymbol{B}$ in 3D-Var. Also, the EnKF does not require linearized or adjoint versions of the model, which saves a huge computing overhead. Variational and ensemble methods have brought benefits to DA, but each has its own unsolved limitations[6]. In order to maximize the benefits and eliminate the inadequacies of two separate methods, there appears a bunch of techniques to incorporate them. The coupling of these methods named as the hybrid data assimilation. Hybrid DA has been developed to accentuate the best features of each source of background-error information, including the statistically robust background-error covariances of Var and the flow-dependency of EnKF. The alliance of two background-error covariances to get a recalibrated and more robust $\boldsymbol{B}$ has been experimentally proved to be effective[7, 8]. That is, Hybrid DA methods are competitive with pure Var or ensemble methods. Nevertheless, different contribution ratios of $\boldsymbol{B}$ from Var and ensemble DA directly affect the quality of the "analysis field"[9]. Dealing with the uncertainty and huge numbers of degrees of freedom $\boldsymbol{B}$(typically $\geq$ O(10^7)) in DA systems takes great efforts, although, national NWP centers have the capability to implement Hybrid DA in operational systems[10], they use empirical proportions to ease tremendous computational cost yet sacrifice the best accuracy. The scheme of Var and ensemble $\boldsymbol{B}$ accounting for half of each was adopted by the tasks of the UK Met Office[11], while using one-third of the former and two-thirds of the latter was tested by the GSI assimilation system in the United States[12].

Machine learning (ML) is a data-driven method, which can contribute to solving the space-time related problems of Earth system science to some extent[13]. Among them, the functions of the "physical" sub-model, which simulate the evolution of the Earth system, refers to that based on physical theory, empirical parameters are adopted, and semi-empirical property exists, so this "physical" sub-model can be replaced by a ML model. Naturally, the nonlinear neural networks could link to dynamical models to form hybrid neural–dynamical models to be a new class of variational data assimilation[14]. Based on a neural network method, Liaqat et al.[15] developed a method, which is proposed for approximating dynamical systems via weak constraint data assimilation formulation and completed the simulation in the weakly and highly nonlinear cases of the Lorenz system. Cintra et al.[16–19] presented an approach for employing artificial neural networks (NNs) to emulate the local ensemble transform Kalman filter (LETKF) as a method of DA. One data assimilation cycle used MLP-DA and LETKF with the same synthetic observations to test the SPEEDY model, an atmospheric general circulation model. Resulting in a similar assimilation quality, the CPU-time cycle assimilation with MLP-DA is 90 times faster than LETKF.

However, the ultimate pursuit of DA is to obtain the optimal "analysis field" quickly and accurately to serve NWP. By using a deluge of meteorological data, the explored neural network DA methods, which are mentioned previously, have only accelerated the assimilation process by simulating the existing DA methods without promoting the performance. The principal reason is that they just focused on the statistical features of data but ignored the physical properties of the atmospheric state itself. The assignment of this article is using Cache-MLP neural network to simulate and optimize the traditional DA methods——3D-Var and EnKF employing the physical laws of atmospheric motion and the temporal characteristics of atmospheric state variables. Among them, the Lorenz 63 and the Lorenz 96 classic chaotic models are used for testing. Furthermore, utilizing a Backtracked-MLP neural network to couple 3D-Var and EnKF through learning adaptive combinative weight matrix. In comparison with pure 3D-Var and EnKF, also the experimental hybrid method, this Hybrid DA optimal approach can efficiently enhance performance, which is measured by root-mean-square error (RMSE) between the "analysis field" and the true state of forecast models.

The structure of this article is as follows. In section 2 we briefly introduce the related theories of the paper: 3D-Var, EnKF, MLP. In section 3 we describe the methodology of a Hybrid DA approach which is based on a Cache-Backtracked neural network. In section 4 we test the validity of the Hybrid DA approach by Lorenz model. Finally, conclusions and discussions are given.

II. Related Theories

A. Three-dimensional variational data assimilation (3D-Var)

Based on Bayes' theorem, Var has been the workhorse of DA in NWP for many years before 1990s[20]. In general, 3D-Var estimates a single initial state of an NWP model. The "analysis field" is found by minimizing a cost function J:

$$\begin{aligned} J(\boldsymbol{x}) =& J_b + J_o \\ =& \frac{1}{2}[(\boldsymbol{x}-\boldsymbol{x^b})^T\boldsymbol{B}^{-1}(\boldsymbol{x}-\boldsymbol{x^b}) \\ &+ (\boldsymbol{y^o}-h(\boldsymbol{x}))^T\boldsymbol{R}^{-1}(\boldsymbol{y^o}-h(\boldsymbol{x}))] \end{aligned} \tag{1}$$

The interpretation of this function J is as follows.

- J_b measures the deviation between $\boldsymbol{x}$ and $\boldsymbol{x^b}$ to optimize the fit of the initial state to $\boldsymbol{B}$.
- J_o measures the deviation between $\boldsymbol{y^o}$ and $\boldsymbol{H(x)}$ to optimize the fit of the model's version of the observation to the real observation by the observation-error covariance matrix ($\boldsymbol{R}$).

Using the nonlinear conjugate gradient method to find the minimum of eq.(1), then the Jacobian vector of eq.(2) can be verified as:

$$\bigtriangledown J(\boldsymbol{x}) = \boldsymbol{B}^{-1}(\boldsymbol{x}-\boldsymbol{x^b}) - \boldsymbol{H^T}\boldsymbol{R}^{-1}(\boldsymbol{y^o}-h(\boldsymbol{x})) = 0 \quad (2)$$

where $\boldsymbol{H}$ is defined as:

$$\boldsymbol{H} = \frac{\partial h}{\partial x} \quad (3)$$

According to eq.(2), $\boldsymbol{x^a}$ is the state that minimizes J and is given as the solution to

$$[\boldsymbol{B}^{-1} + \boldsymbol{H^T}\boldsymbol{R}^{-1}h]\boldsymbol{x^a} = \boldsymbol{B}^{-1}\boldsymbol{x^b} + \boldsymbol{H^T}\boldsymbol{R}^{-1}\boldsymbol{y} \quad (4)$$

So, the best "analysis field" is:

$$\begin{aligned}\boldsymbol{x^a} &= [\boldsymbol{B}^{-1} + \boldsymbol{H^T}\boldsymbol{R}^{-1}h]^{-1}[\boldsymbol{B}^{-1}\boldsymbol{x^b} + \boldsymbol{H^T}\boldsymbol{R}^{-1}\boldsymbol{y}] \\ &= [\boldsymbol{B}^{-1} + \boldsymbol{H^T}\boldsymbol{R}^{-1}h]^{-1}[\boldsymbol{B}^{-1}\boldsymbol{x^b} + \boldsymbol{H^T}\boldsymbol{R}^{-1}h(\boldsymbol{x^b}) \\ &- \boldsymbol{H^T}\boldsymbol{R}^{-1}h(\boldsymbol{x^b}) + \boldsymbol{H^T}\boldsymbol{R}^{-1}\boldsymbol{y}] \\ &= \boldsymbol{x^b} + [\boldsymbol{B}^{-1} + \boldsymbol{H^T}\boldsymbol{R}^{-1}h]^{-1}\boldsymbol{H^T}\boldsymbol{R}^{-1}(\boldsymbol{y}-h(\boldsymbol{x^b}))\end{aligned} \quad (5)$$

This minimization could be regarded as an inverse problem and it is hard to find the inverse of high-dimensional matrix $\boldsymbol{B}$ and $\boldsymbol{R}$. Instead of directly solving eq.(5), using optimization methods (such as Newton method, quasi-Newton method, BFGS method, LBFGS method, etc.) to get the approximate minimum value.

B. Ensemble Kalman filter (EnKF)

The forecast step is

$$\begin{aligned}\boldsymbol{x^f}(t_i) &= M_{i-1}[\boldsymbol{x^a}(t_{i-1})] \\ \boldsymbol{P^f}(t_i) &= L_{i-1}\boldsymbol{P^a}(t_{i-1})L_{i-1}^T + \boldsymbol{Q}(t_{i-1})\end{aligned} \quad (6)$$

The analysis step is

$$\begin{aligned}\boldsymbol{x^a}(t_i) &= \boldsymbol{x^f}(t_i) + \boldsymbol{K}_i\boldsymbol{d}_i \\ \boldsymbol{P^a}(t_i) &= (\boldsymbol{I} - \boldsymbol{K}_i\boldsymbol{H}_i)\boldsymbol{P^f}(t_i)\end{aligned} \quad (7)$$

where

$$\begin{aligned}\boldsymbol{d}_i &= \boldsymbol{y_i^o} - h(\boldsymbol{x^f}(t_i)) \\ \boldsymbol{K}_i &= \boldsymbol{P^f}(t_i)\boldsymbol{H_i^T}[\boldsymbol{R}_i + \boldsymbol{H_i}\boldsymbol{P^f}(t_i)\boldsymbol{H^T}]^{-1}\end{aligned} \quad (8)$$

$\boldsymbol{P}^f$ is estimated by an ensemble of N data assimilation systems, which is carried out simultaneously. That is, after completing the ensemble of analyses at time t_{i-1} and the N forecasts $\boldsymbol{x_n^f}(t_i)$, $\boldsymbol{P}^f$ can be obtained approximately by

$$\boldsymbol{P}^f \approx \frac{1}{N-1}\sum_{n=1}^{N}(\boldsymbol{x_n^f} - \overline{\boldsymbol{x}}^{\boldsymbol{f}})(\boldsymbol{x_n^f} - \overline{\boldsymbol{x}}^{\boldsymbol{f}})^T \quad (9)$$

where the overbar in eq.(9) indicates the sample average.

A summary of notation of this article is provided in Table 1.

TABLE I
Summary of key notation used in this article.

Symbol	Description
$\boldsymbol{x}$	State vector(^{b}background,^{a}analysis,$_{(i)}$ensemble member)
$\boldsymbol{y^o}$, $\boldsymbol{y}$	Observations
h	Linear observation operator
$\boldsymbol{B}$, $\boldsymbol{P^f}$	Background-error covariance matrix
$\boldsymbol{R}$	Observation-error covariance matrix
J, J_b, J_o	Cost function($_b$background,$_o$observation)
M	(Forecast) Model
L	Linear tangent model
$\boldsymbol{P^a}$	Analysis-error covariance matrix
$\boldsymbol{Q}$	Model-error covariance matrix
$\boldsymbol{K}$	Kalman gain or weight matrix
d	Innovation or observation increment
N	Ensemble size

C. Multilayer Perceptron (MLP)

Based on the simulation of the human brain, Rosenblatt[21] designed the first neural network——the perceptron through the simple abstraction of biological neural cells. The perceptron is a linear classifier, that is, for the linear separable data, the perceptron can learn to fit the better linear equation. However, the expression capability of single neuron perceptron model is not enough to solve more complex problems, especially for high nonlinear numerical problems. Multilayer Perceptron, which is also called a feed-forward artificial neural network (ANN), has a full connection structure, except the input layer, each node can be seen as a processing unit with a nonlinear activation function. If each activation function is a linear function, each node of MLP in any layer can be simplified as a perceptron. Namely, MLP is a kind of classical deep neural network, which is a universal approximator. That is, providing sufficiently many hidden units, MLP with one hidden layer using appropriate activation functions are capable of approximating any measurable function $f(x)$[22, 23].

The structure of the perceptron and an example of MLP are shown in Fig.1.

III. Methodology

A. The Cache-MLP

In operational systems, using short-range forecasts as a first guess ($\boldsymbol{x^b}$) and incorporating the information from the fresh observations ($\boldsymbol{y^o}$) to obtain the "analysis field" ($\boldsymbol{x^a}$) is called an "analysis cycle"(Fig.2).

In 3D-Var or EnKF, it defines the problem of DA to find the best weight(ω) to couple information from two sources at time t as in eq.(10).

$$\boldsymbol{x_t^a} = \boldsymbol{x_t^b} + \boldsymbol{\omega} * (\boldsymbol{y_t^o} - h(\boldsymbol{x_t^b})) \quad (10)$$

The Cache-MLP is based upon the traditional "analysis cycle" and captures the training data set which consists of the first guess and observations to optimize the "analysis field". Instead of relying on the data of only one single time,

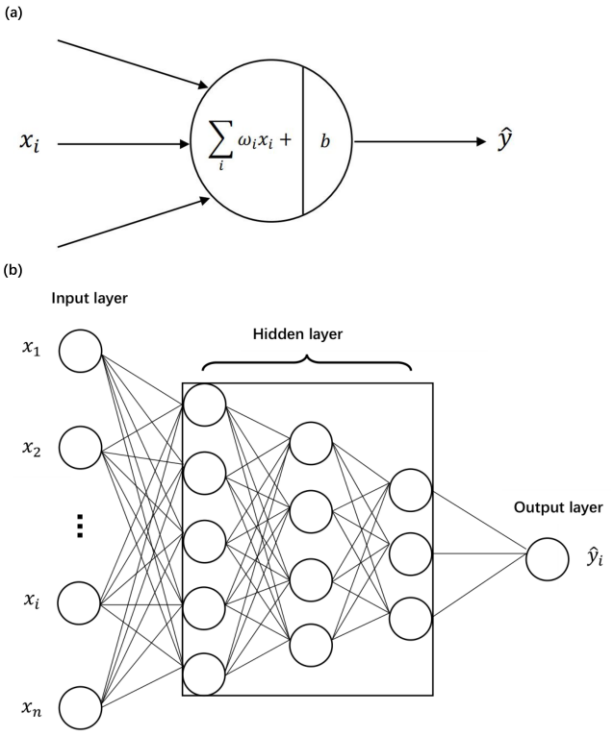

Fig. 1. (a)The structure of the perceptron.(b)An example of a MLP

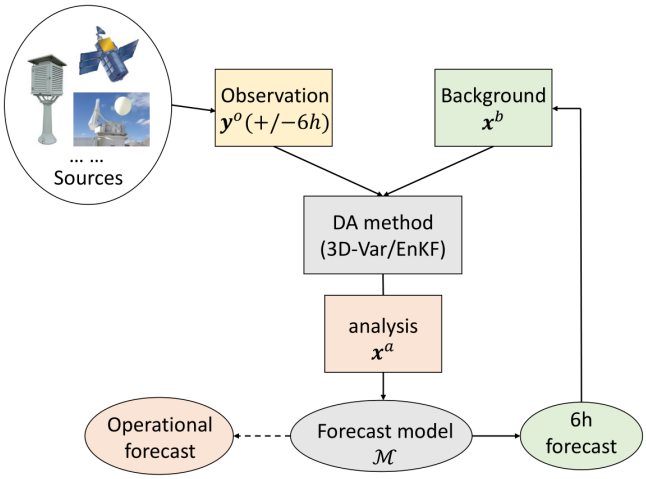

Fig. 2. Data assimilation cycle.

in accordance with the flow-dependency of atmospheric variables, $\boldsymbol{x^a}$ at time t can be affected by at least the previous five time steps data ($\boldsymbol{x^a_{t-5}}$,$\boldsymbol{x^a_{t-1}}$). Together with data at time t, it constitutes the input variables of the Cache-MLP, as a means of introducing the temporal features contained in meteorological data itself which are exactly crucial. Hereafter, the Cache-MLP solves the optimization problem of DA by eq.(11).

$$\boldsymbol{x_t^a} = \sum_{i=0}^{5} \boldsymbol{\omega_{t-i} x_{t-i}^b} + \sum_{i=0}^{5} \boldsymbol{\nu_{t-i} y_{t-i}^o} + bias \qquad (11)$$

where $\boldsymbol{\omega}$ and $\boldsymbol{\nu}$ are weight matrixes for $\boldsymbol{x^b}$ and $\boldsymbol{y^o}$, respectively. The structure of the Cache-MLP is shown in Fig.4.

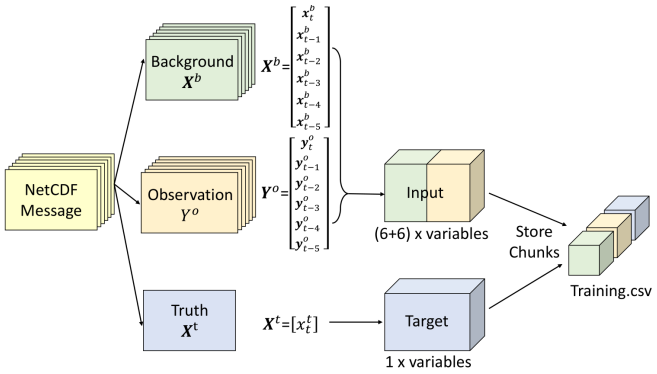

Fig. 3. The structure of dataset preparation.

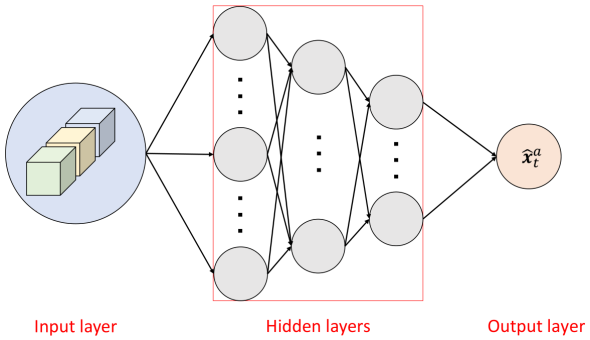

Fig. 4. Schematic diagram of the Cache-MLP.

The Cache-MLP configuration was defined by empirical tests, and we found the following characteristics:

1) One input node for a vector, which is obtained by caching the observations and the background information at 6 times;
2) One output node for the analysis vector results at time t;
3) Three hidden layer and the number of neurons in each hidden layer is obtained by numerical experimentation;
4) The rectifier linear unit (ReLU) as the activation function to obtain the nonlinear results;
5) Learning rate η=0.0005;
6) The loss function is mean square error(MSE);
7) Training stops when the loss reaches 10^{-5}.

B. The Backtracked-MLP

In traditional hybrid DA methods, researchers exploded various techniques to couple pure Var and ensemble DA method, and the quality of the "analysis field" was improved indeed by experimental confirmation. Inspired by the idea of hybrid, the Backtracked-MLP uses the results from the trained Cache-MLPs as the training data set. The structure of the Backtracked-MLP is shown in

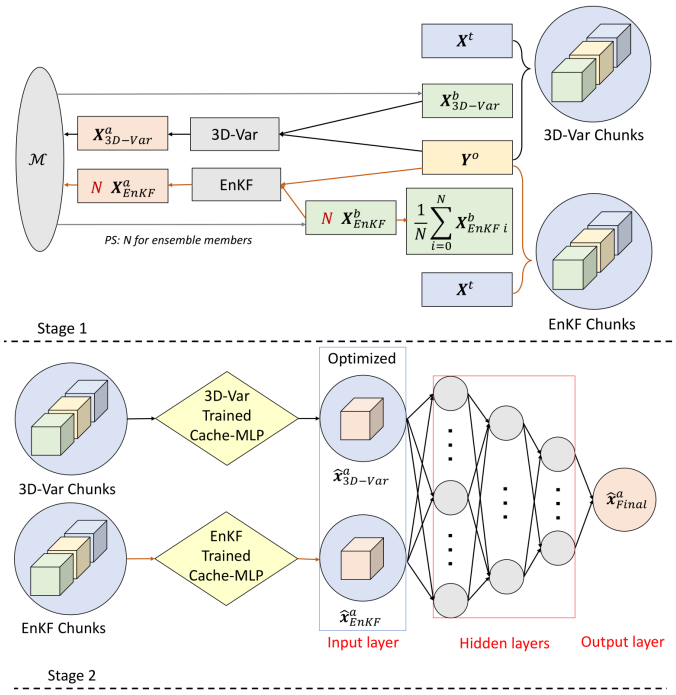

Fig. 5. Schematic diagram of the Backtracked-MLP.

Fig.5. The explanation of the operational steps is as follows.

1) Stage 1: recalling the training data sets, which are used to train the Cache-MLP neural networks, and concentrating the 3D-Var chunks and the EnKF chunks;
2) Stage 2: updating the 3D-Var chunks and the EnKF chunks, that is, regarding them as the "new" inputs of the completed and fixed Cache-MLP to get the output vectors——the preliminary optimizations of the "analysis fields" respectively; Stacking the optimized $\hat{\boldsymbol{x}}^{\boldsymbol{a}}$ as the input vectors of the Backtracked-MLP;
3) The structure of the Backtracked-MLP is similar to the Cache-MLP, also the number of neurons of each hidden layer is designed by empirical tests.

Finally, the mathematical expression of the hybrid data assimilation approach to calculate $\boldsymbol{x}^{\boldsymbol{a}}$ can be defined simply as

$$\begin{aligned} \hat{\boldsymbol{x}}^{\boldsymbol{a}}_{\boldsymbol{3Dvar}} &= CacheMLP(3Dvar(\boldsymbol{x}^{\boldsymbol{b}}_{\boldsymbol{i}}, \boldsymbol{y}^{\boldsymbol{o}}_{\boldsymbol{i}})) \\ \hat{\boldsymbol{x}}^{\boldsymbol{a}}_{\boldsymbol{EnKF}} &= CacheMLP(EnKF(\boldsymbol{x}^{\boldsymbol{b}}_{\boldsymbol{i}}, \boldsymbol{y}^{\boldsymbol{o}}_{\boldsymbol{i}})) \\ \boldsymbol{x}^{\boldsymbol{a}}_{\boldsymbol{t}} &= BacktrackedMLP(\hat{\boldsymbol{x}}^{\boldsymbol{a}}_{\boldsymbol{3D-Var}}, \hat{\boldsymbol{x}}^{\boldsymbol{a}}_{\boldsymbol{EnKF}}) \end{aligned} \tag{12}$$

where $i \in [t-5, t]$.

IV. Experiment and result

In this section, we utilized two classical nonlinear chaotic dynamical systems, which have been used as verification models in many researches[10], to test the validity of the proposed the new hybrid DA method based on MLP.

A. Lorenz-63——A low dimensional chaotic system

Lorenz-63 system is a simplified mathematical model for atmospheric convection[24], and offers the advantages of strong non-linear interactions among three variables. The evolution of states in Lorenz-63 system can be given by three ordinary differential equations:

$$\begin{aligned} \frac{dx}{dt} &= \sigma(y-x) \\ \frac{dy}{dt} &= x(\rho - z) - y \\ \frac{dz}{dt} &= xy - \beta z \end{aligned} \tag{13}$$

where the parameters are set to the standard values to produce a chaotic regime:σ=10,β=$\frac{8}{3}$,ρ=28.

4.1.1 Data preparation

It is hypothesized here that Lorenz-63 system is perfect, there is no model error so that model-error covariance matrix $\boldsymbol{Q}$=0; given the initial state $\boldsymbol{x_0}$=1.508870,$\boldsymbol{y_0}$=-1.537121,$\boldsymbol{z_0}$=25.46091 and the Lorenz-63 model is integrated in time using a fourth-order Runge-Kutta time difference scheme with a time step dt=0.01, the number of analog integration steps 111000, we can obtain the "true fields" ($\boldsymbol{x^t}$) by integrating the equations in eq.(13) without data assimilation. In real atmospheric models, the ground truths of variables generally represent $\boldsymbol{x^t}$. Observation errors chosen randomly from a Gaussian distribution are added to the true state to gather the approximate observations. Besides, we set linear observation operator $h = I$ and observation-error covariance matrix $\boldsymbol{R} = I$, where I is the third order unit matrix. Similarly, with the same initial conditions, the data assimilation system by using conventional DA method 3D-Var (or EnKF) completes 111000 times assimilation "analysis cycles" and generate the "analysis fields" ($\boldsymbol{x^a}$). Among them, the Lorenz-63 system is integrated for the first 1000 time steps of spinning up, with the aim of reaching steady-state and to obtain the simulated atmosphere; the middle 100000 steps are used to prepare the training sets for the Cache-MLP, and the last 10000 steps are the test data.

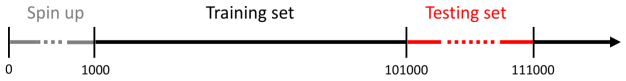

Fig. 6. Schematic diagram of data preparation.

Additionally, known the initial state, which is a point in the state of Lorenz-63 chaotic system, the static background error covariance $\boldsymbol{B}$ for 3D-Var can be simplified calculated by the NMC method.

Before the experiment, it must be mentioned that considering the computational cost, the number of ensemble members is too small in comparison with the number of numerical model variables in real Numerical Weather Prediction systems. In empirical verify, the number of actual model variables is over 10^7, while the ensemble size

is set to 80-100 in EnKF normally. Therefore, the statistics will not guarantee to be representatives of the real states. That is, because of the limited ensemble members, EnKF would bring the sampling error that cannot be ignored in practice[25]. These problems can be lessened with a procedure called ensemble inflation[26] by multiplicative processes[27]. Sensitivity experiments were conducted to find the optimal size of ensemble and the multiplicative factor for EnKF to generate the training data set with the least error for the Cache-MLP. Hint, RMSE is widely used in data assimilation to measure the similarity between $\boldsymbol{x^t}$ and $\boldsymbol{x^a}$, which is defined as:

$$RMSE = \sqrt{\frac{1}{NT}\sum_{i=0}^{T}\sum_{j=1}^{N}((\boldsymbol{x_j^r})_i - (\boldsymbol{x_j^a})_i)^2} \qquad (14)$$

where i is the ith state, j is the jth ensemble member.

The results of sensitivity experiments are shown as follows.

As shown in Fig.7(a), keeping the multiplicative factor equals to 1.01, and changing the number of ensemble members, it can be seen that with the increasement of ensemble members, the sampling error is relieved significantly. In consideration of there are only three variables in Lorenz-63 system and the efficiency of the further experiment, it is reasonable to choose 20 as the size of ensemble members (N=20); In Fig.7(b), keeping N=20 and changing the multiplicative factor from 1 to 1.1 with an interval of 0.01. It can be seen that when the factor is 1.01, RMSE reaches the minimum value 0.1406. After sensitivity experiments, the parameters of EnKF N=20 and factor=1.01 are selected ultimately for subsequent experiments.

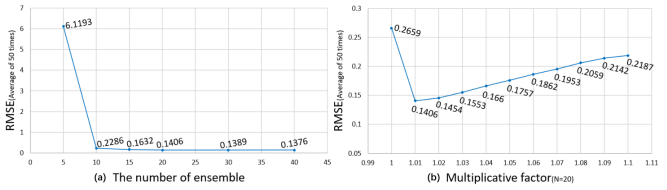

Fig. 7. The results of sensitivity experiments for selecting (a)ensemble member. (b)multiplicative factor.

4.1.2 The optimal and simulative process of the Cache-MLP

Setting the same initial state for 3D-Var and EnKF data assimilation systems, which is x_0'=0.7055775, y_0'=2.264438504, z_0'=20.89017307, that is the 101001st chaos state obtained after 101000 steps of integration with the initial state as mentioned in the process of data preparation. Then, we continue to execute 10000 more "analysis cycles" 50 times to get ready for testing data sets for the Cache-MLP and calculate the RMSE for each data set. For the purpose of avoiding the randomness of the experiment, we repeat 50 times. Fig.8 shows the differences of various DA methods. Comparing Fig.8(a) and Fig.8(c), it is clear that having 20 ensemble members and taking the expansion factor 1.01, EnKF has much better performance than 3D-Var, whose mean RMSE is 0.1454 while 0.4705 in 3D-Var. Among them, utilizing the Cache-MLP, in Fig.8(a) and Fig.8(b), the performance of 3D-Var can be improved 39.85 percentage and the optimized mean RMSE for 50 times is 0.2830. Although, the Cache-MLP, its performance of optimization is inconspicuous when processing the EnKF data set, but it is able to simulate EnKF well and gets almost the same result according to Fig.8(c) and Fig.8(d).

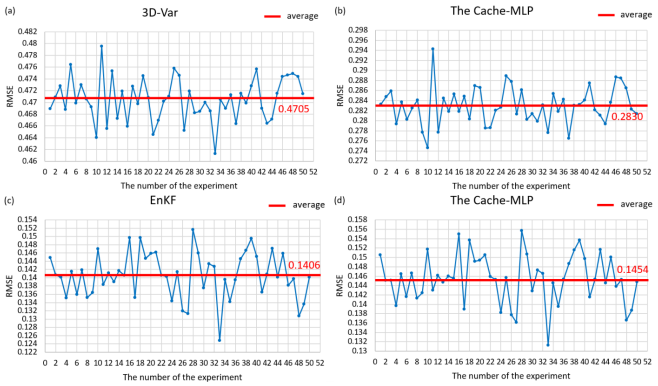

Fig. 8. The RMSE of 50 times experients with various DA methods for Lorenz-63.(a)Pure 3D-Var.(b)The Cache-MLP for 3D-Var.(c)Pure EnKF.(d)The Cache-MLP for EnKF.(The solid red line and value represent the average RMSE of 50 times)

Furthermore, in order to figure out the main improved behavior in the Cache-MLP, the trajectories of the "analysis fields" from conventional 3D-Var DA and DA with NNs optimizations are shown. First of all, from a rough view on the Fig.9, the trajectories of either 3D-Var DA or the Cache-MLP DA(MLP+DA) can track the real state in all three axes (x,y,z) on a 10000-long timestep. Then, choosing one axis and its three partial sequences of trajectories in the beginning (from 0 to 300), in the middle (from 5000 to 5300) and in the end (from 9700 to 10000) of the timestep respectively to have a sharper image, we discover that both 3D-Var DA and MLP+DA perform well in the monotonically increasing or decreasing parts(Fig.9(a) and Fig.9(e)). However, MLP+DA is superior to 3D-Var DA in where the derivative of the trajectory is 0, colloquially, at the corner(Fig.9(b), Fig.9(c) and Fig.9(d)). In other words, MLP+DA can seize the unusual tendency much more sensitive.

4.1.3 The hybrid DA based on the Backtracked-MLP

After training the Cache-MLP model, using it to optimize its training data sets to prepare for the Backtracked-MLP and redoing the optimal process, the benefit is obvious and can be seen in the following statistic pictures. Previously, a simple way to decide the empirical weight to couple the "analysis fields" of 3D-Var and EnKF is giving those two methods a fixed proportion, which is traversing from 0.1 to 1 with the increase rate 0.1, the

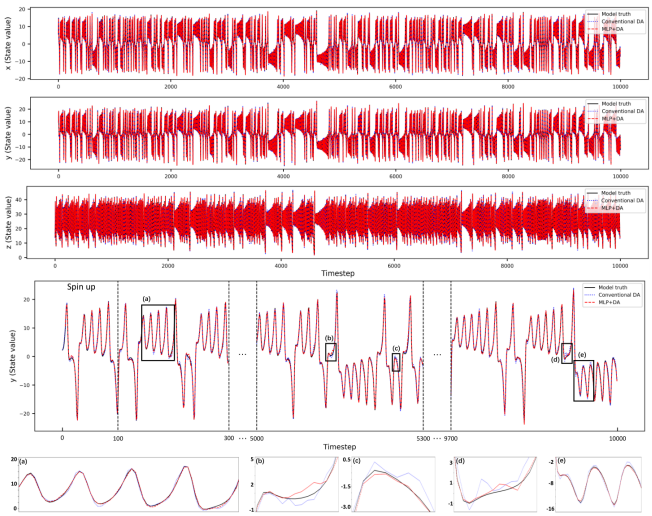

Fig. 9. The trajectories of state values in Lorenz-63.(The black solid line is the model x^t, the blue dotted line is the x^a of pure 3D-Var , and the red dashed line is the x^a of the Cache-MLP for 3D-Var.)

summation of the total rate is 1 in the meantime, that is rate(3DVar)+rate(EnKF)=1. In accordance with the figure below, when the ratio of $\boldsymbol{x^a}$ from 3D-Var is 10 percent and the rest of $\boldsymbol{x^a}$ from EnKF is 0.9, we can get the experimental minimization mean of RMSE, which is 0.1386(Fig.10).

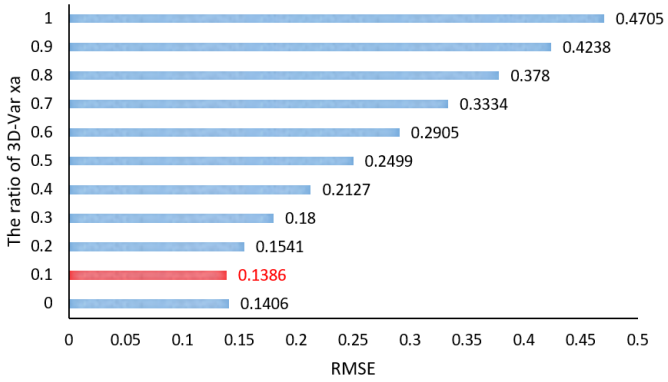

Fig. 10. The results of the empirical method for Lorenz-63.

But in comparison to Fig.11(a) and Fig.11(b), the Backtracked-MLP defeats the best performance of mathematical statistics by learning new optimal weight matrix through neural networks. And the result of the mean RMSE is 0.1243, which improves performance by 10.32 percent.

B. Lorenz-96——A high dimensional chaotic system

Compared with Lorenze63, Lorenz-96[28] is a much more complicated nonlinear dynamical system. Its function can be defined as eq.(15). Among them, the first term on the right-hand side is an advection term, while the second term represents damping. Besides, F represents an external forcing constant and is set to 8, which is a

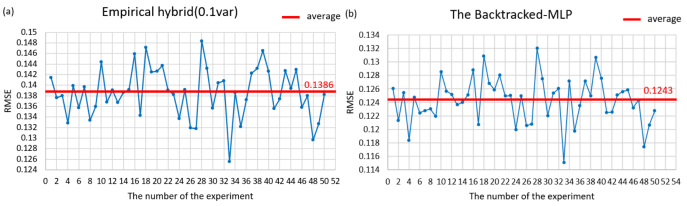

Fig. 11. The RMSE of 50 times experiments comparison between the empirical method and the Backtracked-MLP for Lorenz-63.(The solid red line and value represent the average RMSE of 50 times)

general value known to cause chaotic behavior. Lorenz-96 is a periodic system which is used to simulate the time evolution of atmospheric variables commonly.

$$\frac{dx_j}{dt} = (x_{j+1} - x_{j-2})x_{j-1} - x_j + F, j = 1, 2, ...J \quad (15)$$

where J is the number of state variables.

4.2.1 Data preparation

Also, assuming that the Lorenz-96 system is perfect. With the same setting as Lorenz-63 system, running 111000 analysis cycles to create the training data sets for the Cache-MLP, except that there are 40 state variables in Lorenz-96 while just 3 in Lorenz-63. That is J=40, F=8, dt=0.01. One primary change is to figure out another appropriate size of ensemble members, which has numerous effects on the accuracy of the "analysis field" and the execution time of the whole "analysis cycles", therefore a preliminary experiment has been completed and the result is shown in Table II and Fig.12.

TABLE II
The results of the preliminary experiment.

Ensemble numbers(N)	RMSE	time (s)	Performance Improvement (%)	Time increment (%)
10	0.5216	1547	/	/
20	0.3745	1860	28.20	20.23
40	0.3446	2531	7.98	36.08
60	0.3360	3474	2.50	37.26

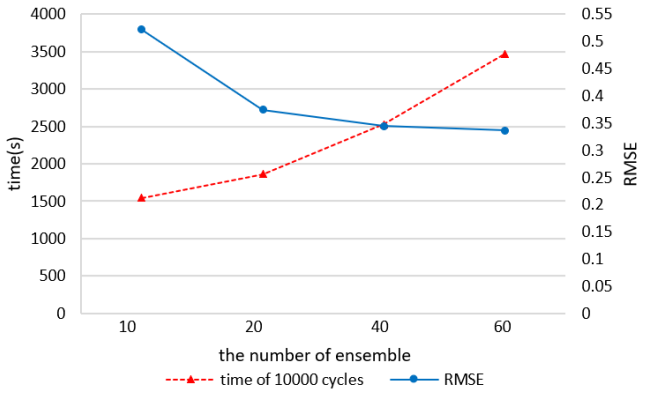

Fig. 12. The variations of RMSE by the time-consuming.

With the same default setting, namely, having the same initial state and using a fourth-order Runge-Kutta time difference scheme with a time step dt=0.01, apart from designing various size of ensemble members for the EnKF DA scheme, we finished 50 replicated experiments for each matching setting and got the following statistical table. Thereinto, the third column of the following table shows the mean of the total CPU time when achieving 10000 "analysis cycles". The fourth and fifth columns are calculated between two adjacent lines. For instance, comparing with N=40, performance improves 2.5% but time-consuming increases 37.26% when N=60. Considering the computational cost and the accuracy of the "analysis fields", N=40 is an advisable choice for the coming experiments, which can obtain comparatively lower RMSE with an acceptable time-consuming.

4.2.2 The optimal process of the Cache-MLP

The configuration of the Cache-MLP for the Lorenz-96 system is defined by empirical tests. Comparing Fig.13(a) with Fig.13(b) and Fig.13(c) with Fig.13(d) respectively, it is showing that regardless of different DA methods (3D-Var and EnKF), the Cache-MLP has an improvement of RMSE to some extent.

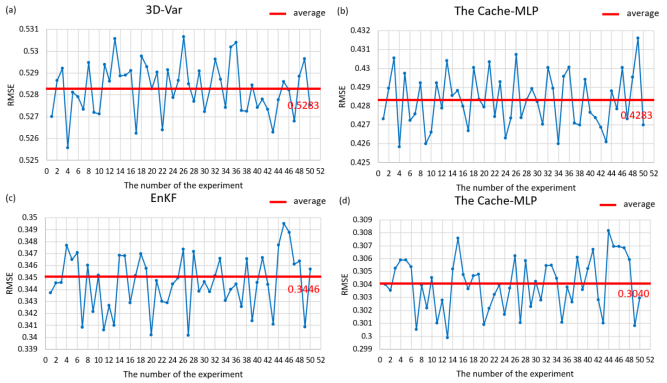

Fig. 13. The RMSE of 50 times experients with various DA methods for Lorenz-96.(a)Pure 3D-Var.(b)The Cache-MLP for 3D-Var.(c)Pure EnKF.(d)The Cache-MLP for EnKF.(The solid red line and value represent the average RMSE of 50 times)

Different from the Lorenz-63 system, there are 40 variables in the Lorenz-96 system, instead of showing the trajectories of all variables, we take 4 for instance, and draw only some small parts of the trajectories of the "analysis fields" from conventional EnKF DA and MLP+DA to go further analyze. It is obvious that each variable behaves badly in the spin up period in the Lorenz-96 system, while this circumstance is undiscovered in the Lorenz-63 system, this may be because with the increase of the number of variables, the uncertainty of the atmospheric state increases dramatically so that the Lorenz-96 system would take longer time to spin up and reach the balanced condition. After that, as shown above, the capability of the Cache-MLP verges on the EnKF DA, namely, their tracks almost coincide, just slightly away from the real state in a few same places. However, the mean RMSE of the traditional EnKF DA is 0.3446, comparatively, the Cache-MLP DA promotes a little with 0.3040. Also, from the track circled by a rectangle in each picture, having a more robust representation at the turning point may play a key role in the Cache-MLP DA to gain a small increment in performance.

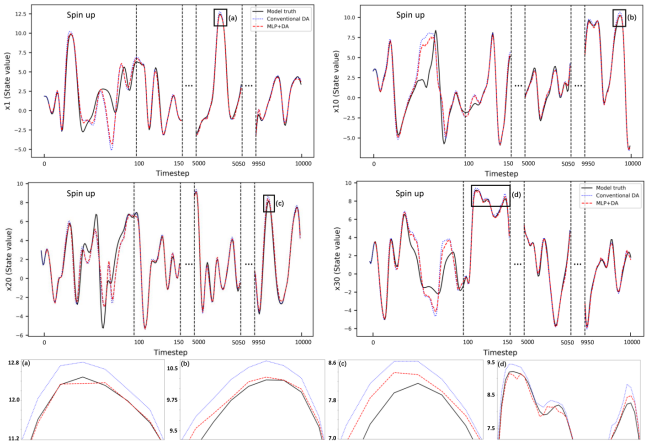

Fig. 14. The trajectories of state values in Lorenz-96.(The black solid line is the model x^t, the blue dotted line is the x^a of pure 3D-Var , and the red dashed line is the x^a of the Cache-MLP for 3D-Var.)

4.2.3 The hybrid DA besed on the Backtracked-MLP

In traditional and empirical methods, which is based on mathematical principle rate(3DVar)+rate(EnKF)=1, to capture the best performance of the "analysis field", it can be seen that the value of RMSE is the least while giving the weight of 30 percent for 3D-Var and the scalar is 0.2943(Fig.15).

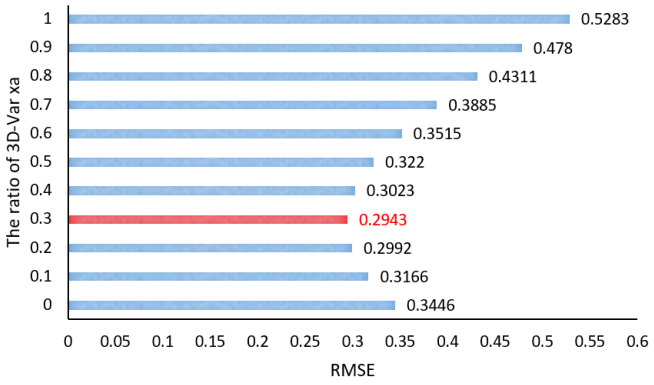

Fig. 15. The results of the empirical method for Lorenz-96.

However, taking advantage of the Backtracked-MLP, we can get a lower value of the mean RMSE for 50 times experiments, which is 0.2530, and it is promoted by 14.03 percent according to the Fig.16(a) and Fig.16(b).

The following Table III and Table IV summarize the experimental results.

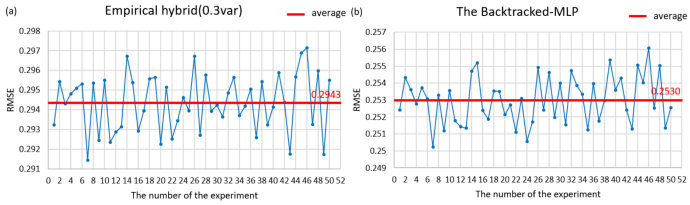

Fig. 16. The RMSE of 50 times experients comparison between the empirical method and the Backtracked-MLP for Lorenz-96.(The solid red line and value represent the average RMSE of 50 times)

TABLE III
Performance comparison between pure conventional DA methods and the Cache-MLP.

Model	DA method	Conventional RMSE	The Cache-MLP RMSE	Improvement (%)
Lorenz -63	3D-Var	0.4705	0.2830	39.85
	EnKF	0.1406	0.1454	-3.41
Lorenz -96	3D-Var	0.5283	0.4283	18.93
	EnKF	0.3446	0.3040	11.78

V. Conclusion and discussion

This paper proposes and verifies a hybrid 3DVar-EnKF DA approach based on MLP. In this research, DA can be defined as a problem of temporal sequences depends on historical meteorological records containing the observations and the first guess. Guided by spatial-temporal peculiarities, the Cache-MLP, the first part of the hybrid NNs, utilizes the 3D-Var chunks and EnKF chunks, which are stacked by six-time information as its training datasets. The model has fulfilled the mission of optimizing or simulating the results of the conventional 3D-Var and EnKF DA methods in two classical nonlinear dynamic models. Among them, in comparison with pure 3D-Var and EnKF DA in Lorenz-96 system, the Cache-MLP reaches the improvement of the accuracy of 18.93% and 11.78% respectively; while in Lorenz-63 system, the Cache-MLP has a significant promotion by 39.85% compared with pure 3D-Var DA, besides it is capable to imitate pure EnKF DA well only with slightly less 3.41%. Noticeably, the Cache-MLP has a more robust representation of turning points in the whole trajectory. In addition, the computational efficiency for all methods are shown in Table V, thereinto, the Cache-MLP is faster than others in varying degrees, since it just requires the calculation of forward propagation. Furthermore, the second part of the new hybrid NNs, the Backtracked-MLP challenges the existing mathematical hybrid DA method, uses the optimized "analysis fields", which are the results of the previous Cache-MLP as its training dataset to couple the "analysis fields" from 3D-Var and EnKF. The final Backtracked-MLP increases the accuracy by 10.32% in the Lorenz-63 system and 14.03% in the Lorenz-96 system.

TABLE IV
Performance comparison between empirical hybrid DA method and the Backtracked-MLP.

Model	Empirical hybrid RMSE	The Backtracked-MLP RMSE	Improvement (%)
Lorenz-63	0.1368	0.1243	10.32
Lorenz-96	0.2943	0.2530	14.03

TABLE V
Run time of 10000 analysis cycles(In seconds).

Model	The Cache-MLP	EnKF	3D-Var
Lorenz-63	0.02	15.12	46.06
Lorenz-96	0.37	1500.26	175.23

The effectiveness and success of the new hybrid DA approach based upon MLP in optimizing the "analysis fields" have proved the correctness of the idea of caching the previous information——the observations and the background fields to obtain the final analysis fields in DA. In the future, this creative thought and the Cache-MLP and the Backtracked-MLP neural networks can be applied in more extensive and complicated atmospheric models.

Acknowledgment

This research is partially supported by the National Natural Science Foundation of China, Grant Number(41605070).

References

[1] V. Bjerknes, "Das problem der wettervorhers-age, betrachtet vom standpunkte der mechanik und der physik," Meteor. Z., vol. 21, pp. 1–7, 1904.

[2] A. C. Lorenc, "Analysis methods for numerical weather prediction," Quarterly Journal of the Royal Meteorological Society, vol. 112, no. 474, pp. 1177–1194, 1986.

[3] C. Leith, "Numerical models of weather and climate," Plasma physics and controlled fusion, vol. 35, no. 8, p. 919, 1993.

[4] J. M. Lewis and J. C. Derber, "The use of adjoint equations to solve a variational adjustment problem with advective constraints," Tellus A, vol. 37, no. 4, pp. 309–322, 1985.

[5] P. Houtekamer, L. Lefaivre, J. Derome, H. Ritchie, and H. L. Mitchell, "A system simulation approach to ensemble prediction," Monthly Weather Review, vol. 124, no. 6, pp. 1225–1242, 1996.

[6] A. C. Lorenc, "The potential of the ensemble kalman filter for nwp—a comparison with 4d-var," Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography, vol. 129, no. 595, pp. 3183–3203, 2003.

[7] B. J. Etherton and C. H. Bishop, "Resilience of hybrid ensemble/3dvar analysis schemes to model error and

ensemble covariance error," Monthly weather review, vol. 132, no. 5, pp. 1065–1080, 2004.

[8] J. Poterjoy and F. Zhang, "Systematic comparison of four-dimensional data assimilation methods with and without the tangent linear model using hybrid background error covariance: E4dvar versus 4denvar," Monthly Weather Review, vol. 143, no. 5, pp. 1601–1621, 2015.

[9] M. Zhang and F. Zhang, "E4dvar: Coupling an ensemble kalman filter with four-dimensional variational data assimilation in a limited-area weather prediction model," Monthly Weather Review, vol. 140, no. 2, pp. 587–600, 2012.

[10] R. Bannister, "A review of operational methods of variational and ensemble-variational data assimilation," Quarterly Journal of the Royal Meteorological Society, vol. 143, no. 703, pp. 607–633, 2017.

[11] N. E. Bowler, A. Arribas, K. R. Mylne, K. B. Robertson, and S. E. Beare, "The mogreps short-range ensemble prediction system," Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography, vol. 134, no. 632, pp. 703–722, 2008.

[12] M. Demirtas, D. Barker, Y. Chen, J. Hacker, X.-Y. Huang, C. Snyder, and X. Wang, "A hybrid data assimilation (wrf-var and ensemble transform kalman filter) system based retrospective tests," in 2009-06-01]. http://www. mmm. ucar. edu/wrf/users/workshops/WS2009/abstracts/2A-10. pdf.

[13] M. Reichstein, G. Camps-Valls, B. Stevens, M. Jung, J. Denzler, N. Carvalhais et al., "Deep learning and process understanding for data-driven earth system science," Nature, vol. 566, no. 7743, pp. 195–204, 2019.

[14] W. W. Hsieh and B. Tang, "Applying neural network models to prediction and data analysis in meteorology and oceanography," Bulletin of the American Meteorological Society, vol. 79, no. 9, pp. 1855–1870, 1998.

[15] A. Liaqat, M. Fukuhara, and T. Takeda, "Applying a neural network collocation method to an incompletely known dynamical system via weak constraint data assimilation," Monthly weather review, vol. 131, no. 8, p. 1696, 2003.

[16] R. S. Cintra, H. Campos Velho, and H. C. Furtado, "Neural network for performance improvement in atmospheric prediction systems: data assimilation," in 1st BRICS Countries & 11th CBIC Brazilian Congress on Computational Intelligence. Location: Recife, Brasil. Porto de Galinhas Beach, 2013.

[17] R. Cintra, H. de Campos Velho, J. Anochi, and S. Cocke, "Data assimilation by artificial neural networks for the global fsu atmospheric model: Surface pressure," in 2015 Latin America Congress on Computational Intelligence (LA-CCI). IEEE, 2015, pp. 1–6.

[18] R. Cintra, H. de Campos Velho, and S. Cocke, "Tracking the model: Data assimilation by artificial neural network," in 2016 International Joint Conference on Neural Networks (IJCNN). IEEE, 2016, pp. 403–410.

[19] R. S. Cintra and H. F. de Campos Velho, "Data assimilation by artificial neural networks for an atmospheric general circulation model," Advanced Applications for Artificial Neural Networks, p. 265, 2018.

[20] F. Rawlins, S. Ballard, K. Bovis, A. Clayton, D. Li, G. Inverarity, A. Lorenc, and T. Payne, "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, vol. 133, no. 623, pp. 347–362, 2007.

[21] F. Rosenblatt, The perceptron, a perceiving and recognizing automaton Project Para. Cornell Aeronautical Laboratory, 1957.

[22] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," Neural networks, vol. 2, no. 5, pp. 359–366, 1989.

[23] G. Cybenko, "Approximation by superpositions of a sigmoidal function," Mathematics of control, signals and systems, vol. 2, no. 4, pp. 303–314, 1989.

[24] E. N. Lorenz, "Deterministic nonperiodic flow," Journal of the atmospheric sciences, vol. 20, no. 2, pp. 130–141, 1963.

[25] P. L. Houtekamer and H. L. Mitchell, "A sequential ensemble kalman filter for atmospheric data assimilation," Monthly Weather Review, vol. 129, no. 1, pp. 123–137, 2001.

[26] J. L. Anderson and S. L. Anderson, "A monte carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts," Monthly Weather Review, vol. 127, no. 12, pp. 2741–2758, 1999.

[27] J. S. Whitaker, T. M. Hamill, X. Wei, Y. Song, and Z. Toth, "Ensemble data assimilation with the ncep global forecast system," Monthly Weather Review, vol. 136, no. 2, pp. 463–482, 2008.

[28] E. N. Lorenz, "Predictability: A problem partly solved," in Proc. Seminar on predictability, vol. 1, no. 1, 1996.