Atmospheric Research 183 (2017) 1–16

Contents lists available at ScienceDirect

Atmospheric Research

journal homepage: www.elsevier.com/locate/atmosres

Improving precipitation forecast with hybrid 3DVar and time-lagged ensembles in a heavy rainfall event

Yuanbing Wang [a], Jinzhong Min [a,*], Yaodeng Chen [a], Xiang-Yu Huang [b], Mingjian Zeng [c], Xin Li [c]

[a] Key Laboratory of Meteorological Disaster of Ministry of Education (KLME)/Joint International Research Laboratory of Climate and Environment Change (ILCEC)/Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disasters, Nanjing University of Information Science & Technology, Nanjing 210044, China
[b] Centre for Climate Research Singapore, Meteorological Service, Singapore 537054, Singapore
[c] Jiangsu Research Institute of Meteorological Sciences, Nanjing 210009, China

ARTICLE INFO

Article history:
Received 14 December 2015
Received in revised form 21 June 2016
Accepted 23 July 2016
Available online 26 July 2016

Keywords:
Data assimilation
Hybrid
Time-lagged ensemble
Precipitation forecast

ABSTRACT

This study evaluates the performance of three-dimensional variational (3DVar) and a hybrid data assimilation system using time-lagged ensembles in a heavy rainfall event. The time-lagged ensembles are constructed by sampling from a moving time window of 3 h along a model trajectory, which is economical and easy to implement. The proposed hybrid data assimilation system introduces flow-dependent error covariance derived from time-lagged ensemble into variational cost function without significantly increasing computational cost. Single observation tests are performed to document characteristic of the hybrid system. The sensitivity of precipitation forecasts to ensemble covariance weight and localization scale is investigated. Additionally, the TLEn-Var is evaluated and compared to the ETKF(ensemble transformed Kalman filter)-based hybrid assimilation within a continuously cycling framework, through which new hybrid analyses are produced every 3 h over 10 days. The 24 h accumulated precipitation, moisture, wind are analyzed between 3DVar and the hybrid assimilation using time-lagged ensembles.

Results show that model states and precipitation forecast skill are improved by the hybrid assimilation using time-lagged ensembles compared with 3DVar. Simulation of the precipitable water and structure of the wind are also improved. Cyclonic wind increments are generated near the rainfall center, leading to an improved precipitation forecast. This study indicates that the hybrid data assimilation using time-lagged ensembles seems like a viable alternative or supplement in the complex models for some weather service agencies that have limited computing resources to conduct large size of ensembles.

1. Introduction

The development of society and economy has great requirement for numerical weather prediction (NWP), especially precipitation forecast. Accuracy of initial condition is one of the main conditions influencing forecast skill. The operational NWP centers usually enhance initial condition with data assimilation technique. In order to introduce effective synoptic information into initial condition, it is necessary to frequently assimilate recent observations for short-range NWP (Benjamin et al., 2004).

Several data assimilation methods have been proposed and applied in NWP, such as three-dimensional variational (3DVar), four-dimensional variational (4DVar), ensemble Kalman filter (EnKF) and the hybrid 3DVar or 4DVar (Hybrid) methods. Among these methods, 3DVar has been most widely used because of its advantage in computational cost and convenience to implement (Wu et al., 2002; Barker et al., 2004). However, 3DVar assumes that background error covariance is time-invariant, homogeneous and isotropic, which conflicts with the reality of "error of the day", i.e. flow-dependence. 4DVar allows implicit evolution of the background error covariance along with the adjoint model, but the use of static background error covariance at the start of each 4DVAR assimilation window represents a major limitation (Huang et al., 2009; Zhang et al., 2014). Furthermore, the computational cost of 4DVar is huge because of the adjoint and tangent linear models and development and maintenance of them is also a hard work. The EnKF method, in which the background error covariance is estimated from an ensemble of short-term forecasts, provides an alternative to variational data assimilation systems for its flow-dependence as well as convenience to implement (Evensen, 1994; Anderson, 2001; Bishop et al., 2001; Whitaker and Hamill, 2002; Hunt et al., 2007). For smaller ensembles, however, the EnKF is rank deficient and its background error covariance estimation suffers from a variety of sampling errors, including spurious correlations for widely separated locations (Hamill and Snyder, 2000).

The hybrid data assimilation method that couples ensemble-based and variational data assimilation systems has emerged as an alternative method (Hamill and Snyder, 2000) and become one of the research

* Corresponding author at: Nanjing University of Information Science & Technology, Nanjing 210044, China.
E-mail address: minjz@nuist.edu.cn (J. Min).

http://dx.doi.org/10.1016/j.atmosres.2016.07.026
0169-8095/

focuses in data assimilation field (Wang et al., 2008a,b; Zhang et al., 2013; Schwartz and Liu, 2014). The hybrid assimilation method incorporates flow-dependent background error covariance derived from ensemble into the variational cost function, so that the new background error covariance comes from a combination of traditional static error covariance with ensemble error covariance which can be used to update fields of observed model variables as well as fields of unobserved model variables (Keppenne et al., 2014). The hybrid covariance takes advantage of the relative strengths of both the EnKF and variational methods and ameliorates the problem of rank deficiency caused by limitation of ensemble size as well as isotropous, homogeneous and static covariance caused by assumptions in 3DVar. Hybrid data assimilation has been shown to be more robust than conventional ensemble data assimilation schemes, especially when ensemble size is small or the model error is large (Wang et al., 2007; Zhang et al., 2013) and has been implemented in several variational assimilation systems such as the Weather Research and Forecasting (WRF) model data assimilation system (WRFDA) (Barker et al., 2012) and the Gridpoint Statistical Interpolation (GSI) system which has become operational at National Centers for Environmental Prediction (NCEP) since 2012 (Wang et al., 2013). However, the cost is significantly higher than the one by none ensemble methods. For some weather service agencies, it is probably difficult to afford the cost of ensemble model integrations, thus implementation of hybrid assimilation may compromise between size of the ensemble and the resolution of the model.

Therefore, to avoid huge cost of ensemble integrations in hybrid assimilation system which assimilates observations and outputs forecast products with high frequency, the time-lagged forecasts of short range interval which are initiated from different past analysis times but verify at a same forecast time are directly pulled together from history storage to construct an ensemble in this study, namely time-lagged ensemble (Zhou et al., 2010). The time-lagged ensemble initially proposed as an alternative to Monte Carlo ensemble method (Hoffman and Kalnay, 1983) can be interpreted as forecasts obtained from a set of perturbed initial conditions. The initial conditions which initialize the time-lagged forecasts, the observations at different analysis times, the integration time and the lateral boundary conditions are all different for the time-lagged ensemble members, thus the flow-dependent forecast error can be a result of those conditions which cause the uncertainties (Lu et al., 2007; Vogel et al., 2014). This kind of ensemble is built at very low computational cost which does not require multiple integrations of the numerical model and holds promise for high-resolution applications. It has been widely used in many research and operational ensemble forecast systems (Yuan et al., 2008, 2009; Mittermaier, 2007; Trilaksono et al., 2012; Y. Chen et al., 2013; M. Chen et al., 2013; Jie et al., 2014, 2015). It is noted that the traditional and standard EnKF-based hybrid method should be a better choice in the presence of sufficient computing resource. However, if the computational cost associated with data assimilation like EnKF and ensemble integrations is not affordable, a compromise has to be made between the efficiency and accuracy. In such scenario, a hybrid approach by merging the time-lagged ensemble and 3DVar can be a choice because of its efficiency and flow-dependent feature.

In this study, we construct a hybrid data assimilation system based upon WRFDA using the time-lagged ensembles. To evaluate the effectiveness and flow-dependence of the background error covariance derived from time-lagged ensembles in precipitation forecast, this system is applied and tested in a heavy rainfall event occurred in east China and compared with ETKF-based hybrid assimilation within a continuously cycling framework over 10 days. Details are present in the rest of the paper which is organized as follows. In Section 2, the basic methodology of the WRFDA-based hybrid assimilation using time-lagged ensembles ("TLEn-Var" for short) is introduced. Section 3 describes the rainfall event and Section 4 details the model and data assimilation configurations as well as the experiment design. Results are presented in Section 5 before we conclude in Section 6.

2. Methodology

The cost function of hybrid data assimilation in WRFDA is defined as

$$J(\delta\mathbf{x}_1,\boldsymbol{\alpha}) = \beta_1\frac{1}{2}\delta\mathbf{x}_1^T\mathbf{B}^{-1}\delta\mathbf{x}_1 + \beta_2\frac{1}{2}\boldsymbol{\alpha}^T\mathbf{A}^{-1}\boldsymbol{\alpha} + \frac{1}{2}\left(\mathbf{y}^{o\prime}-\mathbf{H}\delta\mathbf{x}\right)^T\mathbf{R}^{-1}\left(\mathbf{y}^{o\prime}-\mathbf{H}\delta\mathbf{x}\right) \tag{1}$$

In Eq. (1), the first term of right hand is the background term associated with the static covariance $\mathbf{B}$. The second term is associated with the ensemble covariance. $\boldsymbol{\alpha}$ is the ensemble extended control variable. $\mathbf{A}$ defines the spatial covariance of $\boldsymbol{\alpha}$. The third term is the observation term, $\mathbf{y}^{o\prime} = \mathbf{y}^o - H\mathbf{x}_b$ is the innovation, $\mathbf{y}^o$ denotes the observation, $\mathbf{x}_b$ is the background forecast, and H is the nonlinear observation operator. $\mathbf{H}$ is the linearized observation operator, and $\mathbf{R}$ is the observation error covariance. Factors β_1 and β_2 respectively define the weights placed on the static background error covariance and the ensemble covariance. β_1 and β_2 are constrained by $1/\beta_1 + 1/\beta_2 = 1$ to conserve the total background error variance.

The analysis increment of the hybrid is a sum of two terms, defined as

$$\delta\mathbf{x} = \delta\mathbf{x}_1 + \sum_{n=1}^{N}\left(\boldsymbol{\alpha}_n \circ \mathbf{x}_{n,b}^e\right) \tag{2}$$

where $\delta\mathbf{x}_1$ is the increment associated with the static background covariance, and the second term is the increment associated with the flow-dependent ensemble covariance. N is the ensemble size. For traditional (EnKF-based) hybrid assimilation, $\mathbf{x}_{n,b}^e$ is the n_{th} ensemble perturbation normalized by $\sqrt{N-1}$:

$$\mathbf{x}_{n,b}^e = \left(\mathbf{x}_{n,b} - \overline{\mathbf{x}_b}\right)/\sqrt{N-1} \tag{3}$$

where $\mathbf{x}_{n,b}$ is the n_{th} ensemble member and $\overline{\mathbf{x}_b}$ is the ensemble mean, provided by the ensemble forecast that is usually initialized by an EnKF data assimilation system.

However, for the hybrid assimilation using time-lagged ensembles, the flow-dependent ensemble error covariance are computed by differences of N previous instances of the model state vectors sampled from the recent history of the current model run. The differences between the time-lagged forecasts launched at different analysis times but verify at the same leading time are calculated and normalized by $\sqrt{N-1}$:

$$\mathbf{x}_{n,b}^e = \left(\mathbf{x}_j - \mathbf{x}_i\right)/\sqrt{N-1},\ 0 < i < j \le N \tag{4}$$

In Eq. (4), $\mathbf{x}_i$ and $\mathbf{x}_j$ are time-lagged ensemble members, the total number of $\mathbf{x}_{n,b}^e$ is calculate by $\sum_{i=1}^{N-1} i$. For example, in this study the deterministic forecast range is 48 h with an output interval of 3 h, then we can obtain 16 time-lagged ensemble members at each analysis time. The differences between time-lagged ensemble members are $\mathbf{x}_2 - \mathbf{x}_1$, $\mathbf{x}_3 - \mathbf{x}_1, \ldots, \mathbf{x}_{16} - \mathbf{x}_1, \mathbf{x}_3 - \mathbf{x}_2, \mathbf{x}_4 - \mathbf{x}_2, \ldots, \mathbf{x}_{16} - \mathbf{x}_2, \ldots$. Thus $\sum_{i=1}^{15} i = 120$ differences are generated. Next these differences can be used as the ensemble perturbations in a hybrid assimilation run (Fig. 1). The perturbations introduced here come from the time-lagged differences in a single model integration and the main goal of this method is to create ensembles with low computational cost, which are used for the calculation of flow-dependent error covariance.

The underlying assumption in the TLEn-Var is that forecast errors in data assimilation are primarily phase errors in time (Keppenne et al., 2014). The time-lagged forecasts in an ensemble are launched at different analysis times after assimilating different observations but verifies at the same leading time, and the integration time and lateral boundary between members are also different, taking the time evolution of forecast error of different time-lagged ensemble members into account.