https://doi.org/10.5194/egusphere-2026-394
Preprint. Discussion started: 22 April 2026


# DCU-accelerated 3DVAR data assimilation with automatic differentiation for WRF-Chem

Hancheng Ye [1], Zengliang Zang[1], Wei You[1], Yiwen Hu[1], Ning Liu[1], Yi Li[1]

[1]College of Meteorology and Oceanography, National University of Defense Technology, Changsha, China

**Correspondence**: Wei You (ywlx_1987@163.com) and Zengliang Zang (zzlqxxy@163.com)

**Abstract**: This study developed a PyTorch-based three-dimensional variational (3DVAR) data assimilation system (Py3DVAR) for the Weather Research and Forecasting model coupled to Chemistry (WRF-Chem), which integrates automatic differentiation (AD) to replace traditional manual gradient derivation and adopts Deep Computing Unit (DCU) acceleration for high computational efficiency. Py3DVAR enables the simultaneous assimilation of gaseous pollutants ($SO_2$, $NO_2$, CO, $O_3$) and particulate matter ($PM_{2.5}$, $PM_{10}$) and supports flexible deployment on both Central Processing Unit (CPU) and DCU computing platforms. To evaluate its performance and efficiency, idealized and real-case assimilation experiments (27 km and 9 km grid resolutions) were conducted, compared against a traditional CPU-parallelized Fortran-based 3DVAR system (Fortran-3DVAR). Idealized results show Py3DVAR effectively propagates observation information, generating increment fields consistent with Fortran-3DVAR. In real-case experiments, Py3DVAR substantially improves the model initial field quality: at 27 km resolution, correlation coefficients (CORR) for $SO_2$, $NO_2$, CO, $O_3$, $PM_{2.5}$, and $PM_{10}$ increased by 0.77, 0.51, 0.71, 0.98, 0.60, and 0.69, respectively; corresponding improvements at 9 km resolution are 0.78, 0.98, 0.66, 0.96, 0.63, and 0.78. The root mean square error (RMSE) and mean absolute error (MAE) are also significantly reduced, with analysis field accuracy comparable to Fortran-3DVAR. In terms of computational efficiency, Py3DVAR shows remarkable advantages: on the same CPU platform, the total iteration time at 27 km resolution is only 7.1 s, approximately 8.8 times faster than Fortran-3DVAR (62.5 s); on the DCU platform, the speedup reaches 32.7 times at 27 km and 40.3 times at 9 km. A 24-hour forecast test shows that the improved initial fields have sustained positive effects on short-term forecasts: the improvements persist for over 24 hours for $SO_2$, CO, $PM_{2.5}$, and $PM_{10}$, and for over 6 hours for $NO_2$ and $O_3$. This study confirms that Py3DVAR achieves order-of-magnitude gains in computational efficiency while maintaining accuracy equivalent to traditional assimilation algorithms, providing a flexible new technical pathway for operational atmospheric chemical data assimilation and future intelligent assimilation systems.

https://doi.org/10.5194/egusphere-2026-394
Preprint. Discussion started: 22 April 2026


## 1. Introduction

Air quality models (AQMs) are crucial tools for atmospheric pollution simulation and forecasting (Li et al., 2013; Wang et al., 2022; Gao et al., 2024). Errors in their initial conditions can accumulate during model integration, ultimately leading to significant simulation biases. Data assimilation technology, which merges multi-source observations (e.g., from surface stations and satellites) with model background fields using specific weighting, generates a theoretically optimal analysis field. This effectively improves the model's initial conditions and is a core technique for enhancing simulation accuracy, widely used in both research and operational forecasting (Engelen et al., 2009; Ma et al., 2025).

Compared to the long-standing development of meteorological data assimilation, atmospheric chemical data assimilation started relatively late, and its theoretical framework and operational applications are still under continuous improvement. Early explorations in atmospheric chemical data assimilation were largely constrained by limitations in observational capabilities and the ability of atmospheric chemistry models to simulate complex mixed aerosol components. Consequently, research initially focused on gaseous precursors like $SO_2$, $NO_2$, $O_3$, and key chemical components like $HNO_3$, which have relatively well-understood chemical mechanisms. The assimilation of particulate matter (PM) components was less common due to the high complexity of their sources and chemical mechanisms. Elbern et al. (1997) used the RADM2 (Second-Generation Regional Acid Deposition Model) gas-phase chemical mechanism and a variational method to assimilate ozone observations, successfully optimizing the initial concentrations of its precursors NOx and VOCs, and enabling quantitative analysis of ozone chemical sensitivity. In earlier work, Austin et al. (1992) assimilated LIMS satellite measurements of $HNO_3$ and $O_3$ into a 2D chemical model, effectively identifying and correcting systematic biases in the original chemical model, significantly improving the analysis accuracy of stratospheric chemical composition, and demonstrating the key role of four-dimensional assimilation in refining chemical models. Frydendall et al. (2009) developed a data assimilation algorithm based on statistical interpolation and coupled it with the Danish Eulerian Operational Model (DEOM), conducting assimilation studies on surface ozone and effectively improving forecast performance by optimizing assimilation parameters.

With advancements in observational technology and the continuous improvement of atmospheric chemistry modeling systems, the focus of atmospheric chemical data assimilation has gradually expanded to include particulate pollutants with more complex compositions and sources. In the early stages, limited by insufficient simulation of aerosol physicochemical processes, assimilation studies often focused on constraining models using total concentration observations like $PM_{2.5}$ and $PM_{10}$, employing relatively simple methods. For example, Denby et al. (2008) applied both Statistical Interpolation (SI) and Ensemble Kalman Filtering (EnKF) data assimilation methods to an atmospheric chemistry model. They found that both methods significantly corrected the model's systematic underestimation of PM10 concentration, with the SI method reducing the RMSE of daily

https://doi.org/10.5194/egusphere-2026-394
Preprint. Discussion started: 22 April 2026


average concentration from 16.7 μg $m^{-3}$ to 9.2 μg $m^{-3}$, and noted that the improvement from SI was significantly better than that from EnKF. Tombette et al. (2009) used the Optimal Interpolation method to assimilate surface $PM_{10}$ observations from France into a regional chemical transport model, effectively enhancing the short-term forecast accuracy of total aerosol mass ($PM_{10}$ and $PM_{2.5}$) across Europe, demonstrating the potential application of this method in operational forecasting.

As a core method in data assimilation, the 3DVAR approach has a clear theoretical foundation, mature computational framework, and stable system performance. In terms of implementation, its computational structure is clear, facilitating code development, and the overall computational cost is acceptable with modern computing resources (Jiang et al., 2013; Seo et al., 2023; Chen et al., 2024; Zhou et al., 2025). Given these advantages, it has been widely adopted by major international operational forecasting systems such as ECMWF and NCEP. Pagowski et al. (2010) introduced surface ozone and fine particulate matter (PM2.5) observations into the WRF-Chem model using the GSI (Grid-point Statistical Interpolation) 3DVAR system. Results showed that while assimilation significantly reduced forecast errors, these errors grew rapidly in the early forecast hours.

With the advancement of observational techniques and the refinement of model mechanisms, simply correcting total aerosol mass often fails to fundamentally optimize the simulation of its internal chemical composition. The target variables for aerosol assimilation are shifting from total mass concentration towards a multivariate system encompassing key chemical components such as sulfate, nitrate, and organic carbon. Li et al. (2013) integrated a 3DVAR algorithm with the WRF-Chem model employing the MOSAIC aerosol scheme. By assimilating PM2.5 and its chemical component observations, they achieved a comprehensive analysis of the concentrations and size distributions of eight key species, including black carbon, organic carbon, sulfate, and nitrate, significantly improving the simulation and forecast accuracy of these components. Jiang et al. (2013) implemented the assimilation of PM10 observations using the 3DVAR method, providing an improved initial field for the WRF-Chem model by optimizing various aerosol components. However, the assimilation effect diminished rapidly due to vertical mixing and advective processes, highlighting the importance of optimizing the aerosol vertical initial state. By designing various assimilation scenarios, Zang et al. (2016) confirmed that introducing aircraft-measured aerosol data into traditional surface-based aerosol data assimilation significantly optimized aerosol mass concentration, vertical distribution, and chemical composition. They also found that the forecast skill improvement from vertical observations lasted longer than that from surface data, underscoring the important influence of vertical resolution on forecast performance.

To more directly optimize the three-dimensional distribution characteristics of aerosols, the focus of aerosol data assimilation research has gradually shifted towards more comprehensive aerosol optical parameters (Zhang et al., 2011; Ye et al., 2021). Aerosol extinction coefficient and aerosol optical depth (AOD), as key optical parameters characterizing the vertical distribution and total column burden of aerosols, have become important observational data for the new generation of

https://doi.org/10.5194/egusphere-2026-394
Preprint. Discussion started: 22 April 2026


aerosol data assimilation (Chen et al., 2017; Pang et al., 2020). Aerosol extinction coefficient profiles directly constrain the vertical structure, while AOD provides a key constraint on the total aerosol load in the atmospheric column. Therefore, assimilating aerosol extinction coefficient and AOD data is regarded as an important approach for optimizing the vertical structure of aerosols and enhancing air quality forecasting capability (Pang et al., 2018; Ye et al., 2022). Liang et al. (2020) developed observation operators and their adjoints for aerosol extinction coefficient based on the IMPROVE (Interagency Monitoring of Protected Visual Environments) equation, and constructed a 3D-Var assimilation system around this core. By simultaneously assimilating surface station particulate mass concentrations and lidar aerosol extinction coefficient profiles, they validated the effectiveness of multi-source observation collaborative assimilation in improving $PM_{2.5}$ simulation and forecasting. To simultaneously improve the vertical distribution of aerosols and the near-surface forecast effect, Ye et al. (2020) developed a collaborative assimilation scheme using the WRF-Chem model and the 3D-Var method. Building upon the assimilation of surface $PM_{2.5}$ and $PM_{10}$ observations, they further integrated CALIPSO satellite extinction coefficient data. Comparative analysis of results using ground-based lidar and AOD data showed that incorporating aerosol three-dimensional profile data on top of surface data improved both the aerosol vertical distribution and the forecast of surface pollutants. To investigate the improvement of multi-source data assimilation on the three-dimensional representation of aerosols, Zang et al. (2020) found that after introducing AOD data into the 3D-Var system, the simulated aerosol vertical structure aligned better with CALIPSO satellite observations, revealing the key role of AOD data in constraining the model's aerosol vertical distribution. Wang et al., based on Mie scattering theory and the 3D-Var method, constructed tangent linear and adjoint models for aerosol optical properties, enabling the direct assimilation of satellite AOD. This effectively improved the simulation accuracy of AOD and PM2.5, particularly correcting the model's underestimation during heavy pollution events.

However, with the increasing richness of observational data, the ongoing refinement of model grids, and the detailed depiction of complex physicochemical processes and the continuous expansion of model variables in atmospheric chemistry models, the state variables and matrix dimensions in assimilation systems have increased significantly (Liang et al., 2020; Wang et al., 2022). This leads to a sharp expansion in the computational scale of traditional data assimilation, causing traditional methods to face the problem of excessive computational load. This not only prolongs the assimilation time but also makes it difficult to meet the timeliness requirements for real-time rapid air pollution forecasting.

In recent years, the rapid advancement of artificial intelligence (AI) technology has provided new ideas for addressing this challenge. GPUs (Graphics Processing Units) and DCUs employ a massively parallel architecture composed of thousands of computing cores. This design characteristic makes them particularly suitable for handling highly parallel operations like high-dimensional matrix computations, providing powerful computing support for large-scale data processing, thereby

https://doi.org/10.5194/egusphere-2026-394
Preprint. Discussion started: 22 April 2026


achieving orders-of-magnitude improvement in computational efficiency at the hardware level (Boudier et al., 2023; De et al., 2021; He et al., 2025). On the other hand, various neural network models possess powerful capabilities for nonlinear fitting and feature learning, enabling them to directly learn complex physical-chemical relationships from vast amounts of data, thus avoiding the tedious process of constructing tangent linear and adjoint models required in traditional methods. Leveraging these advantages, an increasing number of studies in recent years have begun to introduce AI methods into the assimilation process, exploring their potential for enhancing the performance and revolutionizing the process of traditional assimilation algorithms through various neural networks. Intelligent assimilation has thus gradually developed into an emerging research direction aimed at optimizing or replacing traditional assimilation algorithms (Zhang et al., 2021; Li et al., 2024; Xu et al., 2024; Lu et al., 2025).

It should be noted that the early application of AI in assimilation was essentially a local enhancement of the traditional processes, rather than a systematic replacement. Dong et al. (2022) bypassed the explicit development of complex tangent linear and adjoint models required in traditional four-dimensional variational (4DVAR) assimilation by introducing a Bilinear Neural Network (BNN), constructing a purely data-driven ML-4DVAR framework. They validated on the Lorenz-96 model that this framework could significantly improve computational efficiency while ensuring assimilation accuracy. To address the issue of sampling error caused by insufficient ensemble members due to the high computational cost of the classical Ensemble Kalman Filter in high-dimensional systems, Chattopadhyay et al. (2023) proposed a hybrid ensemble Kalman filter. This approach uses deep learning surrogate models to inexpensively generate large ensembles, thereby more accurately estimating the background error covariance and effectively improving the analysis quality of the initial field. Howard et al. (2024) combined machine learning with the Ensemble Kalman Filter (EnKF) and successfully achieved effective assimilation of high-resolution observational data in the Lorenz-96 system. This enhanced assimilation method not only reduced the RMSE of the initial field by about 37% but also significantly improved the accuracy and reliability of forecasts up to 10 days. Addressing the issues of high computational cost and limited forecast improvement in traditional Chemical Transport Model (CTM) assimilation methods, Li et al. (2025) proposed an efficient machine learning self-evolving assimilation system (OIRF-LEnKF v1.0). Tests showed that the assimilation time of this system was only 11.41–16.60% of that of traditional methods. While significantly improving computational efficiency, its self-evolution mechanism notably reduced forecast errors. Compared to the quality of existing reanalysis datasets, the reanalysis dataset generated by this method was of higher quality. Beauchamp et al. (2025) constructed a neural variational assimilation system based on 4DVarNet. This system used a Variational Autoencoder (VAE) with embedded self-attention mechanism as the prior model, replacing the static background error covariance relied upon in traditional assimilation, achieving deep integration of the variational assimilation framework and deep learning. Results indicated that this system significantly outperformed existing operational methods in terms of analysis accuracy,

https://doi.org/10.5194/egusphere-2026-394
Preprint. Discussion started: 22 April 2026


convergence speed, and the ability to resolve small-scale ocean features.

In recent years, AI assimilation research is shifting towards an end-to-end paradigm. This paradigm aims to use neural networks to directly establish a nonlinear mapping from the observation field and background field to the model forecast field, fundamentally bypassing the complex processes in traditional assimilation such as background error covariance estimation, observation operator construction, and iterative optimization. Xu et al. (2025) proposed the FuXi-DA end-to-end deep learning assimilation framework, which circumvented the complex processes of traditional assimilation and achieved direct mapping from raw satellite observations to an optimized analysis field. This framework employed multi-branch U-nets and a unified fusion neural network, effectively avoiding complex steps like observation operator construction and observation thinning in traditional assimilation. By assimilating observations from the Advanced Geosynchronous Radiation Imager (AGRI) aboard the Fengyun-4B satellite, the framework was proven to effectively optimize the analysis field, enhance forecast skill, and the results demonstrated good physical plausibility.

However, although AI-driven assimilation methods significantly improve computational efficiency, they sacrifice physical interpretability due to the lack of dynamical constraints from physical models and the "black-box" nature of neural networks. This not only limits their in-depth analysis of pollution formation mechanisms but also poses potential risks for their application in operational forecasting requiring high reliability. In contrast, the traditional 3D-Var assimilation method, based on a strict variational optimization theoretical framework, has a clear physical meaning and mathematical traceability for the background field correction process, yielding transparent and highly credible results. However, traditional methods are typically implemented using CPU clusters and traditional languages like Fortran. When facing high-dimensional atmospheric chemical assimilation tasks, they are often constrained by computational efficiency bottlenecks, making it difficult to meet the timeliness requirements of real-time forecasting.

To meet the dual demands of physical interpretability and high-performance computational efficiency, this study constructs a three-dimensional variational assimilation system based on the PyTorch deep learning framework, with its core computational workflow can be simultaneously deployed on both CPU and DCU. The core innovation of this system lies in the deep integration of "principle inheritance" and "implementation innovation": At the theoretical level, it strictly adheres to the variational optimization theory of 3DVAR, fully retaining the core logic of cost function construction and gradient descent, ensuring the physical transparency of the assimilation process and the interpretability of the results. At the technical implementation level, it fully leverages the automatic differentiation capability of the PyTorch framework and the native parallel computing advantages of DCU, modernizing the core modules of assimilation. Key steps such as cost function tensor operations, gradient calculation, and matrix multiplication are entirely migrated to the PyTorch framework for execution, utilizing the massive parallel computing power of DCU to break through computational efficiency bottlenecks, ultimately achieving a leap in computational

https://doi.org/10.5194/egusphere-2026-394
Preprint. Discussion started: 22 April 2026


efficiency while maintaining physical consistency. This design strategy preserves the physical reliability of traditional assimilation methods while fully utilizing the extreme computing power of the deep learning hardware ecosystem, establishing a fused assimilation solution for the large-scale, rapid assimilation of multiple pollutants including $SO_2$, $NO_2$, CO, $O_3$, $PM_{2.5}$, and $PM_{10}$, which possesses timeliness, transparency, and strict physical interpretability.

## 2. Methods and data

### 2.1 Air quality model

The Weather Research and Forecasting model with Chemistry (WRF-Chem), was jointly developed by the National Center for Atmospheric Research (NCAR), the National Oceanic and Atmospheric Administration (NOAA), and other institutions. Its core strength lies in the fully online, bidirectional coupling between meteorological and chemical fields, enabling a precise description of pollutant transport, transformation, and removal under complex meteorological conditions. The meteorological fields drive the spatiotemporal distribution and transformation of chemical species through physical processes such as dynamic advection, turbulent diffusion, and dry/wet deposition. Conversely, the chemical fields can feedback on meteorological elements through mechanisms like aerosol radiative forcing and cloud condensation nuclei effects. This online coupling ensures spatiotemporal consistency and physical compatibility between atmospheric dynamic processes and chemical transformations.

The WRF-Chem used in this study integrates the Aerosol Module 5 (AERO5) from the Community Multiscale Air Quality Modeling System (CMAQ). It also features a flexible gas-phase chemistry module that can utilize multiple chemical mechanisms, including the Statewide Air Pollution Research Center (SAPRC), Carbon Bond IV (CB-IV), and Regional Acid Deposition Model version 2 (RADM2) mechanisms (Li et al., 2010, 2011a, 2011b, 2012).

Through over two decades of development, this model version incorporates advanced chemical reactions and processes crucial for atmospheric particulate formation. For example, it includes irreversible aqueous-phase processes of glyoxal and methylglyoxal on particle surfaces, which improves the simulation of secondary organic aerosol (SOA) over China (Xing et al., 2018). A parameterization for heterogeneous nitrous acid (HONO) processes has been implemented, enhancing the model's capability to simulate atmospheric oxidizing capacity and SOA (Feng et al., 2005, 2016). The suppression effect of organic coatings on the heterogeneous hydrolysis uptake of dinitrogen pentoxide ($N_2O_5$) is also parameterized, leading to better simulation of nitrate aerosols. Furthermore, the model incorporates sulfate formation via transition-metal catalysis and Criegee intermediate reactions, improving sulfate simulation. Additionally, a module accounting for aerosol–radiation–photolysis interactions has been developed, enabling assessment of aerosol–cloud interactions, radiative effects of brown carbon aerosols, and direct and indirect effects of various aerosol components.

https://doi.org/10.5194/egusphere-2026-394
Preprint. Discussion started: 22 April 2026


To comprehensively evaluate the computational efficiency and assimilation performance of the Py3DVAR system, this study designed two single-nested grid simulation schemes based on the WRF-Chem model. Both schemes share the same vertical configuration of 34 layers, differing only in horizontal resolution. Scheme 1 employs a coarse-resolution grid with dimensions of 210 (meridional) × 168 (zonal) grid points and a horizontal resolution of 27 km. This setup is suitable for rapid verification and benchmark testing. Scheme 2 utilizes a high-resolution grid, sized 600 × 600 grid points with a 9 km horizontal resolution, approaching the scale of operational assimilation tasks. It is designed to assess the system's computational performance under high-resolution conditions. The simulation domains are shown in Fig. 1.

To maintain focus and systematically evaluate the core capabilities of the Py3DVAR framework, this study primarily conducted experimental design and result analysis using the coarse 27 km grid (210×168). The results for the background error covariance, idealized experiments, and forecast experiments are all presented and discussed at this 27 km resolution. As a supplementary assessment, a real-case assimilation experiment was performed on the 9 km grid. This serves as a comparative analysis against the corresponding 27 km results, aiming to examine the scalability of the Py3DVAR system's computational efficiency and assimilation effectiveness with increasing grid resolution. Through this "primary-supplementary" experimental design, the study provides an in-depth analysis of the system's core performance while preliminarily exploring its potential for high-resolution applications.

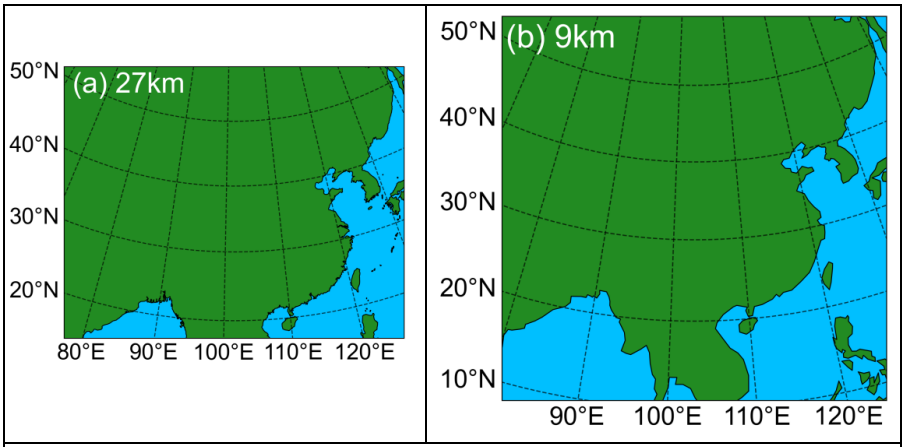



**Figure 1.** Distribution of the WRF-Chem simulation domain at (a) 27 km and (b) 9 km grid resolutions.

## 2.2 Data

The meteorological fields for the WRF-Chem model were driven by the NCEP FNL (Final Operational Global Analysis, https://gdex.ucar.edu/datasets/d083002/) data, which has a horizontal

https://doi.org/10.5194/egusphere-2026-394
Preprint. Discussion started: 22 April 2026


resolution of 1° × 1° and a temporal resolution of 6 hours. Provided by NCAR and NCEP, the FNL dataset assimilates multi-source global observations from surface stations, radiosondes, satellites, etc. It supplies physically consistent meteorological fields and serves as crucial input for generating reliable initial and boundary conditions for the model. This dataset has been widely used in various mesoscale numerical simulations.

Anthropogenic emissions were derived from the 2013 Multi-resolution Emission Inventory for China (MEIC, http://www.meicmodel.org/index.html), provided by Tsinghua University. This inventory has a spatial resolution of 6 km, covers most regions of China, and includes emission data for the gaseous pollutants and aerosol precursors required in this study.

Chemical boundary conditions for the WRF-Chem model were obtained from the Whole Atmosphere Community Climate Model (WACCM, https://www.acom.ucar.edu/waccm/download.shtml) data released by NCAR. This dataset provides the global spatiotemporal distribution of various chemical components, including ozone and aerosol precursors, ensuring reasonable atmospheric chemical boundary conditions for the regional air quality simulations.

Observation data were sourced from the China National Environmental Monitoring Centre (CNEMC, https://air.cnemc.cn:18007/), primarily comprising hourly surface concentrations of major air pollutants: $SO_2$, $NO_2$, CO, $O_3$, $PM_{2.5}$, and $PM_{10}$. The locations of observation sites follow the distribution described in He et al. (2025). To ensure data quality and suitability for 3D-Var assimilation, the following procedures were applied to the raw observational data:

(1) Removal of Invalid Values: All concentration records with negative values for any pollutant were identified as invalid and removed.

(2) Extreme Value Control: To prevent interference from anomalously high values, maximum concentration thresholds were set for each pollutant (units: μg $m^{-3}$ for all except CO in mg $m^{-3}$). The specific thresholds were: 500 μg $m^{-3}$ for $SO_2$, $NO_2$, and $O_3$; 600 μg $m^{-3}$ for $PM_{2.5}$; 1200 μg $m^{-3}$ for $PM_{10}$; and 500 mg $m^{-3}$ for CO. Observations exceeding their respective thresholds were discarded.

(3) Grid-based Thinning and Averaging: To match the irregularly distributed station observations to the model grid, a grid-averaging method was employed for data thinning. First, all station observations falling within the same WRF-Chem model grid cell were grouped. Then, for the valid observation sequence within each grid cell, an internal consistency check using the 3-standard-deviation criterion was applied to remove outliers that deviated excessively from the mean. Finally, the arithmetic mean was calculated from all observations that passed the quality control in each grid cell, resulting in a single representative concentration value for that grid. This process generated a regular, gridded observational dataset consistent with the model's grid resolution.

### 2.3. Control variable design

https://doi.org/10.5194/egusphere-2026-394
Preprint. Discussion started: 22 April 2026


The WRF-Chem model employed in this study incorporates aerosol variables that encompass a variety of fine and coarse particle components. The fine particle components include: elemental carbon (ECJ), ammonium (NH4AJ), sulfate (SO4AJ), nitrate (NO3AJ_A/B), organic aerosols (ORGAJ, ORGBAJ, SOA_RXT_G/M), nine categories of secondary organic aerosols generated from the semi-volatile gases of primary organic aerosol (POA) (CPO1AER–CPO9AER), chloride (CLAJ), and accumulation-mode fine particles (P25AJ), among others. The coarse particle components include: dust (DUST_1–DUST_5), anthropogenic coarse aerosol (ANTHA), and coarse-mode soil-derived particles (SOILA). Furthermore, the model also includes gaseous pollutants such as $SO_2$, $NO_2$, CO, and $O_3$. The first 11 categories of fine particle components collectively constitute $PM_{2.5}$. Adding dust along with the anthropogenic and soil-derived coarse-mode aerosols to this forms the complete $PM_{10}$ composition.

If all the aforementioned aerosol components were directly used as control variables for assimilation, it would provide detailed compositional resolution but would also lead to a drastic increase in computational scale, significantly raising storage demands and reducing computational efficiency. To balance assimilation accuracy with computational cost, this study rationally merges specific components based on their physicochemical properties and particle size characteristics: the dual nitrate components are merged into a single variable; the four organic aerosol components are combined into one group; the nine secondary organic aerosols originating from POA semi-volatile gases are uniformly merged; the dust components are optimally merged according to particle size differences—DUST_1 is retained as an independent control variable, while DUST_2 to DUST_5 are merged into a single variable. Through this merging process, the number of aerosol control variables is reduced from the original numerous components to 12. Gaseous pollutants ($SO_2$, $NO_2$, CO, $O_3$) are directly used as control variables for assimilation. Consequently, the assimilation system constructed in this study utilizes a total of 16 control variables. Table 1 provides the detailed mapping between the WRF-Chem model variables and the 3DVAR control variables.

https://doi.org/10.5194/egusphere-2026-394
Preprint. Discussion started: 22 April 2026


| Model variables | 3DVAR control variables |
|---|---|
| ECJ | EC |
| NH4AJ | NH4 |
| SO4AJ | SO4 |
| NO3AJ_A NO3AJ_B | NO3 |
| ORGAJ ORGBAJ SOA_RXT_G SOA_RXT_M | ORG |
| CPO1AER CPO2AER … CPO9AER | CPO |
| CLAJ | CL |
| P25AJ | P25 |
| DUST_1 | DUST1 |
| DUST_2 DUST_3 DUST_4 DUST_5 | DUST2 |
| ANTHA | ANTHA |
| SOILA | SOILA |
| SO2 | SO2 |
| NO2 | NO2 |
| CO | CO |
| O3 | O3 |

**Table 1.** Mapping between WRF-Chem model variables and 3DVAR control variables.

### 2.4 Three-dimensional variational assimilation theory

The three-dimensional variational assimilation method is grounded in probability and statistical theory, treating both the simulated background field and observations as useful information containing errors. This method constructs a cost function that incorporates the weighted errors of the background field and the observations, and seeks the state that minimizes this function, known as the analysis field. Statistically, this analysis field is considered the optimal estimate of the true atmospheric state. The 3D-Var objective function is:

$$J(x) = \frac{1}{2}(x - x_b)^T B^{-1}(x - x_b) + \frac{1}{2}(Hx - y)^T R^{-1}(Hx - y) \quad (1)$$

where x is the analysis field, representing the set of control variables at all grid points in the model. Its dimensionality is jointly determined by the number of control variables, vertical layers, and horizontal grid points. $x_b$ is the background field, provided by the WRF-Chem model simulation, and has the same dimensions as x. B is the background error covariance matrix, representing the statistical characteristics of the background field errors. H is the observation operator that maps from model space to observation space. y is the observation vector, containing the quality-controlled observational data, with dimensions of Obs_Station×6, where Obs_Station represents the number of valid observation sites, and 6 corresponds to the six types of pollutants. R is the observation error covariance matrix, representing the uncertainty of the observational data, and has dimensions consistent with y. Following Li et al. (2013), the observational error standard

https://doi.org/10.5194/egusphere-2026-394
Preprint. Discussion started: 22 April 2026


deviation in this study is set to half of the background error standard deviation.

The background error covariance matrix B has a very large dimension. Storing and performing iterative computations with it directly would pose serious challenges to DCU memory and computational efficiency. To optimize the solution of the cost function, this study refers to the method of Li et al., factorizing and structurally simplifying the B matrix.

The specific processing is as follows: First, the B matrix is decomposed into the form $B = DCD^T$, where D is a diagonal matrix composed of the background error standard deviations, and C is the background error correlation matrix. In practical assimilation, the analysis field x is typically not solved for directly. Instead, the incremental field $\delta x = x - x_b$ is introduced to solve for the analysis field indirectly, and the control variable is transformed to $\delta z$, where the control variables satisfy $\delta x = DC^{1/2}\delta z$. Defining the transformation operator $F = DC^{1/2}$, the cost function can be rewritten in terms of $\delta z$ as:

$$J(\delta z) = \frac{1}{2}\delta z^T \delta z + \frac{1}{2}(HF\delta z - d)^T R^{-1}(HF\delta z - d) \quad (2)$$

where $d = y - Hx_b$ is the innovation vector (observation increment). After minimizing $J(\delta z)$ to obtain the optimal $\delta z$, the final analysis field $x_a$ is obtained through the following transformation:

$$x_a = x_b + F\delta z \quad (3)$$

### 2.5 Gradient calculation method

During the minimization of the cost function, both the Py3DVAR system under the Python framework and the Fortran-3DVAR system under the Fortran framework employ the Limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) algorithm to achieve minimization. Specifically, the Py3DVAR system directly utilizes the built-in torch.optim.LBFGS optimizer from the PyTorch framework, which is optimized for accelerated tensor operations. In contrast, the Fortran-3DVAR system implements the L-BFGS algorithm manually in Fortran, employing multi-CPU-thread parallel computing. While the optimization objective is identical for both assimilation systems, they differ fundamentally in their underlying code architecture and parallelization strategies.

The Py3DVAR system, built upon the PyTorch deep learning framework, fully leverages its automatic differentiation capability for gradient computation. When minimizing the cost function $J(\delta z)$, the system automatically calculates the gradient $\partial J/\partial \delta z$ via the backpropagation algorithm. This method constructs a computational graph, recording all tensor operations during the forward pass, and then automatically computes the gradients of the objective function with respect to each parameter during the backward pass. This automatic differentiation mechanism not only avoids complex analytical derivation, simplifying the code development process, but also ensures both numerical accuracy and computational efficiency in gradient calculations.

https://doi.org/10.5194/egusphere-2026-394
Preprint. Discussion started: 22 April 2026


In comparison, the traditional Fortran-based Fortran-3DVAR system employs manually derived analytical gradient formulas for computation. This approach requires pre-deriving the analytical expression for the gradient of the cost function with respect to the control variable δz through mathematical analysis:

$$\nabla J(\delta z) = \delta z + F^{T}H^{T}R^{-1}(HF\delta z - d) \quad (4)$$

This derived gradient formula is then explicitly implemented in the program. Although this method has been widely used in traditional scientific computing, the derivation process is complex and prone to error, and it struggles to fully exploit the advantages of parallel computing.

While the two gradient calculation methods are mathematically equivalent, they exhibit significant differences in their implementation approach and computational efficiency.

## 2.5 Background error covariance modeling based on tensor product decomposition and the NMC Method

Furthermore, the background error correlation matrix C is decomposed into the tensor product of correlation matrices along the longitude, latitude, and vertical directions, i.e., $C = C_x \otimes C_y \otimes C_z$. For the horizontal correlation matrices $C_x$ and $C_y$, horizontal isotropy is assumed. Consequently, they can be represented by the same decay function, differing only in their dimensions. Their physical meaning is to describe the horizontal correlation between different grid points in the longitudinal or latitudinal direction, specifically fitted using a Gaussian kernel function:

$$C(x_1, x_2) = \exp[\ -(x_1 - x_2)^2 / (2L^2)\ ] \quad (5)$$

Here, L is the horizontal correlation scale parameter for the background error. The method for obtaining the specific value of L will be introduced later. $x_1$ and $x_2$ represent horizontal grid points.

The key components of the background error covariance matrix B—namely, the background error standard deviation matrix D, the vertical background error correlation matrix $C_z$, and the horizontal correlation scale parameter L—are statistically estimated using the NMC method (Parrish et al., 1992). The specific procedure is as follows: First, 24-hour and 48-hour forecast experiments are initiated daily at 00:00 UTC for 30 consecutive days. The forecast difference field $\varepsilon$ is then constructed from forecasts valid at the same time, calculated as $\varepsilon = X_{48h} - X_{24h}$.

The background error standard deviation matrix D is a diagonal matrix. The values for each control variable in D are computed from the standard deviation of the 30-day forecast difference field:

$$D = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(\varepsilon_i - \bar{\varepsilon})^2} \quad (6)$$

where N is the number of days (N=30 in this study). The vertical profiles of the background error standard deviation for each control variable are shown in Figure 2.

https://doi.org/10.5194/egusphere-2026-394
Preprint. Discussion started: 22 April 2026


Vertical Correlation Matrix C_z: This is obtained by calculating the correlation coefficients of the same variable between different model layers. Its matrix elements can be expressed as:

$$[\mathbf{C}_z]_{mn} = \frac{\mathrm{cov}(m,n)}{\sigma_{\varepsilon_m} \cdot \sigma_{\varepsilon_n}} \tag{7}$$

Here, $\varepsilon_m$ and $\varepsilon_n$ represent the forecast differences at the m-th and n-th layers, respectively, cov denotes the covariance, and $\sigma$ represents the standard deviation. The distribution of the vertical autocorrelation matrix for the background error of each control variable is shown in Figure 3.

Horizontal Correlation Scale Parameter L: This is estimated by analyzing the decay of correlation for the same variable across different horizontal grid distances. The correlation coefficient $\rho(d)$ between all pairs of horizontal grid points is computed as a function of their distance d. L is defined as the maximum grid distance d at which the horizontal correlation coefficient $\rho(d)$ decays to $e^{-1/2}$:

$$L = \max\{\, d \mid \rho(d) \geqslant e^{-1/2} \,\} \tag{8}$$

The horizontal correlation scale parameter L quantitatively describes the decay characteristics of background error correlation between different grid points with increasing distance. Its physical meaning is that when the distance between two horizontal grid points exceeds L, the correlation coefficient between their background errors falls below $e^{-1/2}$ (approximately 0.6065), indicating that their statistical correlation has significantly weakened. This parameter quantifies the decay rate of horizontal correlation and is a key parameter for describing the horizontal structure of the background error covariance. The computed values of L are used to construct the horizontal correlation matrices $C_x$ and $C_y$, which characterize the spatial horizontal correlation structure and directly determine the shape of the Gaussian correlation kernel function. The decay of the horizontal background error correlation coefficient with distance and the corresponding L value for each control variable are shown in Fig. 4.

To achieve efficient computation, the three correlation matrices $C_x$, $C_y$ and $C_z$ are further decomposed via Cholesky decomposition. Each is expressed as the product of a lower triangular matrix and an upper triangular matrix (e.g., $C_x = C_{x\_low} \cdot C_{x\_up}$), where these two triangular matrices are transposes of each other (i.e., $C_{x\_up} = C_{x\_low}^T$, where $T$ denotes transpose). During the actual assimilation process, only the upper triangular matrices for these three directions are retained and used, thereby avoiding direct manipulation of the full B matrix.

https://doi.org/10.5194/egusphere-2026-394
Preprint. Discussion started: 22 April 2026

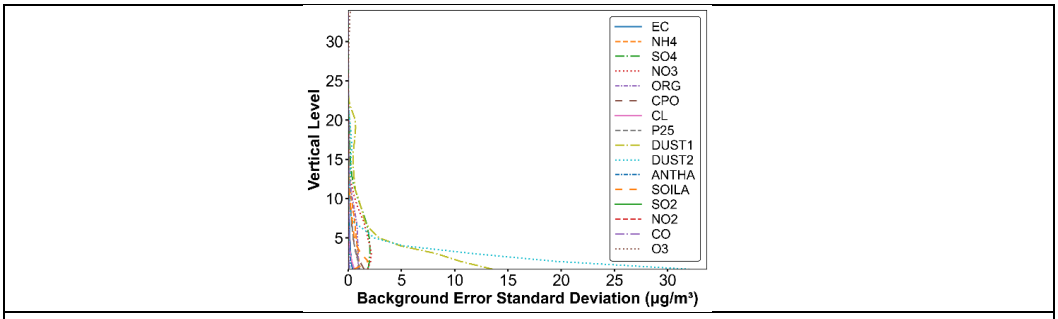




**Figure 2.** Vertical profiles of background error standard deviation for each control variable.

Fig. 2 shows the vertical profiles of the background error standard deviation, which quantifies the uncertainty of the model background field across different vertical layers and for different variables. In the 3D-Var assimilation process, the D matrix determines the relative distribution of analysis increments, induced by observational information, among the control variables. A larger background error standard deviation for a variable indicates higher uncertainty in the background field simulation, allowing for greater adjustment by observational information during assimilation. Conversely, a smaller standard deviation implies higher confidence in the background field, resulting in more moderate increments. Fig. 2 reveals significant differences in the near-surface values among the variables. The background error standard deviations for DUST1 (exceeding 10 μg $m^{-3}$) and DUST2 (exceeding 30 μg $m^{-3}$) are the largest, substantially higher than those of other variables (all below 5 μg $m^{-3}$). With increasing height, the background error standard deviations for all variables generally exhibit a decaying trend, and the differences between them gradually diminish. In terms of profile shape and decay rate, DUST1 and DUST2 decay most rapidly, while the decay of the other variables is relatively gradual.

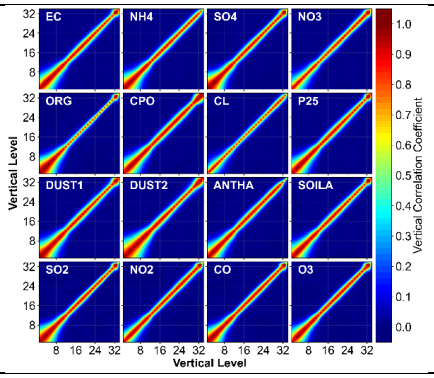



**Figure 3.** Distribution of the vertical autocorrelation matrix of background error for each control variable.

Fig. 3 displays the vertical autocorrelation matrices of the background error for each control

https://doi.org/10.5194/egusphere-2026-394
Preprint. Discussion started: 22 April 2026


variable. These matrices play a key role in the assimilation system by constraining the propagation and distribution of observational information across the model's vertical layers. Analyzing their spatial structure shows that, although the detailed distributions differ among variables, each exhibits strong near-surface correlations, and these correlations decrease with increasing height. This indicates that observational information from the surface layer plays a more dominant role in adjusting and correcting the lower-level state of the model background field, while adjustments to higher levels are more strongly constrained.

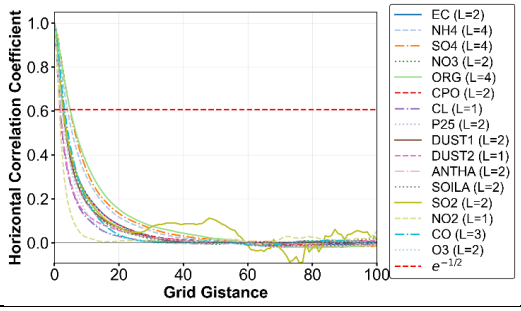



**Figure 4.** Decay of horizontal background error correlation coefficient with distance for each control variable.

Fig. 4 presents the decay curves of the horizontal background error correlation coefficient with grid distance for each control variable, along with their corresponding horizontal correlation scale parameter L. This correlation coefficient determines the horizontal spread and weight allocation of observational information. The overall pattern shows that the horizontal correlation coefficients for all variables decay with increasing grid distance, characterized by an initially fast then slower decay rate. This is consistent with the physical mechanism represented by the Gaussian kernel function: "strong correlation at short distances and weak correlation at long distances." The decay rates differ significantly among the variables, leading to calculated L values distributed between 1 and 4. This reflects the heterogeneity in the horizontal spatial correlation of background errors for different pollutants. Notably, the decay curve for $SO_2$ shows noticeable fluctuations beyond a grid distance of 20, exhibiting less smoothness compared to other variables. This phenomenon may stem from the following reason: $SO_2$, primarily emitted from point sources (e.g., power plants, industrial zones), has a background error field that exhibits stronger spatial heterogeneity and discontinuity. When the distance exceeds the effective propagation range of the observational information, residual local error signals cannot be smoothly fitted, ultimately manifesting as curve fluctuations.

## 3. Experimental design and evaluation metrics

### 3.1 Experimental design

To systematically evaluate the performance and effectiveness of the Py3DVAR assimilation system, this study designed three sets of numerical experiments: an idealized experiment, real case assimilation experiments, and a short-term forecast experiment.

https://doi.org/10.5194/egusphere-2026-394
Preprint. Discussion started: 22 April 2026


First, an idealized twin-observation experiment was designed based on the 27 km grid. Its purpose is to test the system's capability to generate physically reasonable increment fields under observations of varying magnitudes. Two observation sites were positioned far apart to avoid spatial correlation interference, ensuring the independence of constraints from different observation magnitudes. The low-concentration observation site (37.2104°N, 116.0108°E) had prescribed values of 0.001 μg $m^{-3}$ for $SO_2$, $NO_2$, $O_3$, $PM_{2.5}$, and 0.0015 μg $m^{-3}$ for $PM_{10}$, and 0.001 mg $m^{-3}$ for CO. The high-concentration site (37.8666°N, 92.3785°E) had values of 300 μg $m^{-3}$ for $SO_2$, $NO_2$, $O_3$, 500 μg $m^{-3}$ for $PM_{2.5}$, 700 μg $m^{-3}$ for $PM_{10}$, and 300 mg $m^{-3}$ for CO. By comparing the analysis field with the background field, the ability of the assimilation system to produce physically sound increment fields under different observation intensities can be verified.

In addition to the twin-point idealized experiment, a real case assimilation experiment was conducted for the background field at 00:00 UTC on October 27, 2023. This experiment serves as the core comparative test of this study and is the only one covering both grid resolutions. Unlike the idealized experiment, it utilized all quality-controlled real observation data and was executed simultaneously on both the 27 km and 9 km grids. For each grid, three parallel assimilation runs were performed: (1) using the Fortran-3DVAR system on the CPU platform (labeled Fortran-3DVAR); (2) using the Py3DVAR system on the CPU platform (labeled Py3DVAR_CPU); and (3) using the Py3DVAR system on the DCU platform (labeled Py3DVAR_DCU). This resulted in a total of six assimilation experiments across the two grids. The Fortran-3DVAR system runs exclusively on the CPU platform, whereas the Py3DVAR system is deployed on both CPU and DCU platforms, enabling a multi-dimensional comparison of assimilation performance and computational efficiency.

Finally, based on the results of the real case assimilation, a 24-hour short-term forecast experiment was conducted on the 27 km grid. Two types of initial fields were used: the original, non-assimilated background field (serving as the control experiment, labeled Control) and the assimilated analysis fields (from Fortran-3DVAR and Py3DVAR_DCU, respectively). By comparing the forecast results from the control and assimilation experiments, the improvement effect of the assimilation analysis on short-term air quality forecasting and its persistence can be assessed.

To ensure rigorous comparison, all aforementioned experiments used identical background error covariance, background fields, and observation data. All computational processes employed 32-bit floating-point precision to avoid introducing extraneous numerical errors due to precision differences. This approach aims to accurately test the result fidelity of the Py3DVAR system and quantitatively compare the computational efficiency of the two systems and the acceleration effect of the DCU processor.

The specifications of the computing platforms involved in this study are as follows: The CPU computing platform is a single node equipped with two Intel Xeon 7495 processors (128 physical cores @ 3.0 GHz) and 768 GB of memory. All 128 physical cores were utilized for parallel

https://doi.org/10.5194/egusphere-2026-394
Preprint. Discussion started: 22 April 2026


computing during the experiments. The DCU acceleration platform centers on a K500SM_AI DCU with 64 GB of device memory, supporting the parallel accelerated computations for the Py3DVAR_DCU experiments.

### 3.2 Evaluation methods

To quantitatively evaluate the assimilation and forecasting performance of Py3DVAR, this study employs three statistical metrics: the correlation coefficient (CORR), root mean square error (RMSE), and mean absolute error (MAE). CORR measures the linear correlation between simulated and observed values; RMSE is more sensitive to extreme errors and reflects the overall deviation of the simulation results; MAE measures the average absolute error between simulated and observed values and is less sensitive to outliers. The formulas for each metric are as follows:

$$\mathrm{CORR} = \frac{\sum_{i=1}^{N}(M_i-\bar{M})(O_i-\bar{O})}{\sqrt{\sum_{i=1}^{N}(M_i-\bar{M})^2 \sum_{i=1}^{N}(O_i-\bar{O})^2}} \quad (9)$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(M_i - O_i)^2} \quad (10)$$

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|M_i - O_i| \quad (11)$$

Where:

$M_i$ and $O_i$ represent the simulated value and observed value for the i-th sample, respectively.

$\bar{M}$ and $\bar{O}$ are the average values of the simulated and observed values, respectively.

N is the total number of samples.

## 4. Results and evaluation

### 4.1 Evaluation of idealized experiment results

Fig. 5 and 6 present the surface increment fields generated by the Py3DVAR and Fortran-3DVAR assimilation systems, respectively, in the idealized experiment, where red indicates positive increments and blue indicates negative increments. It should be noted that the $PM_{2.5}$ and $PM_{10}$ increment fields shown in the figures are not obtained directly through assimilation. Instead, they are synthesized from the increments of multiple aerosol component control variables: the sum of the increments of the first 8 control variables constitutes the $PM_{2.5}$ increment field, and the sum of the increments of the first 12 control variables constitutes the $PM_{10}$ increment field. In contrast, the increment fields for the four gaseous pollutants ($SO_2$, $NO_2$, CO, $O_3$) are directly derived from their corresponding assimilation control variables.

For all variables in both systems, the increments exhibit an approximately circular distribution

https://doi.org/10.5194/egusphere-2026-394
Preprint. Discussion started: 22 April 2026


centered on the observation point, with the maximum value at the center decaying gradually with increasing distance. This spatial structure aligns with the theoretical characteristic in 3DVAR where observational information spreads horizontally according to the background error horizontal correlation. Further comparison of the increment fields for different pollutants reveals significant differences in their horizontal distribution ranges. This is primarily governed by the horizontal correlation scale parameter L of the background error for each control variable. A larger L value indicates a wider horizontal spread of observational information, resulting in a correspondingly larger horizontal distribution range for the increment field.

From the comparison of the extremum values (retaining two decimal places) of the increment fields presented in Fig. 5 and 6, it is clearly evident that the results from the Py3DVAR and Fortran-3DVAR systems exhibit exceptionally high consistency. Within the bounds of numerical precision, the difference in extremum values for gaseous pollutant CO is 0 mg $m^{-3}$, for gaseous pollutants $SO_2$, $NO_2$, and $O_3$ it does not exceed 0.03 μg $m^{-3}$, and for particulate pollutants ($PM_{2.5}$, $PM_{10}$) it does not exceed 0.04 μg $m^{-3}$. These extremely small differences confirm that although Py3DVAR has undergone fundamental reconstruction in terms of gradient calculation (automatic differentiation) and optimization algorithm (PyTorch L-BFGS), it has not compromised the scientific accuracy of the assimilation. In terms of core assimilation capability, it possesses assimilation performance and numerical reliability comparable to the traditional manually coded Fortran-3DVAR system.

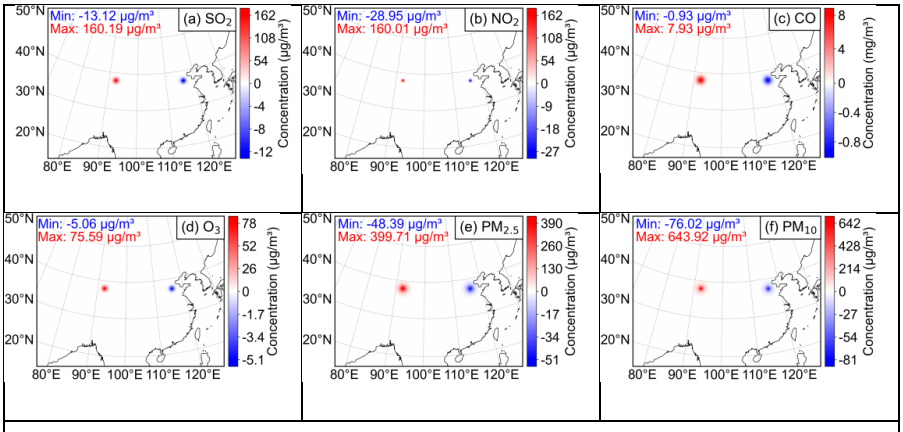



**Figure 5.** Surface increment fields for the six pollutants ((a) $SO_2$, (b) $NO_2$, (c) CO, (d) $O_3$, (e) $PM_{2.5}$, (f) $PM_{10}$) from the Py3DVAR assimilation system in the idealized experiment. Blue and red colors represent negative and positive increments, respectively.

https://doi.org/10.5194/egusphere-2026-394
Preprint. Discussion started: 22 April 2026

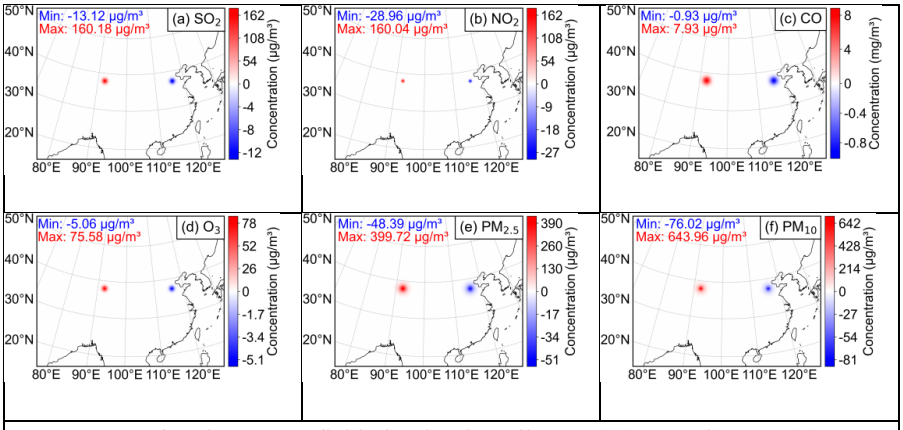




**Figure 6.** Surface increment fields for the six pollutants ((a) $SO_2$, (b) $NO_2$, (c) CO, (d) $O_3$, (e) $PM_{2.5}$, (f) $PM_{10}$) from the Fortran-3DVAR assimilation system in the idealized experiment. Blue and red colors represent negative and positive increments, respectively.

Fig. 7 and 8 present the vertical cross-sections of the increment fields generated by the Py3DVAR and Fortran-3DVAR assimilation systems, respectively, in the idealized experiment, with red and blue denoting positive and negative increments, respectively. The distribution of vertical increments for various pollutants from both systems exhibits a distinct columnar structure. The deepest color and thickest column in the near-surface layer indicate the largest increment values there. As height increases, the color fades and the column narrows, corresponding to a weakening of the increment magnitude. This vertical distribution pattern is fully consistent with the theoretical expectation encoded in the vertical correlation matrix $C_z$ of the background error covariance, namely that error correlation peaks in the near-surface layer and decays with altitude. Specifically, increments above approximately level 10 are near zero, indicating that the effective vertical propagation range of surface observation information is about 10 layers. At higher levels, constrained by the decreasing correlation, the magnitude of increment adjustments becomes minimal.

Comparing the vertical increment fields of different pollutants reveals differences in the thickness of their columnar structures. This thickness essentially reflects the horizontal spread of the observational information; a thicker column represents a larger horizontal influence area. Comparing these with the horizontal increment fields in Fig. 5 and 6 shows a clear correspondence between the column thickness in the vertical cross-sections and the distribution range in the horizontal fields. Both are determined by the horizontal correlation scale parameter L of the background error for each control variable – a larger L value leads to a wider horizontal spread and, consequently, a thicker column in the vertical profile. This further verifies the global constraining effect of the background error's horizontal correlation characteristics on the spatial propagation of observational information.

Comparing the Py3DVAR and Fortran-3DVAR systems shows that the distribution patterns, column thickness, and decay trends of the vertical increment fields for all pollutants are highly

https://doi.org/10.5194/egusphere-2026-394
Preprint. Discussion started: 22 April 2026


consistent between them. This finding corroborates the comparison results of the horizontal increment fields mentioned previously. The consistency in both vertical and horizontal dimensions fully validates the ability of the Py3DVAR assimilation system to maintain physical rationality and numerical accuracy within the three-dimensional space, demonstrating that its reliability and effectiveness are comparable to the mature Fortran-3DVAR assimilation system.

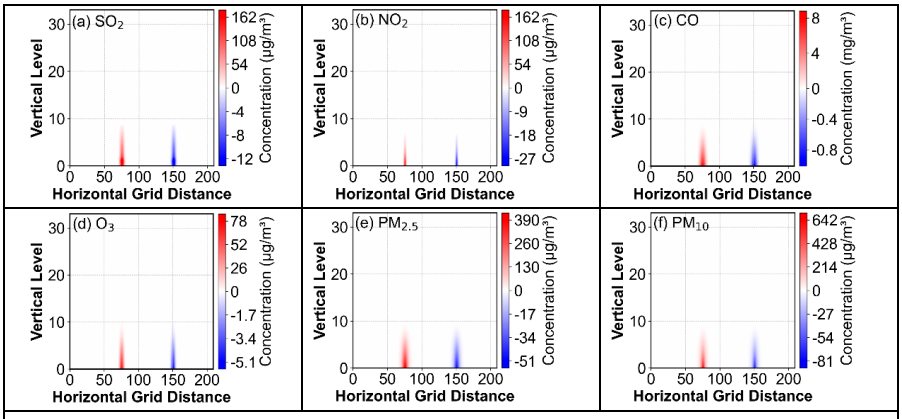



**Figure 7.** Vertical cross-sections of the increment fields for the six pollutants ((a) $SO_2$, (b) $NO_2$, (c) CO, (d) $O_3$, (e) $PM_{2.5}$, (f) $PM_{10}$) from the Py3DVAR assimilation system in the idealized experiment. Blue and red colors represent negative and positive increments, respectively.

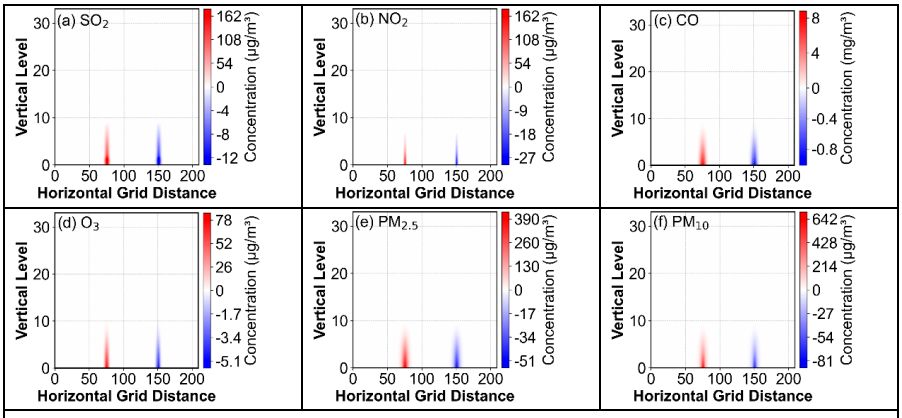


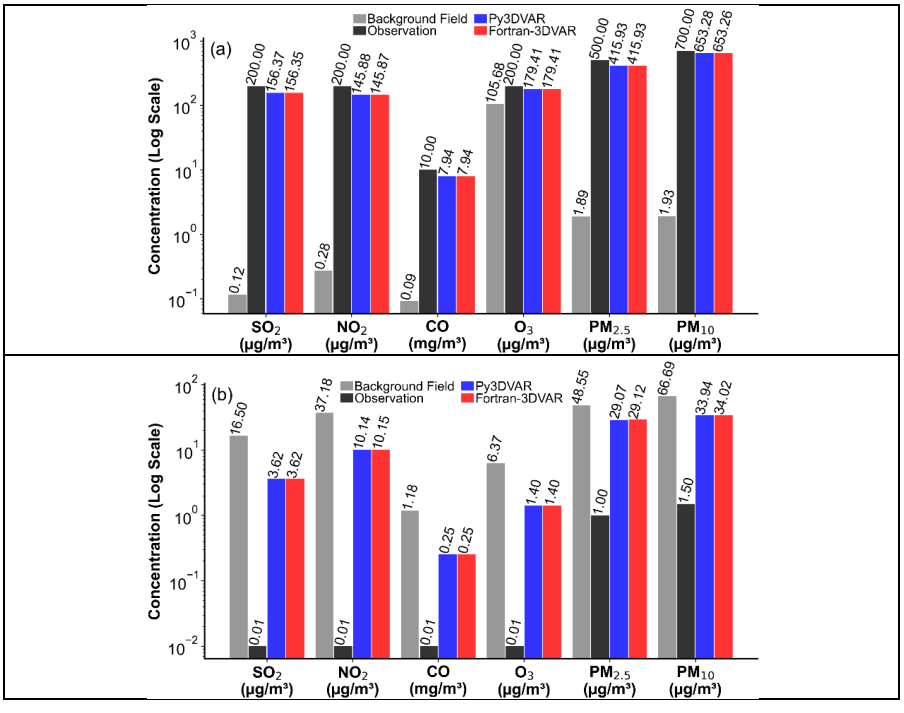

**Figure 8.** Vertical cross-sections of the increment fields for the six pollutants ((a) $SO_2$, (b) $NO_2$, (c) CO, (d) $O_3$, (e) $PM_{2.5}$, (f) $PM_{10}$) from the Fortran-3DVAR assimilation system in the idealized experiment. Blue and red colors represent negative and positive increments, respectively.

To provide a more intuitive assessment of the assimilation performance of Py3DVAR versus Fortran-3DVAR, Fig. 9 compares the background field, the observation field, and the analysis fields

https://doi.org/10.5194/egusphere-2026-394
Preprint. Discussion started: 22 April 2026


from the two assimilation systems under conditions of extremely high (a) and extremely low (b) observed concentrations. The bar chart clearly shows that for both the Py3DVAR and Fortran-3DVAR systems, their respective analysis fields are closer to the observed values than the background field is. This indicates that the assimilation process significantly improves the initial background field. This improvement effect remains stable under both extremely high and extremely low concentration conditions, demonstrating the robustness of the assimilation systems.

By comparing the analysis fields of the ideal experiments of the two assimilation systems (results retained to two decimal places), it can be observed that under extremely high observation concentrations, the differences in $SO_2$, $NO_2$, CO, $O_3$, $PM_{2.5}$, and $PM_{10}$ in the analysis fields of the Py3DVAR and Fortran-3DVAR assimilation systems are 0.02 μg $m^{-3}$, 0.01 μg $m^{-3}$, 0 μg $m^{-3}$, 0 μg $m^{-3}$, 0 μg $m^{-3}$, and 0.02 μg $m^{-3}$ respectively. Under extremely low observation concentrations, the differences in $SO_2$, $NO_2$, CO, $O_3$, $PM_{2.5}$, and $PM_{10}$ in the analysis fields of the two systems are 0 μg $m^{-3}$, 0.01 μg $m^{-3}$, 0 μg $m^{-3}$, 0 μg $m^{-3}$, 0.05 μg $m^{-3}$, and 0.08 μg $m^{-3}$ respectively. It is evident that regardless of extremely high or low observation concentrations, the differences in the analysis fields of various pollutants between the two systems are very small, and this small difference range is highly consistent with the previous comparison results of the increment fields. In general, the Py3DVAR system can effectively use observation information to correct the background field, and its analysis field results are highly consistent with the mature Fortran-3DVAR system, exhibiting reliable assimilation performance.



https://doi.org/10.5194/egusphere-2026-394
Preprint. Discussion started: 22 April 2026


**Figure 9.** Comparison of the background field (gray), observation field (black), Fortran-3DVAR analysis field (red), and Py3DVAR analysis field (blue) under extremely high (a) and extremely low (b) observation concentration conditions.

644

645 **4.2 Analysis of initial field assimilation effects: comparing control and Py3DVAR/Fortran-**
646 **3DVAR assimilation experiments at the 27 km grid resolution**

647 Fig. 10 and 11 present scatter comparisons between simulated results and observations for the
648 control experiment, the Py3DVAR_DCU assimilation experiment, and the Fortran-3DVAR
649 assimilation experiment at the 27 km grid resolution, respectively. In these figures, blue dots
650 represent the control experiment, while red dots represent the assimilation experiments. Although the
651 Py3DVAR assimilation system was tested on both CPU and DCU platforms, the results differ only
652 due to the computational platform. Therefore, the Py3DVAR_DCU assimilation experiment is
653 selected for analysis here. Regarding the distribution pattern, the blue dots from the Control
654 experiment are relatively scattered and deviate considerably from the 1:1 line. In contrast, the red
655 dots from the assimilation experiments show a clear convergence towards the 1:1 line, indicating that
656 both assimilation systems effectively improve the quality of the initial field, with a very pronounced
657 improvement effect.

658 Examining the specific evaluation metrics, the degree of improvement in the initial fields varies
659 among different pollutants. After assimilation, the CORR values for all pollutants increased
660 significantly. The correlation coefficients for $SO_2$ and CO both exceeded 0.8, while those for $NO_2$,
661 $O_3$, $PM_{2.5}$, and $PM_{10}$ reached above 0.9. Concurrently, the RMSE and MAE for all pollutants
662 decreased substantially compared to the Control experiment, further confirming the corrective effect
663 of assimilation on initial field biases.

664 By comparing the assimilation performance of the two systems, Py3DVAR_DCU and Fortran-
665 3DVAR, it can be observed that the red circles are highly close in distribution in the scatter plot,
666 indicating good consistency in the assimilation performance of the two systems. From the
667 perspective of quantitative indicators, for the six pollutants ($SO_2$, $NO_2$, CO, $O_3$, $PM_{2.5}$, and $PM_{10}$), the
668 degree of improvement in CORR, RMSE, and MAE between the two systems is basically equivalent,
669 with only a minor difference of 0.01 $\mu g\ m^{-3}$ in both the MAE of $O_3$ and the RMSE of $PM_{2.5}$. This
670 quantitative result is consistent with the difference characteristics of the increment fields and analysis
671 fields in the previous ideal experiments, further confirming the consistency of the two systems in
672 core assimilation performance.

https://doi.org/10.5194/egusphere-2026-394
Preprint. Discussion started: 22 April 2026

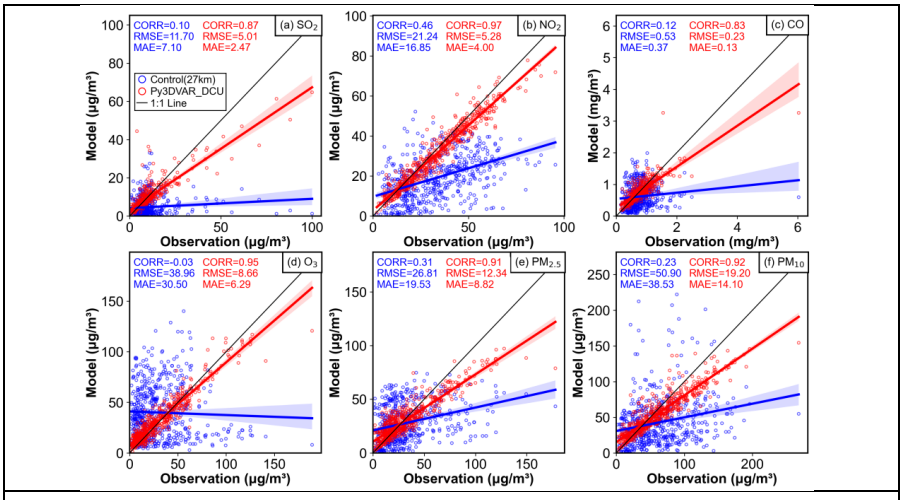




**Figure 10.** Scatter plots of the initial field for the Control and Py3DVAR_DCU assimilation experiments at the 27 km grid resolution. Red circles represent the Py3DVAR_DCU experiment, blue circles represent the Control experiment, and the black line is the 1:1 line.

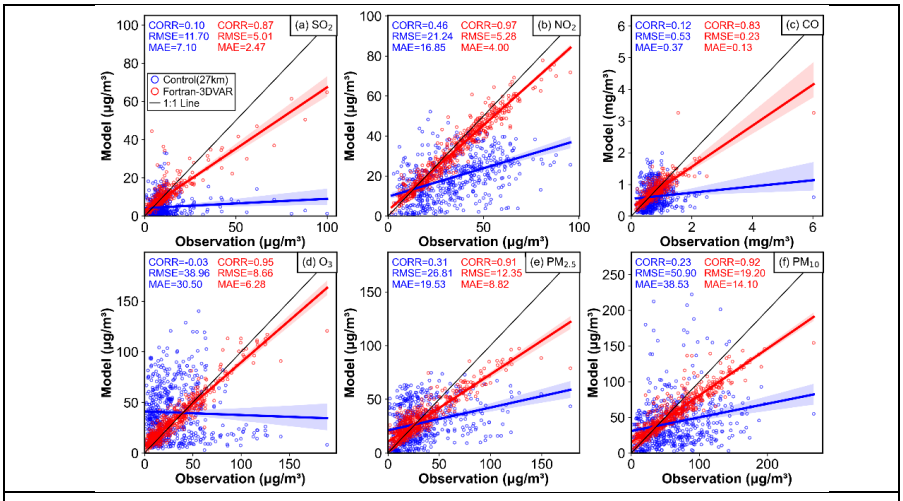



**Figure 11.** Scatter plots of the initial field for the Control and Fortran-3DVAR assimilation experiments at the 27 km grid resolution. Red circles represent the Fortran-3DVAR experiment, blue circles represent the Control experiment, and the black line is the 1:1 line.

### 4.3 Analysis of initial field assimilation effects: comparing control and Py3DVAR/Fortran-3DVAR assimilation experiments at the 9 km grid resolution

Fig. 12 and 13 display scatter comparisons between simulated results and observations for the control experiment, the Py3DVAR_DCU assimilation experiment, and the Fortran-3DVAR

https://doi.org/10.5194/egusphere-2026-394
Preprint. Discussion started: 22 April 2026


assimilation experiment, respectively, at the 9 km grid resolution. In these plots, blue dots represent the control experiment, while red dots represent the assimilation experiments. It should be noted that the 9 km grid experiments serve as a supplement to the 27 km grid experiments, aiming to preliminarily investigate the system's performance in a higher-resolution scenario.

From the control experiment (Control), it can be observed that the model simulations for all six pollutants are generally lower than the observations. Compared to the control experiment at the 27 km grid, certain differences exist, primarily due to variations in model errors at different grid resolutions and inconsistencies in the scale of the emission inventory. The assimilation results show that both Py3DVAR_DCU and Fortran-3DVAR effectively correct the systematic underestimation in the background field. After assimilation, the scatter points of the analysis fields align more closely with the 1:1 line, indicating that the assimilation systems maintain good performance at the 9 km grid. A further comparison between the two assimilation experiments reveals that their scatter distributions are highly overlapping, and the values of various evaluation metrics are essentially consistent. This aligns with the conclusions drawn from Figures 10 and 11 for the 27 km grid, demonstrating that Py3DVAR achieves assimilation effectiveness comparable to the traditional Fortran-3DVAR across different resolutions.

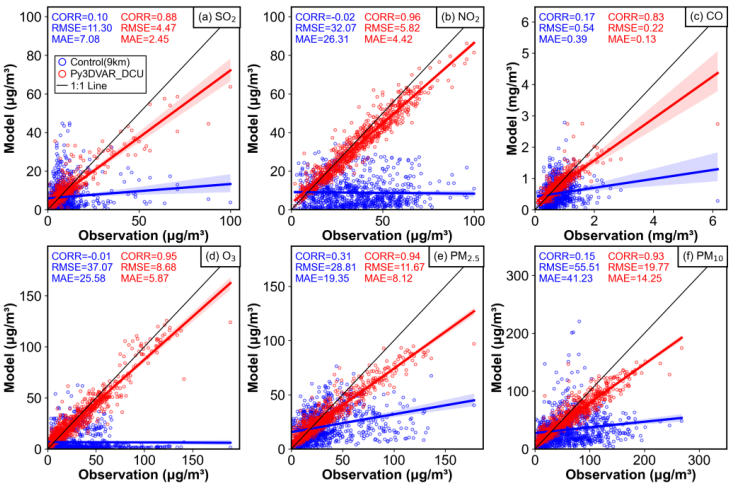



**Figure 12.** Scatter plots of the initial field for the Control and Py3DVAR_DCU assimilation experiments at the 9 km grid resolution. Red circles represent the Py3DVAR_DCU experiment, blue circles represent the Control experiment, and the black line is the 1:1 line.

https://doi.org/10.5194/egusphere-2026-394
Preprint. Discussion started: 22 April 2026

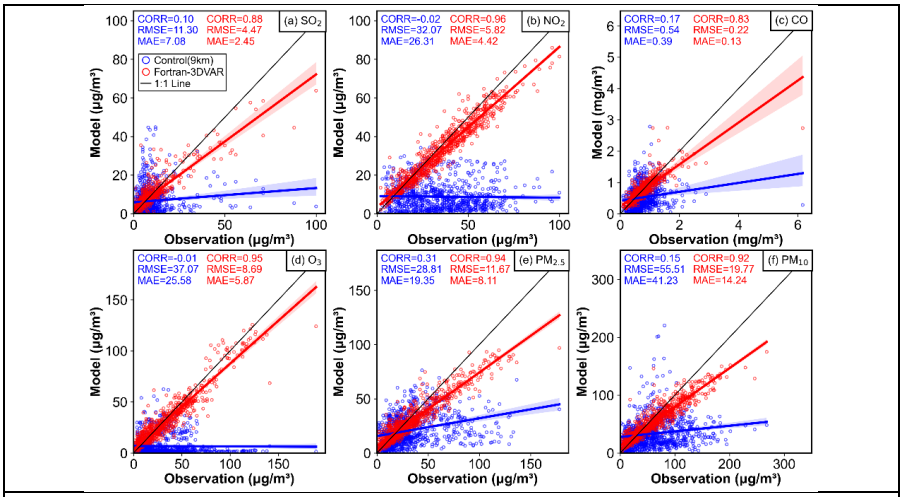




**Figure 13.** Scatter plots of the initial field for the Control and Fortran-3DVAR assimilation experiments at the 9 km grid resolution. Red circles represent the Fortran-3DVAR experiment, blue circles represent the Control experiment, and the black line is the 1:1 line.

### 4.4 Analysis of the Assimilation System Iteration Process and Cost Function Descent Characteristics

Fig. 14 (a) and (b) illustrate the descent trends of the cost function values with iteration counts for different assimilation systems at the 27 km and 9 km grid resolutions, respectively. The figures show that the cost function values for both types of assimilation systems on different computing platforms decrease rapidly as iterations proceed, exhibiting typical nonlinear convergence characteristics: the descent rate is fastest during the initial iterations, gradually narrows as iterations increase, and the curves flatten, indicating that the assimilation systems gradually approach the optimal solution and the convergence rate slows down.

At the 27 km grid resolution, the numerical characteristics of the cost functions across the assimilation systems show a high degree of consistency. For the Py3DVAR system, the initial cost function values on the CPU and DCU platforms are 127,484.29 and 127,484.30, respectively, with final convergence costs of 24,906.77 and 24,906.77. For the Fortran-3DVAR system, the initial cost is 127,483.55 and the final convergence cost is 24,906.90. The comparison reveals that the cost value differences for the Py3DVAR system between computing platforms are minimal, attributable to minor numerical deviations resulting from floating-point operation rule differences between the CPU and DCU platforms. The differences in initial and final costs between Py3DVAR and Fortran-3DVAR are slightly larger than the former but remain within an acceptable range overall. These discrepancies are inherent to numerical computations under different algorithmic frameworks and computing architectures, are reasonable, and align with the pattern of numerical deviations in the

https://doi.org/10.5194/egusphere-2026-394
Preprint. Discussion started: 22 April 2026


assimilation results discussed earlier. In summary, the final convergence costs for the three experiment sets are very close, indicating that despite differences in iteration paths, the systems ultimately achieve equivalent convergence outcomes. This conclusion corroborates the earlier assertion of "highly consistent assimilation effectiveness."

Analyzing the dynamic characteristics of the iteration process reveals that the initial iteration stage is crucial for cost function reduction, and the Py3DVAR system exhibits a significantly higher cost reduction rate than Fortran-3DVAR. Specifically, Py3DVAR achieves a substantial decrease in the cost function value in the very first iteration, whereas the reduction in the first iteration for Fortran-3DVAR is relatively limited. This difference stems from the distinct implementation approaches of the optimization algorithms in the two systems: Fortran-3DVAR employs a manually coded L-BFGS optimization algorithm, while Py3DVAR relies on the built-in gradient-descent-class optimizer from the PyTorch framework. The differences in engineering implementation directly lead to the divergence in convergence rates, further highlighting the optimization efficiency advantage of the Py3DVAR system. Additionally, comparing the Py3DVAR_CPU and Py3DVAR_DCU experiments shows that, although they use identical code and input data, subtle differences exist in the iteration process due to hardware architecture variations between the computing platforms. However, the overall iteration trends are highly consistent, demonstrating good cross-platform stability.

In terms of the number of iterations required for convergence, the Py3DVAR system requires only 15 and 14 iterations to converge on the CPU and DCU platforms, respectively, while Fortran-3DVAR requires 25 iterations, a significantly higher count. The core reason for this disparity is that, although both systems are based on the core principles of the L-BFGS algorithm, there are differences in the engineering details of their implementations—specifically in convergence threshold settings, step-size adaptation strategies, and Hessian matrix approximation update methods. Furthermore, the slower initial cost reduction rate of Fortran-3DVAR ultimately necessitates more iterations to meet the convergence criteria. These results fully demonstrate the more efficient convergence characteristics of the Py3DVAR system.

As a comparative test, the pattern of cost function descent at the 9 km grid resolution is highly consistent with that at the 27 km grid: all assimilation systems exhibit the "rapid initial descent followed by gradual flattening" convergence feature, and the differences in iteration processes between Py3DVAR and Fortran-3DVAR also follow the same pattern observed at the 27 km grid. This indicates that the Py3DVAR assimilation system maintains good numerical stability and convergence consistency across different horizontal resolutions.

https://doi.org/10.5194/egusphere-2026-394
Preprint. Discussion started: 22 April 2026

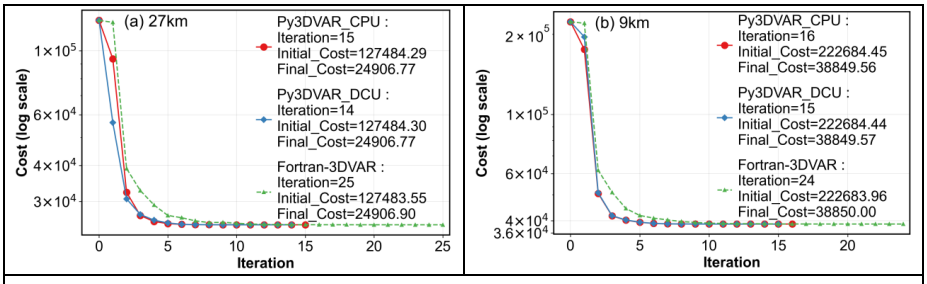




**Figure 14.** Descent of the cost function with iteration number for different assimilation systems at (a) 27 km and (b) 9 km grid resolutions. Red, blue, and green lines represent the Py3DVAR_DCU, Py3DVAR_CPU, and Fortran-3DVAR experiments, respectively.

### 4.5 Analysis of Iteration Time Consumption in the Assimilation System

Fig. 15 (a) and (b) present the characteristics of single-iteration time consumption and cumulative time consumption with iteration counts for the Py3DVAR_CPU and Py3DVAR_DCU experiments at the 27 km and 9 km grid resolutions, respectively. Overall, the iteration time consumption of the Py3DVAR assimilation system on both computing platforms follows a consistent pattern: the first iteration requires significantly more time than subsequent iterations. Specifically, the first iteration time exceeds 0.6 s for the Py3DVAR_CPU experiment and 0.5 s for the Py3DVAR_DCU experiment. In contrast, the time for subsequent iterations decreases markedly and stabilizes. For Py3DVAR_CPU, the single-iteration time remains in the range of 0.4 – 0.5 s, while for Py3DVAR_DCU, it stabilizes around 0.1 s. This phenomenon stems from the computational characteristics of the PyTorch framework: during the first execution, the automatic differentiation mechanism needs to construct and initialize the computational graph, incurring additional overhead. Subsequent iterations can directly reuse the already-built computational graph without repeated initialization, leading to significantly reduced and stable time consumption. It is noteworthy that the acceleration advantage of the DCU computing platform is not fully manifested in the first iteration; its core benefit primarily emerges in subsequent iterations. Compared to the CPU platform, the DCU substantially reduces the single-iteration time for later iterations due to its parallel computing architecture, further highlighting the efficiency enhancement of heterogeneous computing for assimilation.

Analyzing the trend of cumulative time consumption, the cumulative time curves for both experiments exhibit an almost linear increase, indicating that the total computation time is approximately linearly correlated with the number of iterations. This characteristic reflects the good computational stability and predictability of the system during the iteration process.

In the comparative experiments at the 9 km grid, due to the significantly larger grid size and expanded computational scale, both single-iteration time and cumulative time increase substantially.

https://doi.org/10.5194/egusphere-2026-394
Preprint. Discussion started: 22 April 2026


However, the overall pattern of time consumption remains consistent with that observed at the 27 km grid: higher overhead for the first iteration, stable time for subsequent iterations, and an approximately linear growth in cumulative time. This further demonstrates that Py3DVAR maintains stable computational characteristics across different resolutions.

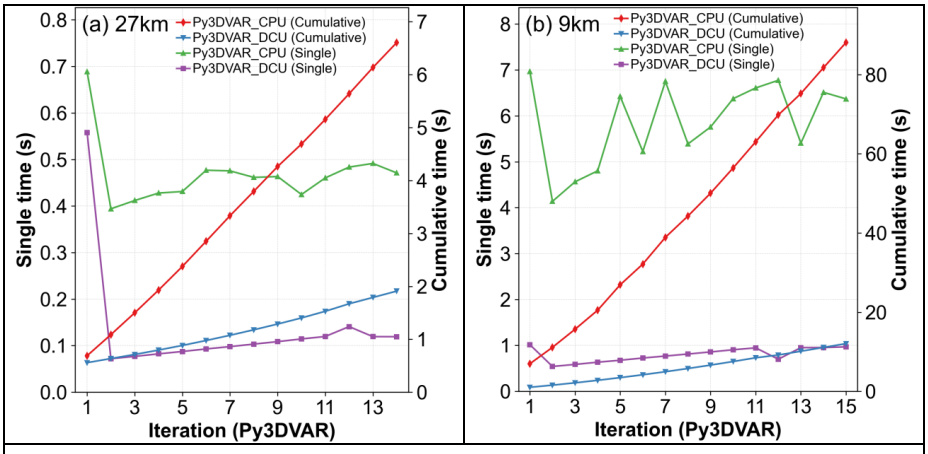



**Figure 15.** Iteration time consumption for different assimilation systems at (a) 27 km and (b) 9 km grid resolutions. Red and green lines represent the cumulative and single-iteration time, respectively, for the Py3DVAR_CPU experiment. Blue and purple lines represent the cumulative and single-iteration time, respectively, for the Py3DVAR_DCU experiment.

## 4.6 Comparison of Computational Efficiency for Assimilation Systems at Different Grid Resolutions

Fig. 16 (a) and (b) present the comparison of total iteration times for the Fortran-3DVAR, Py3DVAR_CPU, and Py3DVAR_DCU experiments at the 27 km and 9 km grid resolutions, respectively.

At the 27 km grid, the total iteration times for Fortran-3DVAR, Py3DVAR_CPU, and Py3DVAR_DCU are 62.5 s, 7.1 s, and 1.9 s, respectively. Although both Fortran-3DVAR and Py3DVAR_CPU run on the same CPU platform, the latter achieves a speedup of 8.8 times. This significant performance improvement stems primarily from fundamental differences in their computational paradigms: Fortran-3DVAR employs loop-based serial computing with manually partitioned parallel regions, making it difficult to fully exploit the parallel potential of multi-core CPUs. In contrast, Py3DVAR leverages the vectorized operations and automatic parallel optimization of the PyTorch framework, enabling efficient utilization of multi-core CPU resources and thereby substantially enhancing computational efficiency. This demonstrates that Py3DVAR represents not only a code-level refactoring but also an innovation in computational paradigm.

Furthermore, Py3DVAR_DCU, running on the DCU platform, achieves a 3.7 times speedup

https://doi.org/10.5194/egusphere-2026-394
Preprint. Discussion started: 22 April 2026


compared to Py3DVAR_CPU, reducing the total time to under 1.9 s and demonstrating preliminary potential for near-real-time assimilation. Compared to the traditional Fortran-3DVAR, Py3DVAR_DCU achieves an even greater speedup of 32.7 times, highlighting its significant performance advantages within the existing hardware architecture.

At the 9 km grid, as the computational scale increases, the time consumption for all systems rises substantially, but the acceleration pattern exhibits new characteristics: the speedup of Py3DVAR_CPU relative to Fortran-3DVAR decreases slightly to 5.1 times, while the speedup of Py3DVAR_DCU relative to Fortran-3DVAR increases further to 40.3 times. The speedup of Py3DVAR_DCU relative to Py3DVAR_CPU also increases to 7.8 times. This indicates that as the problem scale expands, reliance solely on multi-core CPU parallelism gradually reaches a bottleneck, whereas the advantages of massive parallel computing on the DCU platform become more fully realized. With smaller grid scales, the parallel computing units of the DCU may not be fully utilized, resulting in a limited performance gap compared to the CPU. However, when the computational scale increases significantly, the DCU can more efficiently schedule its vast number of computing cores, thereby exhibiting more substantial acceleration effects.

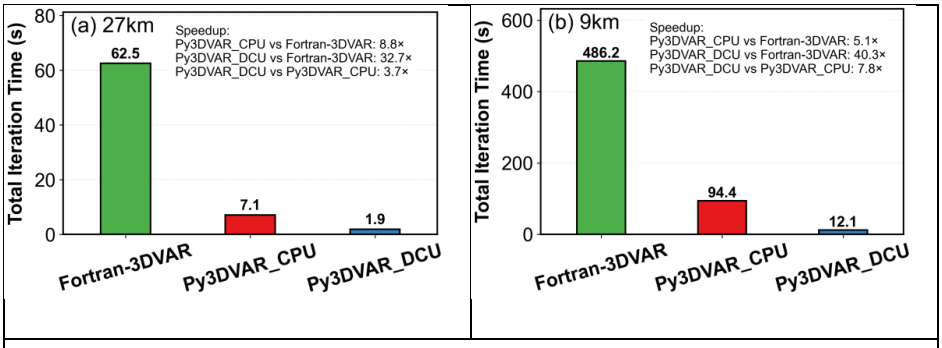



**Figure 16.** Total L-BFGS gradient descent iteration time and speedup ratios of different experiments for assimilation systems at (a) 27 km and (b) 9 km grid resolutions. Green, red, and blue bars represent the Fortran-3DVAR, Py3DVAR_DCU, and Py3DVAR_CPU experiments, respectively.

### 4.7 Comparative analysis of 24-hour forecast performance between Py3DVAR and Fortran-3DVAR at the 27 km grid resolution

Fig. 17 and 18 display the temporal evolution of the correlation coefficient (CORR), root mean square error (RMSE), and mean absolute error (MAE) for various pollutants during the 24-hour forecast period, initialized from the analysis fields of Py3DVAR and Fortran-3DVAR, respectively. Comparing the forecast performance of the two systems reveals that the curves of the forecast metrics for all pollutants are fundamentally consistent in terms of morphological characteristics, temporal trends, and numerical values. This finding corroborates the previous results showing minor

https://doi.org/10.5194/egusphere-2026-394
Preprint. Discussion started: 22 April 2026


differences in idealized experiment increment fields, similar analysis field effectiveness in the real case, and comparable cost function convergence characteristics during iteration. It fully demonstrates that the Py3DVAR system achieves performance consistent with the mature Fortran-3DVAR system not only in initial field assimilation but also in subsequent forecast applications, thoroughly verifying the reliability and practical utility of its assimilation results.

Regarding the evolution of forecast performance, although the initial field is significantly optimized by assimilation, this positive effect decays rapidly as the forecast lead time increases. This decay primarily stems from inherent model uncertainties, spatiotemporal errors in emission inventories, and inaccuracies in the simulated meteorological fields affecting pollutant dispersion and transport. These errors accumulate during the forecast process and gradually become dominant, thereby progressively diminishing the positive effect of the assimilation improvements.

It is noteworthy that the duration of the positive assimilation effect varies significantly among different pollutants. The improvement for $SO_2$, CO, $PM_{2.5}$, and $PM_{10}$ can be maintained for over 24 hours, whereas the positive effects for $NO_2$ and $O_3$ are shorter, lasting approximately 8 hours and 9 hours, respectively. This difference is mainly determined by the respective physicochemical properties and dominant life-cycle processes of the pollutants. $SO_2$, CO, and particulate matter typically have longer atmospheric lifetimes, and their concentration distributions are more influenced by large-scale transport processes. Consequently, improvements in the initial field yield more lasting positive effects. In contrast, $NO_2$ and $O_3$ have strong photochemical reactivity. Their concentrations during daytime are more significantly governed by local photochemical processes and boundary layer variations, causing the initial field improvements from assimilation to dissipate quickly.

In summary, Py3DVAR not only possesses assimilation capabilities comparable to Fortran-3DVAR but also produces positive effects with good forecast persistence. Although this effect decays with increasing forecast lead time, its improvement for short-term forecasts of various pollutants, especially long-lived species, is clear and effective.

https://doi.org/10.5194/egusphere-2026-394
Preprint. Discussion started: 22 April 2026

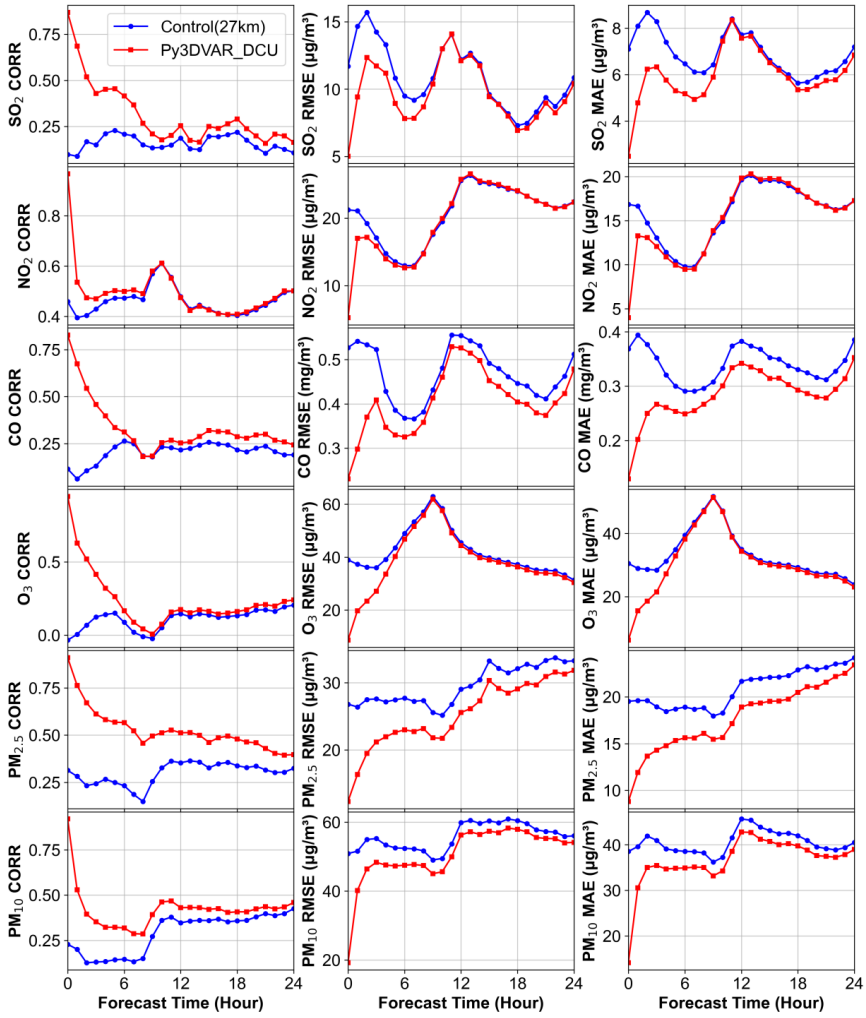




**Figure 17.** Temporal evolution of the 24-hour forecast performance for the Py3DVAR_DCU assimilation experiment and the Control experiment for the six pollutants ($SO_2$, $NO_2$, CO, $O_3$, $PM_{2.5}$, $PM_{10}$). Each column represents one of the three evaluation metrics: correlation coefficient (CORR), root mean square error (RMSE), and mean absolute error (MAE), in sequence. The red lines represent the results of the Py3DVAR_DCU experiment, and the blue lines represent the results of the Control experiment.

846

https://doi.org/10.5194/egusphere-2026-394
Preprint. Discussion started: 22 April 2026

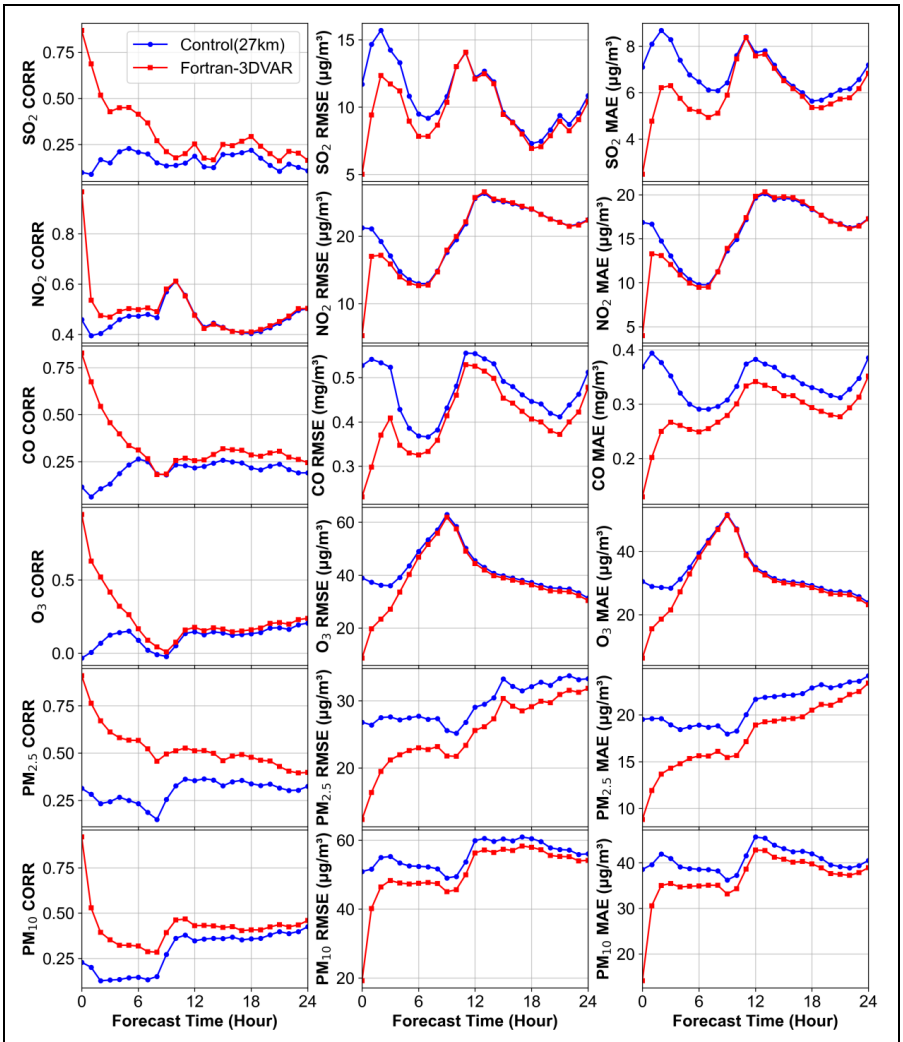




**Figure 18.** Temporal evolution of the 24-hour forecast performance for the Fortran-3DVAR assimilation experiment and the Control experiment for the six pollutants ($SO_2$, $NO_2$, CO, $O_3$, $PM_{2.5}$, $PM_{10}$). Each column represents one of the three evaluation metrics: correlation coefficient (CORR), root mean square error (RMSE), and mean absolute error (MAE), in sequence. The red lines represent the results of the Fortran-3DVAR experiment, and the blue lines represent the results of the Control experiment.

## 5. Summary and outlook

Based on the three-dimensional variational (3D-VAR) assimilation principle and integrated with the PyTorch deep learning framework, this study successfully developed a DCU-accelerated

https://doi.org/10.5194/egusphere-2026-394
Preprint. Discussion started: 22 April 2026


atmospheric chemical three-dimensional variational data assimilation system (Py3DVAR). The system is built upon the CMAQ-AERO5 aerosol scheme coupled with the WRF-Chem model. It consolidates 27 aerosol species into 12 aerosol control variables and directly designates four gaseous pollutants as gaseous assimilation control variables. Simultaneously, the background error covariance matrix is constructed using the NMC method, enabling the simultaneous assimilation of ground-observed concentrations of $SO_2$, $NO_2$, CO, $O_3$, $PM_{2.5}$, and $PM_{10}$. At the algorithm implementation level, the core computations of the system replace the traditional Fortran-based loop structures with Torch tensors. Leveraging PyTorch's automatic differentiation mechanism and its built-in L-BFGS optimizer, the system achieves efficient minimization on DCU hardware.

To comprehensively evaluate the system's performance and computational efficiency, this study conducted an idealized experiment, multi-platform and multi-resolution real-case assimilation experiments, and a short-term forecast experiment, all compared and validated against the traditional Fortran-3DVAR system. The idealized experiment results demonstrate that Py3DVAR can propagate observation information both horizontally and vertically through the background error covariance, generating Gaussian-distributed increment fields. These fields show high consistency with those from Fortran-3DVAR, verifying the physical soundness of the system.

Real-case assimilation experiments further confirm that Py3DVAR significantly optimizes the model's background field, substantially improves the correlation (CORR) between the analysis field and observations, and reduces the root mean square error (RMSE) and mean absolute error (MAE). At the 27 km grid resolution, the Py3DVAR_DCU experiment increased the CORR for $SO_2$, $NO_2$, CO, $O_3$, $PM_{2.5}$, and $PM_{10}$ by 0.77, 0.51, 0.71, 0.98, 0.60, and 0.69, respectively. The RMSE was reduced by 6.69 $\mu g \cdot m^{-3}$ (57.2%), 15.96 $\mu g \cdot m^{-3}$ (75.1%), 0.30 $mg \cdot m^{-3}$ (56.6%), 30.30 $\mu g \cdot m^{-3}$ (77.8%), 14.47 $\mu g \cdot m^{-3}$ (54.0%), and 31.70 $\mu g \cdot m^{-3}$ (62.3%), respectively. The MAE was reduced by 4.63 $\mu g \cdot m^{-3}$ (65.2%), 12.85 $\mu g \cdot m^{-3}$ (76.3%), 0.24 $mg \cdot m^{-3}$ (64.9%), 24.21 $\mu g \cdot m^{-3}$ (79.3%), 10.71 $\mu g \cdot m^{-3}$ (54.8%), and 24.43 $\mu g \cdot m^{-3}$ (63.4%), respectively.

At the 9 km grid resolution, the Py3DVAR_DCU experiment increased the CORR for $SO_2$, $NO_2$, CO, $O_3$, $PM_{2.5}$, and $PM_{10}$ by 0.78, 0.98, 0.66, 0.96, 0.63, and 0.78, respectively. The RMSE was reduced by 6.83 $\mu g \cdot m^{-3}$ (60.4%), 26.25 $\mu g \cdot m^{-3}$ (81.8%), 0.32 $mg \cdot m^{-3}$ (59.3%), 28.39 $\mu g \cdot m^{-3}$ (76.6%), 15.14 $\mu g \cdot m^{-3}$ (56.5%), and 35.74 $\mu g \cdot m^{-3}$ (64.4%), respectively. The MAE was reduced by 4.63 $\mu g \cdot m^{-3}$ (65.4%), 21.89 $\mu g \cdot m^{-3}$ (83.2%), 0.26 $mg \cdot m^{-3}$ (66.7%), 19.71 $\mu g \cdot m^{-3}$ (77.1%), 11.23 $\mu g \cdot m^{-3}$ (58.0%), and 26.98 $\mu g \cdot m^{-3}$ (65.4%), respectively.

The assimilation accuracy at both grid resolutions is comparable to that of Fortran-3DVAR, while Py3DVAR exhibits exceptional computational efficiency. Even on identical CPU hardware, Py3DVAR achieves an 8.8-fold speedup compared to the original loop-based Fortran-3DVAR due to its use of vectorized PyTorch tensor computations, revealing the inherent efficiency advantages of modern computational frameworks. When deployed on the DCU platform, the acceleration benefits

https://doi.org/10.5194/egusphere-2026-394
Preprint. Discussion started: 22 April 2026


are even more substantial. At the 27 km grid, the total iteration time is reduced from 62.5 s for the Fortran-CPU version to 1.9 s, representing a speedup of 32.7-fold. At the larger-scale 9 km grid, the speedup further increases to 40.3-fold, highlighting the immense potential of DCU's massively parallel computing for handling high-dimensional assimilation problems.

The forecast experiment results indicate that the improvements from assimilation have a sustained positive effect on short-term air quality forecasts. The positive effects for $SO_2$, CO, $PM_{2.5}$, and $PM_{10}$ last over 24 hours, while those for $NO_2$ and $O_3$ last over 6 hours. Furthermore, the decay trends of the forecast improvement effects are consistent between Py3DVAR_DCU and Py3DVAR_CPU, further validating the stability and consistency of the system's assimilation effectiveness.

In summary, the Py3DVAR system developed in this study successfully achieves orders-of-magnitude improvements in computational efficiency while preserving the physical interpretability and numerical accuracy of traditional 3D-VAR assimilation. This is accomplished through a synergistic innovation encompassing both the algorithm implementation paradigm (shifting from manual derivation to automatic differentiation) and the computational hardware architecture (transitioning from CPU to DCU). This work not only provides a new solution that combines timeliness and reliability for the operational application of atmospheric chemical data assimilation but also, by being built upon the PyTorch framework, lays a crucial technical foundation for the future integration of deep learning models and the advancement towards intelligent assimilation.

*Code and data availability.* The WRF-Chem model source code can be downloaded at the WRF model download page (https://www.mmm.ucar.edu/models/wrf). The Py3DVAR system was developed by the authors. The exact version of the aerosol Py3DVAR system code and input data for supporting this paper are available at https://doi.org/10.5281/zenodo.18401481 (You, 2026).

*Author contributions.* Hancheng Ye conceived the study, developed the Py3DVAR assimilation system, performed all numerical simulations and assimilation experiments, conducted the formal analysis, and wrote the original manuscript. Zengliang Zang and Wei You contributed to the methodology design and assisted in the validation and evaluation of the assimilation system. Yiwen Hu, Ning Liu and Yi Li discussed the assimilation effects of Py3DVAR, provided supervision for the research, and reviewed and edited the manuscript. All authors have read and agreed to the published version of the manuscript.

*Competing interests.* The contact author has declared that neither they nor their co-authors have any competing interests.

https://doi.org/10.5194/egusphere-2026-394
Preprint. Discussion started: 22 April 2026


*Acknowledgements.* We would like to express our sincere gratitude to the National Centers for Environmental Prediction (NCEP) for providing the FNL Operational Global Analysis dataset, and to the China National Environmental Monitoring Center (CNEMC) for making available the observational data of atmospheric pollutants.

*Financial support.* This research is supported by the National Natural Science Foundation of China (Grant Nos. 42430612, 42475166 and 42405149).

## References

Austin, J.: Toward the four dimensional assimilation of stratospheric chemical constituents, J. Geophys. Res.-Atmos., 97, 2569-2588, https://doi.org/10.1029/91JD02603, 1992.

Beauchamp, M., Karagali, I., Gacitúa, G., Høyer, J. L., Ballarotta, M., and Fablet, R.: Multiscale neural assimilation scheme for high-resolution sea surface temperature reconstruction from satellite observations, Sci Rep, 15, 39928, https://doi.org/10.1038/s41598-025-23682-9, 2025.

Boudier, P., Fillion, A., Gratton, S., Gürol, S., and Zhang, S.: Data Assimilation Networks, J Adv Model Earth Syst, 15, e2022MS003353, https://doi.org/10.1029/2022MS003353, 2023.

Chattopadhyay, A., Nabizadeh, E., Bach, E., and Hassanzadeh, P.: Deep learning-enhanced ensemble-based data assimilation for high-dimensional nonlinear dynamical systems, J. Comput. Phys., 477, 111918, https://doi.org/10.1016/j.jcp.2023.111918, 2023.

Chen, D., Liu, Z., Davis, C., and Gu, Y.: Dust radiative effects on atmospheric thermodynamics and tropical cyclogenesis over the Atlantic Ocean using WRF-Chem coupled with an AOD data assimilation system, Atmos. Chem. Phys., 17, 7917-7939, https://doi.org/10.5194/acp-17-7917-2017, 2017.

Chen, H., Gao, J., Sun, T., Chen, Y., Wang, Y., and Carlin, J. T.: Assimilation of Water Vapor Retrievals from ZDR Columns Using the 3DVar Method for Improving the Short-Term Prediction of Convective Storms, Mon. Wea. Rev., 152, 1077-1095, https://doi.org/10.1175/MWR-D-23-0196.1, 2024.

De Luca, P., Galletti, A., Giunta, G., and Marcellino, L.: Recursive filter based GPU algorithms in a Data Assimilation scenario, J. Comput. Sci., 53, 101339, https://doi.org/10.1016/j.jocs.2021.101339, 2021.

Denby, B., Schaap, M., Segers, A., Builtjes, P., and Horálek, J.: Comparison of two data assimilation

https://doi.org/10.5194/egusphere-2026-394
Preprint. Discussion started: 22 April 2026


methods for assessing PM10 exceedances on the European scale, Atmos. Environ., 42, 7122 – 7134, https://doi.org/10.1016/j.atmosenv.2008.05.058, 2008.

Parrish, D. F. and Derber, J. C.: The National Meteorological Center's Spectral Statistical-Interpolation Analysis System, Mon. Wea. Rev., 120, 1747-1763, https://doi.org/10.1175/1520-0493(1992)120%3C1747:TNMCSS%3E2.0.CO;2, 1992.

Dong, R., Leng, H., Zhao, J., Song, J., and Liang, S.: A Framework for Four-Dimensional Variational Data Assimilation Based on Machine Learning, Entropy, 24, 264, https://doi.org/10.3390/e24020264, 2022.

Elbern, H., Schmidt, H., and Ebel, A.: Variational data assimilation for tropospheric chemistry modeling, J. Geophys. Res.-Atmos., 102, 15967-15985, https://doi.org/10.1029/97JD01213, 1997.

Engelen, R. J., Serrar, S., and Chevallier, F.: Four-dimensional data assimilation of atmospheric CO2 using AIRS observations, J. Geophys. Res.-Atmos., 114, 2008JD010739, https://doi.org/10.1029/2008JD010739, 2009.

Feng, T., G. Li, J. Cao, N. Bei, Z. Shen, W. Zhou, S. Liu, T. Zhang, Y. Li, G., Zhang, R., Fan, J., and Tie, X.: Impacts of black carbon aerosol on photolysis and ozone, J. Geophys. Res., 110, 10.1029/2005jd005898, 2005.

Feng, T., N. Bei, R. Huang, J. Cao, Q. Zhang, W. Zhou, X. Tie, S. Liu, T. Zhang, X. Su, W. Lei, L. T. Molina, and G. Li.: Summertime ozone formation in Xi'an and surrounding areas, China, Atmos. Chem. Phys., 16, 4323-4342, 2016.

Frydendall, J., Brandt, J., and Christensen, J. H.: Implementation and testing of a simple data assimilation algorithm in the regional air pollution forecast model, DEOM, Atmos. Chem. Phys., 9, 5475-5488, https://doi.org/10.5194/acp-9-5475-2009, 2009.

Gao, C., Zhang, X., Xiu, A., Tong, Q., Zhao, H., Zhang, S., Yang, G., Zhang, M., and Xie, S.: Intercomparison of multiple two-way coupled meteorology and air quality models (WRF v4.1.1-CMAQ v5.3.1, WRF-Chem v4.1.1, and WRF v3.7.1-CHIMERE v2020r1) in eastern China, Geosci. Model Dev., 17, 2471-2492, https://doi.org/10.5194/gmd-17-2471-2024, 2024.

He, X., Zang, Z., Li, Y., Liu, N., Hu, Y., Liu, L., and You, W.: Inversion of a Near-Real-Time China Gridded Hourly SO2 Emission Inventory Using Deep Learning Combined With 3D-Variational Assimilation, Geophys. Res. Lett., 52, e2024GL114375, https://doi.org/10.1029/2024GL114375, 2025.

Howard, L. J., Subramanian, A., and Hoteit, I.: A Machine Learning Augmented Data Assimilation Method for High-Resolution Observations, J Adv Model Earth Syst, 16, e2023MS003774, https://doi.org/10.1029/2023MS003774, 2024.

https://doi.org/10.5194/egusphere-2026-394
Preprint. Discussion started: 22 April 2026


Jiang, Z., Liu, Z., Wang, T., Schwartz, C. S., Lin, H., and Jiang, F.: Probing into the impact of 3DVAR assimilation of surface PM10 observations over China using process analysis, JGR Atmospheres, 118, 6738-6749, https://doi.org/10.1002/jgrd.50495, 2013.

Li, C., Fang, H., Cao, X., Duan, D., Xiao, C., Huang, H., Ren, G., Lin, Y., and Cai, Y.: A New Method for Reconstruction of Regional Three-Dimensional Electron Density Distributions Using AI-Based Data Assimilation Method and Incoherent Scatter Radar Measurements, Geophys. Res. Lett., 51, e2024GL112352, https://doi.org/10.1029/2024GL112352, 2024.

Li, G., Lei, W., Zavala, M., Volkamer, R., Dusanter, S., Stevens, P., and Molina, L. T.: Impacts of HONO sources on the photochemistry in Mexico City during the MCMA-2006/MILAGO Campaign, Atmos. Chem. Phys., 10, 6551-6567, 10.5194/acp-10-6551-2010, 2010.

Li, G., Bei, N., Tie, X., and Molina, L. T.: Aerosol effects on the photochemistry in Mexico City during MCMA-2006/MILAGRO campaign, Atmos. Chem. Phys., 11, 5169-5182, 10.5194/acp-11-5169-2011, 2011a.

Li, G., Zavala, M., Lei, W., Tsimpidi, A. P., Karydis, V. A., Pandis, S. N., agaratna, M. R., and Molina, L. T.: Simulations of organic aerosol concentrations in Mexico City using the WRF-CHEM model during the MCMA-2006/MILAGRO campaign, Atmos. Chem. Phys., 11, 3789-3809, 10.5194/acp-11-3789-2011, 2011b.

Li, G., Lei, W., Bei, N., and Molina, L. T.: Contribution of garbage burning to chloride and PM2.5 in Mexico City, Atmos. Chem. Phys., 12, 8751-8761, 10.5194/acp-12-8751-2012, 2012.

Li, H., Yang, T., Kong, L., Zhang, D., Tang, G., and Wang, Z.: OIRF-LEnKF v1.0: A Self-evolving Data Assimilation System by Integrating Incremental Machine Learning with a Localized EnKF for Enhanced PM2.5 Chemical Component Forecasting and Analysis, EGUsphere [preprint], https://doi.org/10.5194/egusphere-2025-3960, 2025.

Li, Z., Zang, Z., Li, Q. B., Chao, Y., Chen, D., Ye, Z., Liu, Y., and Liou, K. N.: A three-dimensional variational data assimilation system for multiple aerosol species with WRF/Chem and an application to PM2.5 prediction, Atmos. Chem. Phys., 13, 4265-4278, https://doi.org/10.5194/acp-13-4265-2013, 2013.

Liang, Y., Zang, Z., Liu, D., Yan, P., Hu, Y., Zhou, Y., and You, W.: Development of a three-dimensional variational assimilation system for lidar profile data based on a size-resolved aerosol model in WRF-Chem model v3.9.1 and its application in PM2.5 forecasts across China, Geosci. Model Dev., 13, 6285-6301, https://doi.org/10.5194/gmd-13-6285-2020, 2020.

Lu, F.: U-Net Kalman Filter (UNetKF): An Example of Machine Learning-Assisted Data Assimilation, J Adv Model Earth Syst, 17, e2023MS003979, https://doi.org/10.1029/2023MS003979, 2025.

https://doi.org/10.5194/egusphere-2026-394
Preprint. Discussion started: 22 April 2026


Ma, X., Liu, H., and Peng, Z.: Improving WRF-Chem PM2.5 predictions by combining data assimilation and deep-learning-based bias correction, Environ. Int., 195, 109199, https://doi.org/10.1016/j.envint.2024.109199, 2025.

Pagowski, M., Grell, G. A., McKeen, S. A., Peckham, S. E., and Devenyi, D.: Three-dimensional variational data assimilation of ozone and fine particulate matter observations: some results using the Weather Research and Forecasting-Chemistry model and Grid-point Statistical Interpolation, Quart. J. Royal. Meteoro. Soc., 136, 2013-2024, https://doi.org/10.1002/qj.700, 2010.

Pang, J., Liu, Z., Wang, X., Bresch, J., Ban, J., Chen, D., and Kim, J.: Assimilating AOD retrievals from GOCI and VIIRS to forecast surface PM2.5 episodes over Eastern China, Atmos. Environ., 179, 288-304, https://doi.org/10.1016/j.atmosenv.2018.02.011, 2018.

Pang, J., Wang, X., Shao, M., Chen, W., and Chang, M.: Aerosol optical depth assimilation for a modal aerosol model: Implementation and application in AOD forecasts over East Asia, Sci. Tot. Environ., 719, 137430, https://doi.org/10.1016/j.scitotenv.2020.137430, 2020.

Seo, M.-G. and Kim, H. M.: Effect of meteorological data assimilation using 3DVAR on high-resolution simulations of atmospheric CO2 concentrations in East Asia, Atmos. Pollut. Res., 14, 101759, https://doi.org/10.1016/j.apr.2023.101759, 2023.

Tombette, M., Mallet, V., and Sportisse, B.: PM10 data assimilation over Europe with the optimal interpolation method, Atmos. Chem. Phys., 9, 57-70, https://doi.org/10.5194/acp-9-57-2009, 2009.

Wang, D., You, W., Zang, Z., Pan, X., Hu, Y., and Liang, Y.: A three-dimensional variational data assimilation system for aerosol optical properties based on WRF-Chem v4.0: design, development, and application of assimilating Himawari-8 aerosol observations, Geosci. Model Dev., 15, 1821-1840, https://doi.org/10.5194/gmd-15-1821-2022, 2022.

Wei You.: A PyTorch-based three-dimensional variational data assimilation system with automatic differentiation and DCU acceleration for atmospheric chemicals: development and application to WRF-Chem, Zenodo [data set], https://doi.org/10.5281/zenodo.18401481, 2026.

Xing, L., Wu, J., Elser, M., Tong, S., Liu, S., Li, X., Liu, L., Cao, J., Zhou, J., El-Haddad, I., Huang, R., Ge, M., Tie, X., Prévôt, A. S. H., and Li, G.: Wintertime secondary organic aerosol formation in Beijing-Tianjin-Hebei (BTH): Contributions of HONO sources and heterogeneous reactions, Atmos. Chem. Phys. Discuss., 2018, 1-25, 10.5194/acp-2018-770, 2018.

Xu, H., Duan, Y., and Xu, X.: Exploring the integration of a global AI model with traditional data assimilation in weather forecasting, Environ. Res. Lett., 19, 124079, https://doi.org/10.1088/1748-9326/ad93e8, 2024.

Xu, X., Sun, X., Han, W., Zhong, X., Chen, L., Gao, Z., and Li, H.: FuXi-DA: a generalized deep learning data assimilation framework for assimilating satellite observations, npj Clim Atmos Sci, 8,

https://doi.org/10.5194/egusphere-2026-394
Preprint. Discussion started: 22 April 2026


156, https://doi.org/10.1038/s41612-025-01039-3, 2025.

Ye, H., Pan, X., You, W., Zhu, X., Zang, Z., Wang, D., Zhang, X., Hu, Y., and Jin, S.: Impact of CALIPSO profile data assimilation on 3-D aerosol improvement in a size-resolved aerosol model, Atmos. Res., 264, 105877, https://doi.org/10.1016/j.atmosres.2021.105877, 2021.

Ye, H., You, W., Zang, Z., Pan, X., Wang, D., Zhou, N., Hu, Y., Liang, Y., and Yan, P.: Observing system simulation experiment (OSSE)-quantitative evaluation of lidar observation networks to improve 3D aerosol forecasting in China, Atmos. Res., 270, 106069, https://doi.org/10.1016/j.atmosres.2022.106069, 2022.

Zang, Z., Li, Z., Pan, X., Hao, Z., and You, W.: Aerosol data assimilation and forecasting experiments using aircraft and surface observations during CalNex, Tellus. B.: Chem. Phys. Meteo., 68, https://doi.org/10.3402/tellusb.v68.29812, 2016.

Zang, Z., You, W., Ye, H., Liang, Y., Li, Y., Wang, D., Hu, Y., and Yan, P.: 3DVAR Aerosol Data Assimilation and Evaluation Using Surface PM2.5, Himawari-8 AOD and CALIPSO Profile Observations in the North China, Remote. Sens., 14, 4009, https://doi.org/10.3390/rs14164009, 2022.

Zhang, J., Campbell, J. R., Reid, J. S., Westphal, D. L., Baker, N. L., Campbell, W. F., and Hyer, E. J.: Evaluating the impact of assimilating CALIOP-derived aerosol extinction profiles on a global mass transport model: CALIOP DATA ASSIMILATION, Geophys. Res. Lett., 38, n/a-n/a, https://doi.org/10.1029/2011GL047737, 2011.

Zhang, J., Spurr, R. J. D., Reid, J. S., Xian, P., Colarco, P. R., Campbell, J. R., Hyer, E. J., and Baker, N. L.: Development of an Ozone Monitoring Instrument (OMI) aerosol index (AI) data assimilation scheme for aerosol modeling over bright surfaces – a step toward direct radiance assimilation in the UV spectrum, Geosci. Model Dev., 14, 27 – 42, https://doi.org/10.5194/gmd-14-27-2021, 2021.

Zhou, Y., Sun, W., Liu, Z., Gao, L., Chen, D., Feng, J., Zhang, T., and Zhou, Z.: Further development and application of the WRFDA-Chem three-dimensional variational (3DVAR) system: Joint assimilation of satellite AOD retrievals and surface observations, Atmos. Res., 316, 107942, https://doi.org/10.1016/j.atmosres.2025.107942, 2025.