Check for updates

JAMES | Journal of Advances in Modeling Earth Systems

RESEARCH ARTICLE
10.1029/2025MS005319

Duo-AttnOPNets: Advancing Global Operational Forecasting for Atmospheric Carbon Monoxide With AI-Empowered 4D-Var

Judongyang Zhou[1,2], Fangxin Fang[1,2], Christopher C. Pain[2], Linfeng Li[2], Jie Zheng[2], Yanghua Wang[1,2], Ionel Michael Navon[3], and Jiang Zhu[4]

[1]Resource Geophysics Academy, Imperial College London, London, UK, [2]Department of Earth Science and Engineering, Imperial College London, London, UK, [3]Department of Scientific Computing, Florida State University, Tallahassee, FL, USA, [4]Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China

Key Points:

- DuoAttnOPNets delivers operational global CO forecasts in seconds with accuracy exceeding the European Center for Medium-Range Weather Forecasts CIFS system
- DuoAttnVarNet enables fast scalable data assimilation without retraining or manual adjoint development
- DuoAttnForeNet improves spatio-temporal learning via dual attention and recurrence architecture

Supporting Information:

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

Correspondence to:

F. Fang,
f.fang@imperial.ac.uk

Citation:

Zhou, J., Fang, F., Pain, C. C., Li, L., Zheng, J., Wang, Y., et al. (2026). Duo-AttnOPNets: Advancing global operational forecasting for atmospheric carbon monoxide with AI-empowered 4D-Var. *Journal of Advances in Modeling Earth Systems*, *18*, e2025MS005319. https://doi.org/10.1029/2025MS005319

Received 2 JUL 2025
Accepted 11 JAN 2026

Author Contributions:

Conceptualization: Fangxin Fang
Data curation: Judongyang Zhou
Formal analysis: Judongyang Zhou
Funding acquisition: Fangxin Fang
Investigation: Judongyang Zhou
Methodology: Judongyang Zhou
Project administration: Fangxin Fang
Resources: Judongyang Zhou
Supervision: Fangxin Fang
Validation: Judongyang Zhou
Visualization: Judongyang Zhou
Writing – original draft: Judongyang Zhou

© 2026 The Author(s). Journal of Advances in Modeling Earth Systems published by Wiley Periodicals LLC on behalf of American Geophysical Union. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Abstract The accuracy of global forecasting of atmospheric composition is essential for protecting public health and advancing climate research. Carbon monoxide (CO), a key pollutant with indirect greenhouse effects, requires timely and accurate prediction. We propose Duo-AttnOPNets, an operational framework that combines a deep learning-based forecasting module Duo-AttnForeNet with a training-free data assimilation module Duo-AttnVarNet. Duo-AttnForeNet employs CSLSTM blocks—built from ConvLSTM cells and dual attention mechanisms—to effectively capture spatio-temporal dynamics. Duo-AttnVarNet leverages automatic differentiation and GPU acceleration to enable efficient four-dimensional variational (4D-Var) assimilation without extensive manual adjoint coding. We evaluate Duo-AttnOPNets against the Integrated Forecasting System for Atmospheric Composition (C-IFS). Results show that Duo-AttnOPNets achieves comparable or superior accuracy in both forecasting and assimilation, while generating 5-day forecasts within seconds on a single GPU. These findings demonstrate its potential for real-time, scalable, and accurate CO forecasting, marking a promising advance in integrating deep learning with traditional variational methods for operational atmospheric modeling.

Plain Language Summary Carbon monoxide (CO) is a harmful air pollutant that can pose serious risks to human health and also influence climate change. Accurate and timely forecasting of CO levels is important for public safety and environmental research. However, current global air quality models require large amounts of computational resources especially when incoperating with real-time observations. In this study, we present a novel framework called Duo-AttnOPNets that uses artificial intelligence (AI) to improve both the speed and accuracy of global CO forecasts. Our system has two components: one predicts future CO levels using weather and pollution data, and the other updates forecasts by blending in real-time satellite observations. This second tool is especially efficient because it uses a modern AI technique called automatic differentiation, which avoids the time-consuming coding normally needed in traditional models. We tested Duo-AttnOPNets and found that it performs as well as—or better than—existing forecasting systems, while running much faster. This research shows how AI can support real-time air quality forecasting, and offers a new path for building flexible, accurate, and fast environmental models that could benefit science, policy, and public health.

1. Introduction

Carbon monoxide (CO), a major air pollutant, can significantly risk human health by competitively binding to hemoglobin, thereby diminishing the blood's oxygen-carrying capacity (Douglas et al., 1912). Acute exposure to elevated CO levels can result in symptoms such as headaches, dizziness, and cognitive impairments, loss of consciousness, or death in extreme cases (R. Chen et al., 2011; Levy, 2015). Additionally, CO plays a crucial role in atmospheric chemistry and climate change (Gaubert et al., 2017). Although CO itself is not a potent greenhouse gas, it significantly influences atmospheric concentrations of methane (CH_4) and tropospheric ozone (O_3), both of which are potent greenhouse gases. CO reacts with hydroxyl radicals (OH) in the atmosphere, reducing their availability to oxidize CH_4, thus prolonging the atmospheric lifetime of methane and enhancing its warming effect (Y. Zhao et al., 2020). In addition, CO contributes to tropospheric ozone formation, further exacerbating the greenhouse effect (Fishman & Seiler, 1983; Kim et al., 2013). Due to increasing direct emissions from anthropogenic activities, more frequent wildfires, and increased use of hydrogen energy (Bertagni et al., 2022),

Writing – review & editing: Judongyang Zhou

atmospheric CO levels are expected to rise, and the atmospheric lifetime of CO can be prolonged. Therefore, understanding the variations of atmospheric CO has become increasingly important.

Various numerical models have been developed to predict atmospheric composition variations, such as the Weather Research and Forecasting Model coupled with Chemistry (WRF-Chem) (Grell et al., 2005), the Goddard Earth Observing System Atmospheric Chemistry Transport Model (GEOS-Chem) (Bey et al., 2001) and the Community Atmosphere Model with Chemistry (CAM-chem) (Lamarque et al., 2012). Earlier developments such as the Brazilian Regional Atmospheric Modeling System and its chemistry-coupled extensions, CAT-BRAMS and CATT-BRAMS, represented the first operational environmental models for atmospheric composition forecasting (Freitas et al., 2000). These physics-based models are constructed by taking into account multiple gas-phase and aerosol chemistry schemes and meteorological dynamics, which are typically formulated as partial differential equations (PDEs). Solving these equations at fine spatial and temporal resolutions requires extensive computational resources. To make these models computationally feasible, various simplifications and parameterization schemes are applied, which often lead to inaccuracies in real-time predictions. Moreover, parameter uncertainty, arising from an incomplete understanding of the Earth system and the difficulty in capturing interactions across different scales, poses a challenge in modeling forecast accuracy (Simmons & Hollingsworth, 2002).

Among these models, the Composition-Integrated Forecasting System (C-IFS) developed by the European Center for Medium-Range Weather Forecasts (ECMWF) represents a state-of-the-art online global-scale operational forecasting system for atmospheric composition including CO (Inness et al., 2015). This operational forecasting system integrates a variational data assimilation technique, using real-time satellite observations to adjust the current state and improve the accuracy of the forecast. However, the data assimilation process increases computational demands, particularly due to the need for gradient computations in the optimization process. Although techniques such as linearization with adjoint methods can reduce computational costs and enhance stability (Janisková & Lopez, 2013), developing and maintaining differentiation routines for complex PDEs requires meticulous coding and a deep understanding of both physical processes and mathematical formulations (Giering & Kaminski, 1998). Each modification of the model requires extensive revisions to the differentiation routines, making the process time consuming and error prone. This challenge poses significant limitations to the future scalability of data assimilation systems (Bauer et al., 2020).

With the rapid advancement of artificial intelligence (AI), AI-based models are increasingly being used for weather predictions (Bi et al., 2023; L. Chen et al., 2023; Lam et al., 2023; Pathak et al., 2022). These models have achieved performance comparable to that of current state-of-the-art physical models and demonstrated significant advantages in computational efficiency (Rasp et al., 2024). However, there are uncertainties in AI-based forecasting systems, such as, initial and boundary conditions, which can result in significant forecasting errors. To improve predictive accuracy and enhance computational efficiency, an increasing number of studies are exploring the integration of data assimilation and AI techniques. One common approach involves coupling traditional data assimilation methods, typically the Ensemble Kalman Filter, with AI-based forecasting models (Brajard et al., 2020; Cai, Fang, & Wang, 2024). An alternative strategy trains deep learning-based data assimilation models to fully replace traditional assimilation systems (Fablet et al., 2021; Farchi et al., 2021; Huang et al., 2024; Wang et al., 2024; Wu et al., 2021). These assimilation models are trained for given forecasting models, using their predictions and available observations as input, and taking a cost function derived from variational data assimilation principles as the training loss. However, a major limitation of such AI-based surrogate assimilation models is that they learn the forecast error distribution through training, so any modification to the forecasting model usually requires retraining the assimilation model.

As a more scalable alternative, a training-free, AI-empowered approach has been proposed. Inspired by four-dimensional variational data assimilation (4D-Var), this method uses automatic differentiation and GPU acceleration to approximate the adjoint model and speed up the optimization process for best initial condition estimation (Dong et al., 2022; Nonnenmacher & Greenberg, 2021). Although these approaches have demonstrated potential, their evaluation has been largely limited to simplified numerical models, such as, Lorenz 96 and a few basic real-world case studies (Li et al., 2024; Maulik et al., 2022). There is a lack of rigorous comparisons between AI-based and traditional data assimilation models. These limitations hinder a full assessment of the capabilities of machine learning in this area. Moreover, global-scale atmospheric composition forecasting remains relatively underexplored, particularly for CO predictions.

19422466, 2026, 2, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025MS005319 by NICE, National Institute for Health and Care Excellence, Wiley Online Library on [25/05/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

19422466, 2026, 2, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025MS005319 by NICE, National Institute for Health and Care Excellence, Wiley Online Library on [25/05/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

In this work, we propose a novel data-driven operational forecasting framework, Duo-AttnOPNets, to provide a more accurate and computationally efficient approach to predict the change of surface-level CO at the global scale. Our method combines two modules: one is the forecasting module, Duo-AttnForeNet, based on a neural network consisting of a pair of encoder and decoder and double attentions added block CSLSTM. Another module, Duo-AttnVarNet, is a GPU-accelerated, manual-coding-free, training-free and 4D-Var-based module, incorporating satellite observations to adjust the initial condition based on posterior estimates. By implementing the auto-differentiation and adjoint state method inherently built in deep learning frameworks, we can significantly reduce the computational cost. This efficiency benefit also comes from constraining the optimization process with a neural network-based surrogate of the forward dynamical model, allowing both forward and backward passes to be executed efficiently.

This paper has made three main contributions:

1. *AI-based operational CO forecast at a global scale:* Duo-AttnOPNets is an AI-based model capable of real-time, global-scale CO forecasting computationally efficiently, with accuracy comparable or superior to leading physics-based model C-IFS. It demonstrates the feasibility of applying AI techniques to atmospheric composition forecasting.
2. *A scalable and efficient data assimilation method:* We propose an AI-empowered 4D-Var-inspired data assimilation model that integrates satellite observations to refine initial conditions. Unlike traditional 4D-Var, Duo-AttnVarNet utilizes GPU acceleration and automatic differentiation to enhance efficiency and eliminate the need for labor-intensive adjoint model development. Moreover, in contrast to other AI-based data assimilation models that rely on learning the forecasting error distribution through training, Duo-AttnVarNet is more scalable, as it does not require retraining when the forecasting model changes.
3. *Improved spatio-temporal learning via dual attention and recurrence:* Leveraging dual attention mechanisms and recurrent structures, our Duo-AttnForeNet effectively captures complex spatio-temporal dependencies in CO variabilities. This includes capturing diffusion driven by meteorological fields, accumulation due to topography, and high-concentration hotspots from intense emissions.

It is worth mentioning that Duo-AttnVarNet implemented in this study opens new avenues for a wide range of scientific and engineering applications. These include solving inverse problems, conducting sensitivity analyzes, performing uncertainty quantification, and enabling optimal control and design. Its ability to deliver fast and accurate gradient evaluations without the need for manual derivation or coding of adjoint models positions it as a powerful and flexible tool for data-driven decision-making and model-driven optimization across disciplines.

2. Methodology

In this work, we proposed Duo-AttnOPNets, which is a computationally efficient operational forecasting framework that integrates the forecasting module Duo-AttnForeNet and the data assimilation module Duo-AttnVarNet. The Duo-AttnForeNet is a neural network with a key recurrent component named Channel-Spatial Attentioned Convolutional Long Short-term Memory (CSLSTM). This component uses ConvLSTM cells as the backbone and dual attention mechanisms, allowing the model to better capture spatial dependencies and temporal evolution, as well as feature relations. From a mathematical perspective, the forecasting task is formulated as a spatiotemporal prediction problem, taking the current state of the atmospheric concentration as input to predict the states for the next few time steps. Meteorological variables and emission data are provided as dynamic drivers for the evolution of atmospheric CO concentrations.

The data assimilation module is an auto-differentiation module constructed in a GPU-accelerated framework. It originates from four-dimensional variational data assimilation, which takes real-time satellite observations to adjust the state variables through maximum a posteriori (MAP) estimation. The optimized value is known as analysis.

Before diving into mathematical formulations, we first clarify the notation of our parameters as follows: We denote the emission rate as e, the meteorological variables as m, the carbon monoxide concentration at the surface level as x, and the satellite observations as y. We use the subscript to represent time. For example, $x_{1:T} = (x_1, x_2, \ldots, x_T)$ is a sequence of state variables from timestep 1 to T.

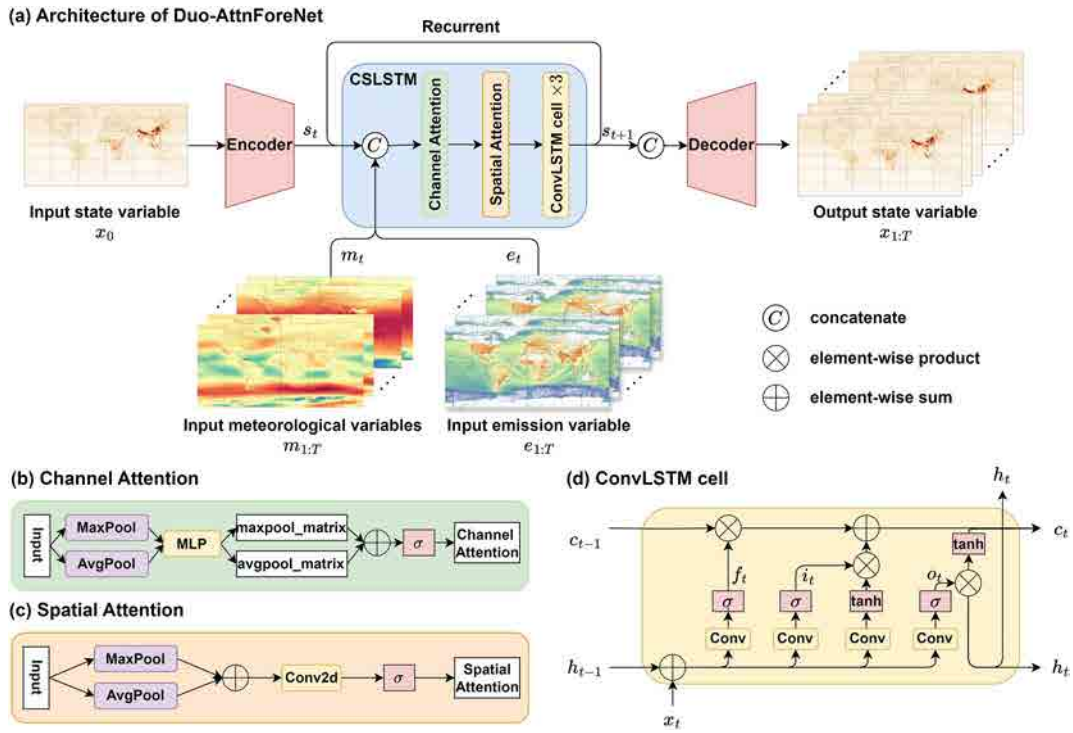


Figure 1. Overall architecture of Duo-AttenForeNet and overview of its components. (a) The architecture of Duo-AttenForecastNet and its inputs and outputs. (b) The channel attention block. (c) The spatial attention block. (d) The ConvLSTM cells.

2.1. Duo-AttnForeNet

The core of our forecasting module Duo-AttnForeNet is a forward model that simulates the evolution of the state variable x over times. In this study, we build a data-driven neural network $\mathcal{M}$ as a surrogate to a traditional physics-and-chemistry based model. This neural network can be formulated as follows:

$$x_{1:T} = \mathcal{M}_\theta(x_0, m_{1:T}, e_{1:T}), \tag{1}$$

which predicts the future state variable $x_{1:T}$ based on current states x_0 and meteorological variables $m_{1:T}$ and emission rate $e_{1:T}$.

The overall architecture of Duo-AttnForeNet can be found in Figure 1. This model consists of three components: a convolutional encoder, a recurrent CSLSTM, and a convolutional decoder. Given any initial state variable x_0 with shape $C_{in} \times T \times H \times W = 1 \times 1 \times H \times W$, where C_{in} is the number of atmospheric compositions, T is the history time steps, and H, W represent the spatial dimensions, we encode it by a 2-dimensional convolution layer into a feature-extracted hidden state s_0 with shape $C_{hidden} \times 1 \times H \times W$, where C_{hidden} denotes the number of hidden feature channels. Then the hidden state s_0 is concatenated with two inputs: meteorological driver m_t and emission driver e_t at the next time-step $t = 1$. The concatenated tensor is passed into the CSLSTM module to generate the next hidden state $s_{t=1}$. The CSLSTM is implemented in a recurrence format for T times. Once the recurrence is complete, the sequence of hidden variables $s_{1:T}$, with shape $C_{hidden} \times T \times H \times W$, is decoded by a three-dimensional convolutional layer to capture a periodic pattern and output final predicted state variables $x_{1:T}$, with shape $1 \times T \times H \times W$.

At the core of Duo-AttnForeNet is the CSLSTM module, which consists of two attention mechanisms—channel and spatial (Park et al., 2018; Woo et al., 2018)—followed by a Convolutional Long Short-Term Memory (ConvLSTM) block (Shi et al., 2015). These components are applied sequentially to enhance feature representation and spatiotemporal learning. The channel attention and spatial attention refine input features by applying a channel-wise and a spatial-wise weight matrix. These two weights represent the different contributions of the

19422466, 2026, 2, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025MS005319 by NICE, National Institute for Health and Care Excellence, Wiley Online Library on [23/05/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

19422466, 2026, 2, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025MS005319 by NICE, National Institute for Health and Care Excellence, Wiley Online Library on [23/05/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

driver variables to the change of state in different spatial positions. As a result, Duo-AttnForeNet is able to prioritize the most informative channels and effectively capture spatial dependencies.

Following the dual attention mechanisms, the processed feature tensor is passed onto ConvLSTM cells. ConvLSTM learns the temporal evolution through three gating mechanisms: input gate i_t, forget gate f_t, and output gate o_t, which together control the memory information c_t and further decide the output h_t. By integrating convolutional operations within these gates, ConvLSTM effectively preserves spatial structure in sequential data. In this study, we use three CSLSTM layers to enable Duo-AttnForeNet to better learn hierarchical spatiotemporal representations, further improving prediction accuracy. Detailed descriptions of these components are provided in the Supplementary Text S1 in Supporting Information S1.

2.2. Duo-AttnVarNet

2.2.1. Mathematics Behind Duo-AttnVarNet

Four-dimensional variational data assimilation (4D-Var) is formulated as an optimization problem based on the MAP estimate of the initial state conditioned on other parameters. Given an initial condition x_0 in the state space $\mathcal{X} \subseteq \mathbb{R}^n$, a sequence of subsequent states $X = \{x_1, \ldots, x_T\}$ is evolved by a forward model $\mathcal{M}$ as follows:

$$X = \mathcal{M}_{0 \mapsto T}(x_0, \Gamma_p) + \varepsilon, \tag{2}$$

where Γ_p corresponds to process parameters such as the boundary conditions, forcing terms, etc., and ε represents the model error.

A set of observations $Y = \{y_1, \ldots, y_T\}$ in the observational space $\mathcal{Y} \subseteq \mathbb{R}^m$ is given at each time step. A function called the observational operator $\mathcal{H} : \mathcal{X} \to \mathcal{Y}$ maps from the state space to observational space as follows:

$$y_t = \mathcal{H}(x_t, \theta_p) + \epsilon, \tag{3}$$

where θ_p represents the observational parameters, such as the instrument parameters, and ϵ is the observational error. A posterior probability distribution is defined proportional to the product likelihood and prior probability, whose density function can be written as:

$$\rho(X, x_0, \Gamma_p, \theta_p | Y) \propto \rho(Y | X, \theta_p) \cdot \rho(X | x_0, \Gamma_p) \cdot \rho(x_0, \Gamma_p, \theta_p), \tag{4}$$

where $\rho(x_0, \Gamma_p, \theta_p)$ is the prior density function based on the prior knowledge about all parameters. The problem of finding the best estimates of the initial state is to maximize the posterior probability $\rho(X, x_0, \Gamma_p, \theta_p | Y)$. Taking log on both sides, Equation 4 is rewritten as:

$$\underset{x_0 \in \mathcal{X}}{\operatorname{argmax}} \ln \rho(X, x_0, \Gamma_p, \theta_p | Y) \propto \underset{x_0 \in \mathcal{X}}{\operatorname{argmax}} \ln \rho(Y | X, \theta_p) + \ln \rho(X | x_0, \Gamma_p) + \ln \rho(x_0, \Gamma_p, \theta_p). \tag{5}$$

With the assumption that the process model is deterministic and the background error is in a Gaussian distribution $\mathcal{N}(x_0 | x_b, B)$ and the observational errors are in the Gaussian distributions $\mathcal{N}\left(y_t | \mathcal{H}(x_t), R_t\right)$, we can rewrite Equation 5 as:

$$\ln \rho(x_0 | Y) \propto \ln \rho(x_0) + \sum_{t=1}^{T} \ln \rho\left(y_t | x_t\right) \tag{6}$$

$$\propto \ln \left(\frac{\exp\left(-\frac{1}{2} \| x_0 - x_b \|^2_{B^{-1}}\right)}{\sqrt{(2\pi)^n |B|}} \right) + \sum_{t=1}^{T} \ln \left(\frac{\exp\left(-\frac{1}{2} \| y_t - \mathcal{H}(x_t) \|^2_{R_t^{-1}}\right)}{\sqrt{(2\pi)^m |R_t|}} \right), \tag{7}$$

19422466, 2026, 2, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025MS005319 by NICE, National Institute for Health and Care Excellence, Wiley Online Library on [23/05/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

where x_b is the background state, B is the background error covariance matrix, R_t represents the observational error covariance, $|\cdot|$ is the determinate of a given matrix, and $\|\cdot\|$ is the Mahalanobis norm.

Given the error covariance matrices B and R_t, the 4D-Var in this study can be formulated as minimizing the cost function $J(x_0)$:

$$\mathcal{J}(x_0) = \frac{1}{2}\|x_0 - x_b\|^2_{B^{-1}} + \frac{1}{2}\sum_{t=1}^{T}\| y_t - \mathcal{H}(x_t)\|^2_{R_t^{-1}}. \tag{8}$$

2.2.2. AI-Empowered Auto-Diff Adjoint Optimization Procedure

Gradient-based approaches are often used to solve this optimization problem. The gradient as the derivative of the cost function $\mathcal{J}(x_0)$ with respect to x_0 can be computed using the chain rule:

$$\nabla_{\mathbf{x}_0}\mathcal{J} = \frac{\partial \mathcal{J}}{\partial X}\frac{\partial X}{\partial x_0} + \frac{\partial \mathcal{J}}{\partial x_0}. \tag{9}$$

However, such a direct computation can be computationally very expensive, as $\frac{\partial \mathcal{J}}{\partial X}\frac{\partial X}{\partial x_0}$ is a summation of multiplication $\sum_{t=1}^{T}\prod_{i=1}^{t}\frac{\partial \mathcal{J}}{x_i}\frac{\partial x_i}{\partial x_0}$ which has many duplicate terms. An efficient way to compute the gradient is to use the adjoint state method, which introduces a set of variables $\Lambda = \{\lambda_0, \ldots, \lambda_T\} \in \mathcal{X}$, called Lagrange multipliers, calculated iteratively backward in time from the end of the assimilation window $t = T$ to the start of the assimilation window $t = 0$. The Lagrange multiplier at the final time step $t = T$ is

$$\lambda_T = \left(\frac{\mathrm{d}\mathcal{H}(x_T)}{\mathrm{d}x_T}\right)^T R_T^{-1}\left(y_T - \mathcal{H}(x_T)\right), \tag{10}$$

and $\{\lambda\}_{t\in\{0,\ldots,T-1\}}$ are computed iteratively backward for each time step by

$$\lambda_t = \left(\frac{\mathrm{d}\mathcal{M}(x_t)}{\mathrm{d}x_t}\right)^T \lambda_{t+1} + \left(\frac{\mathrm{d}\mathcal{H}(x_t)}{\mathrm{d}x_t}\right)^T R_t^{-1}\left(y_t - \mathcal{H}(x_t)\right). \tag{11}$$

Finally, the gradient can be obtained by adding a term related to background loss to the Lagrange multiplier at $t = 0$:

$$\nabla_{\mathbf{x}_0}\mathcal{J} = -B^{-1}(\mathbf{x}_b - \mathbf{x}_0) - \lambda_0. \tag{12}$$

A more detailed derivation of Lagrange multipliers can be found in Text S3 in Supporting Information S1.

In the traditional 4D-Var optimization algorithm, the tangent-linear model and its corresponding adjoint model are essential to calculate the gradient of the cost function, but their development requires significant manual engineering effort. Duo-AttnVarNet proposed here uses the auto-differentiation function built in the deep learning framework. The overall optimization procedure can be found in Figure 2.

An advantage of Duo-AttnVarNet is that its gradients are obtained via reverse-mode automatic differentiation, which provides exact derivatives of the discrete loss with respect to the state; thus, unlike finite-difference approximations (which are $O(h)$ for forward and $O(h^2)$ for central differences), it does not introduce truncation error.

3. Global Surface-Level CO Experiments

3.1. Data Sets

The data sets used in this study are categorized into four types: reanalysis, forecasts, emissions, and satellite observations. Data set selection is aligned with those used in C-IFS to ensure a fair and consistent basis for comparison. A summary of the data sets is provided in Tables 1 and 2.

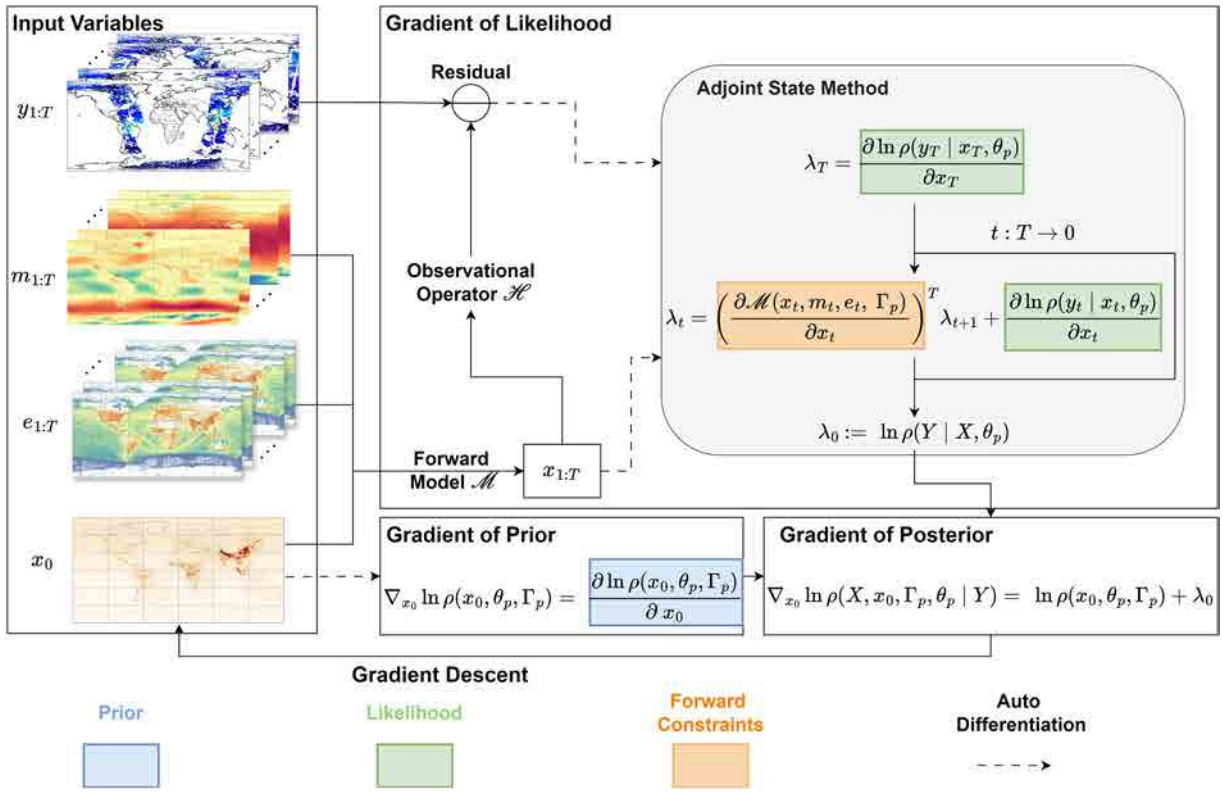


Figure 2. The Optimization procedure for Duo-AttnVarNet. The principle of Duo-AttnVarNet is an optimization problem solved via gradient descent. Given an initial state variable x_0, along with meteorological variables $m_{1:T}$, emission variables $e_{1:T}$ and a forward model $\mathcal{M}$, we first compute the predicted state sequence $x_{1:T}$. These predicted states $x_{1:T}$ are mapped to the observational space using an observational operator $\mathcal{H}$, and the difference between observations $y_{1:T}$ and mapped states is used to compute the gradient of likelihood via adjoint state method. Moreover, the gradient of the prior probability distribution is computed based on the background information. By combining the gradients of likelihood and prior through Bayes' Theorem, we obtain the gradient of posterior probability distribution. Gradient descent is then used to update the initial state x_0. The optimization process continues until the posterior distribution is maximized, yielding the analysis state.

Table 1
Summary of Emission, Reanalysis and Forecast Data Sets Used in This Study

Data set	Category	Variables	Resolution	Coverage	References
CAMS-GLOB-ANT v2.1	Emission	Sum of emission from all sectors	0.1° × 0.1°, monthly	2000–2023	Soulie et al. (2023)
CAMS-GLOB-BIO v3.1	Emission	Biogenic CO flux	0.25° × 0.25°, monthly	2000–2018	Sindelarova et al. (2022)
GFAS v1.2	Emission	Wildfire CO flux	0.1° × 0.1°, daily	2003-present	Kaiser et al. (2012)
EAC4	Reanalysis	10 m u-wind	0.75° × 0.75°, 3-hourly	2003–2024	Inness et al. (2019)
		10 m v-wind			
		2 m temperature			
		Surface geopotential			
		Total column water vapor			
		CO (model level = 60)			
CAMS forecasts	Forecast	10 m u-wind	0.4° × 0.4°, 3-hourly	2015-present	Inness et al. (2015)
		10 m v-wind			
		2 m temperature			
		Surface geopotential			
		Total column water vapor			
		CO (model level = 60)			

19422466, 2026, 2, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025MS005319 by NICE, National Institute for Health and Care Excellence, Wiley Online Library on [23/05/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

19422466, 2026, 2, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025MS005319 by NICE, National Institute for Health and Care Excellence, Wiley Online Library on [25/05/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 2
Summary of Satellite Data Sets Used in This Study

Instrument/Satellite	Space agency	Resolution	Coverage	References
MOPITT/Terra	NASA/V8	22 km at nadir; daily, 8-day, monthly (location dependent)	2000-present	Ziskin (2019)
IASI/MeTop-A	EUMETSAT/CNES	24 km on average; less than one day	2007–2019	August et al. (2012)
IASI/MeTop-B			2013-present	Blumstein et al. (2004)

3.1.1. The ECMWF Atmospheric Composition Reanalysis 4 Data Set

The ECMWF Atmospheric Composition Reanalysis 4 data set (EAC4) is the latest global atmospheric composition reanalysis data set produced by ECMWF (Inness et al., 2019). It provides a comprehensive and consistent representation of atmospheric conditions by assimilating observational data offline into the chemical and physical model C-IFS. Due to its high accuracy, the EAC4 CO concentrations serve as reference values and are used for both model training and evaluation in this study. In addition, five meteorological variables from EAC4–10 m wind components (u and v), 2 m temperature, surface geopotential, and total column water vapor—are employed as meteorological forcings for the forward model. All variables are available at a spatial resolution of $0.75° \times 0.75°$ and a temporal resolution of 3 hr.

3.1.2. CAMS Global Atmospheric Composition Forecasts

Copernicus Atmosphere Monitoring Service (CAMS) global atmospheric composition forecasts data set provides near-real-time predictions of atmospheric pollutants and trace gases, produced using the C-IFS model—an extended version of ECMWF's Integrated Forecasting System with atmospheric composition capabilities (Inness et al., 2015). Forecasts are initialized twice daily and extend up to 5 days from each initialization. In this study, this data set serves as a baseline for evaluating the performance of our model. To ensure a realistic and fair comparison, we use forecasted meteorological variables from this data set for all evaluations, rather than reanalysis data, to reflect real-time operational conditions.

3.1.3. Emission Data Sets

Three global emission data sets are incorporated into the model as input to represent the major sources of CO emissions.

1. CAMS-GLOB-ANT v2.1: The Global Anthropogenic Emissions data set developed under the CAMS, provides monthly averages of emissions for 36 pollutants and greenhouse gases, including CO (Soulie et al., 2023). Anthropogenic emissions from 20 sectors (e.g., transportation, industry, and agriculture) are aggregated as the total anthropogenic emission source in our study.
2. CAMS-GLOB-BIO v3.1: This data set provides global monthly means of biogenic CO emissions, estimated using the Model of Emissions of Gases and Aerosols from Nature (MEGAN v2) (Sindelarova et al., 2022). This model takes into account key environmental factors such as ecosystem distribution, biomass density, and meteorological conditions to generate reliable estimates of biogenic CO emissions from natural processes.
3. Global Fire Assimilation System (GFAS) v1.2: The CAMS GFAS version 1.2 provides daily averaged estimates of wildfire and biomass burning emissions for various atmospheric pollutants and greenhouse gases (Kaiser et al., 2012). These estimates are based on fire radiative power observations from NASA EOS-Terra and EOS-Aqua satellites, ensuring high spatial resolution and accuracy.

3.1.4. Satellite Data Sets

Satellite observations used in this study are derived from two instruments: the Measurements of Pollution in the Troposphere (MOPITT) and the Infrared Atmospheric Sounding Interferometer (IASI). The MOPITT instrument onboard NASA's Terra satellite provides long-term CO measurements via a nadir-viewing gas correlation radiometer, using both thermal infrared and near-infrared spectral bands. This study utilizes Level-2 MOPITT CO Retrievals Version 8, which is produced using an optimal estimation algorithm (Ziskin, 2019). Furthermore, the MetOp-A and MetOp-B satellites from the IASI instrument, are operated by the European Organization for the Exploitation of Meteorological Satellites (EUMETSAT) and the Center National d'Etudes Spatiales (CNES).

19422466, 2026, 2, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025MS005319 by NICE, National Institute for Health and Care Excellence, Wiley Online Library on [25/05/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

These two satellites provide hyperspectral infrared measurements for multiple atmospheric compositions, including CO, with global coverage. We used Level-2 IASI CO retrievals for the entire year of 2018, which offer high temporal and spatial resolution for atmospheric CO concentration observations (August et al., 2012; Blumstein et al., 2004).

3.2. Data Preprocessing

This section details the data preprocessing procedures employed in this study to ensure the stability, convergence, and predictive accuracy of the neural network model as well as the implementation of the whole workflow.

3.2.1. Core Preprocessing Steps for Duo-AttnForeNet Inputs

Three key preprocessing steps were applied to the model inputs:

Base-10 logarithmic transformation: A base-10 logarithmic transformation was applied to the CO concentration values. Since CO concentrations are strictly positive and exhibit a right-skewed, approximately lognormal distribution (similar to water vapor), the log transformation also accounts for the asymptotic behavior near zero and helps stabilize variance across regions. This transformation mitigates the influence of substantial regional variations in CO distribution, which exhibit a highly skewed pattern. From 2015 to 2018, approximately 57.4% of global regions exhibited relatively low annual average CO concentrations $\left(\leq 1 \times 10^{-7}\ \text{kg CO/kg Air}\right)$, while 42.5% showed elevated concentrations (up to $\leq 1 \times 10^{-6}$ kg CO/kg Air), and fewer than 0.1% exceeded $\leq 1 \times 10^{-5}$ kg CO/kg Air. The difference between high- and low-concentration areas spans over two orders of magnitude, which can introduce an inevitable bias when using the MSE loss function. Without transformation, the model tends to prioritize high-concentration areas and underperform in low-concentration regions. The log transformation compresses the dynamic range, enhancing the model's ability to capture variations across all concentration levels.

Min-max Feature Scaling: All input features were normalized using min-max scaling to a predefined range, typically [0, 1]. This ensures that no single feature dominates the learning process and facilitates faster convergence during training. Normalizing the input range also helps mitigate the risk of vanishing or exploding gradients, thus improving training efficiency.

Spatial Resampling: All emission and forecast data are resampled at a spatial resolution of $0.75° \times 0.75°$ to match the reanalysis data, ensuring feasibility for evaluation. Bilinear resampling is adopted in this study, as the choice of resampling method had little impact on the results.

3.2.2. Satellite Data Preprocessing for Data Assimilation

Reliable data assimilation requires strict quality control to prevent the introduction of unreliable or biased satellite measurements into the analysis. Satellite CO products can be affected by cloud contamination, high zenith angles, and challenging surface conditions, all of which reduce retrieval accuracy. Following the data-quality recommendations used in the operational ECMWF/CAMS Composition-IFS system Inness et al. (2015), we apply a series of filtering and resampling steps to retain only high-quality observations and to reduce spatially correlated observation errors. These procedures ensure that the assimilated observations are consistent, robust, and suitable for use in the Duo-AttnOPNets framework.

Quality Filtering: IASI MetOp-A and MetOp-B provide a quality flag to categorize observations into three types: 2 for the most reliable data, 1 for valuable data that should be used with caution, and 0 for data that are not recommended for use. In this study, only the most reliable data are utilized. MetOp-A observations serve as the primary data set, with missing values supplemented using observations from MetOp-B and MOPITT.

Regional Mask: Due to the relatively high observational errors in the polar regions, we mask observations above $65°N$ and below $65°S$, following the methodology applied in the C-IFS 4D-Var system.

Spatial Resampling: Since the spatial resolution of satellite data is higher than that of our model, the data must be thinned to match the model resolution. To achieve this, we employ a random selection strategy to prevent oversampling and mitigate spatially correlated observation errors.

Table 3
Table of Settings of Duo-AttnForeNet

Hyper-parameter settings		
Training and testing size		1,862
Testing size		328
Learning rate		0.001
Optimizer		Adam
Batch size		8
Number of epochs		400
Kernel Size		[(7, 7), (5, 5), (3, 3)]
Number of Hidden Channels		[64, 64, 1]
Loss		MSE
Model structure		
Layer	Output shape	Parameter number
Conv2D encoder	batch ×1 × 241 × 480	10
Channel Attention	batch ×7 × 241 × 480	22
Spatial Attention	batch ×1 × 241 × 480	98
ConvLSTM Layer 1	batch ×64 × 241 × 480	891,072
ConvLSTM Layer 2	batch ×64 × 241 × 480	819,648
ConvLSTM Layer 3	batch ×1 × 241 × 480	2,347
Conv3D decoder	batch ×1 × 8 × 241 × 480	10

Note. Total parameters: 1,713,207, estimated total size: 30,875.84(MB).

3.3. Evaluation Metrics

To assess the performance of our method, we employ two widely used statistical metrics: the root mean square error (RMSE) and the Pearson correlation coefficient (R) (For details, see Text S2 in Supporting Information S1). These metrics offer a comprehensive and robust assessment of model performance—RMSE quantifying overall prediction errors, and R evaluating how well spatial and temporal variations are captured.

3.4. Model Configurations

3.4.1. Configuration for Duo-AttnForeNet

The detailed hyperparameters used for training Duo-AttnForeNet in this study are illustrated in Table 3.

3.4.2. Parameter Configurations for Duo-AttnVarNet

The observation operator, $\mathcal{H}$, consists of two components: location mapping and vertical profiling. The location mapping component is represented as a binary matrix, where a value of 1 indicates the presence of an observation at a given time and location, while a value of 0 masks void values. This ensures that only valid observations are assimilated into the model.

The vertical profiling component is designed to map total column CO measurements to surface-level CO concentrations. Unlike conventional data assimilation systems such as C-IFS or GEOS-Chem, which explicitly resolve three-dimensional CO fields across multiple pressure levels and can therefore integrate model layers to match satellite total column retrievals, the current implementation of Duo-AttnOPNets focuses solely on surface-level CO (corresponding to model level 60 in EAC4). Because the model does not simulate vertical profiles, direct integration to match satellite retrievals is not feasible. Although both IASI and MOPITT provide their partial column vertical profiles to estimate CO concentration at different atmospheric levels, these profiles cannot be directly utilized in this study due to a discrepancy between their lowest retrieval level and the surface level used in EAC4. Specifically, the lowest level in IASI is 0–1 km (Kerzenmacher et al., 2012), whereas, in EAC4, the surface level is defined as 0–10 m. This inconsistency encouraged the use of a custom vertical profile approach, in which the empirical conversion factor was derived by dividing the total CO in the column by the CO at the surface level in EAC4, ensuring consistency with the forecasting framework in this study.

In this study, both the background and observational error covariance matrices B and R are assumed to be univariate and time-independent, following the formulation adopted in the ECMWF/CAMS Composition-IFS system (Inness et al., 2015). Under this assumption, B and R are diagonal. B is estimated from the differences between the EAC4 reanalysis (used as reference) and the corresponding C-IFS initial conditions, while R is estimated from the differences between the preprocessed satellite retrievals and EAC4. These diagonal matrices provide efficient and stable error representations for the Duo-AttnVarNet assimilation framework.

4. Results

All models in this study are evaluated at a global scale throughout 2018. The evaluation is conducted in three stages: First, we assess the forecasting performance of Duo-AttnForeNet by comparing its predictions to those from the C-IFS model, using identical initial conditions, real-time meteorological forecasts, and estimated emission inputs. Second, we evaluate the data assimilation capability of Duo-AttnVarNet by using the same satellite observations as those used by C-IFS. This stage focuses on quantifying the Duo-AttnVarNet's ability to reduce accumulated forecast errors. Furthermore, its resulting analysis (the refined state) is compared with those from C-IFS. Finally, we assess the operational forecast performance of Duo-AttnOPNets by running the two modules once per day, continuously generating analyzes and forecasts from 1 January to 31 December 2018. Again, C-IFS serves as a benchmark throughout this process.

19422466, 2026, 2, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025MS005319 by NICE, National Institute for Health and Care Excellence, Wiley Online Library on [25/05/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

19422466, 2026, 2, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025MS005319 by NICE, National Institute for Health and Care Excellence, Wiley Online Library on [25/05/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

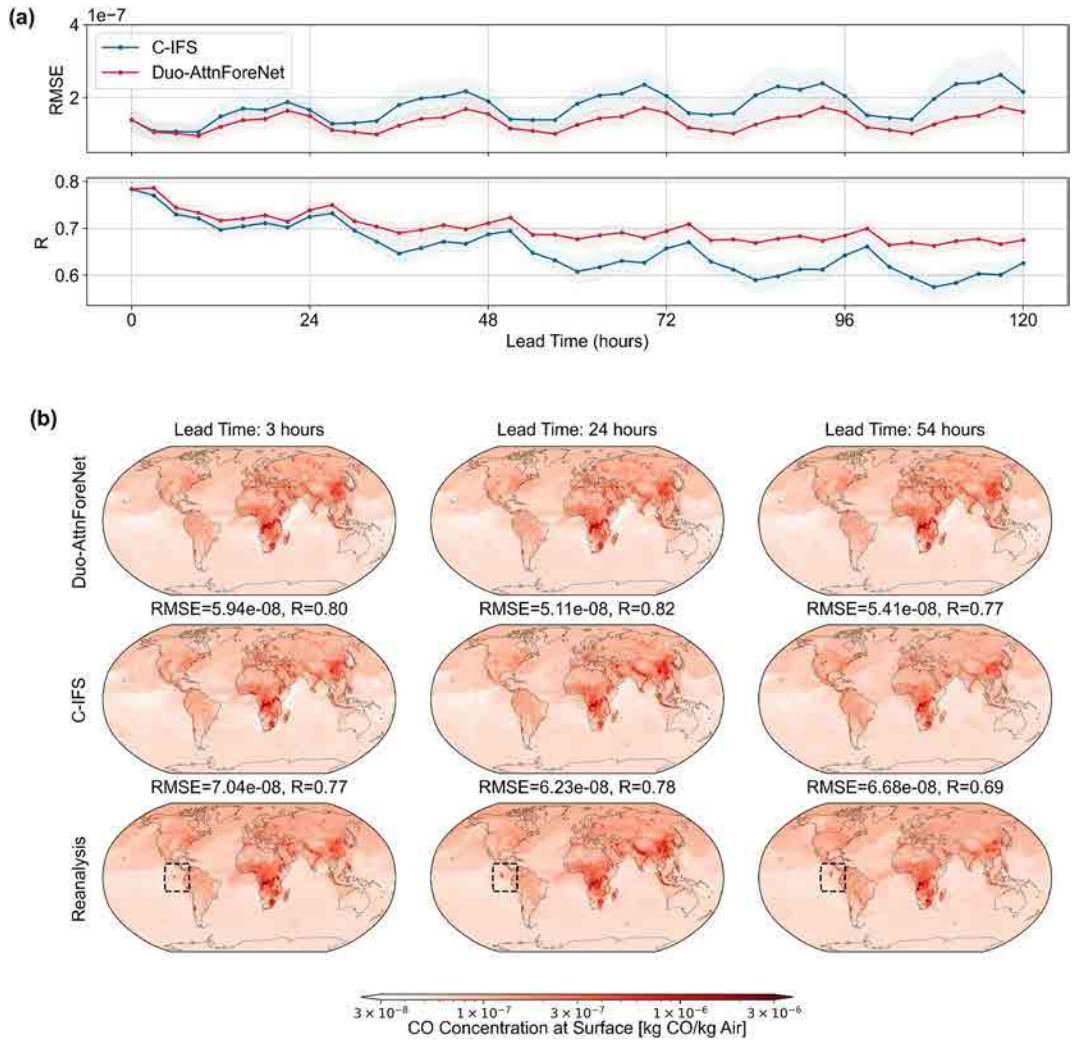


Figure 3. Comparison of forecast skills between C-IFS and Duo-AttnForeNet (a) A comparison for the forecast accuracy in different lead times. The two plots present RMSE (top) and R (bottom) in different lead times. For each metric, the mean value (solid lines) and the 95% confidence interval (shaded regions) are computed at each lead time across all samples from 2018. (b) An illustration of the surface-level CO concentration forecasted by Duo-AttnForeNet (first row) and C-IFS (second row), and the corresponding EAC4 reanalysis (third row) as ground truth. The forecasts are initialized at 0:00 UTC on 17 June 2018. The columns from left to right are the forecasts at 3-, 24-, and 54-hr lead time. The black dashed box on the reanalysis panels is centered on the Galápagos Islands, where a sudden CO plume associated with the La Cumbre Volcano eruption occurred (see Figure S1 in Supporting Information S1 for a zoom-in view).

4.1. Evaluation of Duo-AttnForeNet

In this section, we evaluate the forecasting performance of Duo-AttnForeNet. A series of 5-day forecasts, each initialized at 00:00 UTC daily throughout 2018, were carried out at 3-hr intervals, consistent with the temporal resolution and forecast horizon of C-IFS. For a fair comparison, Duo-AttnForeNet was provided with the same initial conditions x_0, meteorological forcings $m_{1:T}$, and emission sources $e_{1:T}$ as those used in C-IFS.

Figure 3a presents the prediction accuracy of the models with different lead times, evaluated using the rooted mean squared error (RMSE) and the Pearson correlation coefficient (R). Averaged across all samples in 2018, the results show that Duo-AttnForeNet consistently outperforms C-IFS. Notably, Duo-AttnForeNet demonstrates superior performance in cumulative error control. The mean of RMSE remains stable and below 1.74×10^{-7} (kg CO/kg Air) at all lead times, while the mean of R consistently exceeds 0.66. In contrast, C-IFS exhibits an increasing trend in cumulative error over time, with RMSE reaching 2.62×10^{-7} (kg CO/kg Air) and R dropping to 0.57. Furthermore, the narrow confidence interval indicates that Duo-AttnForeNet is stable and well-generalized.

19422466, 2026, 2, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025MS005319 by NICE, National Institute for Health and Care Excellence, Wiley Online Library on [23/05/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

To further evaluate the performance of Duo-AttnForeNet, we analyze a specific case initialized at 00:00 UTC on 17 June 2018. Figure 3b shows the forecast results at 3-, 24-, and 54-hr lead times (from left to right). The predictions from Duo-AttnForeNet (first row) are compared with those from C-IFS (second row) and the ground truth from EAC4 Reanalysis (third row). It can be found that our model not only captures meteorologically driven variations but also accurately represents high-concentration regions influenced by natural and anthropogenic emissions. Moreover, Duo-AttnForeNet exhibits spatial coherence comparable to C-IFS, underscoring its fidelity and reliability.

However, sudden and sporadic events, such as carbon monoxide surges from the eruption of La Cumbre Volcano (see zoom-in in Figure S1 in Supporting Information S1), wildfires in the Congo rainforest in Central Africa (Figure S2 in Supporting Information S1), and uncertainties in anthropogenic emissions, pose significant challenges for accurate forecasting. These unpredictable factors highlight the importance of data assimilation, which corrects the current state with real-time satellite observations and improves the accuracy of subsequent forecasts, contributing to a more reliable prediction system.

4.2. Evaluation of Duo-AttnVarNet

4.2.1. Effectiveness of Duo-AttnVarNet at Global and Regional Scale

In this subsection, we evaluate the effectiveness of Duo-AttnVarNet by comparing the results with and without data assimilation. We refer the refined state variable after data assimilation as analysis. Here, reanalysis is used as "ground truth." Figure 4 presents an example on 17 June 2018. All errors are computed by the model or observation minus the reanalysis; thus, positive values indicate overestimation, while negative values indicate underestimation.

As shown in Figure 4a, the predicted CO concentration level without data assimilation exhibits a relatively high inaccuracy, particularly in regions such as the East Coast of the United States. This inaccuracy is mainly due to uncertainties in anthropogenic emissions. By incorporating satellite observations, Duo-AttnVarNet significantly improves the accuracy of the current state in these regions. Moreover, forecasts over oceanic areas tend to underestimate CO concentrations, while satellite observations indicate an overestimation in the Northern Hemisphere and tropical areas. By assimilating these observations, Duo-AttnVarNet effectively refines the forecasts, yielding a more accurate representation of atmospheric conditions.

Another important example is the carbon monoxide surges from the eruption of La Cumbre Volcano as discussed in Section 4.1. Figure 4b zooms in on the area near the Galápagos Islands, where satellite instruments successfully detected high-concentration CO plumes from volcanic eruptions. Despite partial data loss due to cloud cover, the model effectively recalibrates concentration levels using the available observations, clearly identifying the high-concentration plume in the analysis.

Given the spatial variability of CO concentrations, we further evaluate the performance of data assimilation across different regions. Figure 5 illustrates the improvement in RMSE (a) and R (b) for both analysis and forecast at the data assimilation time window (gray shaded area) and lead times ranging from 3 to 96 hr, evaluated globally and across eight selected regions. The regional classification and naming follow the ECMWF scorecards (see Table S1 in Supporting Information S1). The metric ΔRMSE and $-\Delta$R present the improvements due to data assimilation, defined as:

$$\Delta\mathrm{RMSE} = \mathrm{RMSE}_{\mathrm{forecast}} - \mathrm{RMSE}_{\mathrm{analysis}},$$
$$-\Delta\mathrm{R} = \mathrm{R}_{\mathrm{analysis}} - \mathrm{R}_{\mathrm{forecast}}.$$

As illustrated in Figure 5, the influence of data assimilation is particularly significant during the initial 24-hr period of the forecast. However, beyond this initial window, the benefits begin to diminish. With the accumulation of forecast errors over time, the performance difference between forecasts generated with and without data assimilation gradually decreases. By the 48-hr mark, the advantages of data assimilation become marginal, indicating that its impact is most pronounced in the early stages of the forecast cycle and less influential as the lead time increases.

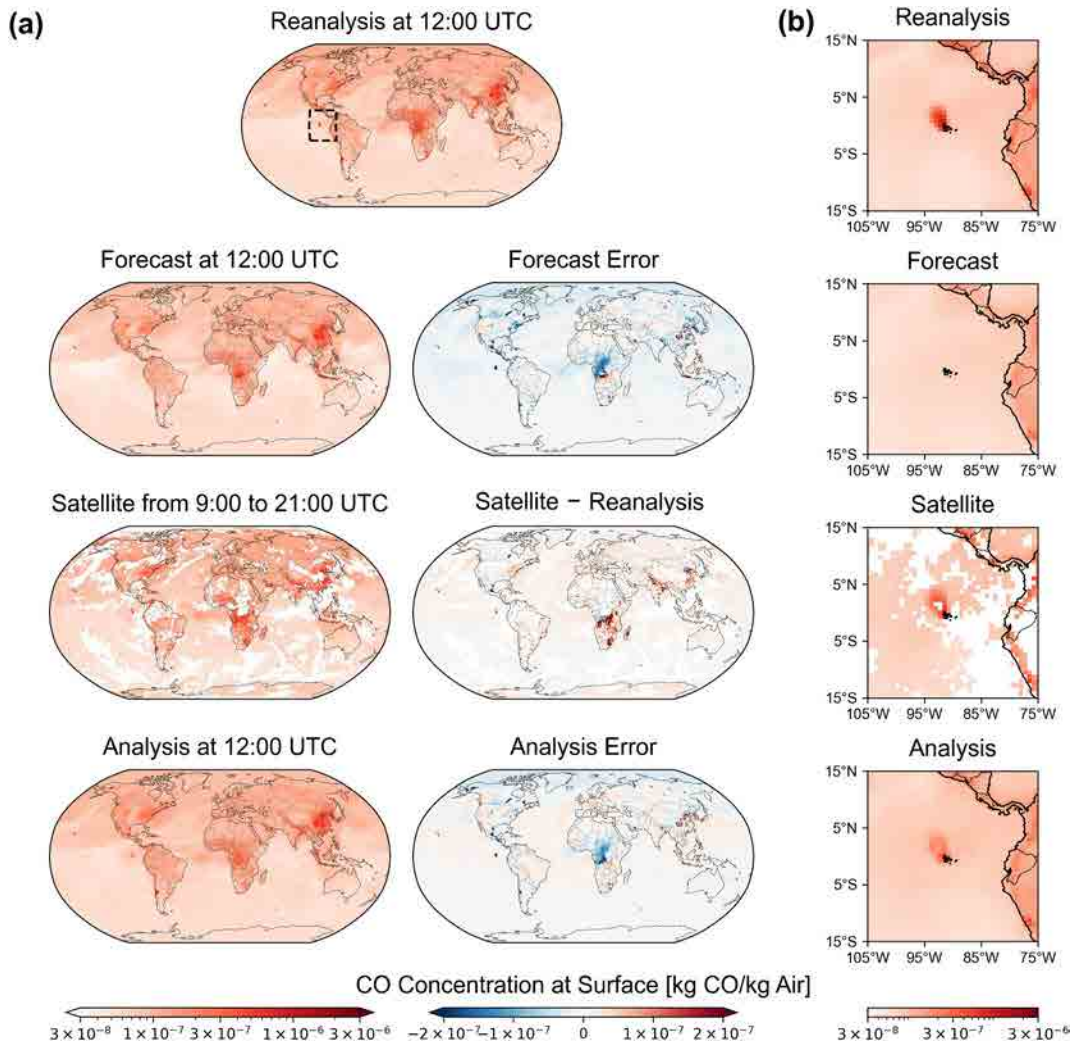


Figure 4. An illustration of integrating data assimilation by Duo-AttnVarNet to adjust initial conditions. (a) The EAC4 reanalysis, Duo-AttnForeNet forecast and corresponding forecast error, the satellite observation and satellite–reanalysis difference, Duo-AttnVarNet results with data assimilation(referred to as the analysis) and the corresponding analysis error on 17 June 2018. All the errors are computed by the model outputs minus the reanalysis. (b) A zoomed-in view of the area centered at the Galápagos Islands.

Globally, Duo-AttnVarNet leads to a reduction in RMSE of approximately 1.42×10^{-8} kg (CO/kg Air) and an increase in R of 0.10. Regionally, the effectiveness of Duo-AttnVarNet varies, showing spatial heterogeneity, with different levels of improvement in geographic areas. Regions with higher CO concentrations, such as North America (n.amer), East Asia (e.asia), Australia and New Zealand (austnz), and the tropics benefit most significantly from data assimilation, showing greater improvements in both RMSE and R compared to the global average. In Europe (europe), the Northern Hemisphere extropics (n.hem), and the Southern Hemisphere extropics (s.hem) the RMSE reduction is comparable to the global mean, while the increase in R is more pronounced. Although the North Atlantic (n.atl) and the North Pacific (n.pac), which has lower CO concentrations, show smaller improvements, with both RMSE reduction and R enhancement falling below the global average, Duo-AttnVarNet still improved accuracy in these areas.

4.2.2. Comparison With C-IFS Analysis

We compare the performance of Duo-AttnVarNet and C-IFS by evaluating the analysis results at 12:00 UTC on each day of 2018. To ensure a fair comparison, we adopt the same assimilation window (09:00–21:00 UTC) and initial conditions and satellite observations as used in C-IFS.

19422466, 2026, 2, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025MS005319 by NICE, National Institute for Health and Care Excellence, Wiley Online Library on [25/05/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

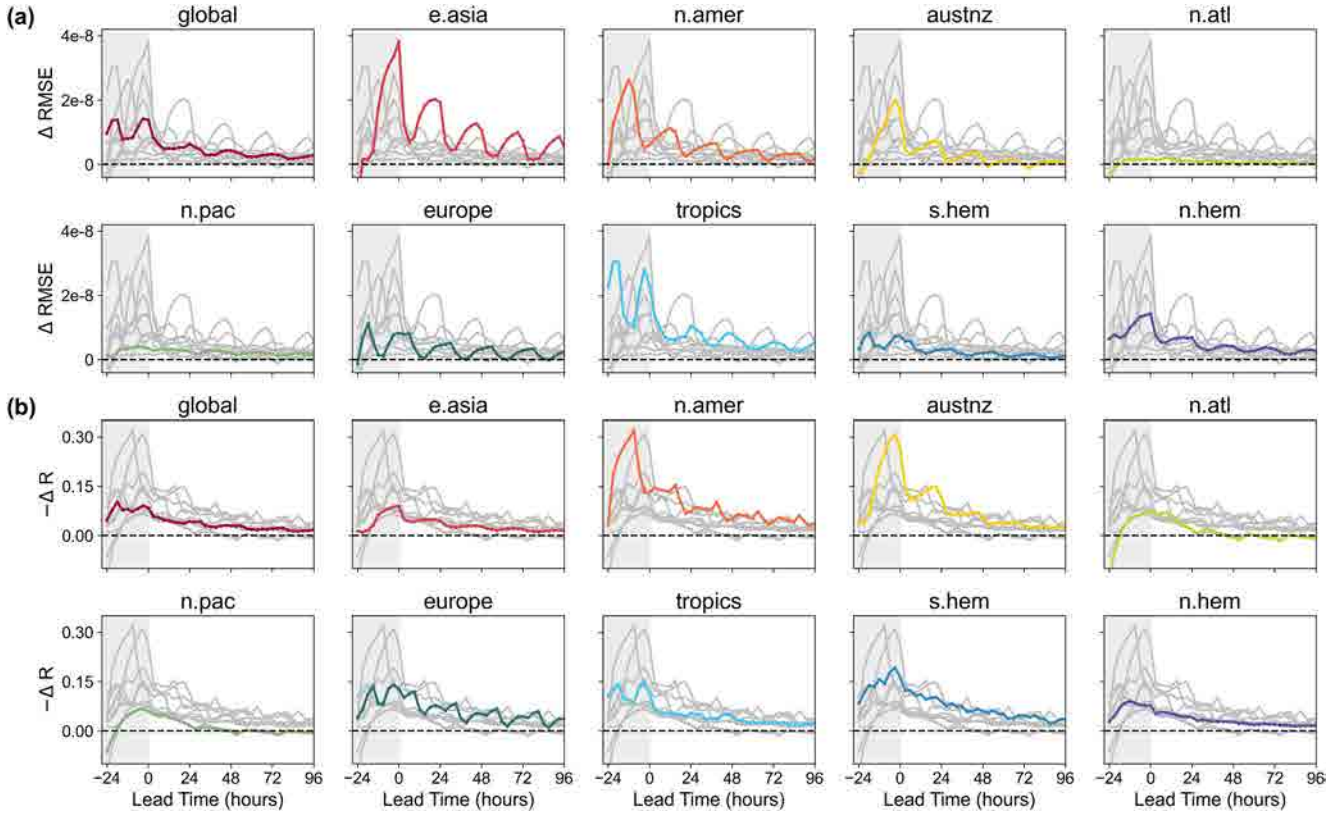


Figure 5. Data Assimilation performance in different subregions at each lead time. (a) Δ RMSE and (b) $-\Delta R$, where Δ is computed by the metrics of previous forecasts without data assimilation minus that of analysis after data assimilation and its further forecasts. The gray shaded areas indicate the assimilation window. The gray lines are the improvements in other regions, which are provided for better comparison.

Figure 6a shows the analysis error, computed by the analysis minus the reanalysis, on 1 June, 5 July, and 15 August 2018. C-IFS generally underestimates CO concentrations these days, while Duo-AttnVarNet tends to overestimate values in several regions. Notably, on 1 June, C-IFS significantly underestimated CO levels across the Northern Hemisphere and Central Africa, whereas Duo-AttnVarNet better refines the state of these areas by effectively assimilating observations.

Figures 6b and 6c display the RMSE and R values for 2018 on daily, monthly, and annual scales, respectively. The result in Figure 6b shows that the analysis resulting from Duo-AttnVarNet consistently achieves lower RMSE and higher correlation on most days, indicating that improvements are not limited to a few extreme cases, but are robust and consistent throughout the year.

When estimating model performance over different months, both models perform worse during the summer months (June–September) and better during the winter months (December–February). We note that Duo-AttnVarNet outperforms C-IFS in every month except January, with particularly significant improvements observed between April and October. This suggests that Duo-AttnVarNet is more effective in challenging conditions, with a stronger ability to incorporate observations and correct substantial initial errors.

On an annual scale, the average RMSEs for Duo-AttnVarNet and C-IFS are 1.12×10^{-7} (kg CO/kg Air) and 1.48×10^{-7} (kg CO/kg Air) respectively, reflecting a 24% improvement in accuracy. The annual average correlation R improves from 0.70 (C-IFS) to 0.79 (Duo-AttnVarNet), marking a relative increase of approximately 11%.

4.3. Evaluation of Duo-AttnOPNets

We finally evaluate the performance of Duo-AttnOPNets, which integrates the forecasting module (Duo-AttnForeNet) and the data assimilation module (Duo-AttnVarNet) iteratively. Specifically, Duo-AttnVarNet

19422466, 2026, 2, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025MS005319 by NICE, National Institute for Health and Care Excellence, Wiley Online Library on [25/05/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

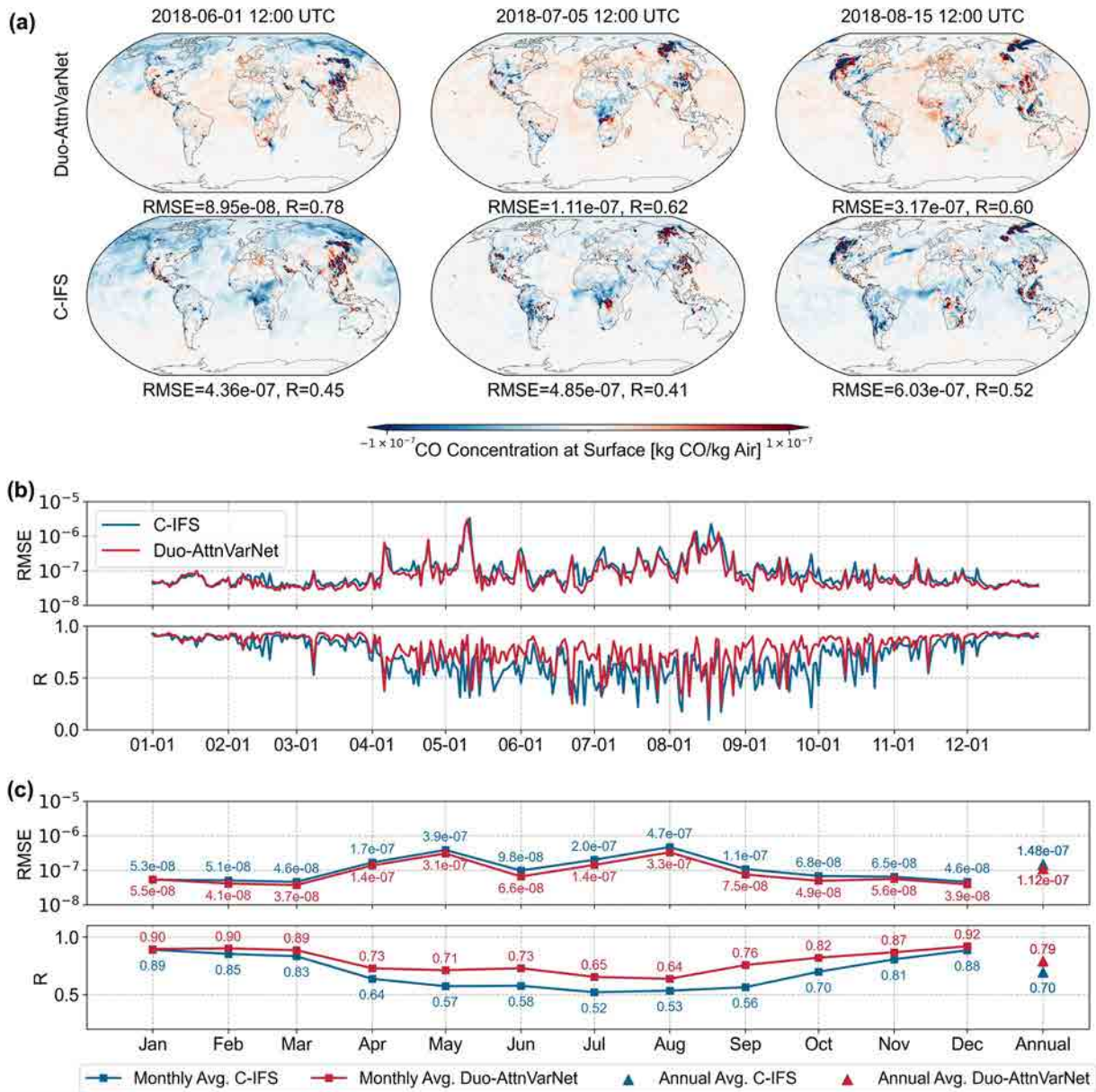


Figure 6. Comparison of 12 UTC analyses between C-IFS and Duo-AttnVarNet. (a) The error analysis obtained by Duo-AttnVarNet(up) and C-IFS(down) at 12:00 UTC on 1 June 2018, 5 July 2018, 15 August 2018 (left to right). All the analyses are computed using an assimilation window of 9:00–21:00 UTC. (b) The RMSE and *R* for the analysis obtained by C-IFS (blue line) and Duo-AttnVarNet (red line) respectively for every day at 12 UTC in the year 2018. (c) The lines with squared markers are the monthly averaged RMSE and *R* for C-IFS (blue) and Duo-AttnVarNet (red), and the upper triangle markers are the annual averaged RMSE and *R* respectively.

assimilates near-real-time satellite observations to adjust the current atmospheric state, and the refined state (analysis) is then used as the initial condition for Duo-AttnForeNet to generate future forecasts. The background fields for each assimilation cycle are provided by the previous forecasts from Duo-AttnForeNet, making the system fully self-cycling, where both the forecast and the analysis are produced by the AI components without reliance on external numerical model outputs. As new observations become available, the assimilation process is repeated, followed by an updated forecast. This cyclical procedure constitutes operational forecasting.

In this section, Duo-AttnVarNet is run once a day, using the period from 00:00 to 24:00 UTC as the assimilation window. The forecasting results for the next 24 hr from Duo-AttnForeNet are then compared with the corresponding forecasts obtained from C-IFS. Duo-AttnOPNets was operated continuously from 1 January to 31 December 2018 to assess its performance of the year-long operational forecast.

19422466, 2026, 2, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025MS005319 by NICE, National Institute for Health and Care Excellence, Wiley Online Library on [25/05/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

19422466, 2026, 2, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025MS005319 by NICE, National Institute for Health and Care Excellence, Wiley Online Library on [23/05/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Figure 7a displays the forecasts from Duo-AttnOPNets and C-IFS models, along with reanalysis, at 00:00 UTC on 03 April 2018, 25 May 2018 and 28 August 2018. It can be found that Duo-AttnOPNets performs comparably to C-IFS in capturing the spatio-temporal variability of CO concentrations driven by meteorological processes. For instance, on 26 May 2018, Duo-AttnOPNets clearly captures the rapid eastward transport of CO from wildfires or industrial emissions in the central United States to Europe via the westerly jet stream along the northern edge of the North Atlantic subtropical high (see blue dashed square). Similarly, on 28 August 2018, it accurately reproduces the elongated band-shaped CO plumes transported eastward from California under the influence of the westerly flow at the periphery of the North Pacific subtropical high (see purple dashed square). In high-concentration hotspot regions, Duo-AttnOPNets also demonstrates strong capability. For example, it identifies the barrier effect of the southern Himalayas, which inhibits CO dispersion and contributes to accumulation. Moreover, through data assimilation, Duo-AttnOPNets effectively accounts for CO variability caused by high-uncertainty emissions, such as those from urban agglomerations in East Asia and wildfires during the dry season in the Sahel region of Africa.

Figures 7b and 7c provide a more intuitive and quantitative comparison of the two models' performances in terms of RMSE and R. Figure 7b presents comparisons on the 3-hr and daily timescales, while Figure 7c illustrates the performance at monthly and annual scales. Across all temporal resolutions, Duo-AttnOPNets consistently outperforms C-IFS. Similarly to previous findings, the model's advantage is somewhat less pronounced during the winter months, when both models perform better overall. On an annual scale, Duo-AttnOPNets achieves an average RMSE of 1.28×10^{-7} (kg CO/kg air) and an averaged R of 0.74, both of which exceed the performance of C-IFS (RMSE: 1.70×10^{-7} (kg CO/kg air), R: 0.68).

4.4. Computational Cost

In this study, both the forecasting and data assimilation processes are executed on a single Nvidia A4500 GPU (20G) device.

The forecasting module Duo-AttnForeNet was trained offline using the 2015–2017 EAC4 reanalysis CO and CO, meteorological variables and emission inventories as supervisory data, with 85% of the period used for training and 15% for validation. The training was performed on one Nvidia A4500 GPU for 400 epochs, taking approximately 66 hr in total. Once trained, Duo-AttnForeNet achieves an average runtime of 0.3 s per forward step, while iterative forecasting for up to 5 days requires only 1.6 s.

For the data assimilation process, the computational cost primarily depends on two key hyperparameters: the learning rate and tolerance threshold, both of which influence the trade-off between computational efficiency and accuracy. The learning rate determines the step size during each iteration of the optimization, while the tolerance threshold defines the minimum decrease in the objective function that triggers early stopping. Lower learning rate and tolerance generally improve accuracy but significantly increase computational time. Therefore, careful tuning of these parameters is essential to balance efficiency and accuracy in relation to actual needs.

Figure 8 presents the computational cost and accuracy of Duo-AttnVarNet under different selections of learning rate and tolerance for the Adam optimizer. The results were obtained by running the data assimilation process once a day from February to April. For each run, we computed the RMSE and R of the analysis at 12:00 UTC, along with the corresponding computation time and number of iterations. The values shown in the figure represent the averages of these daily results over the entire 2-month period. We can find a general trend from this: lower learning rates and tolerances generally lead to improved accuracy but at the cost of increased computational time.

We can see that the results from Duo-AttnOPNets with all tested parameter settings outperform C-IFS in accuracy. Duo-AttnOPNets, with the fastest configuration, only requires 6.3 s to complete the data assimilation process and ensures accurate results (RMSE = 4.213×10^{-8} kg CO/kg Air, $R = 0.875$). When setting a learning rate of 0.0025 and a tolerance of 1, Duo-AttnVarNet can achieve the best balance between accuracy and efficiency, which yields an RMSE of 3.958×10^{-8} kg CO/kg Air and an R of 0.891 with an average running time of 58 s. Further improvements are expected through additional parameter tuning and optimization techniques, such as learning rate scheduling.

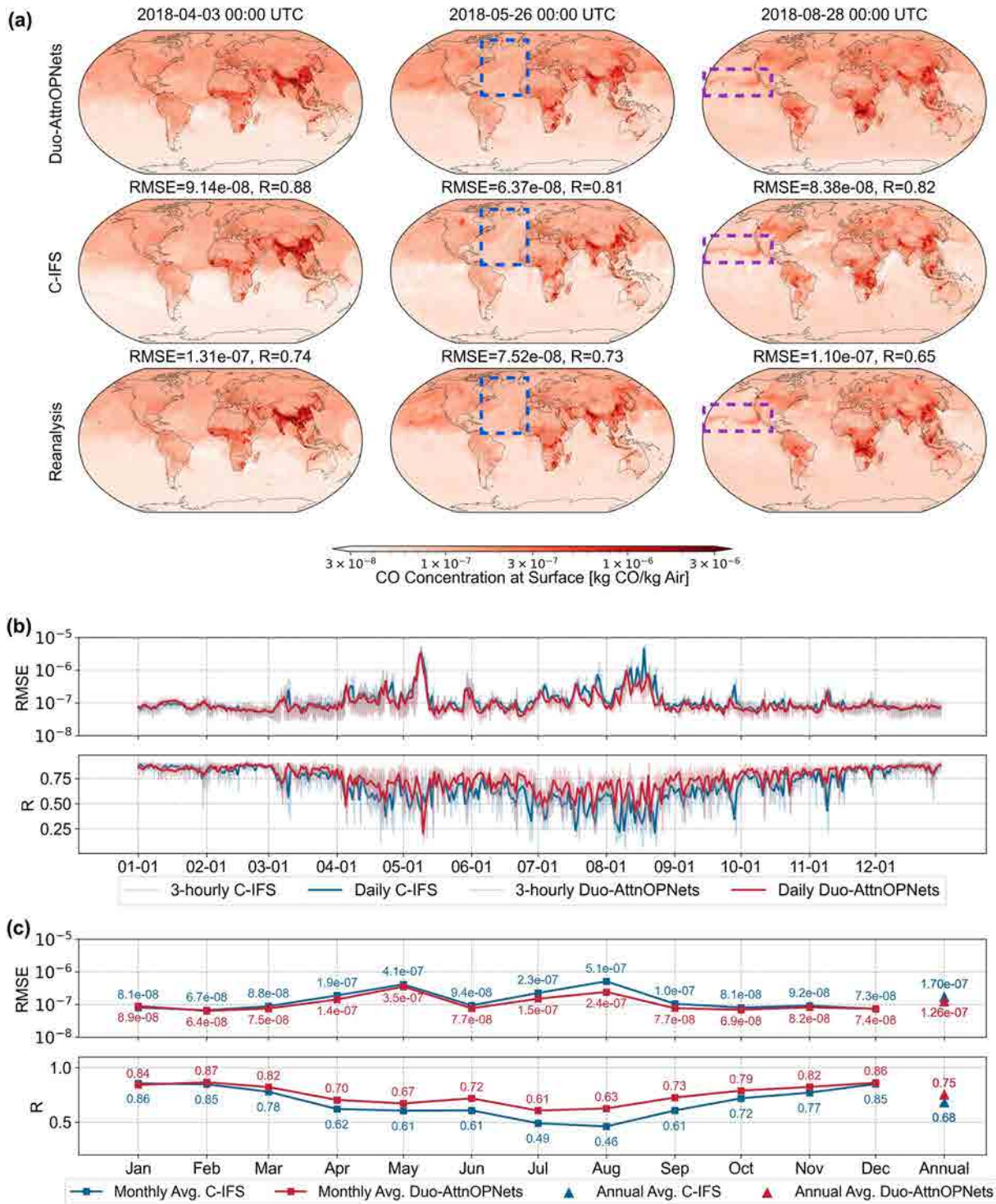


Figure 7. Comparison between Operational Forecasting by C-IFS and by Duo-AttenOPNets. (a) Illustration of CO concentration maps from Duo-AttnOPNets forecasts (top) C-IFS forecasts (middle) and EAC4 reanalysis (bottom), at 00:00 UTC on 3 April 2018, 25 May 2018, 28 August 2018 (left to right). Both operational forecasts were run from 1 January to 31 December 2018. The Duo-AttnOPNets assimilates the near-real-time observations once per day using an assimilation window of 0:00–24:00 UTC. (b) RMSE and *R* for the two models: light red and light blue lines represent 3-hourly forecast errors for Duo-AttnOPNets and C-IFS, respectively, while dark red and dark blue lines show daily averaged errors for forecasts initialized at 00:00 UTC. (c) The lines with squared markers are the monthly averaged RMSE and *R* for C-IFS (blue) and Duo-AttnOPNets (red), and the upper triangle markers are the annual averaged RMSE and *R* respectively.

19422466, 2026, 2, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025MS005319 by NICE, National Institute for Health and Care Excellence, Wiley Online Library on [25/05/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

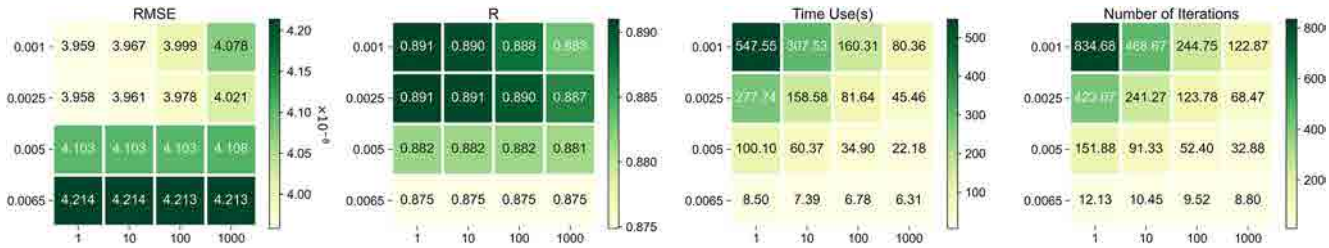


Figure 8. Computational cost and corresponding performance for Duo-AttnVarNet under different learning rates and tolerances of Adam optimizer. Learning rates tested include 0.001, 0.0025, 0.005, and 0.0065, and tolerance values range from 1 to 1,000. The RMSE and R for C-IFS are 4.869×10^{-8} kg CO/kg Air and 0.841 respectively as a reference.

Additionally, the choice of optimizer can influence computational performance. However, by adjusting optimizer parameters, we achieve comparable performance across different optimizers (see Figure S3 in Supporting Information S1 for an example using stochastic gradient descent). This suggests that our framework is not overly sensitive to the optimizer choice.

Overall, Duo-AttnOPNets requires only 7.9 s, on average, to generate a full 5-day forecast with initial analysis on a single GPU. In contrast, C-IFS operates on the ECMWF's supercomputer, Atos Sequana XH2000, which consists of 7,680 nodes (each with 128 cores and 256 GB memory). The C-IFS forecast runs on 32 nodes (4,096 cores in total) with an average runtime of 832 s (±40 s) for forecasting and approximately 2 hr for the complete workflow, including observation data acquisition, preprocessing, assimilation, forecasting, and post-processing. The efficiency of Duo-AttnVarNet highlights the scalability of our approach, making it feasible for real-time operational forecasting without the need for high-performance computing clusters.

5. Discussion

5.1. Summary of Proposed Framework

In this paper, we proposed a novel data-driven operational forecasting framework, Duo-AttnOPNets, that integrates a forecasting module (Duo-AttnForeNet) and a data assimilation module (Duo-AttnVarNet). Duo-AttnForeNet is a neural network based on CSLSTM blocks, that effectively captures the spatiotemporal variations of CO by leveraging the recurrent structure and principles of ConvLSTM, as well as dual attention mechanisms. Duo-AttnForeNet delivers strong predictive performance while remaining highly efficient, requiring only a single GPU for operation. Duo-AttnVarNet, inspired by 4D-Var data assimilation, refines the initial conditions by maximizing the posterior probability of the initial state based on real-time observations along with model dynamics constraints. By utilizing the adjoint state method, built-in automatic differentiation, and GPU acceleration in a deep learning framework, the optimization procedure can be completed at a super-high speed with a slight amount of coding required. Our evaluation demonstrates that Duo-AttnOPNets outperforms the state-of-the-art C-IFS model in both accuracy and efficiency. Specifically, Duo-AttnForeNet maintains a lower cumulative RMSE less than 1.74×10^{-7} (kg CO/kg Air) and higher correlation R greater than 0.66 throughout all lead times, whereas C-IFS experiences an increase in RMSE up to 2.62×10^{-7} (kg CO/kg Air) and a decline in correlation R to 0.57. Additionally, Duo-AttnVarNet enhances initial condition accuracy through satellite data assimilation, with the greatest improvements observed during the first 24 hr, remaining visible up to 48 hr, and diminishing thereafter. The impact also exhibits regional and seasonal heterogeneity, with stronger improvements in high-emission regions such as North America and East Asia, and more significantly from April to October.

5.2. Comparison With Existing Studies

5.2.1. Comparison With Traditional Numerical Models

Traditional numerical forecasting models, such as C-IFS and WRF-Chem, rely on physics-based chemical transport relations such as atmospheric dynamics, thermodynamics, and chemical reactions. These models simulate the evolution of atmospheric compositions by discretizing PDEs and parameterizing unresolved processes. For example, C-IFS, through large-scale parallel computation, provides global CO forecasts with a high spatial resolution of $0.4° \times 0.4°$ in the horizontal. However, such models require extensive high-performance

19422466, 2026, 2, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025MS005319 by NICE, National Institute for Health and Care Excellence, Wiley Online Library on [25/05/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

19422466, 2026, 2, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025MS005319 by NICE, National Institute for Health and Care Excellence, Wiley Online Library on [23/05/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

computing resources, and their computational demands pose significant limitations to advancing forecasts, especially at finer spatial and temporal scales. In contrast, Duo-AttnOPNets achieve comparable or superior accuracy with a dramatically reduced computational burden, making it particularly well-suited for real-time operational forecasting.

5.2.2. Comparison With AI-Based Forecasting Models

Recent advances in AI-driven forecasting models, such as GraphCast (Lam et al., 2023) and FourCastNet (Pathak et al., 2022) have had a significant impact in the field of numerical weather prediction. These models have demonstrated high-resolution predictive capabilities on a global scale, achieving performance comparable to that of traditional physics-based models on various verification metrics (Rasp et al., 2024). However, these models primarily focus on meteorological forecasting (e.g., temperature, wind, precipitation) rather than atmospheric composition modeling. Moreover, Duo-AttnForeNet is a lightweight yet powerful forecasting model, containing orders of magnitude fewer training parameters (∼1.7 M) compared to GraphCast (∼100 M) or FourCastNet (∼1B). To the best of our knowledge, Duo-AttnForeNet is the first global-scale data-driven model for CO forecasting at a high spatial resolution of 0.75° × 0.75° and a temporal resolution of 3 hr. Although previous studies have applied AI models to predict air pollution (Alimissis et al., 2018; Navares & Aznarte, 2020; Noori et al., 2010; Zhang et al., 2023) and greenhouse gas concentration (S. Zhao et al., 2024), most have been limited to regional and subregional scales. For studies that addressed global-scale predictions, forecasts were typically restricted to sparse station-based observations rather than gridded two-dimensional predictions. Moreover, previous research has predominantly used LSTMs and RNNs as model backbones, given their proven ability to capture temporal dependencies. However, Duo-AttnForeNet extends beyond these temporal-only architectures by incorporating spatial dependencies, enhancing its predictive accuracy for spatio-temporal CO variations.

5.2.3. Comparison With AI-Based Data Assimilation Models

Furthermore, AI techniques have been increasingly adopted into data assimilation frameworks due to their advantages in computational efficiency. Most existing AI-based data assimilation approaches approximate the assimilation process by training a neural network as a surrogate model (Cai, Fang, & Wang, 2024; Farchi et al., 2021; Huang et al., 2024; Keller & Potthast, 2024; Wang et al., 2024; Wu et al., 2021). However, these models typically learn the forecast error distribution during training, which means that any change or modification to the forecasting model requires retraining the AI-based assimilation model. In contrast, our proposed Duo-AttnVarNet is more scalable as it is compatible with any forecasting model and requires only a recalculation of the background error covariance matrix. It incorporates the principles of traditional 4D-Var, while taking advantage of deep learning tools to overcome the computational challenges inherent in 4D-Var optimization. The most relevant study by Nonnenmacher and Greenberg (2021) trained neural networks as emulators of the chaotic Lorenz-96 physical system, demonstrating the ability of automatic differentiation to compute the derivatives of these emulators when solving 4D-Var and tuning parameterizations. Subsequent studies have applied similar approaches in realistic case studies (Li et al., 2024; Maulik et al., 2022; Tian et al., 2023). However, these works primarily demonstrated the feasibility and effectiveness of this method, but lacked comprehensive comparisons against established physics-based models.

5.3. Contributions

The main contribution of this work is the development of an AI-empowered 4D-Var operational forecasting framework (Duo-AttnOPNets) that bridges the gap in the application of AI-based models for global-scale forecasting of atmospheric CO. Duo-AttnOPNets demonstrates comparable performance in benchmarking against the well-known numerical model C-IFS. There are two modules in Duo-AttnOPNets: the forecasting module (Duo-AttnForeNet) and the data assimilation module (Duo-AttnForeNet). Duo-AttnForeNet is a lightweight yet powerful forecasting model. It can operate efficiently on a single GPU, demonstrating significantly lower computational requirements than traditional numerical models. This ensures high efficiency and scalable CO forecasting, enabling future extensions to other atmospheric pollutants and applications.

Another key innovation in our approach is the development of an AI-empowered 4D-Var module Duo-AttnVarNet. By utilizing GPU-acceleration and automatic differentiation, it overcomes the limitations of traditional 4D-Var: extensive manual coding and significant computational resources requirements. The accuracy and

19422466, 2026, 2, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025MS005319 by NICE, National Institute for Health and Care Excellence, Wiley Online Library on [25/05/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

robustness of Duo-AttnVarNet are systematically evaluated against C-IFS's data assimilation module in realistic forecasting scenarios, demonstrating its outstanding capability in refining initial conditions. Furthermore, Duo-AttnVarNet employs an adjoint-based approach, whose applications can extend beyond data assimilation to areas such as uncertainty quantification, inverse problems, explainable AI, and optimal control.

5.4. Limitations and Future Work

Despite the strong performance of Duo-AttnOPNets, its forecasting accuracy is inherently constrained by the availability and quality of training data. In this study, Duo-AttnForeNet was trained using the EAC4 reanalysis data set, which limits both its spatial and temporal resolution. While C-IFS provides forecasts at a higher spatial resolution of $0.4° \times 0.4°$ and offers hourly predictions for certain chemical species (e.g., $PM_{2.5}$), Duo-Attn-ForeNet currently operates at a coarser resolution. To further enhance the predictive capability of the model, future work will focus on higher resolution data sets and adopting advanced training strategies, such as super-resolution techniques or physics-informed deep learning approaches.

Moreover, although CO is a chemically reactive gas, the current implementation of Duo-AttnOPNets does not explicitly resolve its chemical reactions. Instead, the framework learns from reanalysis data, which implicitly embeds the combined influences of atmospheric transport, emission, and chemical loss processes. This data-driven formulation allows the model to capture the overall spatiotemporal evolution of CO concentrations without relying on explicit chemical schemes. Nevertheless, the framework is readily extensible to include additional predictors that represent photochemical and oxidation-related processes—such as solar radiation, ozone concentration, or proxies of hydroxyl radicals (OH). Incorporating such variables could further enhance the model's capability to characterize CO's chemical reactivity and improve its predictive accuracy in future applications.

Another limitation of our AI-powered operational framework is its focus on surface-level CO predictions, whereas real-world atmospheric forecasting requires a multi-layer representation of the vertical CO distribution. In contrast, C-IFS simulates CO concentrations across more than 60 vertical levels, later extended to 130 levels in 2019. The ability to resolve vertical transport processes is critical for understanding CO interactions with atmospheric dynamics, chemistry, and long-range transport. In future research, we aim to extend Duo-AttnOPNets to incorporate multi-level CO predictions, enabling a more comprehensive representation of atmospheric composition.

Additionally, integrating fundamental physical principles, such as energy conservation laws and chemical reaction mechanisms, will further enhance the model's applicability to operational air quality and climate monitoring. Expanding the framework to include other atmospheric trace gases and pollutant species will also be a key step toward developing a more robust and generalizable AI-driven forecasting system.

Last but not least, the spatial heterogeneity observed in the assimilation impact reflects differences in regional CO distribution and variability. Regions with strong emission sources often show greater incremental improvement after assimilation, but their absolute forecasting accuracy may remain lower due to the inherent complexity and rapid variability of their CO fields. In contrast, remote oceanic regions with low CO concentrations may exhibit higher absolute accuracy but smaller assimilation impact. One possible way to mitigate this heterogeneity is to train region-specific forecasting and assimilation networks, allowing the models to better capture local emission characteristics, transport processes, and atmospheric regimes. Such region-tailored models may complement the global framework by providing enhanced performance in regions with unique dynamical or emission features.

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Data Availability Statement

All the datasets used in this study can be downloaded from the official websites. The global anthropogenic carbon monoxide emissions dataset CAMS-GLOB-ANT v2.1 (https://eccad.sedoo.fr/#/metadata/479) and biogenic emissions dataset CAMS-GLOB-BIO v3.1 (https://eccad.sedoo.fr/#/metadata/458) are both downloaded from the Global Emissions Initiative data portal (ECCAD). The wildfire emissions GFAS dataset (Copernicus Atmosphere Monitoring Service, 2022; Kaiser et al., 2012) is available in the Copernicus Atmosphere Monitoring Service (CAMS) Atmosphere Data Store, at https://ads.atmosphere.copernicus.eu/datasets/cams-global-fire-emissions-

19422466, 2026, 2, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025MS005319 by NICE, National Institute for Health and Care Excellence, Wiley Online Library on [25/05/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

gfas. The EAC4 reanalysis dataset (Copernicus Atmosphere Monitoring Service, 2020; Inness et al., 2019) was downloaded from https://ads.atmosphere.copernicus.eu/datasets/cams-global-reanalysis-eac4. The global atmospheric composition forecast by IFS (Copernicus Atmosphere Monitoring Service, 2021) is available at https://ads.atmosphere.copernicus.eu/datasets/cams-global-atmospheric-composition-forecasts. Satellite data products for IASI/MeTop-A and IASI/MeTop-B are available on the IASI portal at https://iasi.aeris-data.fr/CO/. The MOPITT dataset used in this study was the Version 8 (V008) product, which was accessible from the NASA Earthdata portal at the time of data acquisition but is no longer available for download. To access the currently available version Version 9 (V009), readers may visit the https://asdc.larc.nasa.gov/project/MOPITT/MOP02J_9.

The source code used in this study is publicly available at Zenodo (https://doi.org/10.5281/zenodo.18320269; (Zhou, 2026)).

Acknowledgments

We are grateful to the sponsors of the Resource Geophysics Academy, Imperial College London, for supporting this research. Funding comes from Engineering and Physical Sciences Research Council (EPSRC) (EP/X029093/1). We appreciate the researchers at ECMWF for their efforts in collecting, archiving, disseminating, and maintaining the EAC4 reanalysis data set, GFAS emission data set, and CAMS global atmospheric composition forecasts dataset. We thank ECCAD (Emissions of atmospheric Compounds & Compilation of Ancillary Data, https://eccad.aeris-data.fr) for providing access to the data used in this study. IASI is a joint mission of EUMETSAT and the Centre National d'Etudes Spatiales (CNES, France). The authors acknowledge the AERIS data infrastructure for providing access to the IASI data in this study, ULB-LATMOS for the development of the retrieval algorithms, and Eumetsat/AC SAF for CO data production.

References

Alimissis, A., Philippopoulos, K., Tzanis, C., & Deligiorgi, D. (2018). Spatial estimation of urban air pollution with the use of artificial neural network models. *Atmospheric Environment*, *191*, 205–213. https://doi.org/10.1016/j.atmosenv.2018.07.058

August, T., Klaes, D., Schlüssel, P., Hultberg, T., Crapeau, M., Arriaga, A., et al. (2012). Iasi on metop-a: Operational level 2 retrievals after five years in orbit. *Journal of Quantitative Spectroscopy and Radiative Transfer*, *113*(11), 1340–1371. https://doi.org/10.1016/j.jqsrt.2012.02.028

Bauer, P., Quintino, T., Wedi, N., Bonanni, A., Chrust, M., Deconinck, W., et al. (2020). *The ECMWF scalability programme: Progress and plans*. European Centre for Medium Range Weather Forecasts.

Bertagni, M. B., Pacala, S. W., Paulot, F., & Porporato, A. (2022). Risk of the hydrogen economy for atmospheric methane. *Nature Communications*, *13*(1), 7706. https://doi.org/10.1038/s41467-022-35419-7

Bey, I., Jacob, D. J., Yantosca, R. M., Logan, J. A., Field, B. D., Fiore, A. M., et al. (2001). Global modeling of tropospheric chemistry with assimilated meteorology: Model description and evaluation. *Journal of Geophysical Research*, *106*(D19), 23073–23095. https://doi.org/10.1029/2001jd000807

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

Blumstein, D., Chalon, G., Carlier, T., Buil, C., Hebert, P., Maciaszek, T., et al. (2004). Iasi instrument: Technical overview and measured performances. *Infrared Spaceborne Remote Sensing XII*, *5543*, 196–207.

Brajard, J., Carrassi, A., Bocquet, M., & Bertino, L. (2020). Combining data assimilation and machine learning to emulate a dynamical model from sparse and noisy observations: A case study with the Lorenz 96 model. *Journal of Computational Science*, *44*, 101171. https://doi.org/10.1016/j.jocs.2020.101171

Cai, S., Fang, F., Tang, X., Zhu, J., & Wang, Y. (2024). A hybrid data-driven and data assimilation method for spatiotemporal forecasting: Pm2.5 forecasting in China. *Journal of Advances in Modeling Earth Systems*, *16*(2), e2023MS003789. https://doi.org/10.1029/2023ms003789

Cai, S., Fang, F., & Wang, Y. (2024). Advancing neural network-based data assimilation for large-scale spatiotemporal systems with sparse observations. *Physics of Fluids*, *36*(9), 096616. https://doi.org/10.1063/5.0228384

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

Chen, R., Pan, G., Zhang, Y., Xu, Q., Zeng, G., Xu, X., et al. (2011). Ambient carbon monoxide and daily mortality in three Chinese cities: The China air pollution and health effects study (capes). *Science of the Total Environment*, *409*(23), 4923–4928. https://doi.org/10.1016/j.scitotenv.2011.08.029

Copernicus Atmosphere Monitoring Service. (2020). *Cams global reanalysis (EAC4)*. Copernicus Atmosphere Monitoring Service (CAMS) Atmosphere Data Store. https://doi.org/10.24381/d58bbf47

Copernicus Atmosphere Monitoring Service. (2021). *Cams global atmospheric composition forecasts*. Copernicus Atmosphere Monitoring Service (CAMS) Atmosphere Data Store. https://doi.org/10.24381/04a0b097

Copernicus Atmosphere Monitoring Service. (2022). *Cams global biomass burning emissions based on fire radiative power (GFAS)*. Copernicus Atmosphere Monitoring Service (CAMS) Atmosphere Data Store. https://doi.org/10.24381/a05253c7

Dong, R., Leng, H., Zhao, J., Song, J., & Liang, S. (2022). A framework for four-dimensional variational data assimilation based on machine learning. *Entropy*, *24*(2), 264. https://doi.org/10.3390/e24020264

Douglas, C., Haldane, J., & Haldane, J. (1912). The laws of combination of haemoglobin with carbon monoxide and oxygen. *Journal of Physiology*, *44*(4), 275–304. https://doi.org/10.1113/jphysiol.1912.sp001517

Fablet, R., Chapron, B., Drumetz, L., Mémin, E., Pannekoucke, O., & Rousseau, F. (2021). Learning variational data assimilation models and solvers. *Journal of Advances in Modeling Earth Systems*, *13*(10), e2021MS002572. https://doi.org/10.1029/2021ms002572

Farchi, A., Bocquet, M., Laloyaux, P., Bonavita, M., & Malartic, Q. (2021). A comparison of combined data assimilation and machine learning methods for offline and online model error correction. *Journal of Computational Science*, *55*, 101468. https://doi.org/10.1016/j.jocs.2021.101468

Fishman, J., & Seiler, W. (1983). Correlative nature of ozone and carbon monoxide in the troposphere: Implications for the tropospheric ozone budget. *Journal of Geophysical Research*, *88*(C6), 3662–3670. https://doi.org/10.1029/jc088ic06p03662

Freitas, S., Dias, M. S., Dias, P. S., Longo, K. M., Artaxo, P., Andreae, M., & Fischer, H. (2000). A convective kinematic trajectory technique for low-resolution atmospheric models. *Journal of Geophysical Research*, *105*(D19), 24375–24386. https://doi.org/10.1029/2000jd900217

Gaubert, B., Worden, H. M., Arellano, A., Emmons, L. K., Tilmes, S., Barré, J., et al. (2017). Chemical feedback from decreasing carbon monoxide emissions. *Geophysical Research Letters*, *44*(19), 9985–9995. https://doi.org/10.1002/2017gl074987

Giering, R., & Kaminski, T. (1998). Recipes for adjoint code construction. *ACM Transactions on Mathematical Software*, *24*(4), 437–474. https://doi.org/10.1145/293686.293695

Grell, G. A., Peckham, S. E., Schmitz, R., McKeen, S. A., Frost, G., Skamarock, W. C., & Eder, B. (2005). Fully coupled "online" chemistry within the WRF model. *Atmospheric Environment*, *39*(37), 6957–6975. https://doi.org/10.1016/j.atmosenv.2005.04.027

Huang, L., Gianinazzi, L., Yu, Y., Dueben, P. D., & Hoefler, T. (2024). Diffda: A diffusion model for weather-scale data assimilation. *arXiv preprint arXiv:2401.05932*. https://doi.org/10.48550/arXiv.2401.05932

Inness, A., Ades, M., Agustí-Panareda, A., Barré, J., Benedictow, A., Blechschmidt, A.-M., et al. (2019). The cams reanalysis of atmospheric composition. *Atmospheric Chemistry and Physics*, *19*(6), 3515–3556. https://doi.org/10.5194/acp-19-3515-2019

Inness, A., Blechschmidt, A.-M., Bouarar, I., Chabrillat, S., Crepulja, M., Engelen, R., et al. (2015). Data assimilation of satellite-retrieved ozone, carbon monoxide and nitrogen dioxide with ECMWF's composition-ifs. *Atmospheric Chemistry and Physics*, *15*(9), 5275–5303. https://doi.org/10.5194/acp-15-5275-2015

Janisková, M., & Lopez, P. (2013). Linearized physics for data assimilation at ECMWF. In S. K. Park & L. Xu (Eds.), *Data assimilation for atmospheric, oceanic and hydrologic applications* (Vol. ii, pp. 251–286). Springer.

Kaiser, J., Heil, A., Andreae, M., Benedetti, A., Chubarova, N., Jones, L., et al. (2012). Biomass burning emissions estimated with a global fire assimilation system based on observed fire radiative power. *Biogeosciences*, *9*(1), 527–554. https://doi.org/10.5194/bg-9-527-2012

Keller, J. D., & Potthast, R. (2024). Ai-based data assimilation: Learning the functional of analysis estimation. *arXiv preprint arXiv:2406.00390*. https://doi.org/10.48550/arXiv.2406.00390

Kerzenmacher, T., Dils, B., Kumps, N., Blumenstock, T., Clerbaux, C., Coheur, P. F., et al. (2012). Validation of Iasi forli carbon monoxide retrievals using FTIR data from NDACC. *Atmospheric Measurement Techniques*, *5*(11), 2751–2761. https://doi.org/10.5194/amt-5-2751-2012

Kim, P. S., Jacob, D. J., Liu, X., Warner, J., Yang, K., Chance, K., et al. (2013). Global ozone–CO correlations from OMI and AIRS: Constraints on tropospheric ozone sources. *Atmospheric Chemistry and Physics*, *13*(18), 9321–9335. https://doi.org/10.5194/acp-13-9321-2013

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

Lamarque, J.-F., Emmons, L., Hess, P., Kinnison, D. E., Tilmes, S., Vitt, F., et al. (2012). Cam-chem: Description and evaluation of interactive atmospheric chemistry in the community earth system model. *Geoscientific Model Development*, *5*(2), 369–411. https://doi.org/10.5194/gmd-5-369-2012

Levy, R. J. (2015). Carbon monoxide pollution and neurodevelopment: A public health concern. *Neurotoxicology and Teratology*, *49*, 31–40. https://doi.org/10.1016/j.ntt.2015.03.001

Li, Y., Han, W., Li, H., Duan, W., Chen, L., Zhong, X., et al. (2024). Fuxi-en4dvar: An assimilation system based on machine learning weather forecasting model ensuring physical constraints. *Geophysical Research Letters*, *51*(22), e2024GL111136. https://doi.org/10.1029/2024gl111136

Maulik, R., Rao, V., Wang, J., Mengaldo, G., Constantinescu, E., Lusch, B., et al. (2022). Efficient high-dimensional variational data assimilation with machine-learned reduced-order models. *Geoscientific Model Development*, *15*(8), 3433–3445. https://doi.org/10.5194/gmd-15-3433-2022

Navares, R., & Aznarte, J. L. (2020). Predicting air quality with deep learning LSTM: Towards comprehensive models. *Ecological Informatics*, *55*, 101019. https://doi.org/10.1016/j.ecoinf.2019.101019

Nonnenmacher, M., & Greenberg, D. S. (2021). Deep emulators for differentiation, forecasting, and parametrization in earth science simulators. *Journal of Advances in Modeling Earth Systems*, *13*(7), e2021MS002554. https://doi.org/10.1029/2021ms002554

Noori, R., Hoshyaripour, G., Ashrafi, K., & Araabi, B. N. (2010). Uncertainty analysis of developed ANN and ANFIS models in prediction of carbon monoxide daily concentration. *Atmospheric Environment*, *44*(4), 476–482. https://doi.org/10.1016/j.atmosenv.2009.11.005

Park, J., Woo, S., Lee, J.-Y., & Kweon, I. S. (2018). Bam: Bottleneck attention module. *arXiv preprint arXiv:1807.06514*. https://doi.org/10.48550/arXiv.1807.06514

Pathak, J., Subramanian, S., Harrington, P., Raja, S., Chattopadhyay, A., Mardani, M., et al. (2022). Fourcastnet: A global data-driven high-resolution weather model using adaptive Fourier neural operators. *arXiv preprint arXiv:2202.11214*. https://doi.org/10.48550/arXiv.2202.11214

Rasp, S., Hoyer, S., Merose, A., Langmore, I., Battaglia, P., Russell, T., et al. (2024). Weatherbench 2: A benchmark for the next generation of data-driven global weather models. *Journal of Advances in Modeling Earth Systems*, *16*(6), e2023MS004019. https://doi.org/10.1029/2023ms004019

Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-K., & Woo, W.-c. (2015). Convolutional LSTM network: A machine learning approach for precipitation nowcasting. In *Advances in Neural Information Processing Systems* (Vol. 28).

Simmons, A. J., & Hollingsworth, A. (2002). Some aspects of the improvement in skill of numerical weather prediction. *Quarterly Journal of the Royal Meteorological Society: A Journal of the Atmospheric Sciences, Applied Meteorology and Physical Oceanography*, *128*(580), 647–677. https://doi.org/10.1256/003590002321042135

Sindelarova, K., Markova, J., Simpson, D., Huszar, P., Karlicky, J., Darras, S., & Granier, C. (2022). High-resolution biogenic global emission inventory for the time period 2000–2019 for air quality modelling. *Earth System Science Data*, *14*(1), 251–270. https://doi.org/10.5194/essd-14-251-2022

Soulie, A., Granier, C., Darras, S., Zilbermann, N., Doumbia, T., Guevara, M., et al. (2023). Global anthropogenic emissions (CAMS-GLOB-ANT) for the Copernicus Atmosphere Monitoring Service simulations of air quality forecasts and reanalyses. In *Earth system science data discussions, 2023* (pp. 1–45).

Tian, X., Conibear, L., & Steward, J. (2023). A neural-network based MPAS—Shallow water model and its 4D-VAR data assimilation system. *Atmosphere*, *14*(1), 157. https://doi.org/10.3390/atmos14010157

Wang, W., Zhang, J., Su, Q., Chai, X., Lu, J., Ni, W., et al. (2024). Accurate initial field estimation for weather forecasting with a variational constrained neural network. *npj Climate and Atmospheric Science*, *7*(1), 223. https://doi.org/10.1038/s41612-024-00776-1

Woo, S., Park, J., Lee, J.-Y., & Kweon, I. S. (2018). Cbam: Convolutional block attention module. In *Proceedings of the European conference on computer vision (ECCV)* (pp. 3–19).

Wu, P., Chang, X., Yuan, W., Sun, J., Zhang, W., Arcucci, R., & Guo, Y. (2021). Fast data assimilation (FDA): Data assimilation by machine learning for faster optimize model state. *Journal of Computational Science*, *51*, 101323. https://doi.org/10.1016/j.jocs.2021.101323

Zhang, A., Fu, T.-M., Feng, X., Guo, J., Liu, C., Chen, J., et al. (2023). Deep learning-based ensemble forecasts and predictability assessments for surface ozone pollution. *Geophysical Research Letters*, *50*(8), e2022GL102611. https://doi.org/10.1029/2022gl102611

Zhao, S., Wu, L., Xiang, Y., & Zhang, F. (2024). Forecasting short-term methane based on corrected numerical weather prediction outputs. *Journal of Cleaner Production*, *460*, 142500. https://doi.org/10.1016/j.jclepro.2024.142500

Zhao, Y., Saunois, M., Bousquet, P., Lin, X., Berchet, A., Hegglin, M. I., et al. (2020). On the role of trend and variability in the hydroxyl radical (OH) in the global methane budget. *Atmospheric Chemistry and Physics*, *20*(21), 13011–13022. https://doi.org/10.5194/acp-20-13011-2020

Zhou, J. (2026). Code for "duo-attnopnets: Advancing global operational forecasting for atmospheric carbon monoxide with ai-empowered 4D-Var" [Software]. *Zenodo*. https://doi.org/10.5281/zenodo.18320269

Ziskin, D. (2019). Measurements of pollution in the troposphere (mopitt) level 2 derived co (near and thermal infrared radiances) (mop02j) v008 [Dataset]. *NASA Atmospheric Sciences Data Center*. https://doi.org/10.5067/TERRA/MOPITT/MOP02J_L2

19422466, 2026, 2, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025MS005319 by NICE, National Institute for Health and Care Excellence, Wiley Online Library on [25/05/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License