npj | climate and atmospheric science

Article

Published in partnership with CECCR at King Abdulaziz University

https://doi.org/10.1038/s41612-026-01410-y

Deep learning-based adjoint modeling for daily CO estimations

Check for updates

Hyun-Jeong Lee[1], Dongjin Cho[2] & Yoo-Geun Ham[1] ✉

In this study, we develop a four-dimensional variational data assimilation (4DVAR)-based adjoint modeling framework using deep learning (DL) for estimating daily carbon monoxide (CO) concentrations over east Asia. The forward model is built on a U-Net architecture, and its adjoint version is formulated by explicitly calculating the gradients of the cost function with respect to the initial condition by leveraging the backpropagation algorithm. The DL-based adjoint model (ADM) was first validated through the adjoint test, the finite-difference gradient test. The idealized experiments using synthetic CO increments at a single grid point over the Korean Peninsula showed that the adjoint sensitivities spread horizontally and vertically to the upstream region along the climatological flows with diffusive features. In addition, the idealized 4DVAR experiments using in-situ CO observations over Korea and Japan during 2020–2022 demonstrated substantial error reduction across the Korean Peninsula, Japan, and northeastern China. This confirms that the DL-based ADM exhibits physically reasonable propagation of the observational information over time and space, offering a promising and efficient alternative to dynamic model-dependent adjoint frameworks constrained by the high computational costs.

Carbon monoxide (CO) is a colorless, odorless, and tasteless gas that is highly toxic[1]. CO is produced through both natural and anthropogenic processes. Nature sources include the oxidation of Methane (CH_4) and Volatile Organic Compounds (VOCs), wildfires and natural combustion, microbial activity in soil and wetlands, or photodegradation of organic matter[2]. Anthropogenic sources are primarily associated with the incomplete combustion of fossil fuels. Once formed, CO undergoes further chemical reactions that pose serious challenges for air quality management[3] and global warming mitigation efforts; CO contributes to the formation of tropospheric ozone, which is harmful to human health, and reduces the concentration of atmospheric hydroxyl radicals (OH), thereby slowing the degradation of other greenhouse gases such as CH_4.

Various types of observations, such as ground-based, airborne-based, and satellite remote sensing, have consistently recognized the East Asia as one of the world's major industrial regions and a significant source of CO[4,5]. In addition, those observations provide crucial information on how the atmospheric CO is transported to the neighboring East Asian countries via large-scale atmospheric circulations[6–8]. Ground- and airborne-based observations offer high accuracy but have limited spatial coverages. On the other hand, Satellite observations using high-resolution sensors, such as Measurement of Pollution in the Troposphere (MOPITT)[9] or Atmospheric Infrared Sounder (AIRS)[10], enable the capture of the three-dimensional CO distribution globally. However, their accuracy is relatively low, and the availability of the observations are limited by meteorological conditions. Thus, Data Assimilation (DA) approaches are essential for integrating various types of observations and short-term global climate model forecasts to produce a complete view of global CO distribution with high accuracy.

Four-dimensional variational data assimilation (4DVAR) is one of the most advanced DA methodologies, as it integrates future observational information over a given assimilation window. This approach enhances the representation of temporal variations of the state vectors, therefore, yields more physically consistent estimates. 4DVAR has been widely applied in numerical weather prediction, reanalysis, and air quality forecasting[11–14].

In addition, due to its ability to seek the linkage between the current states and the future increment (or forecast error), 4DVAR can be applied to the inverse modeling to trace the emission source of pollutant and greenhouses gases[15–17]. Estimation of CO concentrations and emissions using 4DVAR has commonly been conducted by chemical transport models (CTMs) and their adjoint versions. For example, the adjoint model of a Regional Atmospheric Modeling System (RAMS)-based CTM was used to apply a 4DVAR inverse modeling approach to optimize CO emissions and concentrations over East Asia[15]. By assimilating ground-based CO concentration data, prior emission estimates were refined, and it was concluded that the prior emissions derived from the bottom-up inventory were

[1]Department of Environmental Managements, Graduate School of Environmental Studies, Seoul National University, Seoul, Republic of Korea. [2]Environmental Planning Institute, Seoul National University, Seoul, Republic of Korea. ✉e-mail: yoogeun@snu.ac.kr

underestimated mainly in the industrialized eastern part of China. The Goddard Earth Observing System Chemical Transport Model (GEOS-Chem) adjoint system has also been extensively utilized to refine CO emission estimates on both global and regional scales[18,19]. In addition, the implementation of a nested-grid GEOS-Chem adjoint system improved the precision of regional-scale CO estimates[20]. More recently, a hybrid approach integrating the Community Multiscale Air Quality (CMAQ) adjoint-based system with the iterative finite difference mass balance (IFDMB) method significantly improved the stability and temporal representation of CO emission estimates over the Korean Peninsula[21].

While the 4DVAR technique can be employed for estimating CO concentrations and emissions, its application has been largely restricted to a few chemical models equipped with their adjoint versions, making the conclusions drawn from such studies inherently model-dependent. This is due to the fact that the construction of an adjoint operator (i.e., transpose of the tangent linear operator), which computes the linear sensitivities of the cost function with respect to the initial conditions[22,23], based on a dynamical model is challenging for several reasons. First, linearizing highly nonlinear or discontinuous atmospheric processes —such as convective/precipitation schemes and turbulence mixing parameterizations—leads to unavoidable simplifications. Consequently, adjoint versions of the CTMs may struggle to fully capture atmospheric dynamics associated with CO emissions, three-dimensional transport, and sequestration[14,24,25]. Second, the application of 4DVAR incurs a high computational cost, as it requires repeated forward–adjoint iterations in high-dimensional state spaces of dynamical models[15]. Although the Ensemble Kalman Smoother would mitigate these limitations, the limited ensemble size hinders the accurate estimation of flow-dependent background error covariances.

The aforementioned difficulties can be mitigated by Deep Learning (DL) algorithms, which are advantageous not only in effectively learning complex nonlinearities by fitting the nonlinear activation functions to the large amount of the training dataset, but also in that the linearization of the DL-based nonlinear operator is feasible as the linearized form of the nonlinear activation functions are well established[26]. In addition, the back-propagation algorithm enables DL models to compute the gradient of the loss function with respect to the initial conditions using the chain rule of calculus. Furthermore, DL frameworks offer computational advantages, as they are optimized for high-dimensional data processing by leveraging GPU acceleration and parallel computing, significantly reducing the computational burden of adjoint model calculations[27,28].

By leveraging their capabilities, DL algorithms have been employed to build adjoint models within data assimilation frameworks[29,30]. A pioneering study trained a neural network to emulate non-orographic gravity wave drag parameterizations and successfully derived its tangent-linear and adjoint models by differentiating the neural network[29]. Recently, by leveraging the development of DL-based global climate models, complete version of 4DVAR system is built solely based on DL. For example, the FengWu-4DVAR system demonstrated that the automatic differentiation capabilities of DL frameworks not only alleviate the challenges of traditional adjoint model development but also enable practical adjoint-based sensitivity optimization with significantly reduced computational cost[30]. Nonetheless, this fully integrated gradient computation often treats the model as a "black box". That is, the adjoint model is implicitly implemented within the process of calculating the gradient of the 4DVAR cost function, making it difficult to assess the validity and characteristics of the adjoint model through quantitative analysis.

This study aims to enhance the interpretability and analytical utility of adjoint models for three-dimensional atmospheric CO concentrations within the DL framework. To this end, we explicitly calculate the adjoint operator, defined as the transpose of the Jacobian matrix $\partial y/\partial x$, where y denotes the model output and x denotes the state vector, based on a U-Net architecture[31]. By explicitly obtaining the DL-based adjoint operator, we enable quantitative analysis of adjoint sensitivity at each grid point, thereby assess the contribution of each observation component to the total 4DVAR cost function.

This paper is structured as follows. The Data and Methods section describe the datasets and the DL-based modeling framework. The Results section presents the performance of the forward model, the verification of the tangent-linear and adjoint models, and the spatiotemporal patterns of adjoint sensitivities in the DL-based adjoint model. It also includes the results of idealized 4DVAR experiments using ground-based observations. Finally, the Discussion section summarizes and interprets the main findings.

Results

Forward model performance

The predictive performance of the forward model was evaluated by calculating the globally averaged temporal correlation and root mean square error (RMSE) between the predicted and reanalysis CO fields for lead times from 1-day to 10-day (Fig. 1a, b). Forecasts for more than 2-day leads were produced with an autoregressive approach[32], where earlier predictions were recursively used as inputs to extend the forecasts to longer lead times. Temporal correlation coefficients gradually decreased with increasing lead time across all vertical levels. At 985 hPa (near-surface), the correlation remained relatively high (0.62 at a 5-day lead), while it declined more rapidly at upper levels such as 337 hPa (0.47 at a 5-day lead). Consistently, RMSE (Fig. 1b) increased with lead time, with lower values at near-surface levels and higher values at upper levels. To further assess the forward model's ability to capture temporal dynamics, we analyzed the East Asian-averaged (15–55 °N, 70–160 °E) CO anomaly time series at 985 hPa for the evaluation periods in 2020, 2021, and 2022 (Fig. 1c–e). The results indicate that the model effectively tracks the temporal variability and sudden shifts observed in the MERRA-2 data across all evaluation years. The correlation skills, calculated over the entire evaluation period with the respective mean of each year removed, are 0.893 for 1-day lead, 0.637 for 3-day lead, and 0.542 for 5-day lead, which confirms a high degree of temporal consistency. While the amplitude of the anomaly exhibited a negative drift at longer lead times due to error accumulation, the model successfully maintains consistent temporal variations.

These results indicate that the DL-based forward model can reasonably predict CO concentrations in the lower to middle troposphere for approximately ten days, and in the upper troposphere for approximately five days. Based on this evaluation, subsequent sensitivity experiments and idealized 4DVAR experiments were conducted within a 5-day forecast window, during which the forward model retained sufficient accuracy across all vertical levels.

Validity of the tangent linear model and the adjoint model

We first performed a perturbation test to verify the linearity assumption in the tangent linear model (TLM) by ensuring that the nonlinear higher-order terms remained small. The validity of the TLM is quantified by computing the ratio r, which represents the relative difference between the nonlinear Model (NLM, i.e., the DL-based forward model) and the TLM:

$$r = \frac{||\psi_N - \psi_L||^2}{||\psi_N||^2} \tag{1}$$

In Eq. (1), ψ_N denotes the result from the NLM, defined as $f(x + \delta x) - f(x)$. This represents the difference between two nonlinear integrations initialized with and without a perturbation δx. ψ_L denotes the result from the TLM, obtained by prescribing the same perturbation δx as the input[33]. Consequently, as the ratio r approaches zero, it indicates that the higher-order nonlinear terms are negligible and the linear approximation is dominant.

The initial perturbation is defined as the difference between the normalized CO anomaly at time $t + 1$ and that at time $t + 0$ (i.e., $perturbation = CO_{t+1} - CO_{t+0}$). This perturbation can be regarded as the 1-day forecast error of a persistence forecast. A total of 36 cases were selected from January 1st to February 28th during 2020-2022 at 5-day intervals. Note that the boreal winter season (i.e., January and February) was chosen because the climatological CO concentrations around the Korean Peninsula

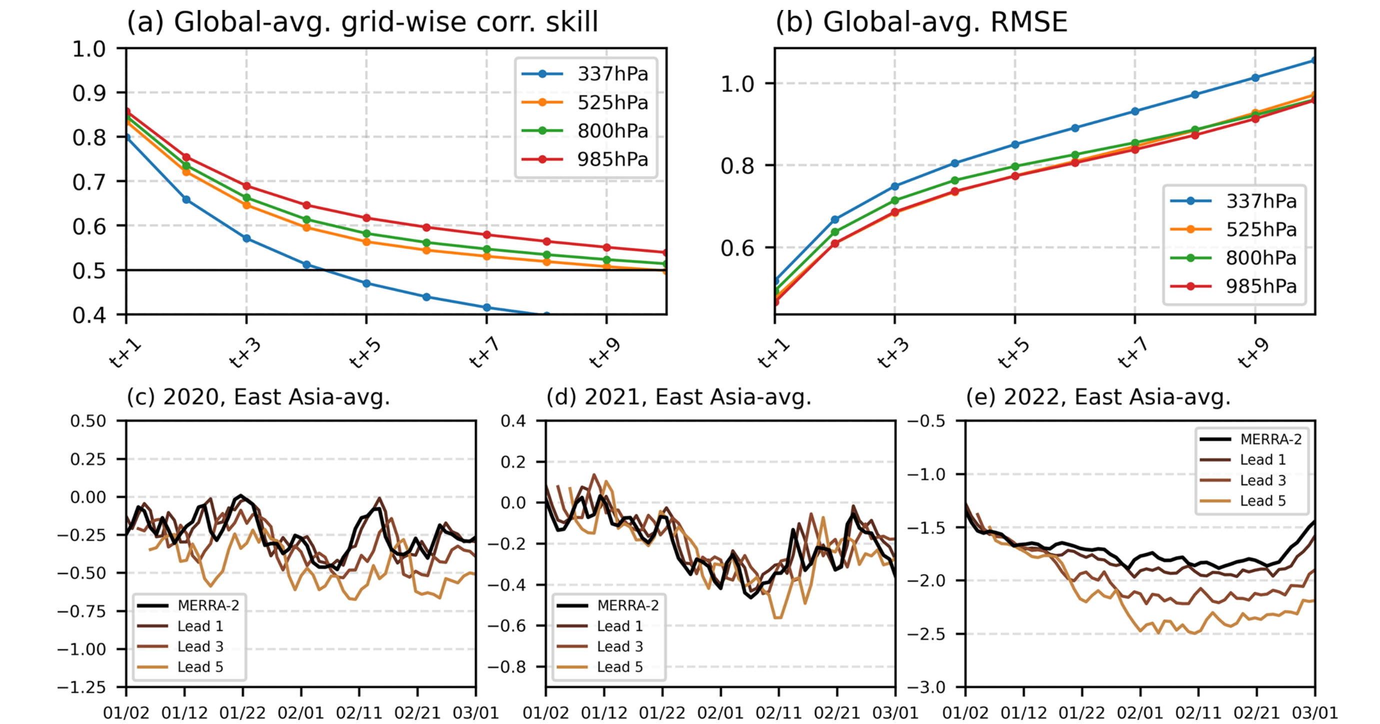

Fig. 1 | Prediction performance of the DL-based forward model. **a** Global-averaged grid-wise temporal correlation skill and (**b**) root mean square error (RMSE) between the predicted CO anomaly and the reanalysis data are presented for lead times from t + 1 to t + 9 across four pressure levels (337, 525, 800, and 985 hPa). Time series of the area-averaged CO anomaly over East Asia (985 hPa, 15–55° N, 70–160° E) for (**c**) 2020, (**d**) 2021, and (**e**) 2022. The black line represents MERRA-2 reanalysis, and the colored lines represent model predictions at lead day 1 (R = 0.893), 3 (R = 0.637), and 5 (R = 0.542).

Table 1 | Results of the perturbation test, adjoint consistency test, and finite-difference gradient test for various perturbation scales evaluated for the 1-day forecast operator

Perturbation Scale	NLM vs TLM	TLM vs ADM	FD vs ADM
10^{0}	1.380850	1.000000	−7.766109
10^{-1}	0.586034	1.000000	−0.076270
10^{-2}	0.016395	1.000000	0.868825
10^{-3}	0.000507	1.000000	0.987382
10^{-4}	0.000036	1.000000	0.998701
10^{-5}	0.000006	1.000000	0.999935

The second column shows the relative squared error between the NLM and the TLM, quantifying the validity of the tangent linear approximation. The third column displays the inner-product consistency between the TLM and the ADM, confirming that the adjoint operator is implemented as the exact transpose of the TLM. The last column reports the agreement between the FD-based directional derivative and the adjoint-derived gradient (FD vs ADM). Values close to unity indicate strong consistency between the FD approximation and the adjoint gradient, while deviations at large perturbation amplitudes reflect the increasing influence of nonlinear effects over the 1-day forecast window.

reach their annual maximum during this period (as will be discussed in the next section). To examine the validity of the TLM for various amplitudes of the initial perturbations, the perturbation amplitudes were scaled by multiplying factors ranging from 10^{0} to 10^{-5}.

As shown in the second column of Table 1, r remained close to zero for perturbations in the range 10^{-5} to 10^{-2}, indicating that the linear approximation held sufficiently within this amplitude range. This demonstrates the validity of using the TLM for perturbation with a magnitude less than 10^{-2} of the forecast error. In contrast, the perturbation amplitude ratio increased to 0.59 and 1.38 for 10^{-1} and 10^{0}, respectively, indicating the linear approximation was not strongly valid in this perturbation amplitude range.

It is noteworthy that the magnitude of increments applied in our 4DVAR experiments fell well within the perturbation range where the linear approximation holds; in the real-observation experiments, the 36 evaluation case-averaged 1-day normalized prediction error at Anmyeondo station (i.e., $R^{-1}|y - f(x)|$) was 0.0945.

Next, the adjoint test was performed to check that the adjoint model (ADM) was correctly implemented as the mathematical transpose of the TLM by calculating the following equation:

$$\left\langle \mathbf{M}(x_t) \cdot \delta x_t,\ \delta x_{t+1} \right\rangle = \left\langle \delta x_t,\ \mathbf{M}^T(x_t) \cdot \delta x_{t+1} \right\rangle \quad (2)$$

where $\mathbf{M}(x_t)$, and $\mathbf{M}^T(x_t)$ is the TLM, and the ADM at time t, respectively, and the angle brackets denote the standard inner product in Euclidean space[34]. The results showed that this identity perfectly held to machine precision across all tested perturbation scales (third column of Table 1), demonstrating that the ADM was implemented exactly as the transpose of the TLM. These validation results demonstrate that the ADM accurately captured the adjoint sensitivity, defined as the gradient of the cost function with respect to the model input, and was mathematically consistent with both the NLM and TLM.

Finally, a systematic finite-difference (FD) gradient test was conducted specifically for the 1-day forecast operator[16,35]. For a cost function J(x), the change induced by applying a small perturbation $\varepsilon\delta x$ to the initial state x can be approximated by the first-order Taylor expansion as

$$\mathrm{J}(x + \varepsilon\delta x) - \mathrm{J}(x) \approx \varepsilon \nabla_x \mathrm{J}(x) \cdot \delta x \quad (3)$$

where $\nabla_x \mathrm{J}(x)$ denotes the gradient of the cost function with respect to the model state, which can be expressed in terms of the ADM as

$$\mathrm{J}(x + \varepsilon\delta x) - \mathrm{J}(x) \approx \varepsilon \left(\frac{\partial y}{\partial x}\right)^{T} \frac{\partial \mathrm{J}}{\partial y} \cdot \delta x \quad (4)$$

The left-hand side of Eq. (11) corresponds to the FD approximation obtained from NLM integrations, while the right-hand side of Eq. (12)

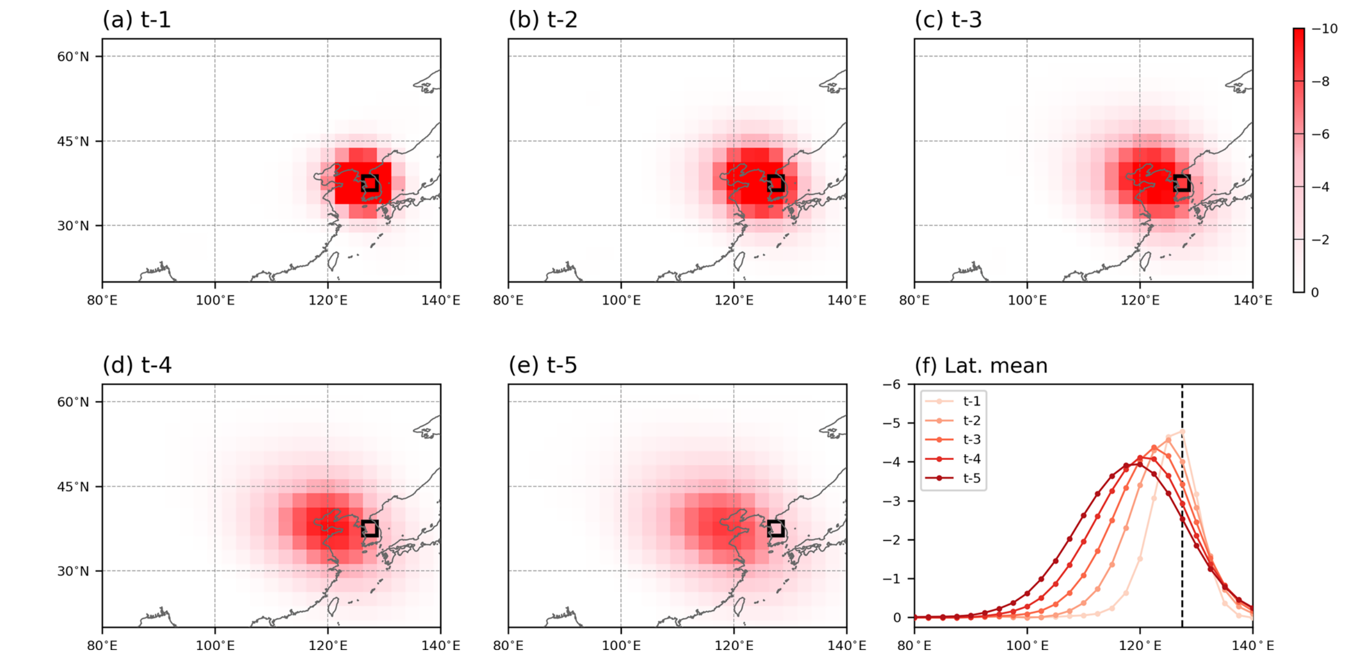

Fig. 2 | Normalized adjoint sensitivities from t − 1 to t − 5, averaged over 36 idealized wintertime cases. **a**–**e** show the spatial distribution of adjoint sensitivities at each time step on the surface (985 hPa), with the observation location marked by a black square (127.5° E, 37.5° N). **f** Zonal averages (20–60° N) of the adjoint sensitivities across different time steps.

represents the directional derivative computed using the ADM (i.e., $\left(\frac{\partial y}{\partial x}\right)^{T}$). Agreement between these two quantities a validity of the ADM implementation.

As shown in the fourth column of Table 1, for the 1-day forecast operator, the ratio of the FD increment to the ADM-based estimate converges toward unity as the perturbation magnitude decreases. Specifically, the ratio reaches approximately 0.87 at $\varepsilon = 10^{-2}$, and tightly converges to near 0.99 for smaller scales ($\varepsilon < 10^{-2}$). In contrast, for larger perturbations ($\varepsilon \geq 10^{-1}$), the ratio deviates substantially and even becomes negative at $\varepsilon = 10^{0}$. This behavior clearly indicates a breakdown of the linear approximation, reflecting the dominance of higher-order nonlinear terms for finite-size perturbations.

The accumulation of forecast errors over an extended period would make the direct numerical verification of the ADM highly challenging. However, we confirmed that the magnitude of the prediction error remains within the verifiable linear regime (near the 10^{-2} scale) throughout the entire assimilation window (i.e., 5-day period). For example, the average normalized prediction error, $R^{-1}|y - f(x)|$, at Anmyeondo station is 0.0968 at Day 5, indicating that the error propagation does not push the model into a highly nonlinear regime. Consequently, since the model state stays within the valid linear range of the TLM, the adjoint sensitivities calculated via our ADM—which has been strictly verified as the exact mathematical transpose of the TLM—are physically reliable and numerically stable for the 5-day window. Therefore, these results confirm that the ADM accurately captures the gradient of the cost function and remains mathematically reliable for both 1-day and 5-day backward sensitivity analyses. This provided a robust foundation for adjoint-based optimization and 4DVAR experiments.

Spatiotemporal pattern of DL-based adjoint sensitivity

To examine the spatiotemporal characteristics of DL-based adjoint sensitivity, we selected 36 cases at 5-day intervals from January 1 to February 28 during 2020–2022. For each case, adjoint sensitivities were computed backward in time for 5 days (i.e., from time t-1 to t-5) by following the experimental setup described in Methods (0.01 at 37.5° N, 127.5° E, 985 hPa).

Figure 2 presents the horizontal pattern of the normalized adjoint sensitivity at 985 hPa from time t-1 to time t-5, averaged over the 36 cases. Note that we normalized the adjoint sensitivity by dividing the standard deviation of the sensitivity over 20–60° N, 80–140° E for each time, to focus on the spatial distribution and temporal propagation characteristics of the sensitivities in time. At time t-1 (Fig. 2a), the maximum sensitivity was centered over the Korean Peninsula, reflecting the direct response to the imposed observation increment. As the adjoint sensitivity iteratively propagated backward in time (Fig. 2b–e), the sensitivity pattern spread northwestward and became more diffuse, reflecting upstream influence through downstream transport processes[36]. This northwestward expansion of the adjoint sensitivity backward in time was consistent with the dominant low-tropospheric northwesterly flow over the East Asia during the boreal winter. The latitudinal mean of the normalized adjoint sensitivities over 20–60° N also confirms that the center of influence systematically propagate to upstream regions with increasing lead time (Fig. 2f).

The spatiotemporal characteristics of adjoint sensitivity was further tested by additional experiments by prescribing an observation increment over a mid-troposphere (i.e., 600 hPa) (Fig. 3). Compared to the near-surface case, the sensitivity fields at 600 hPa exhibited a broader westward spread and extended farther into the inland areas of northern and western China. For example, for the experiment with observation increment at 985 hPa in Fig. 3, the adjoint sensitivity remained strong over upstream regions even five days prior to the observation time, and the signal stayed closer to the surface and was less rapidly advected away. In contrast, for the adjoint sensitivity over 600 hPa, the sensitivity field spread farther westward, reaching deep into northern and western China. Along with the stronger westerly flow aloft that transported the signal more rapidly, the adjoint sensitivity field became more spatially diffuse and weakened more quickly with increasing lead time. The normalized sensitivity decreased below −2 in the farthest upstream regions, whereas near the surface it remained stronger, around −4 even at t − 5.

This enhanced upstream propagation and the more diffusive nature of CO adjoint sensitivity in the mid-troposphere were consistent with the stronger horizontal winds and reduced frictional resistance compared to the surface layer. These findings are consistent with previous observational and modeling studies showing that CO emitted at higher altitudes tends to be transported more efficiently and over longer distances than emissions near the surface, owing to stronger and more persistent westerly winds and weaker vertical mixing at those levels[37,38].

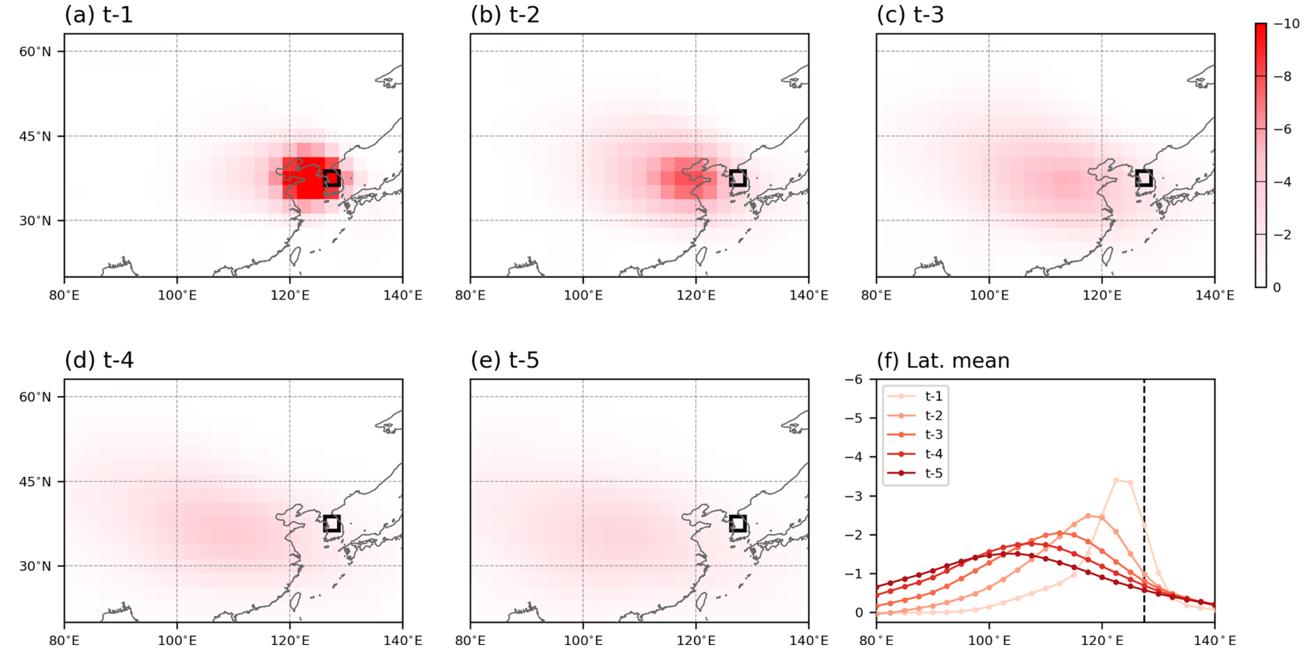

Fig. 3 | Normalized adjoint sensitivities from t − 1 to t − 5 at 600 hPa, averaged over 36 idealized wintertime cases. a–e display the spatial distribution of adjoint sensitivities at each time step, with the observation location indicated by a black square (127.5° E, 37.5° N). **f** Zonal averages (20–60° N) of the adjoint sensitivities across different time steps.

We roughly estimated the potential horizontal transport distance over five days based on climatological winds in the upstream region of the Korean Peninsula (30–60° N, 80–125° E). At 985 hPa, the surface climatological zonal wind speed during December-January-February averaged over the east Asia domain (20–60° N, 80–140° E) is 4.66 m s^{-1}, giving a rough estimated transport distance for 5 days of 2013.12 km. At 985 hPa, sensitivity at time t-5 is shown over 105° E at the westernmost point, meaning possible maximum horizontal extension about 20-degree (i.e., 2000–3000 km) west of Korea. This indicates, to some extent, the simulated extent of adjoint sensitivity and the expected synoptic-scale transport distances under typical winter westerlies support the physical realism of the DL-based ADM.

To examine whether the spatial feature of the adjoint sensitivity is reasonable, we qualitatively compared our results with HYSPLIT[39]. Simulations of 5-day backward air mass trajectories originating from the Korean Peninsula at 600 hPa for wintertime cases (Jan. and Feb., 2020–2022) showed that the mean backward trajectory exhibits a predominant northwestward propagation at early backward time (i.e., t-1 and t-2), then, westward propagation at an earlier time (i.e., earlier than t-3) (not shown). This primary atmospheric transport pathway aligns with the directional progression of the DL-based adjoint sensitivity.

Another key finding from this analysis was the identification of possible geographic regions of anthropogenic CO emissions that influence CO concentrations over Korea and the transport pathways through which these influences occur. When the observation increment was prescribed near the surface (i.e., 985 hPa), the adjoint sensitivities were robust over eastern and parts of north-central China. This relatively confined pattern indicates that, over a 5-day period, surface-level CO anomalies over Korea were affected primarily by emissions or transport from nearby regions. In contrast, the observation at mid-tropospheric levels (e.g., 600 hPa) might imply that CO emitted deeper into inland and northwestern parts of China affect CO concentration over Korea. Such spatial patterns are consistent with previous observational and modeling studies. For example, CO emissions originating near Shanghai in southeastern China have been reported to affect air quality over the Korean Peninsula through long-range atmospheric transport[7]. Under prevailing large-scale synoptic westerlies, pollutants emitted from these regions can cross the Yellow Sea and reach central Korea within a few days. The similarity between these documented transport behaviors and the westward spread of adjoint sensitivities in this study supports the conclusion that the DL-based ADM correctly captures realistic atmospheric transport characteristics.

Next, we examined how adjoint sensitivities propagate vertically over time when the observation increment is applied at various vertical levels. Figure 4 shows the vertical distribution of normalized adjoint sensitivities averaged over 80° E–140° E and 20° N–60° N from time t-1 to t-5. The sensitivities were normalized at each vertical level by dividing them by the standard deviation at the corresponding level to highlight relative vertical variations. When the observation was placed near the surface level (i.e., 985 hPa), sensitivities remained largely near the surface up to 5-day backward integration with minimal vertical extension above 700 hPa (Fig. 4a). For mid-tropospheric observations (800–600 hPa), adjoint sensitivities were distributed both above and below the observation level; the 800 hPa case showed stronger downward propagation, whereas the 600 hPa case exhibited a more evenly distributed sensitivity (Fig. 4b–c). With the observation located at 412 hPa, sensitivities extended primarily upward, with minimal values in the lower troposphere, indicating that its influence is largely confined to the middle and upper troposphere (Fig. 4d). Similarly, in the 208 hPa case (Fig. 4e), sensitivities were concentrated in the upper troposphere, with minimal downward propagation to lower levels.

This analysis demonstrates that adjoint sensitivities propagate both horizontally and vertically, with the vertical level of the observation playing a decisive control on the direction and extent of this propagation. In particular, the results highlight that the observation's vertical level is a key factor in identifying which layers of the model state exert the greatest influence on reducing forecast errors, thereby providing a physically interpretable basis for vertical source attribution in adjoint-based update.

Idealized 4DVAR Experiments Using In-situ Observation Station

4DVAR experiment was conducted to assimilate in-situ observation at Anmyeondo station, aiming to assess how much the DL-based forward and ADM improved forecast accuracy. As described in Methods, the cost function for the idealized 4DVAR included only the observation term, defined based on the observational increments, to focus on the impact of the ADM. Consequently, the spatial pattern of the analysis reflected solely the spatial distribution of the adjoint sensitivity, excluding the background term.

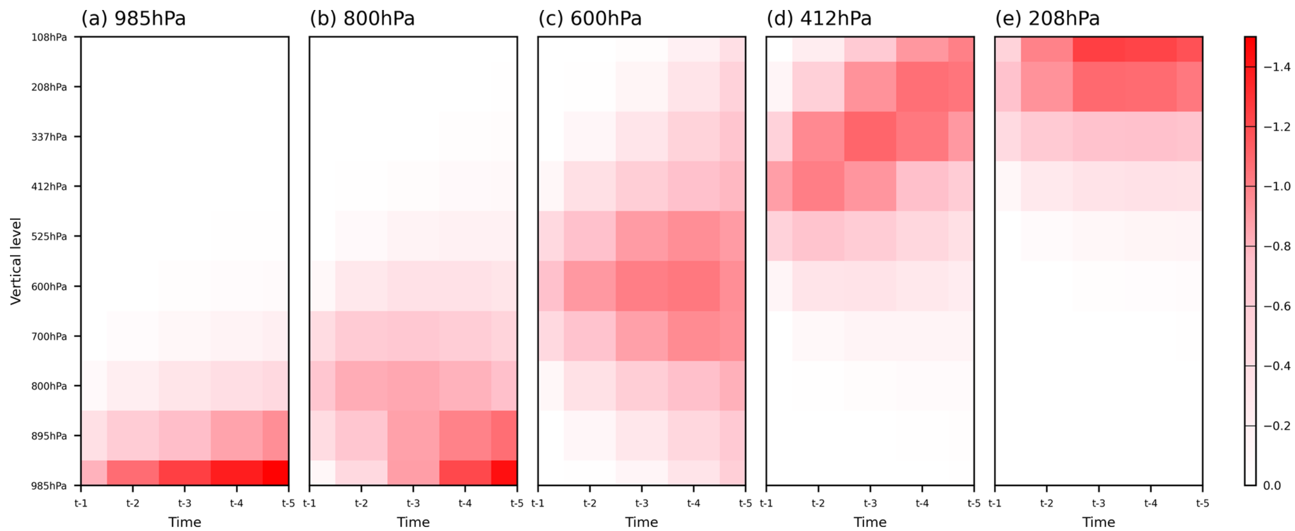

Fig. 4 | Time–vertical level distribution of normalized adjoint sensitivities averaged over the domain 80°–140° E, 20–60° N. Panels show the results when the observation increment is imposed at different pressure levels: **a** 985 hPa, **b** 800 hPa, **c** 600 hPa, **d** 412 hPa, and **e** 208 hPa.

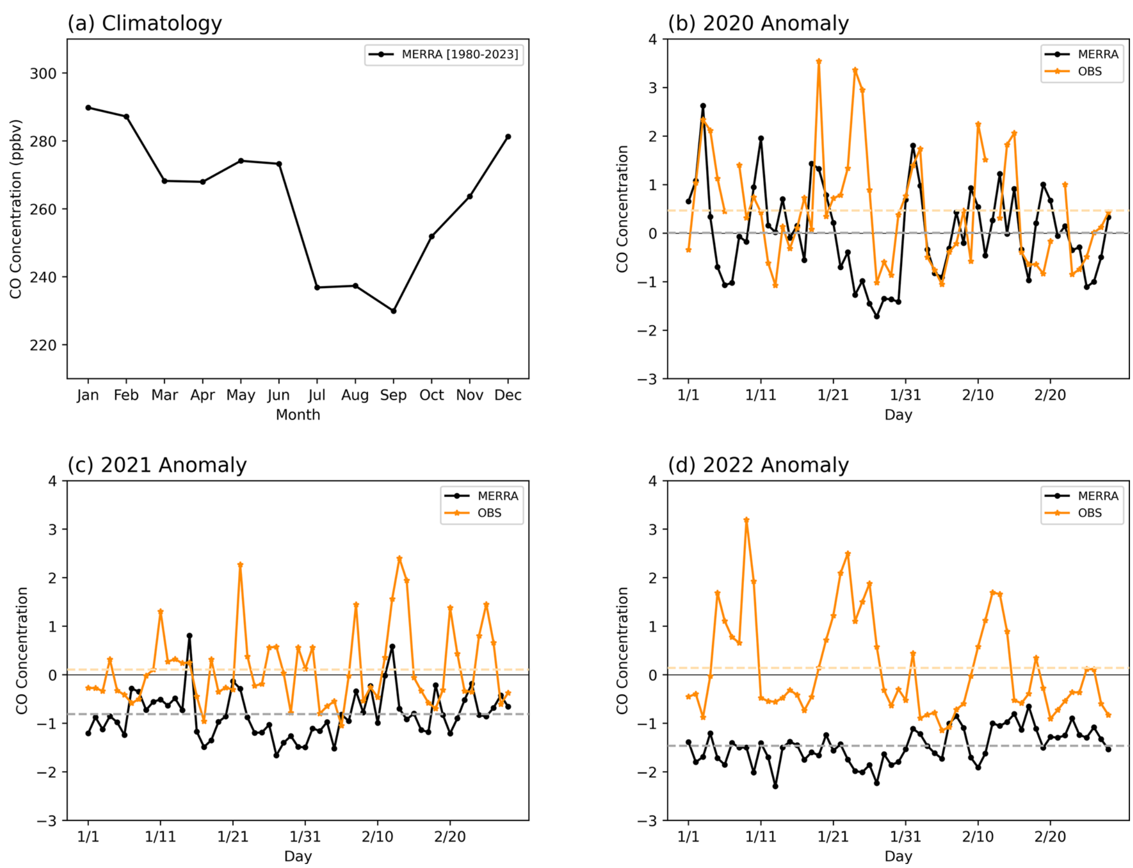

Fig. 5 | Climatology and daily variability of CO concentration over the Korean Peninsula. a Monthly climatology of CO concentration (ppbv) from MERRA-2 averaged over 1980–2023 at the Korean Peninsula grid point (127.5° E, 37.5° N). Daily anomalies of CO concentration at the same grid point during January–February of (**b**) 2020, (**c**) 2021, and (**d**) 2022. Black lines represent MERRA-2 anomalies, and orange lines represent Anmyeondo observation anomalies. Gray dashed lines indicate the mean of MERRA-2 anomalies over the entire period, and yellow dashed lines indicate the mean of Anmyeondo observation anomalies over the same period.

Figure 5 presents the climatological CO concentration from the MERRA-2 reanalysis as a function of calendar month (Fig. 5a) and the daily CO anomalies from both MERRA-2 and in-situ observations at Anmyeondo station for January–February of 2020–2022 (Fig. 5b–d). The comparison of the time series between the reanalysis and the surface observations clearly revealed discrepancies across all three years. In 2020, even both time series exhibited common peaks on January 3, 18, February 2, 15, however, peaks with opposite sign also appeared around January 8, 26, February 10 (Fig. 5b). In 2021, in addition to the discrepancy in day-to-day variations, a pronounced mean bias emerged, particularly after late January (Fig. 5c). The time-averaged in-situ anomaly was approximately +0.9 ppbv higher than the MERRA-2 reanalysis. In 2022, the observed anomalies often exceeded +2

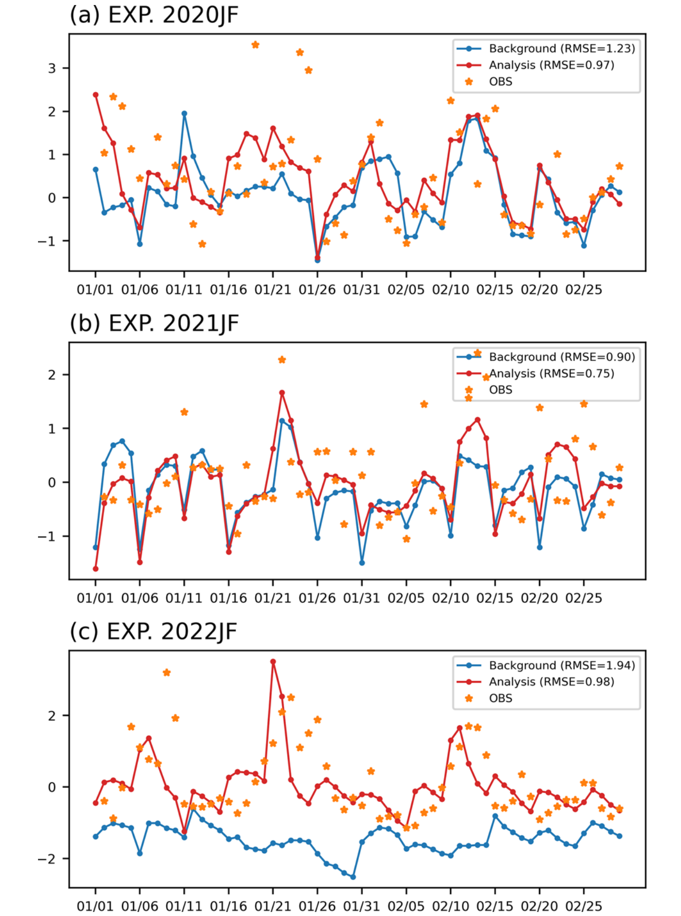

Fig. 6 | Results of idealized 4DVAR experiments using Anmyeondo CO measurements. Panels show daily CO anomalies for January–February in (**a**) 2020, (**b**) 2021, and (**c**) 2022. The blue line indicates the background states (before the 4DVAR experiment), the red line shows the analysis states (after the 4DVAR experiment), and orange stars represent Anmyeondo observations. The MERRA-2 reanalysis field (37.5° N, 127.5° E) is taken from the single grid point closest to the observation site (36.54° N, 126.33°E). RMSE values are provided in the legend.

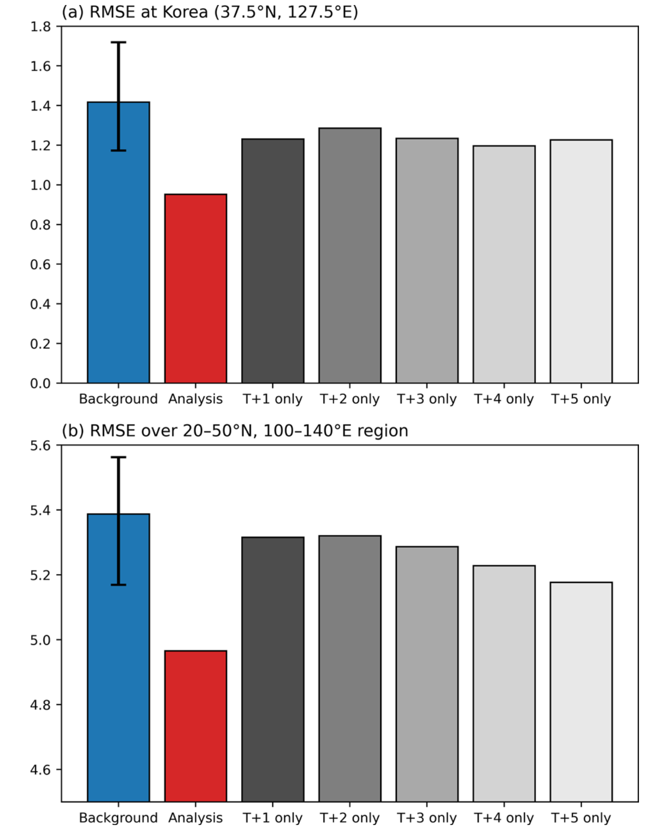

Fig. 7 | RMSE values from the idealized 4DVAR experiments using Anmyeondo observation at different future time steps. a RMSE at Korea (37.5° N, 127.5° E) for January–February 2020–2022. **b** Regional mean RMSE over 20–50° N, 100–140° E for January–February 2022. In both panels, the blue bar indicates the RMSE from the background states (before the 4DVAR experiment), and the red bar shows the RMSE of the analysis states (after the 4DVAR experiment) using all five observations from T + 1 to T + 5. The gray bars, ranging from dark to light, represent RMSEs after the idealized 4DVAR experiment using only a single observation at T + 1, T + 2, …, or T + 5, respectively. The error bar on the blue bar in (**a**) represents the 2.5th–97.5th percentile range derived from 1000 bootstrap samples over 36 cases, while in (**b**) it represents the corresponding range derived from 1000 bootstrap samples over 12 cases, corresponding to the 2022 dataset only.

ppbv, whereas MERRA-2 anomalies remained persistently negative (Fig. 5d). This implies that a clear decline in CO concentrations from 2020 and 2022 was evident in the MERRA-2 reanalysis[40,41], whereas this trend was less pronounced in the in-situ observations. Due to the mean biases and the distinct day-to-day variations, correlation coefficients between the two datasets were only 0.26 for 2020, 0.20 for 2021, and −0.21 for 2022.

In idealized 4DVAR experiment, the initial forecast state was derived from the MERRA-2 reanalysis, and the observation increment was defined as the difference between the observed daily CO concentration at Anmyeondo and the corresponding model forecast. It should be noted that, because MERRA-2 serves both as the training target for the DL-based forward model and as the initial condition for the idealized 4DVAR experiment, the model's original forecast field is expected to reflect the characteristics of the MERRA-2, therefore may inherently exhibit a systematic underestimation of CO concentrations compared to the in-situ measurement.

For each 5-day assimilation window, the initial model input state at $t + 0$ (specifically, the 985 hPa CO concentration at the beginning of the window) was optimized using adjoint sensitivities with respect to future observations from $t + 1$ to $t + 5$. After the optimization, we performed 5-day roll-out predictions initialized from both the original reanalysis state and the adjoint-optimized initial state, and the results are presented in Fig. 6. The idealized 4DVAR analysis was conducted independently for each 5-day period during January–February of 2020–2022. As each window was treated independently, the resulting time series does not represent a single continuous evolution of CO concentrations, but rather a collection of independent 5-day prediction segments.

Figure 6 shows the time series of the background states (blue), the analysis states (red), and the in-situ observation (orange) at the Anmyeondo station. Across all three years, the analysis time-series more faithfully reproduced the observed CO variations than the background states, particularly in capturing the timing and magnitude of peaks and troughs. Before assimilation, numerous high-concentration events—such as those in January 2020, mid-February 2021, and January and February 2022—were inadequately represented in the background states. However, after assimilation, the model reproduced these peaks with greater fidelity, indicating that the adjoint-based update effectively reflects observational information. In particular, largest improvement is shown in 2022 in which the difference between the background states and the observation was greatest. This suggests that when the model's initial state contains substantial bias—such as the systematic underestimation inherited from MERRA-2—the adjoint-based update becomes particularly effective. As a result, in 2022, the RMSE decreased from 1.94 (background) to 0.98 (analysis), corresponding to a 49.5% improvement. Similar reductions were shown in 2020 (from 1.23 to 0.97) and 2021 (from 0.9 to 0.75).

To highlight the role of the ADM in assimilating temporally lagged observations, we conducted sensitivity experiments in which observations at individual time lags were assimilated (Fig. 7). In each experiment, the initial

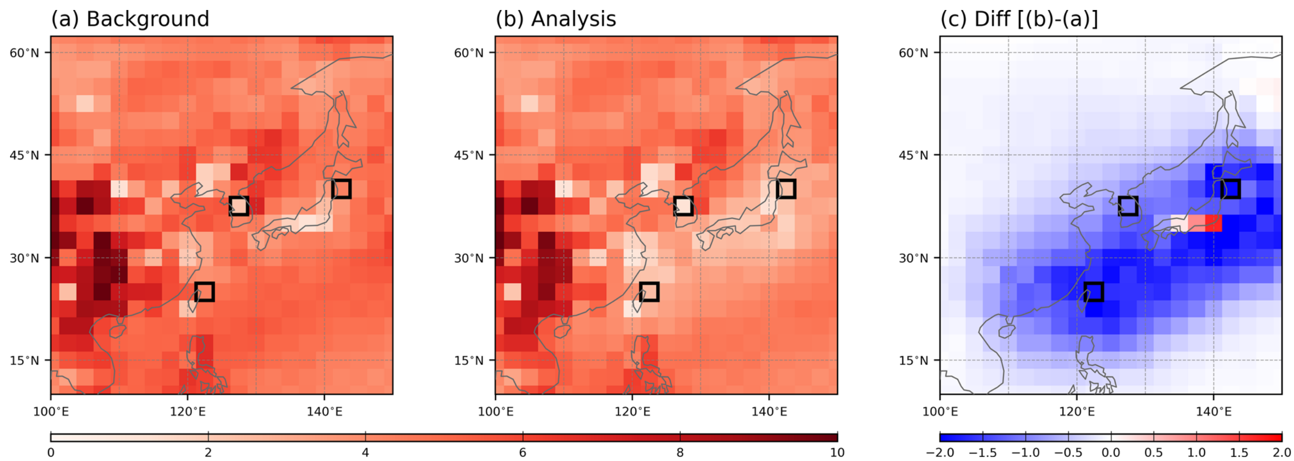

Fig. 8 | Spatial distributions of RMSE between model states and MOPITT satellite observations for January–February 2022. Panels (**a**) and (**b**) show the RMSE fields for the background states (before the 4DVAR experiment) and the analysis states (after the 4DVAR experiment), respectively, using in-situ observations from Anmyeondo, Ryori, and Yonagunijima. Panel (**c**) presents the RMSE difference (analysis minus background), with negative values indicating regions where the RMSE was reduced through the idealized 4DVAR experiment. Observation sites are marked by black squares in all panels.

state was updated using adjoint sensitivities computed solely with respect to a single time lag between t + 1 and t + 5, thereby isolating the influence of single time-lag observation from the others. The results show that observations at all time lags led to a decrease in RMSE compared to the background states over Korea (Fig. 7a). Observations with time lags more than one day also resulted in measurable error reductions. For example, using only the t + 4 or t + 5 observations reduced the RMSE by 15.6% and 13.5%, respectively. This indicates that the ADM effectively propagates observational information from the entire 5-day window back to the initial state, ensuring that even temporally distant data contributes to the optimization.

Further, to evaluate the spatial extent of the idealized 4DVAR impact, we compared the RMSE over the Northeast Asia (20–50° N, 100–140° E) of the results background and analysis states for January–February 2022 (Fig. 7b). As a reference data, we utilized independent satellite CO observations from MOPITT. The area-averaged RMSE averaged over the Northeast Asia confirmed error reductions for all time lags, with the improvements from t + 1 to t + 5 being more evident in the regional average than at the single point. Notably, assimilating t + 5 observation yielded a slightly larger RMSE reduction than assimilating t + 1 observation (t + 1: about 1.3% reduction, t + 5: about 3.9% reduction). This can be explained by the fact that, as the time lag increases, the adjoint sensitivity spreads farther upstream and covers a broader area, thereby updating the background states over a wider domain by reflecting the observational influence. Consequently, relatively larger improvements are manifested in the area-averaged RMSE. This finding indicates that the idealized 4DVAR framework can propagate observational influence backward in time over multiple days, allowing even relatively distant-future observations to enhance the present-day analysis states.

Figure 8 presents the spatial distribution of RMSE in 2022 between the MOPITT satellite observations, the background and the analysis states obtained from the idealized 4DVAR experiment, which was conducted using three ground-based observation sites (i.e., Anmyeondo, Ryori, and Yonagunijima). The results clearly show that RMSE reduction was not confined to the observed points but also extended to surrounding regions. Error reductions were also evident in areas located several hundred kilometers away from the observation sites, underscoring the spatial error propagation capability of the ADM. Note that this spatial influence arises solely from the ADM based on our simplified definition of the cost function. By comparing the analysis states obtained from multi-site ground-based observations with independent high-resolution satellite observations from MOPITT, this experiment demonstrates that integrated localized corrections can significantly improve regional prediction accuracy.

Discussion

This study developed a DL-based adjoint modeling framework for estimating three-dimensional CO concentrations by explicitly constructing the ADM of a DL-based forward model based on the U-Net architecture. Validation experiments using the TLM confirmed that the linear approximation sufficiently held for the wide amplitude ranges of the forecast errors. Furthermore, the additional validity of the ADM was verified through an inner product test using the ADM and TLM (i.e., adjoint test), as well as a FD gradient test.

The DL-based ADM experiments with synthetic CO increments over the Korean Peninsula showed that, the adjoint sensitivities with respect to the grid point over the mid-troposphere (600 hPa) spread over a broader area, extending into northern and western China, while the sensitivity with respect to the grid point over the surface layer (i.e., 985 hPa) persisted relatively longer around the Korean Peninsula. The vertical structure of adjoint sensitivities also varied with the vertical level of the observations: surface-level sensitivities were mainly confined to the near-surface layers, whereas mid- and upper-tropospheric sensitivities were distributed across a broader vertical range.

The idealized 4DVAR experiments using in situ CO observations from the Anmyeondo station during 2020–2022 successfully reduced RMSE at the observation site. Notably, when the experiments were extended to incorporate three observation sites—Anmyeondo, Ryori, and Yonagunijima—the assimilation successfully reduced RMSE not only at the observation points but also across their adjacent areas, such as the Korean Peninsula, Japan, and Eastern China. Further sensitivity experiments confirmed that observations with time lag clearly contributed to error reduction, demonstrating that the ADM can effectively propagate observational influence backward in time over multiple days to optimize the analysis states.

In conclusion, this study demonstrates that DL-based adjoint modeling can overcome the computational and methodological limitations of conventional CTM-based adjoint frameworks by providing physically interpretable gradients, computational efficiency, and scalability. By explicitly constructing the ADM of a DL-based forward model and exploiting the differentiable nature of neural networks with automatic differentiation, the DL-based ADM enables efficient gradient computation and sensitivity analysis within the 4DVAR framework. Unlike previous inverse modeling studies that relied heavily on adjoint versions of dynamical models such as GEOS-Chem or RAMS[15,17], which are computationally expensive and structurally constrained by parameterization schemes, the proposed framework avoids the need to manually linearize highly nonlinear processes. In this sense, the primary contribution of this study is providing an alternative and more flexible framework that enables adjoint sensitivity analysis and 4DVAR experiments without the need for manual linearization of intricate physical processes. Furthermore, the explicit construction of the ADM secures a level of interpretability that is often lacking in end-to-end gradient methods, making it possible to quantitatively diagnose how observational information propagates backward in time and space to update the analysis states. These advantages highlight the potential of DL-based adjoint modeling as a practical and effective tool for advancing inverse modeling and data assimilation in atmospheric sciences, with applications to emission estimation and climate prediction.

Present study adopts an idealized 4DVAR configuration in which the background term is not explicitly included in the cost function. While this choice allows for a more direct analysis of the ADM-driven correction behavior, incorporating background constraints would be necessary to more closely emulate operational data assimilation systems. With constraints by the assumed background error covariance, the analysis would be more strongly constrained toward the background states, leading to smaller-amplitude and spatially smoother correction increments, and the adjoint-driven sensitivities would be partially moderated.

Furthermore, our current framework does not yet explicitly incorporate meteorological processes such as synoptic-scale winds, vertical stability, and boundary-layer dynamics. However, one strength of the DL-based ADM approach is its flexibility: by simply extending the input variables, the framework can be further developed to account for the influence of these meteorological factors on adjoint sensitivities. This highlights the scalability of the method and its potential to quantitatively evaluate the impacts of the day-to-day variations of the additional physical variables, such as horizontal/vertical wind, pressure, vertical stability and so on, underscoring the power of the DL-based ADM framework for future atmospheric data assimilation and inverse modeling applications.

In this study, validation was conducted using a limited number of in situ observation sites over a three-year period. While this setup provided a clear testbed for demonstrating the feasibility of the proposed approach, it inevitably restricts the generality of the findings. Various observation platforms—such as aircraft campaigns, regional ground-based networks, or multiple satellite instruments—could offer independent perspectives on adjoint sensitivities and assimilation performance. The extension to such multi-platform observations can be effectively achieved by adopting a multi-modal deep learning framework. Recent studies in spatiotemporal modeling have demonstrated that integrating heterogeneous data sources—such as combining satellite imagery with point-based meteorological records—can significantly enhance predictive performance by capturing complex cross-modal synergies[42,43]. Following this approach, different observation types, such as satellite-based CO retrievals and in situ surface measurements, can be integrated as independent yet complementary terms within the cost function. This structure naturally supports various combinations of observations enabling a more comprehensive assimilation of the global and local atmospheric states.

Our approach can also be applied to other atmospheric variables, such as ozone (O_3), nitrogen dioxide (NO_2), or particulate matter ($PM_{2.5}$). Moreover, its applicability is not limited to gaseous species; the framework can be extended to the assimilation of general physical variables by augmenting the input layer. For instance, meteorological fields such as wind, temperature, and pressure can be included as additional input channels together with CO, enabling the forward model to better represent transport and mixing processes. In this configuration, the ADM can quantify how variations in these meteorological variables influence the distribution of atmospheric constituents. Furthermore, applying the method to emission estimation, regional air quality forecasting, and climate prediction would expand its utility.

Methods

Dataset

This study uses the Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2) dataset provided by NASA's Global Modeling and Assimilation Office (GMAO)[44]. The dataset spans from January 1, 1980 to December 31, 2023, and includes three-dimensional CO concentration fields with 72 vertical levels, a horizontal resolution of 0.5° latitude × 0.625° longitude, and a temporal resolution of 3 hours. In this study, we used daily values by taking the temporal average of the 3-hourly data and interpolating them onto a regular 2.5° × 2.5° grid using bilinear interpolation. Among the 72 model levels (in the hybrid sigma/pressure coordinate system), we selected ten specific levels — indices 72, 66, 60, 55, 53, 48, 46, 43, 39, and 33—which correspond approximately to pressures of 985, 895, 800, 700, 600, 525, 412, 337, 208, and 108 hPa respectively. These levels were chosen to represent the key vertical structure relevant to CO transport and observational sensitivity.

CO concentrations have relatively large absolute values, often exceeding several hundred ppb. However, commonly used activation functions in deep learning (e.g., Rectified Linear Unit (ReLU), Hyperbolic Tangent (Tanh), or Sigmoid Linear Unit (SiLU)) are most effective within a limited input range (e.g., -1 to 1 or 0 to 1). When provided with large, unnormalized inputs, these functions may become saturated or lead to vanishing gradients, thereby hindering model training. To address this, we subtracted the daily CO climatology from the raw data to remove the seasonal cycle and applied z-score normalization, shifting the mean of each grid point's time series to zero and scaling its variance to one. This preprocessing ensures that the model receives inputs with zero-centered distributions, enhancing numerical stability and training performance.

For the idealized 4DVAR experiments using the DL-based ADM, we utilized daily surface CO observations from the Anmyeondo atmospheric monitoring station (36.54°N, 126.33°E), operated by the Korea Meteorological Administration (KMA)[45]. The data period spanned from January 1, 2020 to February 28, 2022, while the 2023 data were excluded due to a substantial number of missing values. In addition to the Anmyeondo station, surface CO data from Ryori (39.03°N, 141.82°E) and Yonagunijima (24.47°N, 123.01°E) in Japan, operated by the Japan Meteorological Agency (JMA), were also utilized for the same period[46].

In addition, to independently validate the results before and after 4DVAR experiments, we used satellite-based CO retrievals from the Measurements of Pollution in the Troposphere (MOPITT) instrument[47]. We focused on surface-level CO retrievals between January 1 and February 28, 2022, and the data were regridded to a regular 2.5° × 2.5° latitude–longitude grid using distance-weighted remapping to match the spatial resolution of the model data. Since the observational period was relatively short, CO anomalies at both the Anmyeondo station and the MOPITT data were normalized by subtracting the 1980–2023 MERRA-2 climatology and then dividing by the MERRA-2 standard deviation at each grid point.

DL-based forward model using U-net

We constructed a DL-based forward model based on the U-Net architecture to predict the three-dimensional CO concentrations at subsequent day. The model takes the three-dimensional CO concentration anomaly field at time t as input (x_t) and predicts anomaly field at time t + 1 (denoted as $\hat{y}_t$). The input and output fields are defined on a 2.5° × 2.5° horizontal grid with 10 vertical levels consistent with the MERRA-2 reanalysis data described in Methods. For the model formulation, total period was divided into three subsets: 1980–2015 for training, 2016–2019 for validation, and 2020–2023 for testing. To better capture the specific characteristics of the winter season, all datasets used in this study were restricted to daily data for January and February of each year.

The forward model was trained to minimize the L2 loss (i.e., $J_t = \frac{1}{2}(y_t^{obs} - y_t)^2$), where y_t^{obs} is the reanalysis-based CO concentration anomaly field at time t + 1. The model consisted of three encoder and three decoder blocks, each comprising 3 × 3 convolutional layers with SiLU activation functions (Fig. 9a). Downsampling was performed via 2 × 2 max pooling layers, and upsampling was achieved using 3 × 3 transposed convolutions. Skip connections were applied between corresponding encoder and decoder levels to preserve spatial information. A bottleneck layer with 512 channels connected the deepest encoder and decoder blocks, and the final output layer used a 1 × 1 convolution to produce the predicted CO anomaly field with 10 vertical levels. The model was trained using the Adam optimizer with a learning rate of 10^{-3}. Note that the DL-based forward model is used not only for prediction but also serves as the foundation for

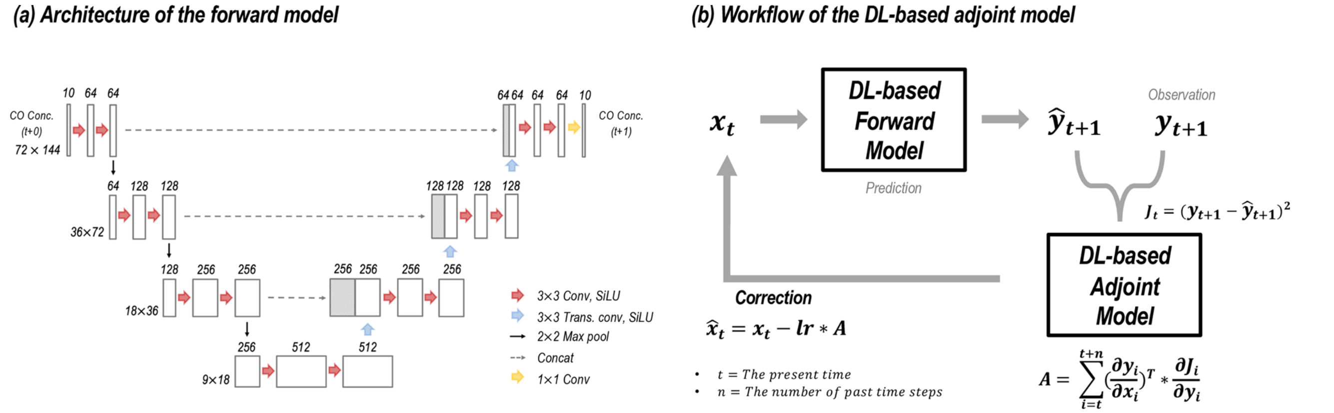

Fig. 9 | DL-based forward model and adjoint modeling framework. **a** Architecture of the DL-based forward model used in the adjoint system. The model follows a U-Net structure composed of convolutional blocks with SiLU activation, max-pooling layers for downsampling, transposed convolution layers for upsampling, and skip connections for feature preservation. The input and output are 3D CO anomaly fields with dimensions [Z = 10, Y = 72, X = 144]. **b** Workflow of the DL-based adjoint modeling system. The forward model predicts the CO concentration at time t + 1 based on the input at time t + 0, while the ADM computes the adjoint sensitivity—defined as the gradient of the cost function J with respect to the input field. The system then performs the state update by propagating the observational increments backward in time through the ADM *A*.

constructing the adjoint model by explicitly computing the Jacobian matrix and its transpose, which will be described in detail below.

DL-based adjoint model

To enable adjoint sensitivity analysis within the DL framework, we explicitly constructed the adjoint operator of the DL-based forward model described in Methods. Since the DL-based forward model is inherently nonlinear, its linearization is required to derive the ADM. This linearized representation is given by the Jacobian matrix $\frac{\partial y}{\partial x} \in \mathbb{R}^{m\times m}$, which describes the sensitivity of the model output $y \in \mathbb{R}^m$ (i.e., predicted CO anomaly at time t + 1) to the model input $x \in \mathbb{R}^m$ (i.e., CO anomaly at time t + 0). This Jacobian matrix serves as the tangent linear model (TLM), where m denotes the total number of grid points, corresponding to the dimensionality of the model state. The ADM is defined as the transpose of this Jacobian matrix, $\left(\frac{\partial y}{\partial x}\right)^{\mathrm{T}} \in \mathbb{R}^{m\times m}$. Therefore, while each column of the TLM denotes the sensitivity of all output locations to a perturbation at a specific input location, each column of the ADM represents the sensitivity of the output at a fixed location to perturbations across all input locations.

More specifically, this study leveraged automatic differentiation to explicitly compute the adjoint sensitivity $\frac{\partial J}{\partial x}$ using the chain rule:

$$\frac{\partial J}{\partial x} = \left(\frac{\partial y}{\partial x}\right)^{\mathrm{T}} \cdot \frac{\partial J}{\partial y} \tag{5}$$

where $()^{\mathrm{T}}$ denotes the transpose.

Cost function $J \in \mathbb{R}$ is a scalar value which is defined by the square summed discrepancy between the model output $y \in \mathbb{R}^m$ and the corresponding observations available at *nobs* locations (*nobs*≪*m*, where *nobs* is a number of observation points). The gradient of the cost function with respect to the output, $\frac{\partial J}{\partial y} \in \mathbb{R}^m$, has the same dimension as the model output but contains nonzero elements only at the observation locations.

Next, the sensitivity of each relevant output value with respect to the input x was computed, corresponding to the columns of the ADM $(\frac{\partial y}{\partial x})^{\mathrm{T}}$, which represent how variations in each input component influence the selected output components. By adding the multiplied values in each column in $(\frac{\partial y}{\partial x})^{\mathrm{T}}$ to the value at the corresponding location in $\frac{\partial J}{\partial y}$ for all observed locations, the contributions of all observations were aggregated to produce the total input gradient $\frac{\partial J}{\partial x}$. The resultant adjoint sensitivity, defined as the gradient of the cost function j with respect to the model input x, quantifies how small changes in the model input affect the cost function. Note that the resulting adjoint sensitivity was spatially smoothed using a Gaussian filter to reduce overly localized sensitivities and to promote spatial continuity. This step improves both the physical plausibility and numerical stability of the update.

Finally, the input is adjusted toward the observations according to the following equation:

$$\hat{x} = x - \alpha \cdot \frac{\partial J}{\partial x} \tag{6}$$

where α is the learning rate. Here, $\hat{x}$ denotes the updated input state after applying the adjoint-sensitivity-based update. The entire procedure was implemented using PyTorch.

Idealized experiments for the adjoint model validation

In this study, the performance of the adjoint model was evaluated through (1) sensitivity experiments and (2) idealized 4DVAR experiments.

In the sensitivity experiments, the observation was assumed to be located at the grid cell over the Korean Peninsula (127.5°E, 37.5°N). A synthetic observation at time t + 1 was to compute the adjoint sensitivities with respect to the preceding time steps (t, t-1, t-2, t-3, …).

As illustrated in Fig. 9b, the NLM $M()$ generates a model prediction with the given input x_t (i.e., $M(x_t) = y_t$), which is then compared with the synthetic observation to define the cost function:

$$J = \frac{1}{2}\left(y_t^{obs} - M(x_t)\right)^2 = \frac{1}{2}\left(y_t^{obs} - y_t\right)^2 \tag{7}$$

The gradient of the cost function by given observation y_t^{obs} with respect to the input state is computed using the ADM. Following the chain rule, the gradient can be expressed as:

$$\left.\frac{\partial J}{\partial x_t}\right|_{y_t} = \left(\frac{\partial y_t}{\partial x_t}\right)^T \cdot \frac{\partial J}{\partial y_t} = \mathbf{M}_t{}^T \cdot \left(y_t^{obs} - M(x_t)\right) \tag{8}$$

where $\mathbf{M}_t{}^T$ denotes the ADM at time t. That is, for the 4DVAR operator, adjoint sensitivity (i.e., matrix multiplication of the ADM $\left(\frac{\partial y_t}{\partial x_t}\right)^T$ and the observational error increment $\frac{\partial J}{\partial y_t}$) is required.

The simplest situation to understand the adjoint sensitivity $\mathbf{M}_t{}^T$ is to define an observation increment $(y_t^{obs} - M(x_t)$, the difference between the observation and the model prediction) at a single location. Once the observation increment amplitude was assumed as 0.01 for the single target location, the spatial pattern of the adjoint sensitivities was determined by the following equation:

$$\hat{x}_t = x_t - \alpha \cdot 0.01 \cdot \mathbf{M}_t{}^T[:, j] \tag{9}$$

Where $\mathbf{M}_t{}^T[i, j]$ denotes the value of the ADM at i^{th} row, j^{th} column, and j is a column corresponds to the target location. and the learning rate α, which is set to 10^{-2} in this study.

The general form of the adjoint sensitivity accumulated for observations for multiple time frames from time t to t + k can be written by using extended chain rule formulation as follows:

$$\left.\frac{\partial J}{\partial x_t}\right|_{y_t, y_{t+1}, \ldots, y_{t+K}} = \sum_{k=1}^{K}[(\prod_{j=k}^{1} \mathrm{M}_{t+j}{}^T) \cdot \frac{\partial J}{\partial y_{t+k}}] \tag{10}$$

k indicates the number of steps ahead the observation is located relative to t. Index j traces backward propagation path in time from t + k to t, as $\mathbf{M}_{t+j}{}^T$ represents the gradient to propagate the cost information from time t + j to time t + j-1.

Idealized 4DVAR experiments were performed by using the DL-based forward and ADM to trace the mismatches between the observation and the model forecasts backward in time, thereby optimizing the input CO values to reduce those mismatches. The 4DVAR cost function in this study omitted terms associated with the background state to ensure that updates were solely driven by the ADM. While omitting the background term carries the risk of over-correction and subsequent systematic degradation of performance, this setup was intentionally adopted to isolate and rigorously evaluate the sensitivity of the DL-based ADM itself, without interference from background constraints. Consequently, the spatial pattern of the analysis was designed to purely reflect the spatial distribution of the adjoint sensitivity. Accordingly, the 4DVAR cost function was simplified as follows.

$$J(x_t) = \frac{1}{2}\sum_{k=1}^{5} \frac{1}{\sigma^2_{obs,\,k}} \left(y^{obs}_{t+k} - M_{t+k-1}(x_t)\right)^2 \tag{11}$$

where y^{obs}_{t+k} is the observed CO anomaly and $M_{t+k-1}(x_t)$ is the model-predicted CO anomaly at 985 hPa over the Korean Peninsula grid at time $t + k$. CO observations from five consecutive days (t + 1 to t + 5) were used to update the model input at time t.

To assimilate independent surface CO observations from three monitoring stations, Anmyeondo, Ryori, and Yonagunijima, each measurement was mapped to their nearest model grid point (37.5 °N, 127.5 °E, 985 hPa for Anmyeondo), (40 °N, 142.5 °E, 985 hPa for Ryori), and (25 °N, 122.5 °E, 985 hPa for Yonagunijima). For the vertical placement, the effective observation altitude of each station (i.e., geological altitude plus observation tower height) was converted to the nearest pressure level (hPa) based on the hydrostatic pressure–height relationship.

The accuracy of the assimilated states is evaluated by using MOPITT dataset as a reference dataset. The assimilation window is set to 5 days, that the gradient of this cost function with respect to the model input at time t was computed using the DL-based ADM, incorporating the contributions from all five observation times via the chain rule (Eq. 10). Optimization was performed over 200 epochs. The learning rate was initially set to 0.5 and decayed to 10^{-5} using a cosine annealing scheduler. Other hyperparameters, including the optimizer, were kept consistent with those used in the idealized sensitivity experiments described earlier.

The standard deviation of the observational error $\sigma_{obs,k}$ was defined by following the specification of the in-situ measurements (i.e., Thermo Scientific Model 60i CO analyzer). According to the instrument's performance documentation, the measurement accuracy is defined as the smaller of 2% of full scale (2,500 ppm) or 5% of the measured value. Therefore, the error was computed as:

$$\sigma_{obs,k} = \min\left(0.02 \times 2500,\ 0.05 \times y^{obs}_{t+k}\right) = \min\left(50,\ 0.05 \times y^{obs}_{t+k}\right) \tag{12}$$

Note that the observational error variance was specified in the cost function after the z-score normalization by dividing the MERRA-2-based CO anomaly variance. This procedure ensures that the observation term was formulated in a dimensionless space consistent with the z-score-normalized state variables.

In both the sensitivity experiment and the idealized 4DVAR experiment, once the spatial distribution of the adjoint sensitivity to any given observed location is calculated, we applied filtering during the backward propagation of adjoint gradients. That is, for each observed location, the adjoint sensitivity field $\frac{\partial J}{\partial x}$ for each vertical level z was normalized across the horizontal domain, and sensitivity with respect to the selected grid points were retained. That is, when the adjoint gradient at level z and grid point (y, x) is denoted as $g_{z,y,x} = \frac{\partial J}{\partial x}{}_{z,y,x}$, for each vertical slice $g_z = (g_{z,y,x})$, we computed the mean and standard deviation and calculated the corresponding z-score at each grid point. Then, grid points satisfying $\left|g'_{z,y,x}\right| > 1$ — where $g'_{z,y,x}$ denotes the standardized adjoint sensitivity —were selected for further backward gradient propagation. This thresholding ensures that only dynamically meaningful regions—where the sensitivity differs notably from the background—are included in the computation, while noisy or weak-signal regions are excluded.

Data availability

MERRA-2 reanalysis data are publicly accessible via the NASA Global Modeling and Assimilation Office (GMAO) data portal (https://gmao.gsfc.nasa.gov/gmao-products/merra-2/data-access_merra-2/). Ground-based CO observations at Anmyeondo station were obtained from the Korea Meteorological Administration (https://data.kma.go.kr/). CO observations at Ryori and Yonagunijima were obtained from the Japan Meteorological Agency (https://www.data.jma.go.jp/env/ghg_obs/en/station/). MOPITT CO satellite products are available via the NASA Earthdata platform (https://search.earthdata.nasa.gov/).

Received: 13 November 2025; Accepted: 31 March 2026;
Published online: 22 April 2026

References

1. Bierhals, J. Carbon monoxide. In *Ullmann's Encyclopedia of Industrial Chemistry.* Vol. 6, 679–693 (2001).
2. Cordero, P. R. F. et al. Atmospheric carbon monoxide oxidation is a widespread mechanism supporting microbial survival. *ISME J.* **13**, 2868–2881 (2019).
3. Long, J. et al. Temporal trends of carbon monoxide poisoning mortality at the global, regional and national levels: a cross-sectional study from the Global Burden of Disease study, 1990 and 2017. *BMJ Open* **11**, e053240 (2021).
4. Tanimoto, H. et al. Diagnosing recent CO emissions and ozone evolution in East Asia using coordinated surface observations, adjoint inverse modeling, and MOPITT satellite data. *Atmos. Chem. Phys.* **8**, 3867–3880 (2008).
5. Buchholz, R. R. et al. Air pollution trends measured from Terra: CO and AOD over industrial, fire-prone, and background regions. *Remote Sens. Environ.* **256**, 112275 (2021).
6. Liang, Q. et al. Long-range transport of Asian pollution to the northeast Pacific: Seasonal variations and transport pathways of carbon monoxide. *J. Geophys. Res. Atmos.* **109**, D23 (2004).
7. Kim, H. S., Tans, P. P. & Novelli, P. C. On the regional background levels of carbon monoxide observed in East Asia during 1991~ 2004. *Air Qual. Atmos. Health* **1**, 37–44 (2008).
8. Turnbull, J. C. et al. Atmospheric observations of carbon monoxide and fossil fuel CO2 emissions from East Asia. *J. Geophys. Res. Atmos.* **116**, D24 (2011).
9. Deeter, M. N. et al. A climate-scale satellite record for carbon monoxide: the MOPITT Version 7 product. *Atmos. Meas. Tech.* **10**, 2533–2555 (2017).
10. McMillan, W. W. et al. Daily global maps of carbon monoxide from NASA's Atmospheric Infrared Sounder. *Geophys. Res. Lett.* **32**, L11801 (2005).
11. Rabier, F. et al. The ECMWF operational implementation of four-dimensional variational assimilation. I: Experimental results with simplified physics. *Q. J. R. Meteorol. Soc.* **126**, 1143–1170 (2000).

12. Elbern, H. et al. Emission rate and chemical state estimation by 4-dimensional variational inversion. *Atmos. Chem. Phys.* **7**, 3749–3769 (2007).
13. Chai, T. et al. Four-dimensional data assimilation experiments with International Consortium for Atmospheric Research on Transport and Transformation ozone measurements. *J. Geophys. Res. Atmos.* **112**, D12 (2007).
14. Bannister, R. N. A review of operational methods of variational and ensemble-variational data assimilation. *Q. J. R. Meteorol. Soc.* **143**, 607–633 (2017).
15. Yumimoto, K. & Uno, I. Adjoint inverse modeling of CO emissions over Eastern Asia using four-dimensional variational data assimilation. *Atmos. Environ.* **40**, 6836–6845 (2006).
16. Henze, D. K., Hakami, A. & Seinfeld, J. H. Development of the adjoint of GEOS. *Chem. Atmos. Chem. Phys.* **7**, 2413–2433 (2007).
17. Kopacz, M. et al. Comparison of adjoint and analytical Bayesian inversion methods for constraining Asian sources of carbon monoxide using satellite (MOPITT) measurements of CO columns. *J. Geophys. Res. Atmos.* **114**, D4 (2009).
18. Kopacz, M. et al. Global estimates of CO sources with high resolution by adjoint inversion of multiple satellite datasets (MOPITT, AIRS, SCIAMACHY, TES). *Atmos. Chem. Phys.* **10**, 855–876 (2010).
19. Tang, Z. et al. The capabilities of the adjoint of GEOS-Chem model to support HEMCO emission inventories and MERRA-2 meteorological data. *Geosci. Model Dev.* **16**, 6377–6392 (2023).
20. Jiang, Z. et al. Regional data assimilation of multi-spectral MOPITT observations of CO over North America. *Atmos. Chem. Phys.* **15**, 6801–6814 (2015).
21. Moon, J. et al. Hybrid IFDMB/4D-Var inverse modeling to constrain the spatiotemporal distribution of CO and NO2 emissions using the CMAQ adjoint model. *Atmos. Environ.* **327**, 120490 (2024).
22. Errico, R. M. What is an adjoint model? *Bull. Am. Meteorol. Soc.* **78**, 2577–2592 (1997).
23. Rabier, F., Thépaut, J.-N. & Courtier, P. Extended assimilation and forecast experiments with a four-dimensional variational assimilation system. *Q. J. R. Meteorol. Soc.* **124**, 1861–1887 (1998).
24. Janisková, M. Physical processes in adjoint models: potential pitfalls and benefits. *Proc. ECMWF Semin. Recent Dev. Data Assim. Atmos. Ocean* **215**, 233 (2003).
25. Bonavita, M., Lean, P. & Holm, E. Nonlinear effects in 4D-Var. *Nonlinear Process Geophys.* **25**, 713–729 (2018).
26. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
27. Paszke, A. et al. PyTorch: An imperative style, high-performance deep learning library. *Adv. Neural Inf. Process. Syst.* **32**, 8024–8035 (2019).
28. Wani, M. A. et al. *Advances in Deep Learning*, *Volume 2* Vol. 12 (Springer Nature, 2025).
29. Hatfield, S. et al. Building tangent-linear and adjoint models for data assimilation with neural networks. *J. Adv. Model. Earth Syst.* **13**, e2021MS002521 (2021).
30. Xiao, Y. et al. Fengwu-4dvar: Coupling the data-driven weather forecasting model with 4d variational assimilation. Preprint at https://arxiv.org/abs/2312.12455 (2023).
31. Ronneberger, O., Fischer, P. & Brox, T. U-net: Convolutional networks for biomedical image segmentation. in *Int. Conf. Med. Image Comput. Comput.-Assist. Interv*. 234–241 (Springer, 2015).
32. Kaur, J., Parmar, K. S. & Singh, S. Autoregressive models in environmental forecasting time series: a theoretical and application review. *Environ. Sci. Pollut. Res.* **30**, 19617–19641 (2023).
33. Mu, M., Guo, H., Wang, J. & Li, Y. The impact of nonlinear stability and instability on the validity of the tangent linear model. *Adv. Atmos. Sci.* **17**, 375–390 (2000).
34. Chrust, M. Tangent linear and adjoint models for variational data assimilation. *ECMWF Training Course* (2020).
35. Blessing, S. *Adjoint Modelling for Assimilation and Diagnosis on Climate Timescales* (PhD thesis, Univ. Hamburg, 2008).
36. Elbern, H. & Schmidt, H. A four-dimensional variational chemistry data assimilation scheme for Eulerian chemistry transport modeling. *J. Geophys. Res. Atmos.* **104**, 18583–18598 (1999).
37. Heald, C. L. et al. Asian outflow and trans-Pacific transport of carbon monoxide and ozone pollution: An integrated satellite, aircraft, and model perspective. *J. Geophys. Res. Atmos.* **108**, 4804 (2003).
38. Liu, H. et al. Transport pathways for Asian pollution outflow over the Pacific: Interannual and seasonal variations. *J. Geophys. Res. Atmos.* **108**, 4646 (2003).
39. Stein, A. F. et al. NOAA's HYSPLIT atmospheric transport and dispersion modeling system. *Bull. Am. Meteorol. Soc.* **96**, 2059–2077 (2015).
40. Worden, H. M. et al. Decadal record of satellite carbon monoxide observations. *Atmos. Chem. Phys.* **13**, 837–850 (2013).
41. Zheng, B. et al. Trends in China's anthropogenic emissions since 2010 as the consequence of clean air actions. *Atmos. Chem. Phys.* **18**, 14095–14111 (2018).
42. Benson, V. et al. Multi-modal learning for geospatial vegetation forecasting. in *Proc*. *IEEE/CVF Conf. Comput. Vis. Pattern Recognit*. (2024).
43. Lilhore, U. K. et al. Advanced air quality prediction using multimodal data and dynamic modeling techniques. *Sci. Rep.* **15**, 27867 (2025).
44. Pawson, S. MERRA-2 inst3_3d_chm_Nv: 3d, 3-Hourly, Instantaneous, Model-Level, Assimilation, Carbon Monoxide and Ozone Mixing Ratio V5.12.4. NASA GSFC DISC HO9OVZWF3KW2 (2015).
45. Lee, H. et al. Observations of atmospheric 14CO2 at Anmyeondo GAW station, South Korea: implications for fossil fuel CO2 and emission ratios. *Atmos. Chem. Phys.* **20**, 12033–12045 (2020).
46. Wada, A. et al. Seasonal variation of enhancement ratios of trace gases observed over 10 years in the western North Pacific. *Atmos. Environ.* **45**, 2129–2137 (2011).
47. Deeter, M. et al. The MOPITT Version 9 CO product: sampling enhancements and validation. *Atmos. Meas. Tech.* **15**, 2325–2344 (2022).

Acknowledgements

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (RS-2024-00442228) and by the Korea Environmental Industry & Technology Institute (KEITI) through the "Project for developing an observation-based GHG emissions geospatial information map," funded by the Korea Ministry of Environment (MOE) (Grant No. RS-2023-00232066).

Author contributions

H.-J.L. and Y.-G.H. contributed to methodology and formal analysis and drafted the manuscript. Y.-G.H. designed the study, supervised the project, and acquired funding. H.-J.L. contributed to data collection and curation. All authors discussed the results and reviewed and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Yoo-Geun Ham.