https://doi.org/10.5194/egusphere-2026-4627
Preprint. Discussion started: 14 September 2026

LARDAS v1.0: a latent-space aerosol–radiance data assimilation system for joint aerosol–surface constraints from satellite shortwave observations

Chongzhao Zhang[1,#], Qiurui Li[1,#], Jun Li[2], Wei Han[3,4], Jing Li[1,5,6,*]
[1] Department of Atmospheric and Oceanic Sciences, School of Physics, Peking University, Beijing, 100871, China
[2] National Satellite Meteorological Center, China Meteorological Administration, Beijing 100081, China.
[3] CMA Earth System Modeling and Prediction Centre (CEMC), China Meteorological Administration, Beijing 100081, China.
[4] State Key Laboratory of Severe Weather Meteorological Science and Technology (LaSW), Beijing 100081, China.
[5] Institute of Carbon Neutrality, Peking University, Beijing 100871, China.
[6] Center for Environment and Health, Peking University, Beijing 100871, China.
[#]These authors contribute equally to this article.

Correspondence to: Jing Li (jing-li@pku.edu.cn)

Abstract. Satellite aerosol data assimilation has traditionally relied on retrieved aerosol products, introducing a "retrieve-then-assimilate" pathway that can propagate retrieval uncertainties and impose assumptions inconsistent with atmospheric models. Direct assimilation of satellite shortwave observations provides a more physically consistent alternative by constraining model states directly in observation space, but its application remains challenging because nonlinear radiative transfer calculations and high-dimensional state optimization are computationally demanding. Here we develop LARDAS (Latent-space Aerosol–Radiance Data Assimilation System), a differentiable latent-space framework that directly assimilates multi-band satellite shortwave reflectance to jointly constrain aerosol optical depth (AOD) and land-surface white-sky albedo. LARDAS combines variational autoencoder-based latent representations of aerosol and surface fields with a differentiable neural radiative transfer emulator, enabling gradient-based optimization from observation space to atmospheric state space. Five-band MODIS top-of-atmosphere reflectances are assimilated through an integrated observation operator consisting of latent decoders, an aerosol refinement network, and band-specific neural radiative transfer surrogates trained against VLIDORT. The proposed framework reduces the control space from approximately 2×5×141×141 physical variables to 256 latent variables while maintaining high reconstruction fidelity (median fraction skill scores >0.95 for aerosol fields and >0.98 for surface albedo). The neural radiative transfer emulator reproduces VLIDORT reflectance simulations with correlations exceeding 0.999 and RMSE below 0.003. Controlled experiments demonstrate that LARDAS can recover aerosol and surface states from multi-band reflectance observations. Real-data assimilation experiments using MODIS observations over eastern China show that LARDAS improves independent AERONET AOD evaluation relative to the model background and aerosol-product assimilation, while achieving comparable skill to physical-space radiance assimilation at approximately three orders of magnitude lower online computational cost under the operationally feasible CPU/GPU deployment configurations used here. Independent CERES validation further demonstrates that joint aerosol–surface optimization improves simulated shortwave radiation, particularly surface upward shortwave flux, where reflectance

https://doi.org/10.5194/egusphere-2026-4627
Preprint. Discussion started: 14 September 2026

assimilation substantially reduces biases compared with aerosol-only assimilation. These results demonstrate that latent-space observation-to-state inference provides an efficient and physically consistent pathway toward next-generation satellite aerosol assimilation systems.

Keywords: Latent space Data Assimilation, Shortwave Radiance Assimilation, Aerosol Data Assimilation, Surrogate Radiative Transfer Model, MODIS, WRF-Chem

1 Introduction

Atmospheric aerosols play important roles in air quality, climate forcing, and Earth's radiation budget through their interactions with radiation, clouds, and atmospheric processes. (Li et al., 2022). However, large uncertainties remain in aerosol simulations due to uncertainties in emissions, chemical transformation, transport, deposition, and aerosol optical properties, which limit the capability of chemical transport models to reproduce observed aerosol variability and quantify aerosol radiative effects (Carslaw et al., 2013; Kahn et al., 2023). Data assimilation (DA), which combines model simulations with observations within a statistical estimation framework, provides an effective approach for improving aerosol analyses and forecasts (Evensen, 1994; Zhang et al., 2008).

Satellite observations are particularly valuable for aerosol data assimilation because of their broad spatial coverage and frequent sampling (Kaufman et al., 1997; Levy et al., 2013). Most existing aerosol assimilation systems rely on retrieved aerosol products, especially aerosol optical depth (AOD), as observational constraints (Adhikary et al., 2008; Cheng et al., 2019; Garrigues et al., 2023; Zhang et al., 2008). This approach has been successfully applied in regional and global aerosol prediction systems using variational and ensemble-based assimilation methods. (Sandu & Chai, 2011). However, retrieved aerosol products are themselves derived from inverse retrieval algorithms and depend on assumptions regarding aerosol properties, surface reflectance, and radiative transfer processes (Levy et al., 2013). These assumptions may differ from those used in atmospheric models, leading to inconsistencies between observation and model spaces (Benedetti et al., 2018; Rodgers, 2000). In addition, retrieval uncertainties are difficult to fully characterize and may introduce additional errors before assimilation (Levy et al., 2013; Sayer et al., 2013; Schutgens et al., 2020). Therefore, the conventional "retrieve-then-assimilate" pathway may limit the information that satellite observations can provide for aerosol state estimation.

Direct assimilation of satellite radiances or reflectances provides an alternative approach by avoiding intermediate retrieval products and constraining model states directly in observation space (Weaver et al., 2007; Xu et al., 2013; Zhang et al., 2026). By linking atmospheric states with satellite measurements through radiative transfer models, this approach provides a more physically consistent framework for aerosol estimation and offers the possibility of jointly constraining aerosol and surface properties (Zhang et al., 2026). Previous studies have demonstrated the feasibility of assimilating satellite radiances into aerosol transport models (Weaver et al., 2007; Xu et al., 2013; Zhang et al., 2026). Nevertheless, direct observation-

https://doi.org/10.5194/egusphere-2026-4627
Preprint. Discussion started: 14 September 2026

space assimilation remains computationally challenging. Radiative transfer calculations are highly nonlinear and expensive when repeatedly evaluated during iterative optimization (Kokhanovsky & Leeuw, 2009; Spurr, 2006) . Moreover, aerosol and surface states are usually represented on high-resolution spatial grids, resulting in control vectors with hundreds of thousands of variables and substantial computational costs (Zhang et al., 2026).

Recent advances in machine learning provide new opportunities to address these challenges. Latent-space data assimilation (LSDA) reduces the dimensionality of complex geophysical systems by representing high-dimensional physical fields using compact latent variables learned from historical data. Assimilation can then be performed in this reduced latent space while maintaining important physical structures encoded by the learned representation (Cheng et al., 2023). Previous studies have demonstrated the potential of latent-space approaches in atmospheric dynamics, ocean modeling, and environmental prediction(Amendola et al., 2021; Fan et al., 2026; Fan et al., 2025; Mohd Razak et al., 2022; Mücke et al., 2024; Peyron et al., 2021; Zheng et al., 2026; Zhong et al., 2023). However, applications of LSDA to atmospheric composition remain limited. In particular, the integration of latent representations with differentiable radiative transfer operators for direct satellite aerosol assimilation has not been fully explored.

Here we develop LARDAS (Latent-space Aerosol–Radiance Data Assimilation System), a retrieval-product-independent latent-space variational assimilation framework that directly constrains aerosol and surface states from satellite shortwave observations. Unlike conventional aerosol-product assimilation, LARDAS assimilates top-of-atmosphere (TOA) reflectance and performs aerosol–surface estimation through a differentiable observation operator. The framework combines (1) variational autoencoder-based latent representations of multi-band aerosol optical depth and surface white-sky albedo, (2) a neural radiative transfer emulator trained against the Vector Linearized Discrete Ordinate Radiative Transfer model (VLIDORT), (3) gradient-based optimization of latent variables through the complete differentiable forward pathway and (4) an analysis-to-model update that maps the optimized aerosol and surface fields back to WRF-Chem restart variables for cycling forecasts.

This study aims to answer three key questions: (1) Can latent representations preserve sufficient aerosol and surface information for effective assimilation while substantially reducing the dimensionality of the optimization problem? (2) Can a neural radiative transfer emulator provide an accurate and efficient observation operator for direct satellite reflectance assimilation? (3) Can latent-space reflectance assimilation improve aerosol analyses and shortwave radiation simulations compared with aerosol-product assimilation? To address these questions, LARDAS is evaluated through three stages. First, the quality of latent representations and the neural radiative transfer emulator are assessed using independent reconstruction and forward-model validation experiments. Second, controlled experiments with known aerosol and surface states are conducted to evaluate the ability of the system to recover prescribed conditions. Third, real MODIS shortwave reflectance observations are assimilated over eastern China and evaluated using independent AERONET aerosol optical depth

https://doi.org/10.5194/egusphere-2026-4627
Preprint. Discussion started: 14 September 2026

measurements and CERES shortwave radiation flux observations. The results demonstrate the potential of latent-space observation-to-state inference as an efficient and physically consistent framework for next-generation satellite aerosol assimilation.

2 Data and methods

This section describes LARDAS as an end-to-end cycling aerosol–radiance data assimilation system. Hereafter, LSDA refers specifically to the latent-space analysis algorithm, whereas LARDAS denotes the complete system linking offline representation learning, neural radiative-transfer emulation, satellite-observation processing, latent-space analysis, model-variable adjustment, forecast cycling, and independent evaluation (Figure 1). Section 2.1 provides a system-level overview. Section 2.2 describes the training of the aerosol and surface representations and the estimation of latent background statistics. Section 2.3 presents the training and implementation of the neural radiative-transfer observation operator. Section 2.4 describes the MODIS reflectance observations and the independent AERONET and CERES evaluation datasets. Section 2.5 specifies the analysis-error assumptions, cycling procedure, model-state update, and experimental configurations.

2.1 Overview of LARDAS

LARDAS integrates two offline learning components with an online cycling analysis–forecast system (Figure 1). Offline, separate variational autoencoders learn compact representations of multi-band aerosol optical depth (AOD) and surface albedo (ALB) (Figure 1a), while a neural radiative-transfer model is trained against VLIDORT to emulate five-band top-of-atmosphere (TOA) reflectance (Figure 1b). Online, WRF-Chem aerosol fields and band-resolved surface albedo provide the background state, and quality-controlled MODIS reflectances provide the observational constraint. The latent-space analysis jointly updates AOD and ALB through the differentiable decoder–radiative-transfer pathway. The resulting aerosol and surface analyses are subsequently mapped back to WRF-Chem prognostic and restart variables, allowing the observational information to propagate through the following forecast cycle (Figure 1c).

https://doi.org/10.5194/egusphere-2026-4627
Preprint. Discussion started: 14 September 2026

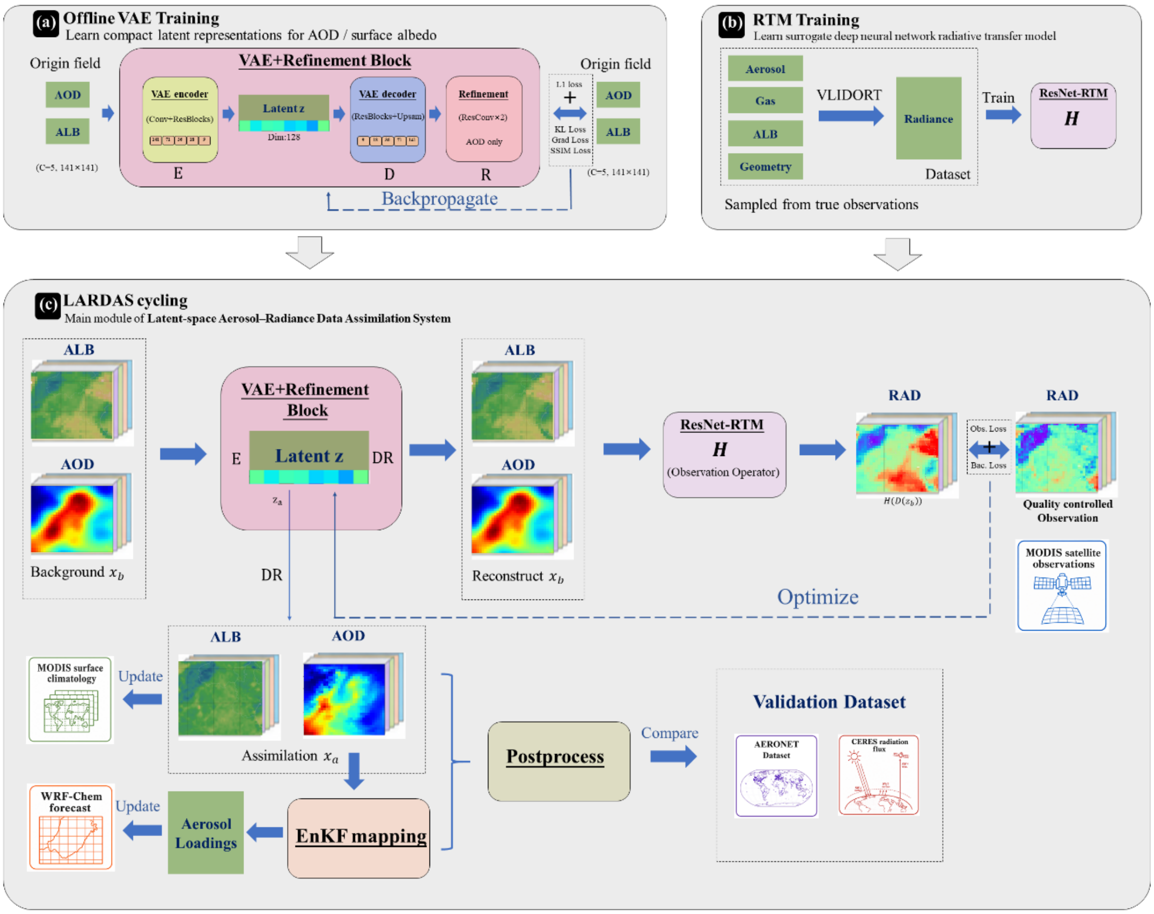

Figure 1: Architecture of the Latent-space Aerosol–Radiance Data Assimilation System (LARDAS). (a) Offline representation learning. Separate variational autoencoders encode five-band AOD and surface-albedo fields into 128-dimensional latent variables and reconstruct the physical fields; an aerosol-only refinement block restores fine-scale AOD structure. (b) Offline radiative-transfer-model training. Aerosol, gaseous, surface, and viewing-geometry inputs are propagated through VLIDORT to construct the reflectance training archive, which is used to train the band-wise differentiable ResNet-RTM observation operator. (c) Online cycling assimilation. Background AOD and ALB are encoded into latent variables, decoded, and mapped to simulated TOA reflectance through the integrated observation operator. The latent variables are optimized by minimizing the background and MODIS-reflectance misfit terms. The optimized latent state is decoded to obtain multi-band AOD and ALB analyses. Analyzed AOD is mapped to MOSAIC aerosol mass and number concentrations through an EnKF-based mapping, while analyzed surface albedo updates the band-resolved surface fields. The updated model state initializes the subsequent WRF-Chem forecast. Postprocessed aerosol and radiation fields are evaluated independently against AERONET and CERES; these validation datasets are not assimilated.

State variables and notation. We consider two physical fields defined on a fixed regional grid of size $H \times W = 141 \times 141$ and five shortwave bands (C = 5) centered near 469, 555,645, 858, and 2100 nm. The aerosol state is the multi-band aerosol optical depth (AOD) field $\boldsymbol{x}_a = (\boldsymbol{x}_{a,1}, \cdots, \boldsymbol{x}_{a,5}) \in \mathbb{R}^{C \times H \times W}$, and the surface state is the multi-band surface

https://doi.org/10.5194/egusphere-2026-4627
Preprint. Discussion started: 14 September 2026

albedo field $\boldsymbol{x_s} = (\boldsymbol{x_{s,1}}, \cdots, \boldsymbol{x_{s,5}}) \in \mathbb{R}^{C\times H\times W}$. Background fields are denoted x_a^b and x_s^b, while the analysis fields are x_a^a and x_s^a. The observation vector $y = (y_1, \ldots, y_5)$ is the gridded multi-band TOA shortwave reflectance after quality control (QC).

(a) Offline latent representation learning. For each field we train a VAE that defines an encoder $E(\cdot)$ and a decoder $D(\cdot)$ (Kingma & Welling, 2014; Rezende et al., 2014). The encoder maps a physical field to the parameters of a diagonal Gaussian posterior over the latent variable, $q(z \mid x) = \mathcal{N}(\mu(x), \sigma^2(x))$, and the latent state is obtained as $z = E(x)$ with dimension $d = 128$. The decoder reconstructs the field, $\hat{x} = D(z)$. Two independent VAEs are trained, one for the aerosol field and one for the surface field, giving the encoders E_a, E_s and decoders D_a, D_s. For the aerosol field, reconstruction based solely on the decoder tends to smooth fine-scale spatial structures. To mitigate this limitation, we introduce an additional refinement block (R) that restores local spatial variability and improves the reconstruction of detailed AOD patterns:

$$z_a = E_a(x_a), \quad z_s = E_s(x_s), \quad \widehat{x_a} = R(D_a(z_a)), \quad \widehat{x_s} = D_s(z_s) \tag{1}$$

with $z_a \in \mathbb{R}^{d_a}$ and $z_s \in \mathbb{R}^{d_s}$ $(d_a = d_s = 128)$. The AOD and surface albedo fields are encoded separately using independent VAEs. This design simplifies the latent representation and allows the aerosol and surface components to be modeled with dedicated networks tailored to their distinct spatial characteristics and physical constraints. The network architecture, the physical constraints imposed at the decoding stage (AOD non-negativity; albedo bounded to [0, 1]), the aerosol-only refinement block R, and the training objective are described in Section 2.2.

(b) Integrated, differentiable observation operator. To assimilate reflectance directly, the latent variables are mapped to observation space through an integrated operator that combines the decoders with a fast surrogate shortwave radiative transfer model (RTM). The surrogate is trained per band, so that for band λ the forward mapping reads

$$\widehat{y_\lambda} = H_\lambda(\widehat{x_a}, \widehat{x_s}) = H_\lambda\big(R_a D_a(z_a), D_s(z_s)\big) =: \mathcal{H}_\lambda(z_a, z_s) \tag{2}$$

where $H_\lambda(\cdot)$ is the band-specific neural-network surrogate RTM (denoted ResNet-RTM in Figure 1b). Stacking the bands yields the full operator $\mathcal{H}(z_a, z_s) \in \mathbb{R}^{5\times H\times W}$. Because every component (decoders, refinement, surrogate RTM) is implemented as a differentiable network, $\mathcal{H}$ is differentiable end-to-end with respect to (z_a, z_s). The construction and training of the surrogate RTM are summarized in Section 2.3.

Innovation. The initial (band-wise) innovation is defined as the observation-minus-background departure in reflectance space,

$$d_\lambda^b = y_\lambda - \mathcal{H}_\lambda(z_a^b, z_s^b) \tag{3}$$

which $z^b = (z_a^b, z_s^b) = (E_a(x_a^b), E_s(x_s^b))$ denotes the background latent state. During the minimization, the observation misfit is evaluated using the current latent state,

$$d_\lambda = y_\lambda - \mathcal{H}_\lambda(z_a, z_s) \tag{4}$$

and is therefore updated at each iteration as the analysis approaches the observations.

https://doi.org/10.5194/egusphere-2026-4627
Preprint. Discussion started: 14 September 2026

(c) Latent space variational assimilation. The analysis is obtained as a Bayesian estimate in the latent space. Given observations y and control vector be $z = (z_a, z_s)$, Bayes' rule gives the posterior

$$p(z_a, z_s|y) \propto p(y|z_a, z_s)p(z_a, z_s|z^b) \quad (5)$$

We seek the maximum a posteriori (MAP) estimate,

$$(z_a^a, z_s^a) = \mathrm{argmax}_{z_a, z_s} \log p(z_a, z_s|y) \quad (6)$$

In physical-space variational assimilation, Gaussian approximations of background error on high-dimensional fields are often poorly justified. Here the state variables are z rather than x: during offline VAE training, the Kullback–Leibler term explicitly regularizes the encoder posterior $q(z \mid x)$ toward a standard normal distribution, so that the learned latent coordinates are approximately Gaussian, decorrelated, and centered. Therefore, for the background, we assume

$$p(z_a, z_s|z^b) \sim \mathcal{N}(z_b, \mathbf{B}), \qquad \mathbf{B} = \mathrm{blockdiag}(\mathbf{B_a}, \mathbf{B_s}) \quad (7)$$

Where $\mathbf{B_a}$ and $\mathbf{B_s}$ are the aerosol and surface latent background error covariances. Cross-covariances between aerosol and surface background errors are neglected, implying that the two components are assumed to be statistically independent in the background state. The background ensemble used to estimate $\mathbf{B_a}$ and $\mathbf{B_s}$ is obtained by encoding WRF-Chem and MCD43C3 surface albedo into the latent space (Section 2.5).

For the observations, we decompose the multi-band vector by wavelength as $y = (y_\lambda)_{\lambda=1}^5$, and likewise $\mathcal{H}(z_a, z_s) = (\mathcal{H}_\lambda(z_a, z_s))_{\lambda=1}^5$. Assuming that observation errors are uncorrelated across bands (and, within each band, are represented by a diagonal covariance $\mathbf{R}_\lambda$), the likelihood factorizes as a product of band-wise Gaussian distributions

$$p(y|z_a, z_s) \approx \prod_{\lambda=1}^5 p(y_\lambda|z_a, z_s), \qquad p(y_\lambda|z_a, z_s) \approx \mathcal{N}(\mathcal{H}_\lambda(z_a, z_s), \mathbf{R}_\lambda) \quad (8)$$

The MAP problem becomes the minimization of the variational cost function.

$$J(z_a, z_s) = -\log p(z_a, z_s|y) + \mathrm{const} = J_s^b + J_a^b + J^o \quad (9)$$

$$J_s^b = \frac{1}{2}(z_s - z_s^b)^T \mathbf{B_s}^{-1}(z_s - z_s^b) \quad (10)$$

$$J_a^b = \frac{1}{2}(z_a - z_a^b)^T \mathbf{B_a}^{-1}(z_a - z_a^b) \quad (11)$$

$$J^o = \frac{1}{2}\sum_{\lambda=1}^5 d_\lambda^T \mathbf{R}_\lambda^{-1} d_\lambda \quad (12)$$

Because J is smooth and each H_λ is differentiable, the gradients with respect to the latent control variables are

$$\nabla_{z_a} J = \mathbf{B_a}^{-1}(z_a - z_a^b) - \sum_{\lambda=1}^5 \left(\frac{\partial \mathcal{H}_\lambda}{\partial z_a}\right)^T \mathbf{R}_\lambda^{-1} d_\lambda \quad (13)$$

$$\nabla_{z_s} J = \mathbf{B_s}^{-1}(z_s - z_s^b) - \sum_{\lambda=1}^5 \left(\frac{\partial \mathcal{H}_\lambda}{\partial z_s}\right)^T \mathbf{R}_\lambda^{-1} d_\lambda \quad (14)$$

The Jacobians $\frac{\partial \mathcal{H}_\lambda}{\partial z_a}$ and $\frac{\partial \mathcal{H}_\lambda}{\partial z_s}$ arise from the composition of the decoders, the aerosol refinement, and the band-wise surrogate RTM. By the chain rule, their transposed Jacobians (adjoints) map sensitivities from the observation space back to the latent space

https://doi.org/10.5194/egusphere-2026-4627
Preprint. Discussion started: 14 September 2026

$$\left(\frac{\partial \mathcal{H}_\lambda}{\partial z_a}\right)^T = \left(\frac{\partial D_a}{\partial z_a}\right)^T \left(\frac{\partial R_a}{\partial \widehat{x_a}}\right)^T \left(\frac{\partial \mathcal{H}_\lambda}{\partial \widehat{x_a}}\right)^T \tag{15}$$

$$\left(\frac{\partial \mathcal{H}_\lambda}{\partial z_s}\right)^T = \left(\frac{\partial D_s}{\partial z_s}\right)^T \left(\frac{\partial \mathcal{H}_\lambda}{\partial \widehat{x_s}}\right)^T \tag{16}$$

which is efficiently computed via the automatic differentiation module of the neural network framework.

The latent variables are optimized iteratively using the Adam algorithm until convergence, yielding the analysis latent states z_a^a and z_s^a. The corresponding analysis aerosol and surface fields are then obtained by decoding (refining) the optimized latent variables through the trained VAE decoders:

$$x_a^a = R(D_a(z_a^a)), \qquad x_s^a = D_s(z_s^a) \tag{17}$$

where x_a^a and x_s^a denote the analyzed multi-band aerosol optical depth and surface white-sky albedo fields, respectively. The decoded fields therefore represent the maximum-a-posteriori estimate in the latent space, constrained jointly by the background information and the multi-spectral reflectance observations through the integrated observation operator.

2.2 Latent representation learning and training

The first offline component of LARDAS (Figure 1a) learns the aerosol and surface representations used by the cycling analysis. A large archive of multi-band AOD and ALB fields is used to train the two VAEs and the aerosol refinement block and to estimate the latent background statistics. These learned components are fixed during online assimilation. The training fields are independent of the radiative-transfer archive used for the neural observation operator in Section 2.3.

WRF-Chem configuration and simulations. WRF-Chem is an online coupled meteorology–chemistry model that integrates atmospheric dynamics, radiation, cloud microphysics, and chemical processes within a unified framework, enabling the explicit simulation of interactions among these components, including aerosol-related processes (Fast et al., 2006; Grell et al., 2005; Powers et al., 2017). We perform regional simulations over the study domain on a 141×141 grid (two nested domains, with the 141×141 inner domain configured at 10 km horizontal resolution) in the north and east part of China. The configuration follows (Zhang et al., 2026)(Supporting Information Text S1 of Zhang et al., 2026) and is summarized in Table 1:

https://doi.org/10.5194/egusphere-2026-4627
Preprint. Discussion started: 14 September 2026

Table 1. WRF-Chem model configurations and emission inventories

Category	Configuration/Data Source	Reference
Microphysics scheme	Morrison 2-moment	(Morrison et al., 2009)
Planetary boundary layer scheme	Yonsei University	(Hong et al., 2006)
Land surface model	Noah	(Chen & Dudhia, 2001)
Radiation scheme	RRTMG(sw/lw)	(Iacono et al., 2008)
Gas-phase chemistry mechanism	CBM-Z	(Zaveri & Peters, 1999)
Aerosol module	MOSAIC-4bins	(Zaveri et al., 2008)
Anthropogenic emissions	MEIC	(Li et al., 2017; Zheng et al., 2018)
Biogenic emissions	MEGAN	(Guenther et al., 2006)
Open burning emissions	FINN	(Wiedinmyer et al., 2011)
Meteorological driving	ERA5	(Hersbach et al., 2020)
Chemical driving	CAM-Chem	(Tilmes et al., 2022)

Simulations from January 2016 to February 2017 were conducted, with one simulation performed for each month (excluding February 2016, which was reserved for the independent assimilation experiment), to construct the latent space training archive. Detailed descriptions of the model configuration can be found in Zhang et al. (2026)

Mixing-Mie optical module. In WRF-Chem, aerosol mass and number concentrations are treated as prognostic variables, while aerosol optical properties are calculated from the simulated aerosol composition and size distribution. Following Zhang et al. (2026), a mixing-Mie module is employed to convert the MOSAIC-predicted size-resolved aerosol composition into spectrally resolved optical properties. The module combines aerosol refractive indices from the OPAC database(Hess et al., 1998) with Mie calculations to estimate aerosol optical properties at the five MODIS shortwave bands (469, 555, 645, 858, and 2100 nm).

Multi-band aerosol optical depth dataset.

The multi-band AOD fields $x_a \in \mathbb{R}^{5\times141\times141}$ derived from WRF-Chem simulations using the Mixing-Mie module calculations, are employed to construct the training dataset for the aerosol VAE. However, the approximately one-year

WRF-Chem simulation provides only a limited number of aerosol state samples, which is insufficient to fully represent the diversity of aerosol spatial patterns required for robust VAE training. To increase the size and variability of the training dataset, a series of data augmentation techniques are applied. For each original AOD sample, the augmentation pipeline retains the unmodified field and generates additional samples through horizontal and vertical reflections, min–max normalization, Gaussian smoothing with multiple spatial scales, rotations over a range of azimuthal angles, and combinations of these transformations (e.g., rotation followed by smoothing or reflection). This procedure substantially enriches the variability of aerosol spatial structures while preserving their physically meaningful characteristics. After data augmentation, 1.2×10^6 samples are used to train the aerosol VAE and refinement block. This augmented corpus is partitioned randomly into training (90%) and validation (10%) subsets. For independent evaluation, 2022 samples from December 2016 through February 2017 are held out without augmentation and used for the reconstruction diagnostics in Section 3.

Multi-band surface albedo dataset.

The surface representation-learning target is multi-band surface albedo $x_s \in \mathbb{R}^{5\times141\times141}$ on the assimilation grid. It is taken from the MODIS MCD43C3 climate-modeling BRDF/albedo product (0.05° resolution) for the year 2017. MCD43C3 is produced as 8-day composites. For each calendar month of 2017, we download the first available composite whose acquisition window falls within that month, yielding 12 monthly maps. Each map is assembled at the five MODIS shortwave bands used in this study (BRDF bands 1, 2, 3, 4, and 7), bilinearly interpolated from the global product onto the regional domain, and gap-filled so that every land pixel has a valid five-band surface albedo value. These 12 monthly maps are then expanded to daily resolution by linear interpolation in day-of-year, with continuity enforced across the year boundary, yielding 365 daily surface albedo states.

Because the intrinsic temporal sampling of the surface fields is far sparser than that of the aerosol fields, we enlarge the surface training ensemble by adding EOF-based perturbations to each of the 365 daily maps. For a given daily map, valid land pixels are decomposed via joint spatial–spectral principal component analysis (PCA). Nine additional realizations are then generated by injecting spatially correlated noise in principal-component space, inverse-transforming to the grid, and clipping albedo to [0, 1]. Together with the unperturbed daily map, this yields ten members per calendar day and a total ensemble of 365 × 10 = 3650 samples per year. The surface VAE is trained on this pool after random partitioning: 3180 samples for training, 350 for validation, and 120 for testing. The test subset comprises all ten EOF members on twelve days fixed in advance (one randomly selected day of year per month); the remaining 353 days supply the training and validation sets in a 90%/10% split. No additional transforms are applied during network optimization.

VAE architecture, refinement, and training objective.

Both VAEs employ a residual convolutional encoder–decoder architecture (ResNet-VAE) with a latent dimension of d = 128. Starting from the five-band input field on the 141 × 141 grid, the encoder applies four down-sampling stages (strided

https://doi.org/10.5194/egusphere-2026-4627
Preprint. Discussion started: 14 September 2026

convolutions, each halving the spatial resolution) with progressively increasing feature channels, resulting in a 256 × 9 × 9 feature map that is projected by fully connected heads to the posterior mean $\mu \in \mathbb{R}^d$ and log-variance $\log \sigma^2 \in \mathbb{R}^d$. Each stage includes a residual block to improve gradient flow. The latent variable is drawn from the encoder posterior using the standard reparameterization,

$$z = \mu(x) + \epsilon \odot \sigma(x), \quad \epsilon \sim \mathcal{N}(0, I) \tag{18}$$

The decoder is symmetric to the encoder: a fully connected layer maps the latent vector back to the feature map, and four up-sampling stages (bilinear interpolation followed by convolution) restore the original resolution while reducing the channel count. Physical-range constraints are enforced at the output through a learnable per-band affine transform and a clipping operation:

$$\hat{x} = \mathrm{clip}(x_{raw} \odot s + t) \tag{19}$$

The clip enforces non-negativity for AOD and the range [0, 1] for surface albedo. Each VAE is trained by minimizing a weighted sum of a reconstruction term and the Kullback–Leibler (KL) regularization:

$$\mathcal{L} = \lambda_1 \mathcal{L}_{L_1} + \lambda_{\mathrm{grad}} \mathcal{L}_{\mathrm{grad}} + \lambda_{\mathrm{ssim}} \mathcal{L}_{\mathrm{ssim}} + \lambda_{kl} \mathcal{L}_{\mathrm{kl}} \tag{20}$$

where the reconstruction term combines a pixel-wise L1 loss, a spatial-gradient loss that penalizes discrepancies in the horizontal and vertical derivatives, and a structural-similarity (SSIM) term. A mild KL regularization is included to encourage a reasonably compact latent distribution; the dominant reconstruction losses ensure that the physical-field content is faithfully preserved in the low-dimensional representation.

For the aerosol field, an additional refinement block R_a consisting of two residual convolution layers is appended after the decoder to restore fine-scale spatial structure lost during compression, giving the refined field $\widehat{x_a} = R(\widehat{x_a})$ that enters the observation operator. It is trained independently after the VAE is frozen, using the same augmented AOD dataset as the aerosol VAE. The training loss combines a pixel-wise weighted L1 term that emphasizes high-AOD regions (with a per-pixel weight $1 + 5\tau$, where τ is the target AOD) and a spatial-gradient loss to preserve sharp edges.

2.3 Neural RTM training and integrated observation operator

The second offline component of LARDAS trains the neural radiative-transfer model used in the online observation operator (Figure 1b). Aerosol, gaseous, surface, and viewing-geometry inputs are first propagated through VLIDORT to generate a multi-band reflectance archive. Five band-specific residual networks are then trained to emulate these calculations. During online cycling (Figure 1c), the trained networks are fixed and combined with the aerosol and surface decoders to form the end-to-end differentiable operator that maps latent variables to MODIS-equivalent TOA reflectance. We summarize the construction here in simplified form and describe the neural-network surrogate used in (Zhang et al., 2026)(Supporting Information Text S2 and training-data flowchart of Zhang et al., 2026).

Training-set generation

https://doi.org/10.5194/egusphere-2026-4627
Preprint. Discussion started: 14 September 2026

The reference forward model is the Vector Linearized Discrete Ordinate Radiative Transfer model (VLIDORT (Spurr, 2006)), a multiple-scattering radiative transfer code that can calculate TOA reflectance. One million input–output pairs are generated at the five MODIS shortwave bands (469, 555, 645, 858, and 2100 nm), with 1% held out for testing and the remainder split 80/20 for training and validation. Each training sample was constructed by combining aerosol, atmospheric, surface, and geometric parameters from multiple data sources to represent a complete radiative-transfer state. Aerosol optical depth (AOD) was sampled within prescribed ranges. Aerosol microphysical properties, including particle size distributions and complex refractive indices, were sampled from the whole records of the Aerosol Robotic Network (AERONET). These properties were used as input to a T-matrix scattering model to derive the full aerosol scattering phase matrix, which was subsequently represented through Legendre polynomial expansion for vector radiative transfer calculations. Vertical-profiles parameters of aerosol extinction were fitted by randomly sampling from an ensemble of aerosol distributions simulated by WRF-Chem over the study domain. This approach preserves the realistic vertical structure and layering characteristics represented by the chemical transport model. Surface reflectance conditions were characterized using Lambertian white-sky albedo values sampled from the MODIS MCD43C3. Solar zenith angle, viewing zenith angle, and relative azimuth angle were sampled from the distributions of observation geometries contained in MODIS Level-1B geolocation products. Gaseous absorption optical depths were derived from Line-By-Line Radiative Transfer Model (LBLRTM) (Clough et al., 2005) reference calculations for each spectral band. The absorption optical depths were scaled according to column water vapor and ozone amounts, where water vapor columns were obtained from the corresponding WRF-Chem simulations and ozone columns were obtained from ERA5 reanalysis data. Rayleigh scattering optical depths were calculated using an empirical pressure–wavelength relationship. Details of the configuration can be found in (Zhang et al., 2026). For each sampled state, VLIDORT integrates the radiative transfer equation and returns the TOA apparent reflectance in each band, producing the reflectance targets used for surrogate training. The resulting archive covers a broad range of aerosol–surface–geometry combinations that encompass those encountered during real assimilation.

Residual-network surrogate (MLRTM)

Zhang et al. (2026) emulated VLIDORT with gradient-boosted trees trained separately for each band and for reflectance and Jacobian outputs (15 models in total). Here we use a fully neural surrogate so that the observation operator is natively differentiable and can be coupled to the VAE decoders in end-to-end latent assimilation. For each band, a compact residual network (ResNet-style feed-forward architecture) maps a per-pixel input vector to reflectance. The input assembles, for that band, the band-specific AOD and single-scattering albedo, Lambertian surface albedo, total and effective gaseous optical depths, solar/viewing geometry, and the seven profile parameters (15 predictors in total, standardized using the training-set statistics). The network output is reshaped to (H, W) during online assimilation. Five independent networks are trained—one per band—mirroring the per-band organization of the companion work but replacing tree models with residual networks. VLIDORT targets and MLRTM outputs are expressed as TOA apparent reflectance on the same scale as the MODIS observations used in Section 2.4. Forward accuracy of the surrogate against the VLIDORT test set is assessed in Section 3.

https://doi.org/10.5194/egusphere-2026-4627
Preprint. Discussion started: 14 September 2026

2.4 Observations and verifications

MODIS Observations and quality control

The primary assimilation observations are top-of-atmosphere (TOA) shortwave reflectance from the Moderate Resolution Imaging Spectroradiometer (MODIS) on board Terra and Aqua. Level-1B calibrated radiances at five visible and near-infrared bands (centered near 469, 555, 645, 858, and 2100 nm) are converted to apparent reflectance and interpolated onto the 141 × 141 assimilation grid, yielding the multi-band observation vector y_λ.

MODIS is a cross-track scanning instrument, and only a subset of the assimilation domain is observed during a given overpass. Additional data gaps are introduced by the quality-control (QC) procedures applied to the MODIS observations. To suppress spurious nonlocal correlations introduced by the LARDAS assimilation procedure, a gap-filling strategy is adopted in which the influence of the background constraint is strengthened in regions lacking observations. At grid cells where valid MODIS reflectance observations are available after quality control, the observation vector y_λ is assigned the measured top-of-atmosphere reflectance. At all remaining grid cells, y_λ is replaced by the background-equivalent reflectance, $\mathcal{H}_\lambda(z_a^b, z_s^b)$, obtained by evaluating the integrated observation operator using the background aerosol and surface latent states. It implies that the reflectance misfit reduces to $\mathcal{H}_\lambda(z_a, z_s) - \mathcal{H}_\lambda(z_a^b, z_s^b)$. In effect, this supplements the latent space background terms with an additional background constraint expressed in observation space over data-void areas, rather than introducing independent observational information. When the global latent controls would otherwise allow nonlocal adjustments to fit limited MODIS footprints, this term helps keep the forward reflectance in unobserved regions close to the background simulation and reduces spurious large-scale corrections.

Throughout this study, the assimilation vector y consists of MODIS gridded top-of-atmosphere (TOA) apparent reflectance (dimensionless), converted from Level-1B radiances using the standard MODIS calibration and geometry. Denoting TOA radiance by L_λ, solar irradiance by $E_{0,\lambda}$, and solar zenith angle by θ_0, the observed reflectance is

$$\rho_\lambda^{\mathrm{obs}} = \frac{\pi L_\lambda}{E_{0,\lambda}\cos(\theta_0)} \tag{21}$$

The MLRTM surrogate is trained against VLIDORT apparent reflectance targets generated under the same definition (Section 2.3). Model-simulated reflectance from the integrated operator is expressed in the same units before comparison with y.

Quality control (QC) is applied prior to assimilation to remove observations affected by clouds, sun glint, snow/ice contamination, and unsuitable surface conditions. In the present implementation, a conservative proxy-based QC strategy is

https://doi.org/10.5194/egusphere-2026-4627
Preprint. Discussion started: 14 September 2026

adopted: a pixel is retained only if it is flagged as suitable for aerosol retrieval in the corresponding MODIS aerosol product. The quality-assurance flags are used solely as indicators of observation quality and do not provide any retrieved aerosol information to the assimilation system. This approach provides a practical means of identifying reliable observations while avoiding the need for a dedicated radiance-level QC procedure. However, it is intentionally conservative and may exclude observations that could be useful for direct reflectance assimilation. Future developments will focus on designing physically based QC procedures that operate directly on Level-1B radiance or reflectance observations, allowing the methodology to be applied consistently across different satellite sensors without reliance on retrieval-product quality flags.

Verification and statistical metric

Independent validation datasets include AERONET direct-sun AOD at 500 nm at selected sites within the domain and, where available, Himawari-8 AOD(Zhang et al., 2018) at 500 nm as an additional satellite-based reference. These follow the preprocessing and collocation procedures in (Zhang et al., 2026)and are used only for verification in Section 3.

Independent evaluation of the simulated shortwave radiation budget was performed using the Clouds and the Earth's Radiant Energy System (CERES) SYN1deg-1Hour gridded radiative flux product. For the present study, daily HDF4 files from the CERES SYN1deg-1Hour Terra–Aqua MODIS Edition4A data stream were obtained for 1–10 February 2016, covering the same multi-day assimilation period used in the AERONET validation. Each file provides globally gridded, hourly estimates of radiative fluxes on a 1° × 1° latitude–longitude mesh(Rutan et al., 2015; Wielicki et al., 1996). The SYN product combines flux estimates from the CERES instruments on the Terra and Aqua satellites and applies a temporal interpolation and normalization procedure to produce continuous hourly fields suitable for model–satellite comparison at synoptic scales.

Assimilation performance is quantified using standard statistical metrics; further details are provided in Zhang et al. (2026). Throughout this section, x_i denotes the assimilated estimate (background or analysis), y_i the reference value, and n the number of valid grid points. The mean absolute bias (MAB) and root-mean-square error (RMSE) are

$$\mathrm{MAB} = \frac{1}{n}\sum_{i=1}^{n} |x_i - y_i| \qquad \mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (x_i - y_i)^2} \tag{22}$$

Systematic bias is

$$\mathrm{Bias} = \frac{1}{n}\sum_{i=1}^{n} (x_i - y_i) \tag{23}$$

The per-pixel error reduction achieved by the assimilation is quantified by the absolute improvement and its relative form:

$$\mathrm{Imp_i} = \left|x_{\{b,i\}} - y_{\mathrm{ref}}\right| - \left|x_{\{a,i\}} - y_{\mathrm{ref}}\right|,\ \ \mathrm{Imp}_{\mathrm{i}}^{r} = \frac{|x_{\{b,i\}} - y_{\mathrm{ref}}| - |x_{\{a,i\}} - y_{\mathrm{ref}}|}{|x_{\{b,i\}} - y_{\mathrm{ref}}|} \tag{24}$$

where $x_{b,i}$ and $x_{a,i}$ are the background and analysis values, and y_{ref} is the reference at grid point i. A positive Imp, indicates that the assimilation has reduced the error.

https://doi.org/10.5194/egusphere-2026-4627
Preprint. Discussion started: 14 September 2026

The Fractions Skill Score (Roberts & Lean, 2008) complements point-wise metrics by evaluating spatial pattern correspondence through sliding-window exceedance fractions

$$FSS = 1 - \frac{\sum_w \left(p_o^w - p_{f/a}^w\right)^2}{\sum_w \left({p_o^w}^2 + {p_{f/a}^w}^2\right)} \tag{25}$$

where p_o^w and $p_{f/a}^w$ are the numbers of grid points within window w exceeding a prescribed threshold. The window size is approximately 5% of the domain dimensions. FSS ranges from 0 to 1 (FSS = 1 is perfect match), and accounts for spatial displacement errors that point-wise metrics such as MAE and RMSE cannot capture.

2.5 Assimilation experiments and error specifications

This subsection completes the online system description in Figure 1c. It specifies the latent background and observation-error covariances, the numerical minimization, the conversion of analyzed AOD and surface albedo into WRF-Chem model variables, and the controlled and real-observation experimental configurations. Two reference assimilation routes used only for the February 2016 intercomparison are documented at the end of the subsection.

Background and observation error covariances

Unlike Zhang et al. (2026), who estimated physical-space background errors with the NMC method for AOD and a multi-day MCD43C3 ensemble for surface albedo, we estimate background statistics directly in latent space: WRF-Chem AOD field for $\mathbf{B_a}$ and MCD43C3-derived surface albedo for $\mathbf{B_s}$. Encoded samples $z_{a,i} = E_a(x_{a,i})$ and $z_{s,i} = E_s(x_{s,i})$ yield ensemble means:

$$\overline{z_a} = \frac{1}{N}\sum_{i=1}^{N} E_a\left(x_{a,i}\right), \qquad \overline{z_s} = \frac{1}{N}\sum_{i=1}^{N} E_s\left(x_{s,i}\right) \tag{26}$$

and diagonal covariances

$$\mathbf{B_a} = \mathrm{diag}\left(\frac{1}{N}\sum_{i=1}^{N}\left(z_{a,i} - \overline{z_a}\right)^2\right), \quad \mathbf{B_s} = \mathrm{diag}\left(\frac{1}{N}\sum_{i=1}^{N}\left(z_{s,i} - \overline{z_s}\right)^2\right) \tag{27}$$

The diagonal approximation is adopted for computational efficiency as the latent variables generated by the VAE are designed to be approximately decorrelated(Fan et al., 2026).

For each band, $\mathbf{R}_\lambda$ is diagonal, i.e. $\mathbf{R}_\lambda = \sigma_{\mathrm{obs}}^2\mathbf{I}$. In the controlled simulation experiment, the pseudo-observations are noise-free forward model output from a different atmospheric state, and the assumed observation error is set to $\sigma_{\mathbf{obs}} = 5 \times 10^{-4}$. Sensitivity to σ_{obs} is examined by scanning values from 10^{-2} to 5 in a separate sweep experiment (Section 3). In real-observation assimilation, a scalar $\sigma_{\mathbf{obs}} = 2\%$ sets the reflectance error standard deviation. When gridded MODIS aerosol optical depth products are assimilated in latent space for the February 2016 intercomparison (below), a separate diagonal AOD observation-error covariance is used with error standard deviation of 0.1.

Controlled simulation-based assimilation

https://doi.org/10.5194/egusphere-2026-4627
Preprint. Discussion started: 14 September 2026

To verify algorithmic consistency before introducing retrieval and sampling errors from real satellites, we conduct a controlled experiment using paired simulation cases from an integrated test archive. The background aerosol and surface fields x_a^b and x_s^b are taken from one case, encoded to (z_a^b, z_s^b), and the multi-band reflectance from a second case serves as the observation vector y (pseudo-observation), is generated by applying the same integrated observation operator used in assimilation to the truth aerosol and surface states of the independent case. Truth fields x_a^t and x_s^t are available from the same archive for evaluation. Because the pseudo-observations are defined on the full 141 × 141 grid, spatial coverage is complete and cloud masks are not applied. The assimilation minimizes J and produces analysis fields x_a^a and x_s^a by decoding the optimized latent state. This experiment tests whether the differentiable operator $\mathcal{H}$ and the latent background term jointly improve aerosol and surface estimates against known truth. We refer to this design as a simulation-based validation rather than a strict observing-system simulation experiment (OSSE), because it does not implement a separate nature run at different resolution.

Real-observation assimilation

In the real-data experiments, the assimilated observations consist of the gridded MODIS reflectance fields described in Section 2.4, with the quality-control and gap-filling procedures applied as described therein. The background aerosol field is derived from vertically integrated multi-spectral aerosol optical thickness contained in the mixing-Mie background dataset, while the background surface field is obtained from monthly climatological surface albedo data. Both background fields are encoded into the latent representations (z_a^b, z_s^b) prior to optimization. Before each forward-model evaluation, radiative-transfer auxiliary variables, including viewing and illumination geometry, gaseous optical depths, aerosol single-scattering albedo, and aerosol profile parameters, are interpolated and reordered to match the wavelength configuration of the assimilated observations. Simulated top-of-atmosphere reflectance is multiplied by π to ensure consistency with the MODIS reflectance product. Following optimization, the analysis aerosol and surface fields are obtained by decoding the optimized latent variables. After assimilation, the analyzed multi-band AOD fields are mapped to WRF-Chem prognostic variables (aerosol mass and number concentrations per MOSAIC size bin) using the same ensemble Kalman filter (EnKF) mapping module as in (Zhang et al., 2026), which preserves the bin-wise composition ratios while scaling the total column loading to match the assimilated AOD spectrum. This step closes the loop from latent space AOD analysis back to the model state vector, enabling the assimilated information to propagate forward during the forecast. The analyzed multi-spectral surface albedo is mapped onto the WRF five-band albedo fields (ALBEDO1–ALBEDO5) and written into the restart files after each cycle. Because the standard WRF-Chem/RRTMG setup uses a single broadband albedo (ALBEDO) for all shortwave bands, the model code was modified so that restart integrations retain and use the band-resolved fields in RRTMG instead of overwriting them with ALBEDO each time step. Forecasts are therefore initialized from both the EnKF-updated aerosol state and the analyzed multi-wavelength surface albedo. Independent evaluation of the analysis results is performed using the AERONET and Himawari-8 datasets described in Section 2.4.

https://doi.org/10.5194/egusphere-2026-4627
Preprint. Discussion started: 14 September 2026

Reference assimilation configurations and intercomparison protocol

For context, we document three online assimilation configurations that share the same WRF-Chem domain, spin-up windows, MODIS reflectance quality control, ensemble Kalman filter mapping from analyzed AOD to MOSAIC concentrations, and short-range forward integration for verification. They differ in analysis space, observation vector, background-error specification, and observation operator (Table 2). The proposed method is latent space assimilation of multi-band MODIS reflectance, as described in the preceding paragraphs and in Section 2.1. The other two are reference configurations used only in February 2016: physical-space variational reflectance assimilation following Zhang et al. (2026), and latent space assimilation of gridded MODIS 555 nm aerosol optical depth products with the same latent optimizer and downstream mapping. Side-by-side AERONET time series and station-wise statistics for all three appear only for 1–10 February 2016.

Table 2. Assimilation configurations for the February 2016 intercomparison. The latent-reflectance configuration is the proposed LARDAS analysis, whereas the other two configurations serve as references for the February 2016 intercomparison.

	Physical-space reflectance	Latent space reflectance	Latent space AOD
Observations	TOA reflectance	TOA reflectance	AOD
Control vector	AOD and ALB	Latent z	Latent z
Minimization	Conjugate Gradient	Adam	Adam
B estimation	NMC & ALB ensemble	Encoded archives	Encoded archives
$\mathcal{H}$	Machine Learning RTM	Integrated $\mathcal{H}$	Decoder

Physical-space reflectance assimilation. This configuration follows the companion shortwave assimilation system of Zhang et al. (2026). The control vector comprises gridded multi-band aerosol optical depth and white-sky albedo on the 141×141 domain. Background errors are estimated in physical space (NMC for aerosol and a multi-day MCD43C3 ensemble for surface albedo). The analysis minimizes a variational cost function with a band-wise machine-learning radiative transfer emulator as the observation operator. The same MODIS reflectance quality control and gap-filling philosophy as in Section 2.4 is applied before optimization. It provides the baseline for computational cost comparisons with latent reflectance assimilation in Section 3.5.

Latent space reflectance assimilation. This is the real-data design in the preceding paragraphs: reflectance observations constrain latent aerosol and surface variables through the integrated differentiable operator in Section 2.1, with background terms given above. The controlled simulation-based experiment uses this configuration only.

https://doi.org/10.5194/egusphere-2026-4627
Preprint. Discussion started: 14 September 2026

Latent space MODIS AOD product assimilation. The observation vector is replaced by gridded MODIS retrieved aerosol optical depth at 555 nm on the assimilation grid; the observation operator maps the decoded aerosol field to AOD at that wavelength, and a diagonal AOD observation-error covariance is assumed. This is not a separate WRF variational product-assimilation system; it tests whether assimilating retrieval products within the LARDAS framework can approach the skill of direct reflectance assimilation under identical model and mapping settings.

For 1–10 February 2016, WRF-Chem aerosol optical depth at 500 nm from the background, from latent reflectance assimilation, and from each reference configuration is collocated to AERONET sites and scored with the metrics in Section 2.4. For February, March, April, and October 2017, monthly archives of latent space reflectance assimilation and the background are evaluated at six AERONET stations on the same protocol. Numerical scores and figures are reported in Section 3.

The WRF-Chem simulation configuration and the experimental time windows follow (Zhang et al., 2026). Latent reflectance assimilation is conducted for 1–10 February 2016, with 28 January–1 February as model spin-up; a short-range forward integration from the analysis state is used to examine persistence of skill, with AERONET 500 nm AOD evaluated over that window. A longer experiment with monthly restarts covers December 2016 through February 2017. For February, March, April, and October 2017, separate monthly initializations with restarts are used to evaluate robustness under varying meteorological and aerosol conditions; in those months only latent space reflectance assimilation and the background are compared against AERONET (Section 3), as noted above.

3 Results

The LARDAS couples three components—latent representation, surrogate forward modeling, and variational minimization whose individual fidelity collectively governs the quality of the analysis. We therefore organize the evaluation from offline component verification to online assimilation performance: first, the reconstruction skill of the trained VAEs and refinement block(Section 3.1); second, the accuracy of the surrogate radiative transfer emulator against the reference physical model (Section 3.2); third, assimilation performance under controlled simulation-based conditions with known truth, including a sensitivity study of the assumed observation error(Section 3.3); and fourth, real-observation experiments assimilating MODIS shortwave reflectance, with independent validation against AERONET ground-based measurements over multiple months(Section 3.4). We close with a computational efficiency assessment (Section 3.5).

3.1 Latent representation quality

In LARDAS, the encoder compresses the high-dimensional aerosol and surface fields ($\mathbb{R}^{5\times141\times141}$) into a 128-dimensional latent state, and the decoder—augmented by a refinement block for AOD—reconstructs the physical fields from this compact representation. Any spatial structure lost in this compression cannot be recovered by subsequent assimilation; the

https://doi.org/10.5194/egusphere-2026-4627
Preprint. Discussion started: 14 September 2026

reconstruction fidelity therefore sets an upper bound on the achievable analysis improvement. We first assess how well the trained representations reproduce the target fields on independent test data.

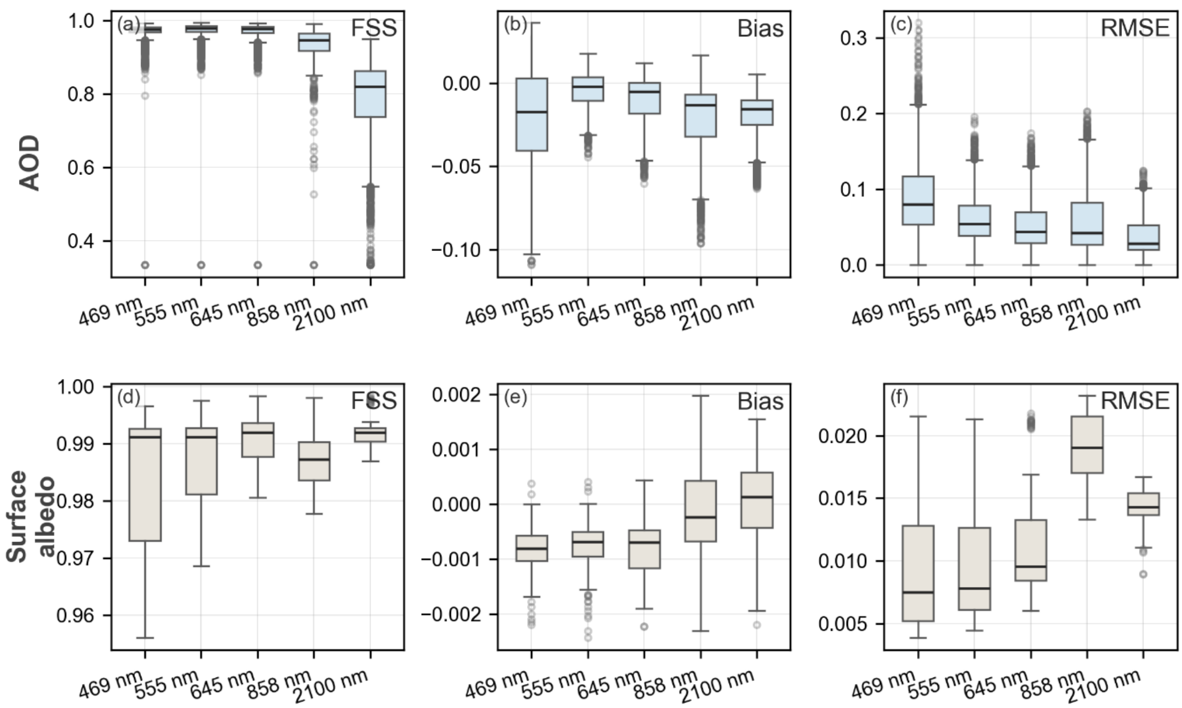

Figure 2. Per-band reconstruction quality of the aerosol VAE+refinement (top row) and the surface VAE (bottom row), evaluated on held-out test samples. Left: fraction skill score(FSS) with threshold at the 50th percentile and a 5×5 neighborhood. Center: root-mean-square error (RMSE). Right: bias. Boxes denote the interquartile range (IQR); whiskers span the 5th–95th percentile; medians are marked by horizontal lines. Wavelengths are shown in ascending order: 469, 555, 645, 858, and 2100 nm.

Figure 2 summarizes the per-band reconstruction metrics across the five shortwave bands. For the aerosol VAE with refinement, the median FSS exceeds 0.95 for the first four shortwave bands, indicating that the latent representation reliably preserves the spatial patterns of aerosol loading. The strong reconstruction performance at these wavelengths can be attributed to the relatively large aerosol optical depth (AOD) in the visible and near-infrared bands, where aerosol signals are more pronounced and spatial structures are more readily captured by the VAE. In contrast, the reconstruction skill decreases for the 2100 nm band, reflecting the weaker aerosol contribution and reduced signal strength at this wavelength. Nevertheless, the median FSS remains above 0.8, demonstrating that the VAE still retains a substantial capability to reconstruct the dominant spatial patterns of aerosol optical properties even in the shortwave infrared band. The

https://doi.org/10.5194/egusphere-2026-4627
Preprint. Discussion started: 14 September 2026

reconstruction bias and RMSE of AOD are also within acceptable ranges across all spectral bands. Even at 469 nm, where aerosol optical depth is generally the largest, the median bias remains below 0.02 and the median RMSE is less than 0.1. These results indicate that the VAE introduces only limited reconstruction errors and is able to accurately preserve the magnitude of aerosol optical properties while maintaining their spatial structures. For the surface albedo VAE, the FSS median exceeds 0.98 in all bands, the median bias is on the order of 10^{-3}, and the RMSE median remains below 0.02. The better performance achieved for surface albedo is likely a consequence of its more stable spatial structure and lower temporal variability relative to AOD, which facilitates the learning and reconstruction of its dominant features by the VAE.

Individual reconstruction examples for aerosol and surface test samples are provided in the Supporting Information (Figs. S1 and S2). The AOD examples confirm that elevated plumes and sharp spatial gradients are generally well captured after refinement, although thin filamentary structures are occasionally smoothed when the peak AOD is modest and the surrounding field is near zero. The surface examples demonstrate that the five-band surface albedo is reproduced with high fidelity. These reconstruction characteristics provide a reference for interpreting assimilation results: corrections beyond the reconstruction error envelope represent genuine observational constraints from the shortwave reflectance.

3.2 Surrogate observation operator accuracy

The accuracy of the observation operator directly affects both the quality of the innovation vector and the gradient information propagated during the minimization process. Therefore, to ensure physically meaningful assimilation results, the surrogate observation operator must achieve a sufficiently high level of accuracy. Figure 3 evaluates the five independent MLRTM emulators against the VLIDORT reference. Across all bands, the emulated reflectance is in close agreement with the physical forward model, with correlation coefficients $R > 0.999$ and RMSE<0.003. The surrogate errors are approximately an order of magnitude smaller than the MODIS reflectance observation error assumed in the assimilation (~2% of reflectance; Section 2.5). Consequently, their impact on the innovation vector is negligible, and the MLRTM emulators provide sufficient accuracy for use within the latent space assimilation framework.

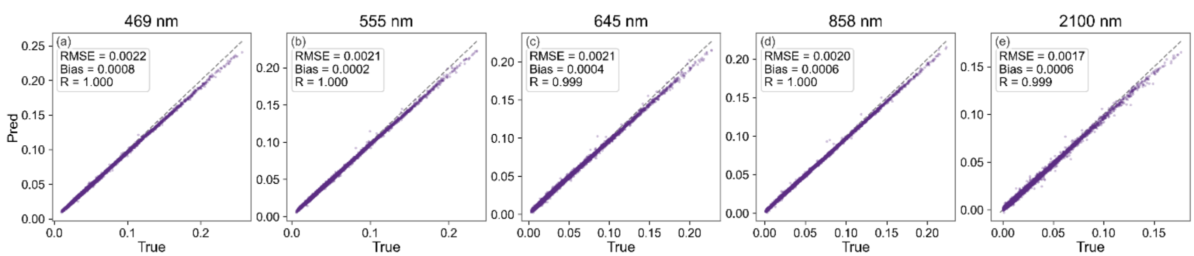

Figure 3. Validation of the band-wise surrogate radiative transfer emulator (MLRTM) against VLIDORT reference simulations. Each panel displays predicted versus reference TOA reflectance for one wavelength, shown left to right

in ascending order (469, 555, 645, 858, and 2100 nm). The black line marks the 1:1 correspondence. Insets report the root-mean-square error (RMSE), mean bias, and Pearson correlation coefficient (R) computed over the held-out test set of 10^4 samples.

3.3 Controlled simulation-based assimilation

Before confronting the spatial incompleteness and retrieval-related uncertainties inherent in real satellite observations, we verify the algorithm under controlled conditions in which the true aerosol and surface states are known. The experiment follows the design in Section 2.5: a background case is paired with pseudo-observations generated from an independent validation case via the integrated operator $\mathcal{H}$, and the assimilation minimizes the cost function. To evaluate the effectiveness of the assimilation framework, an idealized experiment was conducted in the limit of negligible observation error. Under this condition, the assimilation system effectively reduces to an inversion procedure with minimal influence from prior information. The experiment examines whether the resulting posterior distribution can successfully recover the true-state distribution.

Figure 4 presents the 469 nm AOD and surface albedo fields. The background AOD (top left) exhibits a pronounced underestimate over the central–eastern portion of the domain relative to the truth (top right), while the background surface albedo (down left) shows a systematic overestimate over the north portion of the domain. After latent space minimization, the analysis AOD (top middle) markedly reduces the magnitude of the underestimate, and the analysis surface albedo (down middle) is in closer agreement with the truth across most of the domain. The assimilation of multi-band reflectance effectively exploits the distinct spectral signatures of aerosols and surface albedo, enabling the two contributions to be disentangled and constrained simultaneously. Extended results for all five bands appear in the Supporting Information (Figs. S3 and S4 for AOD and surface, respectively). The AOD and surface albedo correction is consistent across bands in both the spatial pattern of the background bias and the structure of the analysis increment.

https://doi.org/10.5194/egusphere-2026-4627
Preprint. Discussion started: 14 September 2026

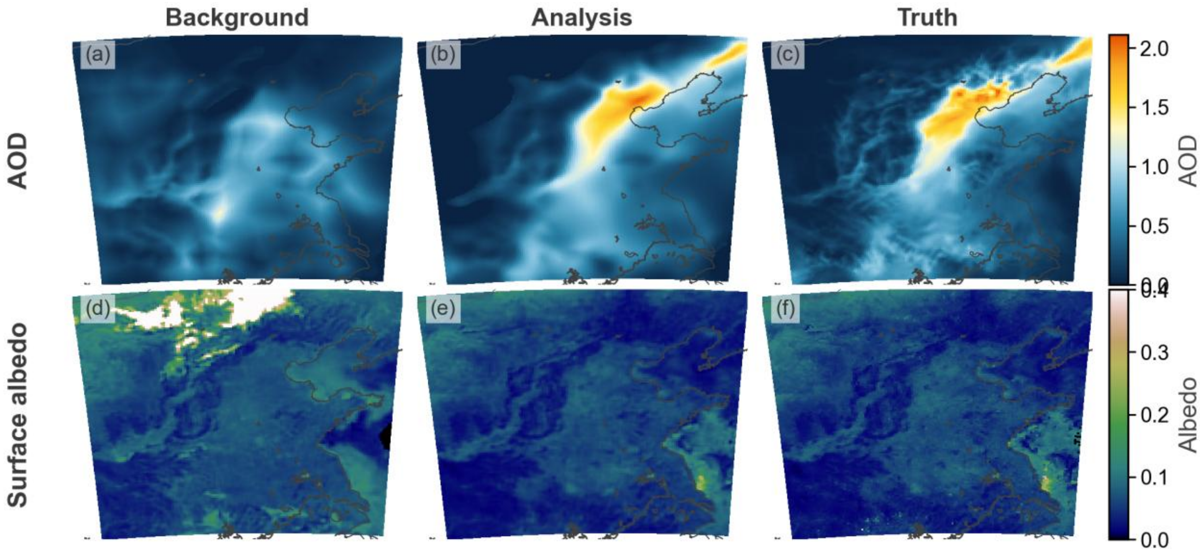

Figure 4. Controlled simulation-based LARDAS at 469 nm. Upper row: aerosol optical depth (AOD). Lower row: surface albedo. Left column: decoded background fields. Middle column: analysis fields after latent space data assimilation. Right column: truth fields.

The relative weight of the background and observation terms in the cost function is controlled by the assumed observation error standard deviation σ_{obs}. To evaluate the robustness of the assimilation system under different assumptions regarding observation uncertainty, a series of idealized experiments was conducted using a range of prescribed observation-error values in the assimilation system. The assimilation performance was then assessed to determine the system's ability to accommodate different observation error conditions. Figure 5 shows the statistic metrics FSS, bias and RMS between analysis-background and analysis-truth which rely on σ_{obs}.

https://doi.org/10.5194/egusphere-2026-4627
Preprint. Discussion started: 14 September 2026

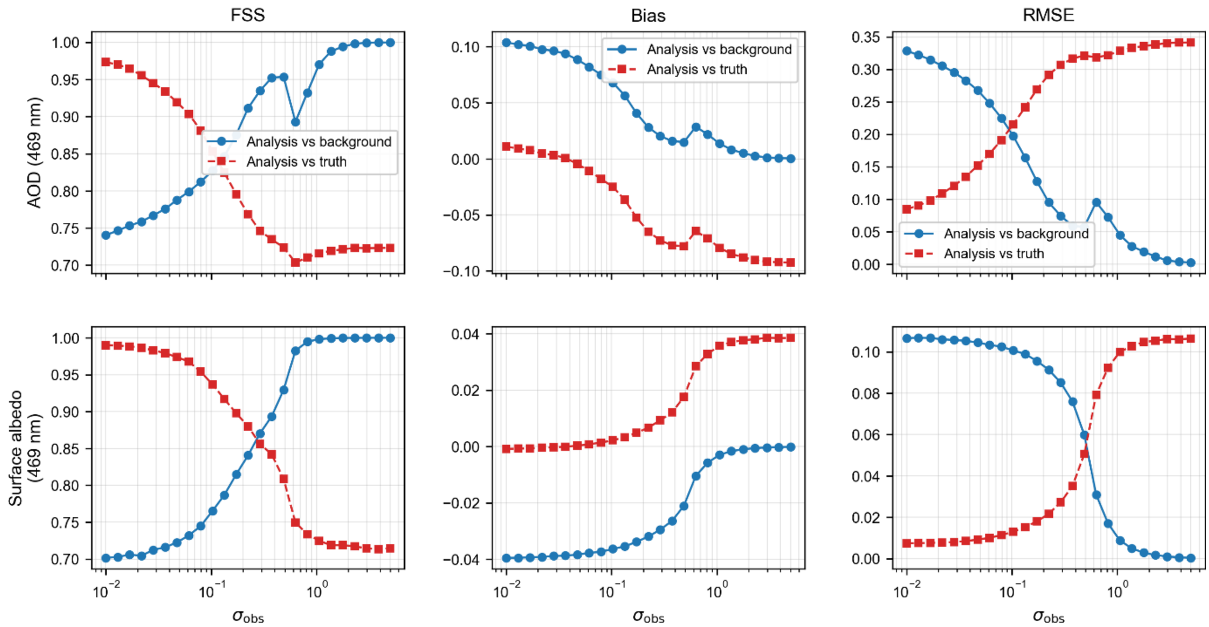

Figure 5. Sensitivity of the assimilated 469 nm AOD (top row) and surface albedo (bottom row) to the assumed observation-error standard deviation, which varies logarithmically from 0.01 to 5. The panels show the Fractions Skill Score (FSS; left, 50th-percentile threshold with a 5 × 5 neighborhood), bias (center), and RMSE (right). Red curves denote the comparison between the analysis and the truth, while blue curves represent the comparison between the analysis and the background state.

Figure 5 illustrates the sensitivity of the assimilation results to the assumed observation-error standard deviation. As expected from the variational formulation, increasing σ_{obs} reduces the relative weight of the observation term in the cost function and increases the influence of the background constraint. Consequently, the analysis gradually transitions from being observation-dominated to background-dominated. This behavior is consistently reflected in all three performance metrics. For both AOD and surface albedo, the analysis exhibits high agreement with the truth when σ_{obs} is small, as indicated by large FSS values and small bias and RMSE. As σ_{obs} increases, the analysis progressively approaches the background state, leading to improved analysis-versus-background statistics but degraded analysis-versus-truth statistics. The smooth crossover between the red and blue curves demonstrates that the assimilation framework responds appropriately to changes in the prescribed observation uncertainty.

The response of the surface-albedo assimilation is largely monotonic with increasing σ_{obs}, reflecting the relatively stable and well-constrained nature of the surface state. In contrast, the AOD assimilation exhibits localized non-monotonic behavior, where temporary fluctuations can be seen in the FSS, bias, and RMSE curves. This behavior suggests that the partitioning of observational information between aerosol and surface corrections is not always optimal in certain regions. Under these

https://doi.org/10.5194/egusphere-2026-4627
Preprint. Discussion started: 14 September 2026

conditions, a disproportionate fraction of the reflectance innovation may be attributed to adjustments in surface albedo, reducing the amount of information used to constrain the aerosol field. Because the reflectance sensitivity to AOD is weaker in these areas, excessive modification of the surface component can lead to oscillations in the retrieved aerosol state and consequently to non-monotonic variations in the AOD assimilation metrics. Nevertheless, the overall trends remain physically consistent, and the analysis retains a clear transition from observation-constrained to background-constrained solutions as the assumed observation error increases.

3.4 Real-observation assimilation

We now apply LARDAS to real MODIS TOA reflectance over the study domain, using the observations and quality control procedures described in Section 2.4. We first present a single-cycle assimilation with observation error ($\sigma_{\mathrm{obs}} = 0.02$) to verify the end-to-end functionality of the LARDAS pipeline on real satellite data, and subsequently validate the analysis against independent AERONET direct-sun AOD measurements over multiple months.

Spatial diagnostics of single assimilation cycles

Figure 6 illustrates a representative assimilation case at 03:00 UTC on 3 February 2016. Compared with the MODIS observations (Figure 6c), the background forward simulation (Figure 6a) systematically underestimates top-of-atmosphere reflectance over the central part of the domain, particularly between approximately 113–118°E and 35–38°N. Following assimilation, the simulated reflectance (Figure 6b) becomes substantially closer to the observed MODIS reflectance pattern, indicating that the analysis successfully reduces the radiance mismatch within the observed swath. The corresponding state adjustments are shown in the lower two rows of Figure 6. Relative to the background AOD field (Figure 6i), the analysis AOD (Figure 6j) exhibits pronounced increases over the regions sampled by MODIS, with the largest increments exceeding 0.4–0.5 (Figure 6l). The spatial distribution of these positive AOD increments closely follows the areas where the observed reflectance exceeds the background simulation, suggesting that the assimilation attributes the reflectance deficit primarily to insufficient aerosol loading in the background state. Outside the observational coverage, only limited adjustments are introduced and the analysis remains close to the background, consistent with the background constraint imposed in unobserved regions. In contrast, the surface albedo adjustments are comparatively small. The analysis surface albedo field (Figure 6f) remains similar to the background estimate (Figure 6e), with changes generally confined to regions constrained by MODIS observations (Figure 6h). These results indicate that the reflectance discrepancy in this case is explained predominantly through corrections to aerosol loading rather than large modifications of the surface state. The ability to simultaneously optimize aerosol and surface variables while maintaining physically plausible adjustments demonstrates that the integrated observation operator effectively separates aerosol and surface contributions to the top-of-atmosphere reflectance signal. Additional assimilation cases with different MODIS overpass geometries are presented in the Supporting Information. These cases show similarly consistent reductions in reflectance mismatch and corresponding improvements in

https://doi.org/10.5194/egusphere-2026-4627
Preprint. Discussion started: 14 September 2026

the analyzed aerosol fields, indicating that the assimilation framework performs robustly under varying observational coverage conditions.

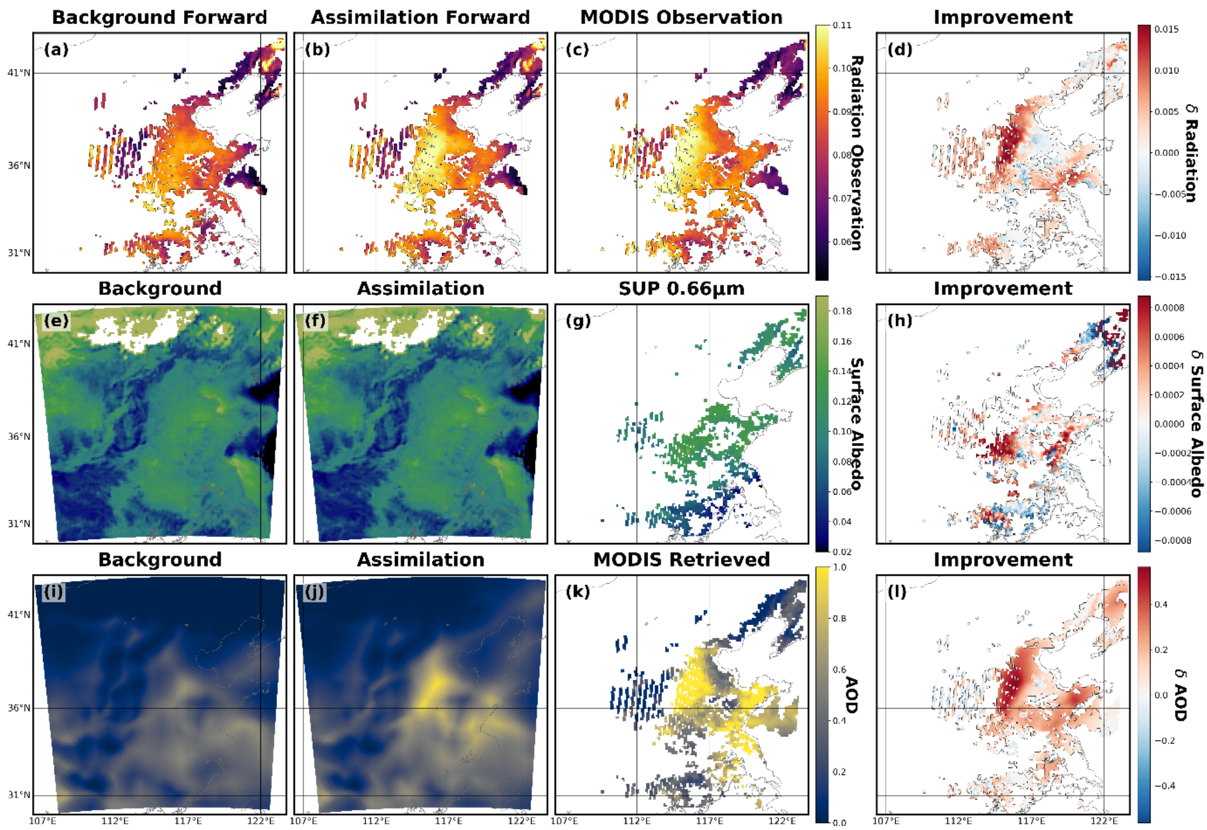

Figure 6. Example assimilation results for the MODIS overpass at 03:00 UTC on 3 February 2016. Panels (a–d) show the top-of-atmosphere reflectance at 555nm from the background forward simulation, assimilation forward simulation, MODIS observations, and the reflectance increment (assimilation minus background), respectively. Panels (e–h) show the background surface albedo, analyzed surface albedo, reference surface albedo, and the corresponding surface albedo increment. Panels (i–l) show the background aerosol optical depth (AOD), analyzed AOD, MODIS-retrieved AOD, and the corresponding AOD increment. Positive increments indicate increases relative to the background state. MODIS data courtesy of NASA LP DAAC.

Time-series validation against AERONET

The single-cycle experiments demonstrate that the proposed LARDAS can successfully assimilate MODIS reflectance observations and generate physically consistent aerosol increments. To evaluate whether these analysis corrections persist under a more realistic error specification, multi-day assimilation experiments were performed and validated against

https://doi.org/10.5194/egusphere-2026-4627
Preprint. Discussion started: 14 September 2026

independent AERONET observations. After each assimilation cycle, the analyzed multi-band AOD fields were mapped to WRF-Chem aerosol concentrations using the EnKF-based update procedure of Zhang et al. (2026), and short-range forecasts were subsequently initialized from the updated state.

Figure 7 compares AOD at 500 nm from the WRF-Chem background and three assimilation configurations—physical-space variational radiance assimilation (Phys_RAD), latent space assimilation of multi-band MODIS top-of-atmosphere reflectance (LARDAS_RAD), and latent space assimilation of gridded MODIS aerosol products (LARDAS_AOD)—against independent AERONET observations at four sites (Beijing, Taihu, Xianghe, and Xuzhou-CUMT) during 1–10 February 2016. Himawari-8 AOD at 500 nm is shown for context but is not assimilated. Model time series are aligned to the background simulation clock and collocated with AERONET before computing verification statistics.

At Beijing, the background RMSE and bias are 0.135 and −0.101, respectively. LARDAS_RAD achieves the best performance (RMSE 0.091, bias +0.014), corresponding to a 32.6% reduction in RMSE relative to the background, while LARDAS_AOD (RMSE 0.143, bias +0.065; +5.9%) and Phys_RAD (RMSE 0.145, bias +0.047; +7.4%) remain slightly worse than the background on RMSE. At Taihu, all three assimilation methods reduce errors relative to the background (RMSE 0.403, bias −0.261): LARDAS_RAD performs best (RMSE 0.301, bias −0.082; −25.3%), followed by Phys_RAD (RMSE 0.322, bias −0.069; −20.1%) and LARDAS_AOD (RMSE 0.332, bias −0.190; −17.6%). At Xianghe, substantial improvements are seen across all methods compared with the background (RMSE 0.221, bias −0.173). Phys_RAD provides the lowest RMSE (0.114, bias −0.035; −48.4%), closely followed by LARDAS_RAD (RMSE 0.119, bias +0.002; −46.2%), while LARDAS_AOD is slightly less accurate (RMSE 0.129, bias +0.051; −41.6%). At Xuzhou-CUMT, Phys_RAD clearly outperforms all other experiments (RMSE 0.206, bias −0.042; −49.5%), substantially improving over the background (RMSE 0.408, bias −0.233). In contrast, both LARDAS_RAD (RMSE 0.472, bias −0.284; +15.7%) and LARDAS_AOD (RMSE 0.495, bias −0.276; +21.3%) degrade performance relative to the background, indicating limited observational constraint from the latent space configurations at this site. Multi-wavelength validation against AERONET direct channels at 440, 675, and 870 nm (Supporting Information Figs. S6–S8) shows a broadly similar pattern. At Taihu and Xianghe, all three configurations generally reduce RMSE; at Beijing, improvement is wavelength-dependent (LARDAS_RAD at 440 nm; all three at 675 and 870 nm in Supporting Information Figs. S6–S8). At Xuzhou-CUMT, Phys_RAD remains the only configuration that consistently improves upon the background across all four wavelengths, whereas LARDAS_RAD and LARDAS_AOD degrade during the 5 February episode—consistent with the limited constraint seen at 500 nm in Figure 7. Overall, LARDAS_RAD shows the strongest performance at Beijing and Taihu, Phys_RAD achieves the lowest RMSE at Xianghe and is the only configuration that improves results at Xuzhou-CUMT, and LARDAS_AOD is generally intermediate between the two radiance-based routes but does not reach the best skill at any site during this window.

https://doi.org/10.5194/egusphere-2026-4627
Preprint. Discussion started: 14 September 2026

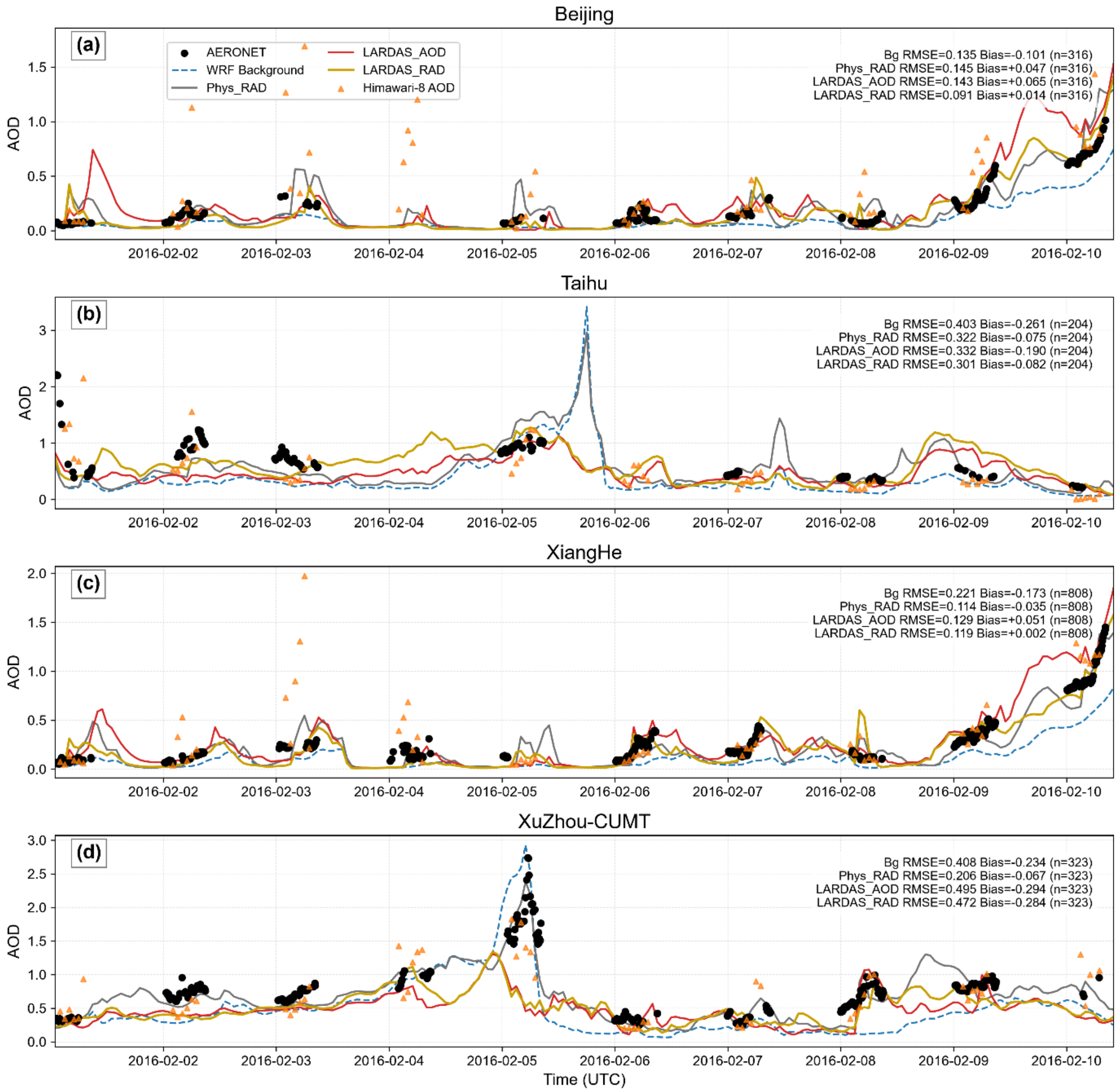

Figure 7. Comparison of aerosol optical depth (AOD) at 500 nm from the 1–10 February 2016 assimilation window with independent AERONET observations at four sites. Panels (a)–(d) show Beijing, Taihu, Xianghe, and Xuzhou-CUMT. Blue dashed lines: WRF-Chem background AOD. Grey solid lines: physical-space variational radiance assimilation (Zhang et al., 2026). Red solid lines: latent space assimilation of gridded MODIS aerosol products. Deep yellow solid lines: latent space assimilation of multi-band MODIS top-of-atmosphere reflectance. Black dots: AERONET direct-sun AOD at 500 nm. Orange triangles: Himawari-8 AOD at 500 nm (shown for context only; not assimilated). Model time series are collocated to AERONET and aligned to the background simulation clock before

https://doi.org/10.5194/egusphere-2026-4627
Preprint. Discussion started: 14 September 2026

plotting. Root-mean-square error (RMSE), mean bias, and sample size n relative to AERONET are listed in the upper-right corner of each panel.

Flux validation against CERES

To assess whether the analyzed aerosol and surface-albedo fields improve shortwave radiative fluxes, the WRF-Chem background and the three assimilation experiments—physical-space radiance assimilation (Phys_RAD), latent reflectance assimilation (LARDAS_RAD), and latent MODIS AOD product assimilation (LARDAS_AOD)—were evaluated against independent CERES SYN1deg hourly observations on the inner nest (d02) during 1–10 February 2016. CERES fields were regridded onto the WRF d02 grid, and model–observation statistics were computed hour by hour over all collocated grid points under all-sky conditions (i.e., without additional cloud screening beyond the CERES product and collocation). Only local daytime samples (08:00–17:00, UTC+8) were retained; nighttime hours are omitted from the statistics.

Figure 8 summarizes the resulting validation using the spatial mean absolute bias (MAB) between each simulation and CERES for surface downward shortwave (SWDNB), surface upward shortwave (SWUPB), and top-of-atmosphere upward shortwave (SWUPT). At each hour, this metric averages |model - CERES| over all collocated d02 cells and therefore characterizes the spatial fidelity of the simulated shortwave fluxes rather than a single domain-mean flux time series. Inset text reports the daytime mean MAB for the background and each assimilation experiment.

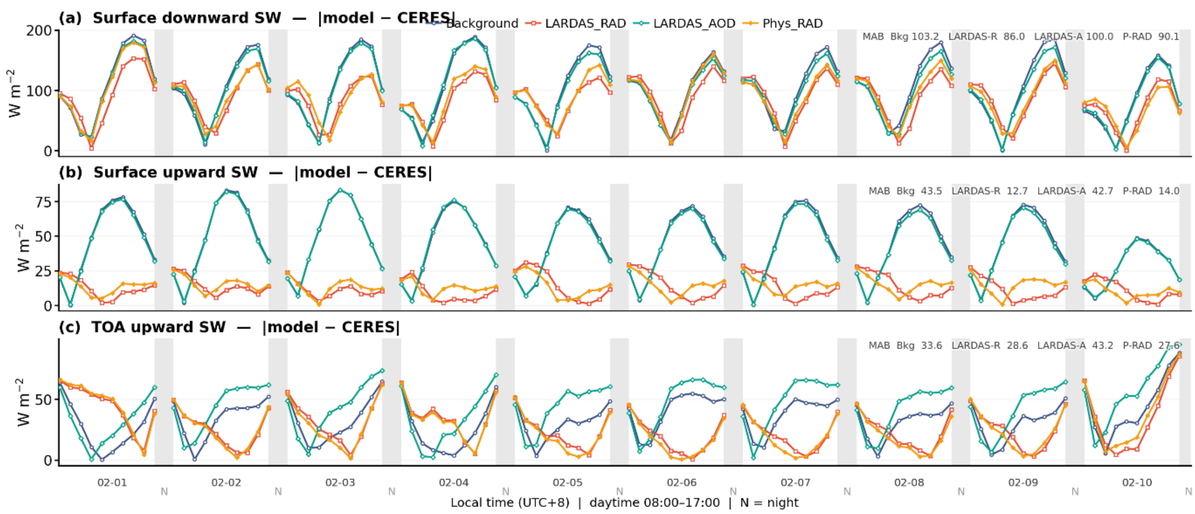

Figure 8. All-sky validation of shortwave radiation on the WRF d02 domain against CERES SYN1deg hourly observations during 1–10 February 2016. At each hour, the spatial mean absolute bias (MAB) between model and CERES, |model - CERES|, is computed over all collocated grid points and plotted for local daytime (08:00–17:00, UTC+8); grey shading marks nighttime (N). (a) Surface downward shortwave (SWDNB), (b) surface upward

https://doi.org/10.5194/egusphere-2026-4627
Preprint. Discussion started: 14 September 2026

shortwave (SWUPB), and (c) top-of-atmosphere upward shortwave (SWUPT). Blue circles: WRF-Chem background; red squares: LARDAS_RAD (latent reflectance assimilation); teal diamonds: LARDAS_AOD (latent MODIS AOD assimilation); orange stars: Phys_RAD (physical-space radiance assimilation). Inset text gives the daytime mean MAB for each experiment (Bkg, L-RAD, L-AOD, P-RAD).

The LARDAS_RAD experiment shows the clearest overall improvement relative to the background and LARDAS_AOD, with Phys_RAD also reducing biases for most components. For surface downward shortwave (SWDNB), the daytime mean all-sky grid MAB decreases from 103.2 W m^{-2} (background) to 86.0 W m^{-2} (LARDAS_RAD), a 16.7% reduction; Phys_RAD yields 90.1 W m^{-2} (12.7% reduction), whereas LARDAS_AOD remains near the background at 100.0 W m^{-2} (3.0% reduction). The largest and most diagnostic gain appears for surface upward shortwave (SWUPB): LARDAS_RAD reduces the grid MAB from 43.5 to 12.7 W m^{-2} (70.8% reduction), and Phys_RAD to 14.0 W m^{-2} (67.8% reduction). By contrast, LARDAS_AOD changes SWUPB MAB only from 43.5 to 42.7 W m^{-2} (1.8% reduction). This contrast is central to interpreting the radiance experiments: because LARDAS_RAD and Phys_RAD assimilate multi-band TOA reflectance and jointly adjust aerosol loading and band-resolved surface albedo, they directly correct the background overestimate of reflected shortwave at the surface; aerosol-only assimilation (LARDAS_AOD) modifies atmospheric scattering but leaves surface reflection largely unchanged, so upward surface fluxes—and therefore SWUPB—remain close to the background.

The SWUPB improvement therefore highlights a practical benefit of RAD-based assimilation beyond AOD skill alone: simultaneous surface optimization is required to bring simulated upward surface shortwave flux into closer agreement with CERES. For top-of-atmosphere upward shortwave (SWUPT), improvements are more modest and less stable hour to hour under all-sky conditions. The grid MAB decreases from 33.6 W m^{-2} (background) to 28.6 W m^{-2} (LARDAS_RAD; 14.9% reduction) and 27.6 W m^{-2} (Phys_RAD; 17.9% reduction), whereas LARDAS_AOD yields a larger departure of 43.2 W m^{-2} (28.6% increase relative to the background). This pattern reflects that TOA fluxes remain sensitive to cloud placement and cloud optical properties, which are not directly constrained by reflectance assimilation; nevertheless, the reflectance experiments still improve TOA statistics relative to both the background and LARDAS_AOD, with Phys_RAD attaining the lowest TOA MAB in this period and LARDAS_RAD nearly as skillful. Together, the results show that reflectance assimilation (LARDAS_RAD and Phys_RAD) improves shortwave radiation most strongly where surface albedo is co-optimized, above all for upward surface flux (SWUPB, ~68–71% MAB reduction), while aerosol-only assimilation produces much smaller surface changes and less reliable TOA statistics. Supplementary Figure S9 provides a complementary domain-mean all-sky view of the same period: hourly SWDNB, SWUPB, and SWUPT are area-averaged over d02 and plotted against CERES, so that readers can inspect the flux evolution directly rather than through the grid MAB metric used in Figure 8.

https://doi.org/10.5194/egusphere-2026-4627
Preprint. Discussion started: 14 September 2026

To test whether the improvement seen in the February 2016 window generalizes to other seasons and aerosol regimes, we evaluate station-wise AOD time series from the LARDAS radiance continuous assimilation configuration at six AERONET sites (three in the Beijing area, Xianghe, Xuzhou, and Taihu) for February, March, April, and October 2017. Figure 9 compares the WRF-Chem background and the post-assimilation model simulation. In all four months and at all six stations, the assimilated series (red) agrees more closely with AERONET (black) than the background (blue), including during elevated-AOD episodes when the background shows the largest departures. Quantitatively, assimilated RMSE and bias in the insets are lower than the background values in every panel, so error reduction is systematic rather than limited to a few days or stations. February–April cover late-winter to spring haze conditions; October represents an autumn aerosol regime. Consistent improvement across months and locations shows that coupling reflectance latent space assimilation updates to aerosol concentrations yields robust, repeatable error reduction under different seasonal and regional conditions, without re-tuning latent assimilation or observation-error settings for each month.

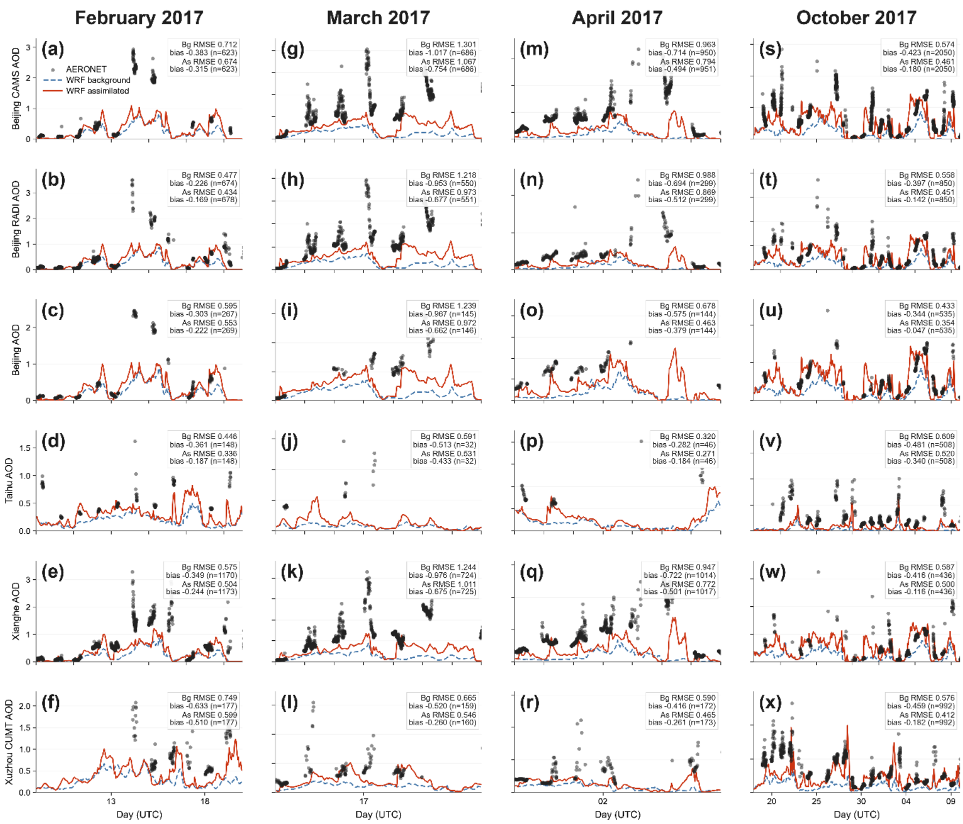

Figure 9. Independent validation of aerosol optical depth (AOD) at 500 nm against AERONET direct-sun measurements at six stations during the continuous assimilation experiment. Rows (top to bottom) correspond to Beijing CAMS, Beijing RADI, Beijing, Taihu, Xianghe, and Xuzhou CUMT; columns (left to right) show February, March, April, and October 2017 (UTC). Subpanels are labeled (a)–(x) in column-major order. Black circles denote

https://doi.org/10.5194/egusphere-2026-4627
Preprint. Discussion started: 14 September 2026

AERONET observations. The blue dashed curve is WRF-Chem background AOD collocated to the station; the red solid curve is WRF-Chem AOD after latent space MODIS shortwave radiance assimilation, EnKF mapping to aerosol concentrations, and short-range forward integration. The symbol legend in panel (a) applies to all panels. Upper-right insets report background (Bg) and assimilated (As) RMSE and bias relative to AERONET, and the number of collocated samples (n).

3.5 Computational efficiency

A primary motivation for performing data assimilation in the latent space is the substantial reduction in computational cost. The dimensionality of the control vector is reduced from 2×5×141×141 = 198,810 state variables in physical space (multi-band AOD and surface albedo at every grid point) to d_a+d_s=256 (latent variables). In addition, the integrated observation operator replaces computationally expensive radiative transfer calculations with a single forward pass through a neural-network surrogate. Figure 10 summarizes the February 2016 three-method intercomparison in terms of assimilation cost, independent AERONET skill, and all-sky shortwave flux validation against CERES; the observation-operator benchmark is discussed separately below because it is not part of that assimilation intercomparison.

Observation operator. Before comparing assimilation configurations, we benchmark the forward model alone. A conventional VLIDORT-based radiative-transfer evaluation (Trad-Obs) requires approximately 0.9 s per case on a single Intel Xeon 6330 CPU core. The neural-network observation operator (NN-Obs) requires only 9.1×10^{-8} s per case on an NVIDIA GeForce RTX 3060 GPU, i.e. a speedup of several orders of magnitude. The NN-Obs timing was obtained by averaging over a very large ensemble (10^6) of forward evaluations executed in batch mode, so that GPU initialization, memory transfer, and other fixed overheads were effectively amortized. This offline acceleration is a prerequisite for affordable online assimilation.

Assimilation cost. When the same NN-Obs is used in the online loop, physical-space variational radiance assimilation requires approximately 24.8 s per iteration on the Xeon 6330 CPU, whereas latent reflectance assimilation requires 0.03 s per iteration and latent MODIS AOD product assimilation requires 0.02 s per iteration on the RTX 3060 GPU—roughly three orders of magnitude faster per iteration than the physical-space configuration. The gain arises primarily from reducing the control vector from $O(10^5)$ physical grid points to 256 latent degrees of freedom and from end-to-end automatic differentiation through the VAE–surrogate chain, which substantially decreases the cost of gradient evaluation and optimization. Latent AOD product assimilation is marginally faster than latent reflectance assimilation because the observation term is restricted to a single AOD band and does not invoke the full five-band MLRTM pass at every iteration. Most computational expense is shifted offline to one-time training of the VAE encoders/decoders, refinement network, and

https://doi.org/10.5194/egusphere-2026-4627
Preprint. Discussion started: 14 September 2026

radiative-transfer surrogate; the online analysis stage remains computationally lightweight and could, in principle, support near-real-time assimilation of multi-spectral satellite reflectance for regional aerosol–surface joint estimation.

4 Discussion and Conclusions

4.1 Discussions

This study presents LARDAS as an end-to-end cycling aerosol–radiance data assimilation system that links learned aerosol–surface representations, a differentiable radiative-transfer observation operator, latent-space analysis, model-variable adjustment, and WRF-Chem forecast cycling. The experiments evaluate both the latent analysis algorithm and its behavior within the complete cycling system. Section 3 presents component validation, controlled experiments, and case-by-case real-observation diagnostics, including AERONET time-series verification and all-sky shortwave flux comparisons against CERES. Section 3.5 quantifies the observation-operator benchmark and the large reduction in online assimilation cost when the control vector is moved from the physical grid to the latent space. Here we synthesize the February 2016 three-method intercomparison, interpret those findings relative to retrieved-product assimilation and to the companion physical-space radiance system of Zhang et al. (2026), note architectural advantages, and discuss principal limitations and extensions.

Overall intercomparison. Figure 10 consolidates the February 2016 evaluation of the WRF-Chem background and the three assimilation experiments—physical-space radiance assimilation (Phys_RAD), latent reflectance assimilation (LARDAS_RAD), and latent MODIS AOD product assimilation (LARDAS_AOD)—in a single view.

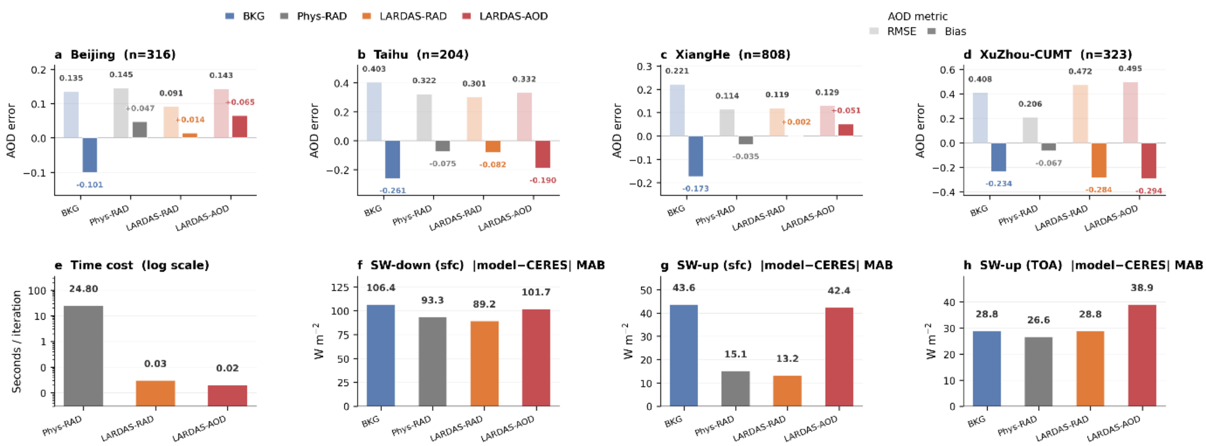

Figure 10. Summary intercomparison of the WRF-Chem background (BKG) and three assimilation experiments—physical-space radiance assimilation (Phys_RAD), latent reflectance assimilation (LARDAS_RAD), and latent MODIS AOD product assimilation (LARDAS_AOD)—for the February 2016 intensive window (1–10 February). (a–

d) Independent AERONET validation of 500 nm AOD at four sites: Beijing (n = 316), Taihu (n = 204), Xianghe (n = 808), and Xuzhou_CUMT (n = 323). Light and dark bars denote RMSE and bias, respectively. (e) Assimilation cost per iteration (seconds; logarithmic axis) when the neural-network observation operator is used in the online loop. (f–h) All-sky validation of shortwave radiative fluxes against CERES SYN1deg hourly observations on WRF d02, expressed as daytime mean spatial grid mean absolute bias (MAB; W m^{-2}): (f) surface downward shortwave (SWDNB), (g) surface upward shortwave (SWUPB), and (h) top-of-atmosphere upward shortwave (SWUPT). Colors denote BKG (blue), Phys_RAD (grey), LARDAS_RAD (orange), and LARDAS_AOD (red) throughout.

Panels (a–d) summarize independent AERONET skill at 500 nm AOD for the four primary validation sites. Latent reflectance assimilation attains the lowest or near-lowest RMSE at Beijing and Taihu, and near-lowest at Xianghe, whereas physical-space radiance assimilation is most skillful at Xianghe and Xuzhou_CUMT; latent AOD product assimilation is generally less competitive than the two radiance experiments and remains close to the background at several sites. Panel (e) places these skill differences in computational context. Phys_RAD requires on the order of tens of seconds per iteration, whereas LARDAS_RAD and LARDAS_AOD require only hundredths of a second—roughly three orders of magnitude faster—so the latent experiments achieve most of the radiance-assimilation benefit at a fraction of the online cost already documented in Section 3.5. Panels (f–h) extend the comparison to all-sky shortwave fluxes against CERES on WRF d02. Both radiance experiments reduce surface-flux biases relative to the background and to LARDAS_AOD; the clearest separation appears for surface upward shortwave (panel g), where LARDAS_RAD and Phys_RAD reduce the mean absolute bias by roughly 65–70%, whereas LARDAS_AOD remains near the background. This confirms that correcting reflected shortwave at the surface requires joint adjustment of aerosol and band-resolved surface albedo, not aerosol updates alone. Surface downward shortwave (panel f) also improves under the radiance experiments, though less dramatically than SWUPB. Top-of-atmosphere upward shortwave (panel h) shows more modest and variable gains, with Phys_RAD slightly best and LARDAS_AOD poorest, reflecting residual sensitivity to cloud placement and cloud optical properties that are not directly constrained by reflectance assimilation under all-sky conditions.

Taken together, Figure 10 shows that LARDAS_RAD offers the most favorable overall balance in this experiment: analysis skill comparable to Phys_RAD at most AERONET sites, a large improvement in surface reflected shortwave, and online cost reduced by roughly three orders of magnitude. Phys_RAD remains the strongest configuration at individual sites and for TOA statistics in this period, while LARDAS_AOD provides only limited benefit for surface radiation and TOA fluxes despite its low computational cost. The near-background performance of LARDAS_AOD for surface upward shortwave (Figure 10g) motivates a closer comparison with retrieved AOD assimilation.

Comparison with retrieved-product (AOD) assimilation. A common route assimilates satellite-retrieved AOD through a simple observation operator, but retrieval assumptions can differ from the chemistry–transport model and add representativeness error. Here, multi-spectral TOA reflectance is assimilated directly through $\mathscr{H}$, jointly constraining aerosol and surface states (Figure 1). As Figure 10 already shows, latent AOD product assimilation is less skillful than the two radiance experiments for both AERONET and surface shortwave fluxes, especially reflected flux at the surface, because aerosol-only updates modify scattering without adequately correcting band-resolved surface albedo. Multi-band reflectance therefore carries more usable information once retrieval-stage assumptions are bypassed.

Comparison with physical-space shortwave assimilation. The main contrast with Zhang et al. (2026) is control space and minimization pathway, not observations or downstream EnKF mapping (Table 2). With the same MODIS reflectance, latent and physical-space radiance assimilation achieve broadly similar AERONET skill (Figure 10a–d), but latent reflectance assimilation reduces online cost by roughly three orders of magnitude (Figure 10e; Section 3.5). The remaining difference is mainly VAE-induced local smoothing, which likely limits gains in some radiative flux components (Figure 10 f, h) even when AOD verification is comparable. LARDAS is thus a practical alternative when online cost is limiting, rather than a fundamentally different observational constraint.

Architectural advantages. The LARDAS pipeline is end-to-end differentiable: reflectance misfits back-propagate through the surrogate RTM, refinement, and decoders to the latent control variables without assembling explicit Jacobians. This property is well suited to coupling with emerging aerosol forecast models implemented as deep networks, and to building latent four-dimensional variational (4D-Var) systems in which the entire analysis step—representation, forward model, and observation terms—resides in a single computational graph. Observation handling in the present implementation is also comparatively simple: Level-1B reflectance is gridded, screened using retrieval-quality flags as a proxy for clear-sky suitability (Section 2.4). Extending the system to richer shortwave observations—for example, polarized radiances or multi-angle measurements from instruments such as POLDER, MISR, DPC, PACE, SPEXONE or future geostationary imagers—does not require reformulating the assimilation algorithm. In principle, one replaces or augments the observation operator while retaining the same latent variational outer loop. The observation operator therefore acts as a modular interface between increasingly complex measurement types and a fixed low-dimensional analysis space.

Limitations.

Several limitations should be borne in mind when interpreting the results.

Representation fidelity. The VAE–refinement representation preserves domain-scale aerosol patterns and high FSS on held-out samples (Figure 2), but thin filaments and sharp local peaks are occasionally smoothed (Supporting Information Figure S1). Assimilation cannot restore detail absent from the decoder manifold, so applications that require neighborhood-scale accuracy—for example, isolating a narrow plume against a clean background—may remain challenging until the representation is improved or augmented with higher-resolution local refinements.

https://doi.org/10.5194/egusphere-2026-4627
Preprint. Discussion started: 14 September 2026

Sparse observations and global latent controls. Latent variables encode spatially aggregated information through the convolutional encoder: each component of z_a and z_s can influence the full 141 × 141 domain after decoding. Classical localization techniques that damp long-range correlations in physical space do not transfer directly to this coordinate system. When MODIS reflectance is sparse in space, this can introduce spurious large-scale adjustments in an attempt to fit limited observational footprints. LARDAS is therefore less naturally suited to extremely sparse observation networks than physical-space schemes with explicit horizontal localization, unless future work introduces localized latent bases, regional tiling, or observation-selective masking of gradient pathways.

Diagonal latent background-error covariance. The present implementation approximates the aerosol and surface latent background-error covariances as diagonal matrices and neglects cross-covariances between the two components. Although KL regularization encourages compact and approximately decorrelated latent coordinates, it does not guarantee that forecast errors are statistically independent in latent space. The diagonal approximation may therefore omit dynamically or physically meaningful error relationships, including correlations among spatial scales and between aerosol loading and surface properties. Future work should evaluate structured, low-rank, ensemble-derived, or flow-dependent latent covariance models.

AOD-to-model-state mapping. LARDAS analyzes column-integrated multi-band AOD, whereas WRF-Chem forecasts size- and composition-resolved aerosol mass and number concentrations. The EnKF-based mapping used here distributes the AOD increment among MOSAIC variables according to ensemble covariances and the existing vertical, size-bin, and compositional structures. Because different aerosol profiles and mixtures can produce similar column AOD, this mapping is not unique and may introduce uncertainty into the initialized chemical state. Future studies should assess sensitivity to the assumed aerosol vertical distribution and composition and should evaluate the resulting forecast increments against vertically resolved and speciated aerosol observations.

Regional and temporal scope. The real-observation experiments focus on eastern China, with the intensive three-method comparison restricted to February 2016 and the continuous evaluation covering four months in 2017 at six AERONET stations. These experiments sample several aerosol and meteorological regimes but do not establish general performance across other regions, seasons, surface types, or aerosol environments. In particular, performance over bright surfaces, biomass-burning regions, dust-dominated conditions, and areas with persistent cloud cover remains uncertain. Broader multi-year and multi-region experiments are required to assess the transferability and operational robustness of LARDAS.

Distributional distinctness of cycling analyses and background. February 2016 was withheld from the aerosol VAE training archive and reserved for the independent assimilation experiment (Section 2.2). To further establish that the cycling LARDAS_RAD analyses are not drawn from the same spatial distribution as the concurrent WRF-Chem background, we compared daily-mean column AOD at 400 nm over domain d02 for 28 January–10 February 2016 (Supporting Information

https://doi.org/10.5194/egusphere-2026-4627
Preprint. Discussion started: 14 September 2026

Fig. S10). Diagnostics were first aggregated to daily means (n=14) and then assessed with ordinary t-tests. The two AOD fields retain moderate large-scale similarity under shared meteorological forcing (daily-mean spatial correlation r=0.722), as expected for simulations on a common domain. However, their high-AOD cores and overall loadings differ systematically: Pattern FSS at 10 km for the highest 15%, 10%, and 5% regions averaged 0.526, 0.445, and 0.321, each significantly below the corresponding useful-skill benchmark (one-sided t-tests: p=0.024,0.001, and p<0.001); hotspot intersection-over-union averaged only 0.367, 0.298, and 0.205 (all p<0.001 versus 0.5); and domain-mean AOD differed by 0.259 (paired t=10.89,p<0.001; LARDAS_RAD 0.427 versus background 0.168). These results show that, although the analyses and background share a common meteorological envelope, their AOD amplitude and high-value spatial patterns are not identically distributed. This interpretation is consistent with held-out VAE reconstruction, where residual discrepancies are likewise concentrated in high-AOD structures (Section 3.1; Supporting Information Fig. S1). Correspondingly, a key limitation of the present algorithm is that assimilation remains confined to the VAE–refinement manifold: high-AOD cores can be shifted and amplified relative to the background, yet sharp local peaks and thin filaments that are poorly represented in latent space cannot be fully recovered.

4.2 Conclusions

We have developed the Latent-space Aerosol–Radiance Data Assimilation System (LARDAS), an end-to-end cycling framework for jointly constraining multi-band aerosol optical depth and surface albedo from MODIS shortwave reflectance and propagating the aerosol analysis into WRF-Chem forecasts.

1. The VAE-based representation, together with the aerosol refinement network, compresses five-band aerosol and surface fields on the 141 × 141 grid into 128-dimensional latent variables per component while retaining the dominant spatial and spectral structures. Median FSS values exceed 0.95 for AOD in the first four bands and 0.98 for surface albedo across all bands.
2. The band-wise MLRTM emulators reproduce VLIDORT reflectance with correlation coefficients of at least 0.999 and RMSE below 0.003. Controlled simulation-based experiments further demonstrate that the differentiable latent analysis can recover prescribed aerosol and surface states and exhibits the expected transition from observation-dominated to background-dominated solutions as the assumed observation error increases.
3. In the February 2016 real-observation experiment, LARDAS_RAD produces physically consistent joint aerosol–surface corrections, improves independent AERONET AOD verification at most sites relative to the background and LARDAS_AOD, and remains broadly competitive with Phys_RAD. Independent CERES evaluation indicates that the two reflectance-assimilation configurations improve the simulated surface shortwave radiation budget, with the clearest benefit appearing in surface upward shortwave flux because aerosol loading and band-resolved surface albedo are adjusted jointly.

https://doi.org/10.5194/egusphere-2026-4627
Preprint. Discussion started: 14 September 2026

4. Continuous LARDAS_RAD experiments for February, March, April, and October 2017 consistently reduce AERONET RMSE and bias relative to the WRF-Chem background. These results suggest temporal transferability across the sampled late-winter, spring, and autumn conditions within the eastern China study region, although broader spatial and seasonal evaluation remains necessary.
5. By combining low-dimensional analysis, a differentiable observation operator, EnKF-based aerosol-state adjustment, band-resolved surface updates, and WRF-Chem forecast cycling, LARDAS provides a computationally practical route for direct multi-band reflectance assimilation. Under the operational CPU/GPU configurations used here, the measured online wall-clock cost of LARDAS_RAD is approximately three orders of magnitude lower than that of Phys_RAD; this comparison represents end-to-end deployment performance rather than a hardware-normalized algorithmic benchmark.

Overall, LARDAS provides a computationally practical route for direct multi-band reflectance assimilation while retaining skill broadly comparable to physical-space radiance assimilation in the present regional experiments. Its principal advantage lies in integrating low-dimensional analysis, a differentiable observation operator, aerosol–surface joint estimation, model-state adjustment, and forecast cycling within a unified system. Further evaluation with broader regions and additional aerosol optical variables is required before operational application.

Code and data availability

The LARDAS v1.0 latent-space assimilation software—including pretrained neural-network weights, controlled OSSE case files, analysis scripts, WRF-Chem V4.4.2 source materials and configuration files, EnSRF / mixing-Mie modules for mapping analyzed AOD to MOSAIC concentrations, and example observation NetCDF files—is archived at Zenodo (Zhang, 2026a). A README in the deposit describes the contents of each file. The WRF-Chem physics, chemistry, and mixing-Mie workflow are further documented in Zhang et al. (2026); a copy of that companion preprint is included in the same Zenodo deposit. MODIS Terra/Aqua Level-1B calibrated radiances (MOD021KM/MYD021KM; https://doi.org/10.5067/MODIS/MOD021KM.061; https://doi.org/10.5067/MODIS/MYD021KM.061) and MODIS aerosol Level-2 products used for quality-assurance screening (MOD04_L2/MYD04_L2; https://doi.org/10.5067/MODIS/MOD04_L2.061; https://doi.org/10.5067/MODIS/MYD04_L2.061) are available from the NASA LAADS DAAC (https://ladsweb.modaps.eosdis.nasa.gov/). The MODIS MCD43C3 BRDF/albedo product is available from the NASA LP DAAC (https://doi.org/10.5067/MODIS/MCD43C3.061). AERONET Version 3 direct-sun AOD observations are available from https://aeronet.gsfc.nasa.gov/. CERES SYN1deg-1Hour Terra–Aqua Edition4A radiative fluxes can be obtained from the CERES ordering portal (https://ceres.larc.nasa.gov/data/; https://doi.org/10.5067/TERRA+AQUA/CERES/SYN1DEG-1HOUR_L3.004A).

https://doi.org/10.5194/egusphere-2026-4627
Preprint. Discussion started: 14 September 2026

Author contributions

Chongzhao Zhang contributed to the study conceptualization, developed the methodology, prepared the data and software, performed debugging and packaging, wrote the original draft, and reviewed and edited the manuscript. Qiurui Li contributed to the methodology, data preparation, and software. Jun Li and Wei Han reviewed and edited the manuscript. Jing Li contributed to the study conceptualization and methodology and reviewed and edited the manuscript.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgements

We thank Yueming Dong and Liang Chang for helpful discussions during the development of this work. This work is also supported by the High-performance Computing Platform of Peking University. Large language model tools were used to assist with English language polishing and to check for typographical and grammatical errors; the authors carefully reviewed and take full responsibility for the scientific content of the manuscript.

Financial support

This research has been supported by the National Natural Science Foundation of China (grant no. 42425503).

References

Adhikary, B., Kulkarni, S., Dallura, A., Tang, Y., Chai, T., Leung, L. R., et al. (2008). A regional scale chemical transport modeling of Asian aerosols with data assimilation of AOD observations using optimal interpolation technique. Atmospheric Environment, 42(37), 8600-8615.

Amendola, M., Arcucci, R., Mottet, L., Casas, C. Q., Fan, S., Pain, C., et al. (2021). Data assimilation in the latent space of a convolutional autoencoder. In M. Paszynski, D. Kranzlmüller, V. V. Krzhizhanovskaya, J. J. Dongarra, & P. M. A. Sloot (Eds.), Computational Science – ICCS 2021 (Lecture Notes in Computer Science, Vol. 12746, pp. 373–386). Springer. https://doi.org/10.1007/978-3-030-77977-1_30

Benedetti, A., Reid, J. S., Knippertz, P., Marsham, J. H., Di Giuseppe, F., Remy, S., et al. (2018). Status and future of numerical atmospheric aerosol prediction with a focus on data requirements. Atmospheric Chemistry and Physics, 18(14), 10615-10643.

https://doi.org/10.5194/egusphere-2026-4627
Preprint. Discussion started: 14 September 2026

Carslaw, K. S., Lee, L. A., Reddington, C. L., Mann, G. W., & Pringle, K. J. (2013). The magnitude and sources of uncertainty in global aerosol. Faraday Discussions, 165, 495-512.

Chen, F., & Dudhia, J. (2001). Coupling an Advanced Land Surface–Hydrology Model with the Penn State–NCAR MM5 Modeling System. Part I: Model Implementation and Sensitivity. Monthly Weather Review, 129(4), 569-585. https://journals.ametsoc.org/view/journals/mwre/129/4/1520-0493_2001_129_0569_caalsh_2.0.co_2.xml

Cheng, S., Quilodrán-Casas, C., Ouala, S., Farchi, A., Liu, C., Tandeo, P., et al. (2023). Machine learning with data assimilation and uncertainty quantification for dynamical systems: a review. IEEE/CAA Journal of Automatica Sinica, 10(6), 1361-1387.

Cheng, Y., Dai, T., Goto, D., Schutgens, N. A. J., Shi, G., & Nakajima, T. (2019). Investigating the assimilation of CALIPSO global aerosol vertical observations using a four-dimensional ensemble Kalman filter. Atmospheric Chemistry and Physics, 19(21), 13445-13467.

Clough, S. A., Shephard, M. W., Mlawer, E. J., Delamere, J. S., Iacono, M. J., Cady-Pereira, K., et al. (2005). Atmospheric radiative transfer modeling: a summary of the AER codes. Journal of Quantitative Spectroscopy and Radiative Transfer, 91(2), 233-244. https://www.sciencedirect.com/science/article/pii/S0022407304002158

Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. Journal of Geophysical Research: Oceans, 99(C5), 10143-10162.

Fan, H., Bai, L., Fei, B., Xiao, Y., Chen, K., Liu, Y., et al. (2026). Physically consistent global atmospheric data assimilation with machine learning in latent space. Science Advances, 12(1), eaea4248.

Fan, H., Liu, Y., Huo, Z., Liu, Y., Shi, Y., & Li, Y. (2025). A novel latent space data assimilation framework with autoencoder-observation to latent space (AE-O2L) network. Part I: The observation-only analysis method. Monthly Weather Review, 153(8), 1335-1348.

Fast, J. D., Gustafson Jr, W. I., Easter, R. C., Zaveri, R. A., Barnard, J. C., Chapman, E. G., et al. (2006). Evolution of ozone, particulates, and aerosol direct radiative forcing in the vicinity of Houston using a fully coupled meteorology-chemistry-aerosol model. Journal of Geophysical Research: Atmospheres, 111(D21). https://doi.org/10.1029/2005JD006721

Garrigues, S., Ades, M., Remy, S., Flemming, J., Kipling, Z., Laszlo, I., et al. (2023). Impact of assimilating NOAA VIIRS aerosol optical depth (AOD) observations on global AOD analysis from the Copernicus Atmosphere Monitoring Service (CAMS). Atmospheric Chemistry and Physics, 23(18), 10473-10487.

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://www.sciencedirect.com/science/article/pii/S1352231005003560

Guenther, A., Karl, T., Harley, P., Wiedinmyer, C., Palmer, P. I., & Geron, C. (2006). Estimates of global terrestrial isoprene emissions using MEGAN (Model of Emissions of Gases and Aerosols from Nature). Atmos. Chem. Phys., 6(11), 3181-3210. https://acp.copernicus.org/articles/6/3181/2006/

https://doi.org/10.5194/egusphere-2026-4627
Preprint. Discussion started: 14 September 2026

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020). The ERA5 global reanalysis. Quarterly Journal of the Royal Meteorological Society, 146(730), 1999-2049. https://doi.org/10.1002/qj.3803

Hess, M., Koepke, P., & Schult, I. (1998). Optical Properties of Aerosols and Clouds: The Software Package OPAC. Bulletin of the American Meteorological Society, 79(5), 831-844. https://journals.ametsoc.org/view/journals/bams/79/5/1520-0477_1998_079_0831_opoaac_2_0_co_2.xml

Hong, S.-Y., Noh, Y., & Dudhia, J. (2006). A New Vertical Diffusion Package with an Explicit Treatment of Entrainment Processes. Monthly Weather Review, 134(9), 2318-2341. https://journals.ametsoc.org/view/journals/mwre/134/9/mwr3199.1.xml

Iacono, M. J., Delamere, J. S., Mlawer, E. J., Shephard, M. W., Clough, S. A., & Collins, W. D. (2008). Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models. Journal of Geophysical Research: Atmospheres, 113(D13). https://doi.org/10.1029/2008JD009944

Kahn, R. A., Andrews, E., Brock, C. A., Chin, M., Feingold, G., Gettelman, A., et al. (2023). Reducing aerosol forcing uncertainty by combining models with satellite and within-the-atmosphere observations: A three-way street. Reviews of Geophysics, 61(2), e2022RG000796.

Kaufman, Y. J., Tanré, D., Remer, L. A., Vermote, E. F., Chu, A., & Holben, B. N. (1997). Operational remote sensing of tropospheric aerosol over land from EOS moderate resolution imaging spectroradiometer. Journal of Geophysical Research: Atmospheres, 102(D14), 17051-17067.

Kingma, D. P., & Welling, M. (2014). Auto-encoding variational Bayes. In Proceedings of the 2nd International Conference on Learning Representations (ICLR). https://arxiv.org/abs/1312.6114

Kokhanovsky, A. A., & Leeuw, G. (2009). Satellite aerosol remote sensing over land: Springer.

Levy, R. C., Mattoo, S., Munchak, L. A., Remer, L. A., Sayer, A. M., Patadia, F., & Hsu, N. C. (2013). The Collection 6 MODIS aerosol products over land and ocean. Atmospheric measurement techniques, 6(11), 2989-3034.

Li, J., Carlson, B. E., Yung, Y. L., Lv, D., Hansen, J., Penner, J. E., et al. (2022). Scattering and absorbing aerosols in the climate system. Nature Reviews Earth & Environment, 3(6), 363-379. https://doi.org/10.1038/s43017-022-00296-7

Li, M., Liu, H., Geng, G., Hong, C., Liu, F., Song, Y., et al. (2017). Anthropogenic emission inventories in China: a review. National Science Review, 4(6), 834-866. https://doi.org/10.1093/nsr/nwx150

Mohd Razak, S., Jahandideh, A., Djuraev, U., & Jafarpour, B. (2022). Deep Learning for Latent Space Data Assimilation in Subsurface Flow Systems. SPE Journal, 27(05), 2820-2840. https://doi.org/10.2118/203997-PA

Morrison, H., Thompson, G., & Tatarskii, V. (2009). Impact of Cloud Microphysics on the Development of Trailing Stratiform Precipitation in a Simulated Squall Line: Comparison of One- and Two-Moment Schemes. Monthly Weather Review, 137(3), 991-1007. https://journals.ametsoc.org/view/journals/mwre/137/3/2008mwr2556.1.xml

Mücke, N. T., Bohté, S. M., & Oosterlee, C. W. (2024). The deep latent space particle filter for real-time data assimilation with uncertainty quantification. Scientific Reports, 14(1), 19447.

Peyron, M., Fillion, A., Gürol, S., Marchais, V., Gratton, S., Boudier, P., & Goret, G. (2021). Latent space data assimilation by using deep learning. Quarterly Journal of the Royal Meteorological Society, 147(740), 3759-3777.

Powers, J. G., Klemp, J. B., Skamarock, W. C., Davis, C. A., Dudhia, J., Gill, D. O., et al. (2017). The Weather Research and Forecasting Model: Overview, System Efforts, and Future Directions. Bulletin of the American Meteorological Society, 98(8), 1717-1737. https://journals.ametsoc.org/view/journals/bams/98/8/bams-d-15-00308.1.xml

Rezende, D. J., Mohamed, S., & Wierstra, D. (2014). Stochastic backpropagation and approximate inference in deep generative models. In Proceedings of the 31st International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 32, pp. 1278–1286). PMLR. https://proceedings.mlr.press/v32/rezende14.html

Roberts, N. M., & Lean, H. W. (2008). Scale-Selective Verification of Rainfall Accumulations from High-Resolution Forecasts of Convective Events. Monthly Weather Review, 136(1), 78-97. https://journals.ametsoc.org/view/journals/mwre/136/1/2007mwr2123.1.xml

Rodgers, C. D. (2000). Inverse methods for atmospheric sounding: theory and practice (Vol. 2): World scientific.

Rutan, D. A., Kato, S., Doelling, D. R., Rose, F. G., Nguyen, L. T., Caldwell, T. E., & Loeb, N. G. (2015). CERES Synoptic Product: Methodology and Validation of Surface Radiant Flux. Journal of Atmospheric and Oceanic Technology, 32(6), 1121-1143. https://journals.ametsoc.org/view/journals/atot/32/6/jtech-d-14-00165_1.xml

Sandu, A., & Chai, T. (2011). Chemical data assimilation—An overview. Atmosphere, 2(3), 426-463.

Sayer, A. M., Hsu, N. C., Bettenhausen, C., & Jeong, M. J. (2013). Validation and uncertainty estimates for MODIS Collection 6 “Deep Blue” aerosol data. Journal of Geophysical Research: Atmospheres, 118(14), 7864-7872.

Schutgens, N., Sayer, A. M., Heckel, A., Hsu, C., Jethva, H., de Leeuw, G., et al. (2020). An AeroCom–AeroSat study: intercomparison of satellite AOD datasets for aerosol model evaluation. Atmospheric Chemistry and Physics, 20, 12431–12457. https://doi.org/10.5194/acp-20-12431-2020

Spurr, R. J. D. (2006). VLIDORT: A linearized pseudo-spherical vector discrete ordinate radiative transfer code for forward model and retrieval studies in multilayer multiple scattering media. Journal of Quantitative Spectroscopy and Radiative Transfer, 102(2), 316-342.

Tilmes, S., Emmons, L. K., Buchholz, R. R., & The, C. D. T. (2022). CESM2.2 CAM-chem as Boundary Conditions. Retrieved from: https://doi.org/10.5065/XS0R-QE86

Weaver, C., da Silva, A., Chin, M., Ginoux, P., Dubovik, O., Flittner, D., et al. (2007). Direct insertion of MODIS radiances in a global aerosol transport model. Journal of the atmospheric sciences, 64(3), 808-827.

Wiedinmyer, C., Akagi, S. K., Yokelson, R. J., Emmons, L. K., Al-Saadi, J. A., Orlando, J. J., & Soja, A. J. (2011). The Fire INventory from NCAR (FINN): a high resolution global model to estimate the emissions from open burning. Geosci. Model Dev., 4(3), 625-641. https://gmd.copernicus.org/articles/4/625/2011/

Wielicki, B. A., Barkstrom, B. R., Harrison, E. F., Lee, R. B., Smith, G. L., & Cooper, J. E. (1996). Clouds and the Earth's Radiant Energy System (CERES): An Earth Observing System Experiment. Bulletin of the American Meteorological

https://doi.org/10.5194/egusphere-2026-4627
Preprint. Discussion started: 14 September 2026

Society, 77(5), 853-868. https://journals.ametsoc.org/view/journals/bams/77/5/1520-0477_1996_077_0853_catere_2_0_co_2.xml

Xu, X., Wang, J., Henze, D. K., Qu, W., & Kopacz, M. (2013). Constraints on aerosol sources using GEOS-Chem adjoint and MODIS radiances, and evaluation with multisensor (OMI, MISR) data. Journal of Geophysical Research: Atmospheres, 118(12), 6396-6413.

Zaveri, R. A., Easter, R. C., Fast, J. D., & Peters, L. K. (2008). Model for Simulating Aerosol Interactions and Chemistry (MOSAIC). Journal of Geophysical Research: Atmospheres, 113(D13). https://doi.org/10.1029/2007JD008782

Zaveri, R. A., & Peters, L. K. (1999). A new lumped structure photochemical mechanism for large-scale applications. Journal of Geophysical Research: Atmospheres, 104(D23), 30387-30415. https://doi.org/10.1029/1999JD900876

Zhang, C.: LARDAS v1.0: latent-space aerosol–radiance data assimilation software, WRF-Chem source materials, and related archives, Zenodo [code], https://doi.org/10.5281/zenodo.22279326, 2026a.

Zhang, C., Li, Q., Li, J., Han, W., Zhang, Y., Chen, S., et al. (2026). Direct assimilation of satellite visible and near-infrared radiances to improve aerosol simulations. ESS Open Archive. https://doi.org/10.22541/essoar.177212682.27022185/v1

Zhang, J., Reid, J. S., Westphal, D. L., Baker, N. L., & Hyer, E. J. (2008). A system for operational aerosol optical depth data assimilation over global oceans. Journal of Geophysical Research: Atmospheres, 113(D10).

Zhang, W., Xu, H., & Zheng, F. (2018). Aerosol optical depth retrieval over East Asia using Himawari-8/AHI data. Remote Sensing, 10(1), 137.

Zheng, B., Tong, D., Li, M., Liu, F., Hong, C., Geng, G., et al. (2018). Trends in China's anthropogenic emissions since 2010 as the consequence of clean air actions. Atmos. Chem. Phys., 18(19), 14095-14111. https://acp.copernicus.org/articles/18/14095/2018/

Zheng, Q., Shao, Q., Han, G., Li, W., Li, H., & Wang, X. (2026). Generating unseen nonlinear evolution in the ocean using deep learning-based latent space data assimilation model. Ocean Modelling, 200, 102677. https://www.sciencedirect.com/science/article/pii/S1463500326000016

Zhong, C., Cheng, S., Kasoar, M., & Arcucci, R. (2023). Reduced-order digital twin and latent data assimilation for global wildfire prediction. Nat. Hazards Earth Syst. Sci., 23(5), 1755-1768. https://nhess.copernicus.org/articles/23/1755/2023/