

Advancing operational global aerosol forecasting with machine learning

<https://doi.org/10.1038/s41586-026-10234-y>

Received: 12 August 2025

Accepted: 4 February 2026

Published online: 4 March 2026

Open access

 Check for updates

Ke Gui^{1,12}, Xutao Zhang^{1,12}, Huizheng Che^{1,✉}, Lei Li¹, Yu Zheng¹, Linchang An², Yucong Miao¹, Hujia Zhao³, Oleg Dubovik⁴, Brent Holben⁵, Jun Wang⁶, Pawan Gupta⁵, Elena S. Lind⁵, Carlos Toledano⁷, Hong Wang¹, Zhili Wang¹, Yaqiang Wang⁸, Xiaomeng Huang⁹, Kan Dai², Xiangao Xia¹⁰, Xiaofeng Xu¹¹ & Xiaoye Zhang^{1,✉}

Aerosol forecasting is important for air-quality management, health risk assessment and climate change mitigation^{1,2}. However, it is more complex than weather forecasting, owing to the interactions between aerosol physicochemical processes and atmospheric dynamics, resulting in high uncertainty and computational costs^{3,4}. Here we develop a machine-learning-driven Global Aerosol–Meteorology Forecasting System (AI-GAMFS), which provides reliable 5-day, 3-hourly forecasts of aerosol optical components and surface concentrations. AI-GAMFS combines a vision transformer and U-Net in a backbone network, robustly capturing the complex aerosol–meteorology interactions via global attention and spatiotemporal encoding. Trained on 42 years of aerosol reanalysis data and initialized with Global Earth Observing System Forward Processing (GEOS-FP) analyses, AI-GAMFS delivers operational 5-day forecasts in 1 minute. Evaluation with independent ground-based observations suggests improved performance compared with the Copernicus Atmosphere Monitoring Service⁵ and regional dust models^{6–9} in forecasting aerosol optical depth and dust components. Compared with GEOS-FP¹⁰, it has a lower root-mean-square error for global aerosol optical depth, with comparable dust forecasting skill and improved surface aerosol component forecasts over the USA and China. Our results provide a step forward in leveraging machine learning to refine aerosol forecasting and may help warn against aerosol pollution events such as dust storms and wildfires.

Atmospheric aerosols have a critical role in Earth's climate system, affecting radiative forcing, cloud microphysics and atmospheric chemistry^{1,2}. Owing to their diverse optical and microphysical properties, combined with complex chemical compositions, aerosols influence weather and climate in various ways^{11,12}. Key components, such as black carbon (BC) and dust, show considerable variability in terms of radiative forcing, making aerosols a major source of uncertainty in climate change assessments^{13,14}. In addition, the complex chemical reactivity and wide particle size ranges of aerosols can degrade air quality^{15,16}, posing health risks that include respiratory, cardiovascular and neurological diseases¹⁷. Accurate forecasting of aerosol distributions and compositions is therefore essential for improving air-quality management, protecting public health and mitigating climate change.

However, aerosol forecasting presents markedly greater complexity and cost than weather forecasting owing to the need to account for diverse aerosol sources and types, intricate chemical reactions, physical processes, and multiscale interactions with weather systems^{3,4}.

These complexities result in nonlinear and highly variable processes for aerosol generation, transport, transformation and removal, contributing substantially to forecast uncertainty¹⁸. To enable short- to medium-term aerosol forecasting, traditional physics-based forecasting systems, such as the Copernicus Atmospheric Monitoring Service (CAMS) from the European Centre for Medium-Range Weather Forecasts⁵ and NASA's Global Earth Observing System Forward Processing (GEOS-FP)¹⁰, couple numerical weather prediction (NWP) models with atmospheric chemical transport models. These systems must simultaneously resolve atmospheric dynamics and compute thousands of aerosol-related chemical reactions and microphysical interactions, further intensifying the already high computational cost of NWP^{19,20}. Recent advances in machine learning have opened new avenues of investigation, leading researchers to explore advanced neural networks as complementary tools for NWP^{21–25} and its downstream tasks, such as oceanic variables^{26,27}. These neural network models have shown considerable promise in enhancing computational efficiency and accuracy in weather forecasting; however,

¹State Key Laboratory of Severe Weather Meteorological Science and Technology (LaSW) and Key Laboratory of Atmospheric Chemistry of CMA, Chinese Academy of Meteorological Sciences, Beijing, China. ²National Meteorological Center, CMA, Beijing, China. ³Institute of Atmospheric Environment, China Meteorological Administration, Shenyang, China. ⁴Université de Lille, CNRS, UMR 8518–Laboratoire d'Optique Atmosphérique, Lille, France. ⁵NASA Goddard Space Flight Center, Greenbelt, MD, USA. ⁶Department of Chemical and Biochemical Engineering, The University of Iowa, Iowa City, IA, USA. ⁷Group of Atmospheric Optics, University of Valladolid, Valladolid, Spain. ⁸Institute of Artificial Intelligence for Meteorology, Chinese Academy of Meteorological Sciences, Beijing, China. ⁹Department of Earth System Science, Tsinghua University, Beijing, China. ¹⁰Laboratory of Middle Atmosphere and Global Environment Observation, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China. ¹¹China Meteorological Administration, Beijing, China. ¹²These authors contributed equally: Ke Gui, Xutao Zhang. ✉e-mail: chehz@cma.gov.cn; xiaoye@cma.gov.cn

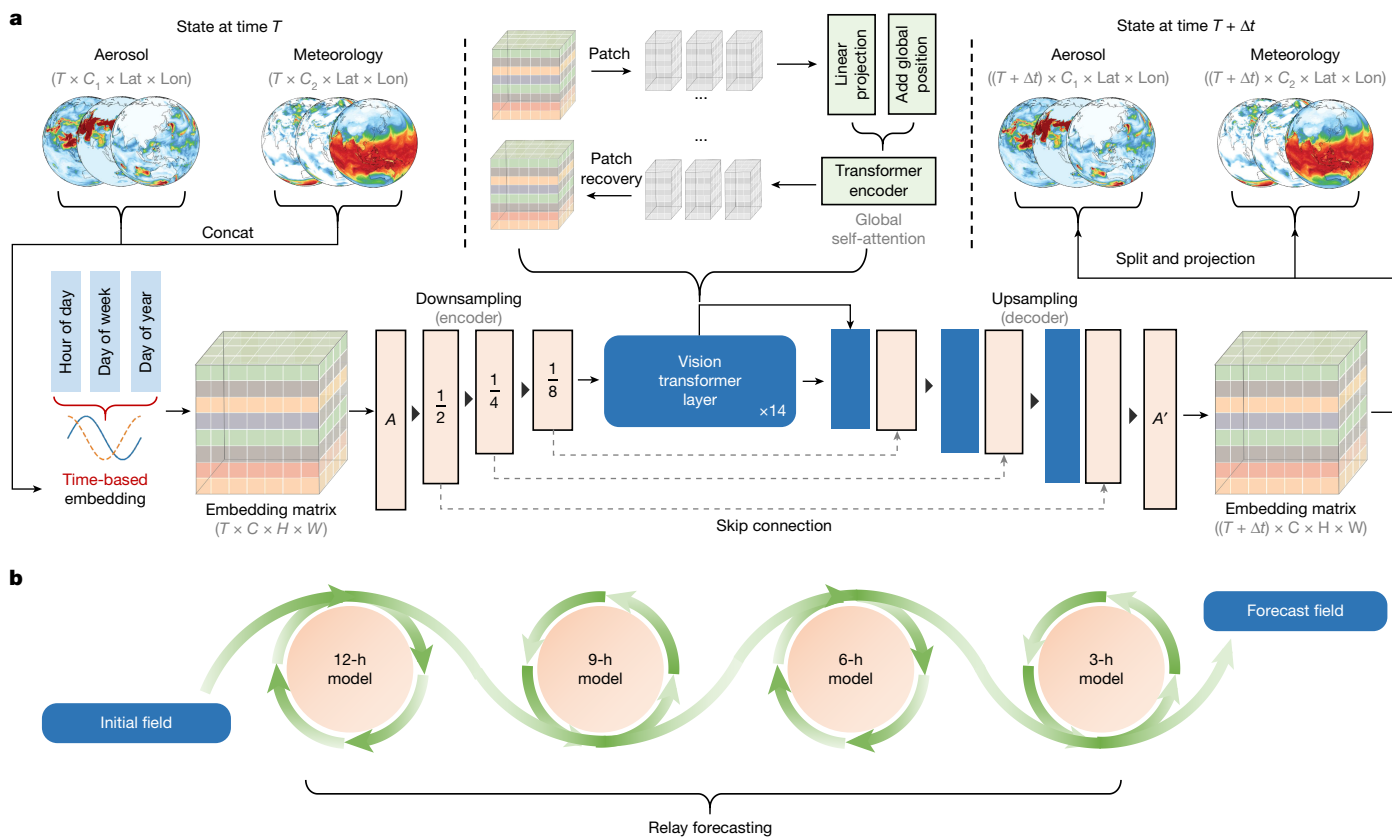


Fig. 1 | Architecture of the machine-learning-driven AI-GAMFS. a, The AI-GAMFS model consists of three primary components: cube embedding, a vision transformer and cube unembedding. T , initial time; Δt , forecast lead time; C_1 and C_2 , channels for aerosol and meteorological variables, respectively; H and W , height and width of the embedded feature tensor, respectively; A ,

embedded feature tensor before downsampling; A' , feature tensor after upsampling, which has the same dimensions as A . **b**, The temporal aggregation strategy used in AI-GAMFS for relay forecasting at specified lead times, achieved by integrating four pretrained models—3-h, 6-h, 9-h and 12-h models—each trained with identical configurations.

machine-learning research specifically targeting global aerosol forecasting remains notably underdeveloped. Although recent studies have begun applying deep learning to aerosol forecasting on both global and regional scales^{28,29}, these efforts depend largely on NWP inputs and are often restricted to single aerosol metrics such as total aerosol optical depth (AOD). The operational integration of machine-learning models for simultaneous global-scale aerosol component and meteorological forecasting remains incomplete, particularly because of the challenges in representing coupled aerosol–weather processes, generalizing across diverse aerosol types and addressing computational constraints.

Here we present a machine-learning-driven Global Aerosol–Meteorology Forecasting System (AI-GAMFS), designed to rapidly simulate complex aerosol–meteorology interactions across spatial and temporal scales. AI-GAMFS was trained on 42 years of Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2)³⁰ atmospheric reanalysis data. Evaluation with global Aerosol Robotic Network (AERONET)³¹ and Chinese Aerosol Remote Sensing Network (CARSNET)³² observations suggests improved performance of the operational AI-GAMFS, compared with the state-of-the-art CAMS and regional dust models, in forecasting AOD and dust components. Compared with GEOS-FP, it also provides improved global AOD forecasts, comparable dust forecasting skill and improved key surface aerosol component forecasts over the USA and China, with an order-of-magnitude reduction in computational cost.

AI-GAMFS

AI-GAMFS is designed to provide global 5-day aerosol–meteorology forecasts at approximately 50-km spatial resolution and 3-hourly

temporal intervals (01:30, 04:30, 07:30, ..., 22:30 UTC), forecasting AOD, the optical properties and surface concentrations of key aerosol components—including sulfate, dust, BC, organic carbon (OC) and sea salt (SS)—as well as surface and upper-level meteorological variables (Supplementary Table 1) that govern aerosol lifecycle dynamics. Its architecture comprises three core modules (Fig. 1a): (1) cube embedding, which extracts three-dimensional spatiotemporal features from the input feature matrix; (2) a vision transformer, which uses a multiheaded self-attention mechanism to process and understand complex relationships between features; and (3) cube unembedding, which reconstructs high-dimensional features back to the original spatial resolution using deconvolution and upsampling techniques. To ensure the accuracy and fidelity of the forecasts, skip connections are incorporated. Working synergistically, these modules accurately forecast the spatial fields of aerosol and meteorological states at the next time step, using the previous time step as input.

We trained four base models separately with a forecast lead time of 3 h, 6 h, 9 h and 12 h. Each base model was trained for 80 epochs using the same framework and settings, containing approximately 1.2 billion parameters, and was trained for 10 days on 8 L40 graphics processing units (GPUs). To mitigate error accumulation owing to long-term iterations in a single model, a temporal aggregation strategy was used to perform relay forecasting with the four base models (Fig. 1b). Once pretraining and relay connection are completed, the final AI-GAMFS model generates 5-day operational forecasts in approximately 39 s on a single L40 GPU, using real-time GEOS-FP analysis fields as input. This represents computational speed that is approximately 360-times faster than that associated with conventional GEOS-FP forecasts (which require approximately 4–6 h).

Relay forecasting reduces accumulation errors

Building on this relay architecture designed to curb error growth, we proceeded to systematically evaluate and optimize its configuration. Using 4 pretrained base AI-GAMFS models with forecast lead times of 3 h, 6 h, 9 h and 12 h, we designed 4 progressive forecasting schemes to identify the optimal relay forecasting strategy: the 3-h single model, the 3- and 6-h relay, the 3-h, 6-h and 9-h relay, and the 3-h, 6-h, 9-h and 12-h relay. Extended Data Fig. 1a illustrates the frequency with which the 4 pretrained base models (with lead times of 3 h, 6 h, 9 h and 12 h) are invoked under these 4 forecasting strategies. Notably, for forecasts with a specific lead time, when at least two pretrained base models are used in the relay, we prioritize models with longer lead times, iteratively using their forecast results as inputs for the next forecast time step, thereby minimizing the number of iterations as much as possible.

We compared the global 5-day forecasting accuracy of AI-GAMFS—initialized by MERRA-2 reanalysis and run daily at 22:30 UTC—for all 12 aerosol variables using different relay forecasting strategies, with the 2022 MERRA-2 data (test set) as a baseline. The aerosol variables include AOD, total scattering AOD (TSAOD), sulfate, dust, BC, OC and SS AOD (SUAOD, DUAOD, BCAOD, OCAOD and SSAOD, respectively), and sulfate, dust, BC, OC and SS surface mass concentration (SUSMC, DUSMC, BCSMC, OCSMC and SSSMC, respectively). Extended Data Fig. 1b,c shows time series of the global spatial correlation coefficient (R) and latitude-weighted root-mean-square error (RMSE) for these aerosol variables, respectively. The results indicate that within a 24-h forecast horizon, the performance of the 3-h, 6-h, 9-h and 12-h relay model is similar to that of the 3-h single model and the other relay models. However, for forecast lead times beyond 24 h, the 3-h, 6-h, 9-h and 12-h relay model shows superior accuracy for nearly all aerosol variables, in terms of both R and RMSE. For example, for all aerosol variables at a 120-h lead time, the average RMSE value for the 3-h, 6-h, 9-h and 12-h relay model is typically 15.1%, 5.6% and 3.2% lower than that of the 3-h single model, the 3-h and 6-h relay model, and the 3-h, 6-h and 9-h relay model, respectively. This advantage is also evident in global forecasts for various meteorological variables (Extended Data Fig. 2). The use of 4 base models in the relay strategy generally yields results comparable to or slightly better than those from the 3-h and 6-h relay and the 3-h, 6-h and 9-h relay models, but that substantially outperform those of the 3-h single model. However, we note that while the aggregation strategy helps alleviate short- to medium-term error accumulation, the improvement tends to plateau as the number of base models increases. Therefore, we ultimately selected the 3-h, 6-h, 9-h and 12-h relay model strategy as the final AI-GAMFS model, which was used in all subsequent evaluations and analyses.

Enhanced global aerosol forecasts

Globally, AOD is one of the most widely observed atmospheric aerosol parameters and is used extensively in climate change research, air-quality monitoring and environmental assessments. As a key component of AOD, DUAOD serves as an essential metric for monitoring the global dust cycle and its impacts. This study provided comprehensive evaluation of the 5-day, 3-hourly global AOD and DUAOD forecasts generated by AI-GAMFS, initialized daily at 22:30 UTC, utilizing MERRA-2 evaluation data from 2023. The performance of operational AI-GAMFS, initialized by GEOS-FP analyses, is compared with that of CAMS (run daily at 00:00 UTC), one of the leading global aerosol forecast models, as illustrated in Fig. 2a. Meanwhile, forecasts from AI-GAMFS initialized with MERRA-2 reanalysis are also presented to evaluate the impact of initial conditions. Over the 0–120-h forecast period, operational AI-GAMFS consistently outperforms CAMS in forecasting both AOD and DUAOD, as measured by both R and RMSE. Specifically, AI-GAMFS shows a clear advantage during the 0–2-day period, improving the average R value by 11.5% and 13.8%, and reducing the average RMSE by 22.3% and 37.3%, for

AOD and DUAOD relative to CAMS, respectively. However, as the forecast lead time increases, the advantage of operational AI-GAMFS diminishes. Nevertheless, at a 120-h lead time, operational AI-GAMFS still produces a lower RMSE than CAMS, with reductions of approximately 11.3% and 25.2% for AOD and DUAOD, respectively. Our analysis also shows that the impact of initial conditions on AI-GAMFS is most pronounced within the first 48 h, with little to no effect thereafter.

Given the differences in initial conditions between AI-GAMFS and CAMS, AI-GAMFS might benefit from using MERRA-2 as the reference data for evaluation. To ensure a fairer comparison, we additionally used Level-2.0 instantaneous global aerosol observations from AERONET in 2023 to evaluate the 5-day, 3-hourly AOD and DUAOD forecast performance of both operational AI-GAMFS and CAMS, for which DUAOD was evaluated against the AERONET coarse-mode AOD (AODc). Figure 2b presents time series of the R and RMSE values, calculated from all matched global samples for 2023, at each forecast lead time. Operational AI-GAMFS shows high forecasting skill against AERONET observations, albeit with a predictable degradation over time. Specifically, the model maintains reasonable accuracy throughout the forecast period (days 1–5), with mean AOD R ranging from 0.57 to 0.78 (RMSE 0.12 to 0.15) and DUAOD R ranging from 0.65 to 0.73 (RMSE 0.04 to 0.06). Consistent with the evaluation using MERRA-2 as the reference, operational AI-GAMFS also provides more accurate AOD and DUAOD forecasts than CAMS. Statistically, across all 40 forecast steps (3-h intervals), operational AI-GAMFS outperforms CAMS at 31 and 36 steps for R , and at 37 and 40 steps for RMSE, for AOD and DUAOD, respectively. Importantly, this consistent accuracy between the MERRA-2-driven and operational (GEOS-FP-driven) configurations affirms the reliability and effectiveness of AI-GAMFS in a real-world operational environment.

Figure 2c further illustrates the spatial distribution of the average RMSE for each step of the 5-day AOD and DUAOD forecasts (a total of 40 steps) from operational AI-GAMFS at each AERONET site, alongside the RMSE difference between CAMS and operational AI-GAMFS. Overall, for AOD, operational AI-GAMFS shows lower RMSE values than those of CAMS at 61.6% of AERONET sites, located primarily in the USA, Europe, Africa and Southeast Asia. Given that China is one of the regions with considerable aerosol loading, yet critically lacks AERONET coverage, we conducted a complementary evaluation using continuous AOD observations from 26 CARSNET sites in 2023 (Extended Data Fig. 3a,b). The results show that operational AI-GAMFS provides acceptable forecasting skill over China, with the mean R ranging from 0.44 to 0.65 and the mean RMSE ranging from 0.26 to 0.34 throughout the forecast period (days 1–5). This evaluation also confirms the robust superiority of operational AI-GAMFS compared with CAMS, achieving a higher R at 63.3% of the forecast steps and a lower RMSE at 61.5% of the sites. Moreover, for global DUAOD forecasting, operational AI-GAMFS shows a clear advantage over CAMS, achieving a lower RMSE at 86.0% of the sites worldwide (Fig. 2c). These results robustly demonstrate the superior performance of operational AI-GAMFS compared with CAMS in terms of global AOD and DUAOD forecasting.

Regional dust storms

East Asia is one of the regions affected most severely by dust storms, highlighting the critical need for accurate forecasts of such events. The operational AI-GAMFS model forecasts both DUAOD and DUSMC, thereby presenting the opportunity to assess its performance relative to several well-established physics-based dust forecasting models. For this evaluation, we used East Asia dust forecast products for 2023 derived from forecasts of CAMS and four physics-based dust forecasting models deployed at the Sand and Dust Storm Warning Advisory and Assessment System (SDS-WAS) Asian Regional Centre. These models include SILAM from the Finnish Meteorological Institute (FMI-SILAM)⁶, CUACE/Dust from the China Meteorological

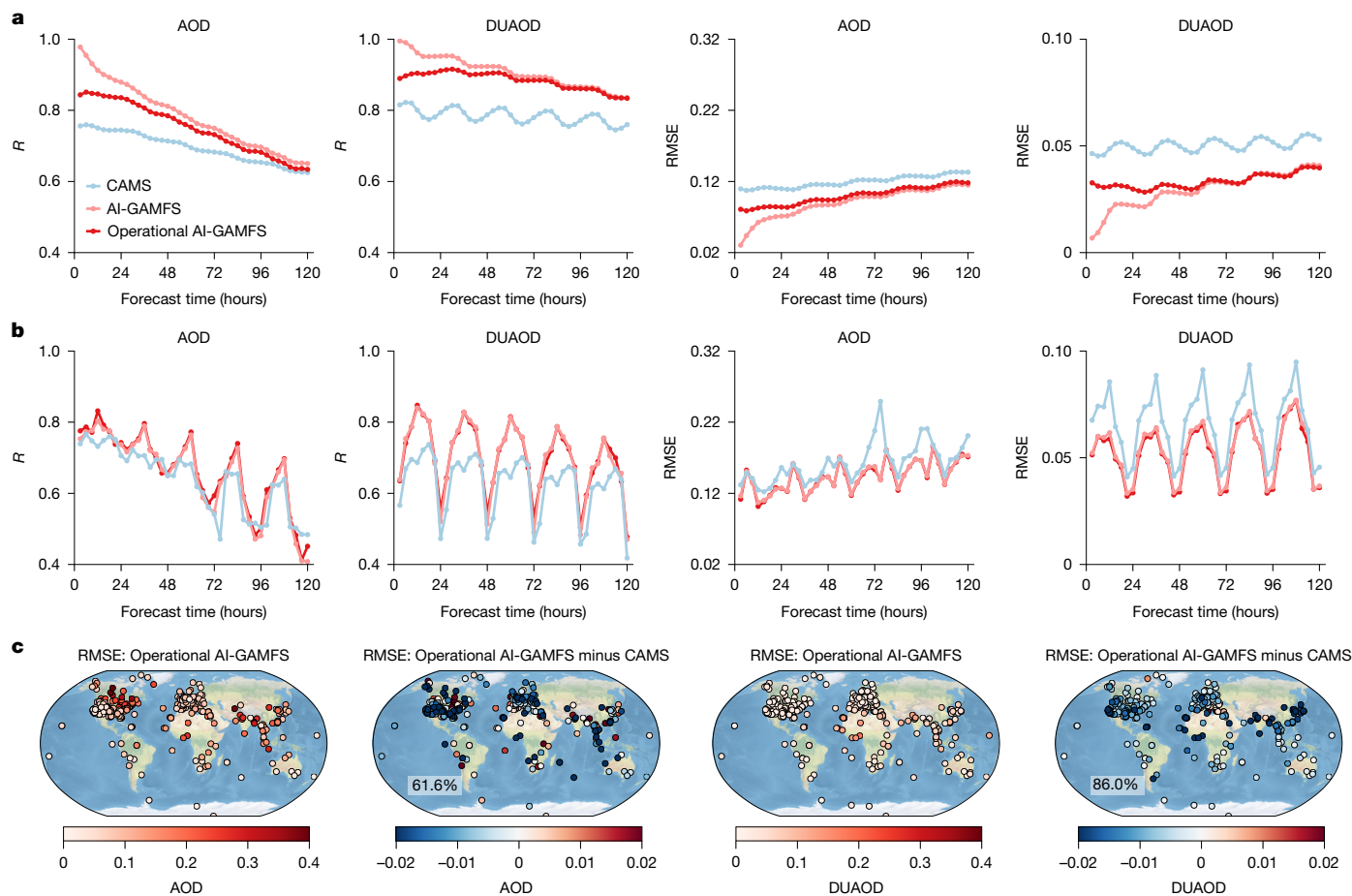


Fig. 2 | Superior performance of AI-GAMFS in global AOD and DUAOD forecasting throughout 2023. **a, b,** Global 5-day forecast accuracy for AOD and DUAOD at 3-hourly intervals from AI-GAMFS (initialized by MERRA-2 reanalysis), operational AI-GAMFS (initialized by GEOS-FP analyses) and CAMS. Performance was evaluated against MERRA-2 (**a**) and AERONET observations (**b**) using R and RMSE as metrics. For AERONET, evaluation metrics for each step

were calculated by aggregating all matched samples globally. **c,** Spatial distribution of the average RMSE for each step of the 5-day AOD and DUAOD forecasts (a total of 40 steps) from operational AI-GAMFS, alongside the RMSE difference between CAMS and operational AI-GAMFS. Percentages in the lower-left corners of the RMSE difference panels indicate the proportion of sites where operational AI-GAMFS showed lower RMSEs than those of CAMS.

Administration (CMA-CUACE/Dust)⁷, MASINGAR from the Japan Meteorological Agency (JMA-MASINGAR)⁸ and ADAM3 from the Korea Meteorological Agency (KMA-ADAM3)⁹. We evaluated the 5-day forecast accuracy of DUAOD and DUSMC from operational AI-GAMFS (initialized daily at 22:30 UTC), CAMS, FMI-SILAM, CMA-CUACE/Dust and KMA-ADAM3 (initialized daily at 00:00 UTC), relative to MERRA-2 data from 2023. Because JMA-MASINGAR initializes daily at 12:00 UTC and provides 3-day forecasts, we adjusted the initialization time for AI-GAMFS to 10:30 UTC for comparison. In addition, owing to differences in forecast coverage areas and temporal resolutions across models, we conducted the evaluation only for the overlapping East Asia region (Extended Data Fig. 4a) and applied temporal interpolation to the different models.

Extended Data Fig. 4b,c shows the time series of spatial R and latitude-weighted RMSE for these models. Consistent with its global performance, operational AI-GAMFS notably outperforms the 5 physics-based dust forecast models over East Asia across all forecast periods at 72 h (JMA-MASINGAR; Extended Data Fig. 4c) and 120 h (other 4 models; Extended Data Fig. 4b). Specifically, the spatial R for DUAOD at a 72-h lead time is improved by 12.0%, 21.4%, 34.2%, 105.1% and 199.7% relative to FMI-SILAM, CAMS, JMA-MASINGAR, CMA-CUACE/Dust and KMA-ADAM3, respectively. At the 120-h lead time (that is, 5 days), the improvement of operational AI-GAMFS is 4.9%, 16.9%, 90.4% and 133.5% relative to FMI-SILAM, CAMS, CMA-CUACE/Dust and KMA-ADAM3, respectively. For DUSMC, operational AI-GAMFS

has a latitude-weighted RMSE of $82.5 \mu\text{g m}^{-3}$ at a 72-h lead time, which is approximately 34.4%, 42.7% and 60.3% lower than that of FMI-SILAM, KMA-ADAM3 and CMA-CUACE, respectively; likewise, a more substantial reduction of approximately 74.1% is observed against JMA-MASINGAR. This regional advantage of operational AI-GAMFS over physics-based dust models is further confirmed by 1-year records of AODc from the AERONET Beijing-CAMS site and AOD from four CARS-NET sites in the northwestern desert region of China (Supplementary Figs. 1 and 2a).

Taking the mega dust storm in northern China in April 2023 as an example, we found that operational AI-GAMFS can reliably reproduce the entire dust transport process, including the affected areas and the intensity (Extended Data Fig. 5). This is further confirmed by better statistical metrics compared with those of other models. More importantly, operational AI-GAMFS not only forecasts dust transport paths within 1–2 days, but also forecasts enhanced dust emissions in the Gobi Desert up to 3–4 days in advance. Typically, it is a challenge for regional dust forecasting models to capture such features.

Aerosol component forecasting

In addition to forecasting AOD and dust-related properties, AI-GAMFS simultaneously forecasts TSAOD, the optical properties of other aerosol components (that is, sulfates, BC, OC and SS) and their surface concentrations. These component forecasts enable precise assessments

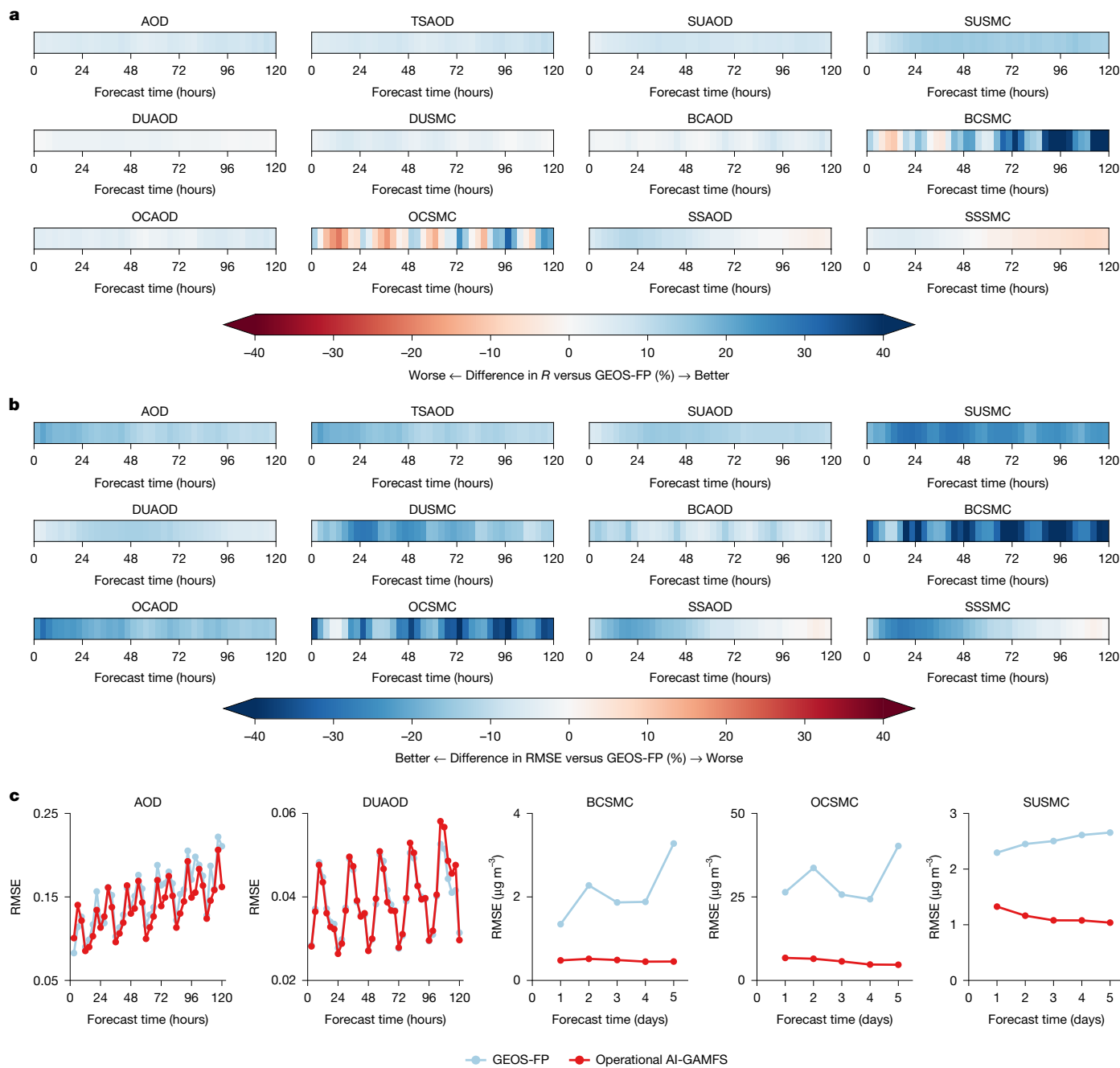


Fig. 3 | Comparison of global aerosol component forecast accuracy between operational AI-GAMFS and GEOS-FP. a, b, Scorecards of spatial *R* (a) and latitude-weighted RMSE (b) for 5-day, 3-hourly global forecasts of 12 aerosol variables from operational AI-GAMFS and GEOS-FP, evaluated against MERRA-2 reanalysis during July–August 2024. **c,** The 5-day forecast accuracy (measured by RMSE) of operational AI-GAMFS and GEOS-FP, evaluated against independent

aerosol observations from July–August 2024: AOD and DUAOD were evaluated against global AERONET observations, and BCSMC, OCSMC and SUSMC were evaluated against the IMPROVE network observations in the USA. For comparison with the daily IMPROVE data, 3-hourly forecasts were averaged to daily means.

of their specific impacts on climate, air quality and public health. We used conventional GEOS-FP as a reference baseline because it represents state-of-the-art atmospheric aerosol component forecasting and provides output configurations fully consistent with operational AI-GAMFS. Using MERRA-2 data collected from July to August 2024 as reference, we evaluate accuracy using the spatial *R* and latitude-weighted RMSE values, as shown in Fig. 3a,b. Additional metrics for surface and upper-level meteorological variables are provided in Extended Data Fig. 6.

The scorecards indicate that operational AI-GAMFS delivers exceptional forecasting performance across all 12 aerosol variables. For the

first 1–3 days, operational AI-GAMFS outperforms GEOS-FP in terms of all variables and at all lead times, except for BCSMC and OCSMC at specific time points (based on the *R* value). At longer lead times, AI-GAMFS consistently outperforms GEOS-FP, except for two SS-related variables: SSAOD and SSSMC. Aerosol component forecasts are highly sensitive to the accuracy of weather forecasts. Although operational AI-GAMFS does not surpass GEOS-FP in terms of the accuracy of certain meteorological variables, such as wind speed, sea-level pressure and temperature, improvements in the forecast accuracy of key variables—such as specific humidity and precipitation—that influence aerosol emissions, transformation and deposition, enable AI-GAMFS

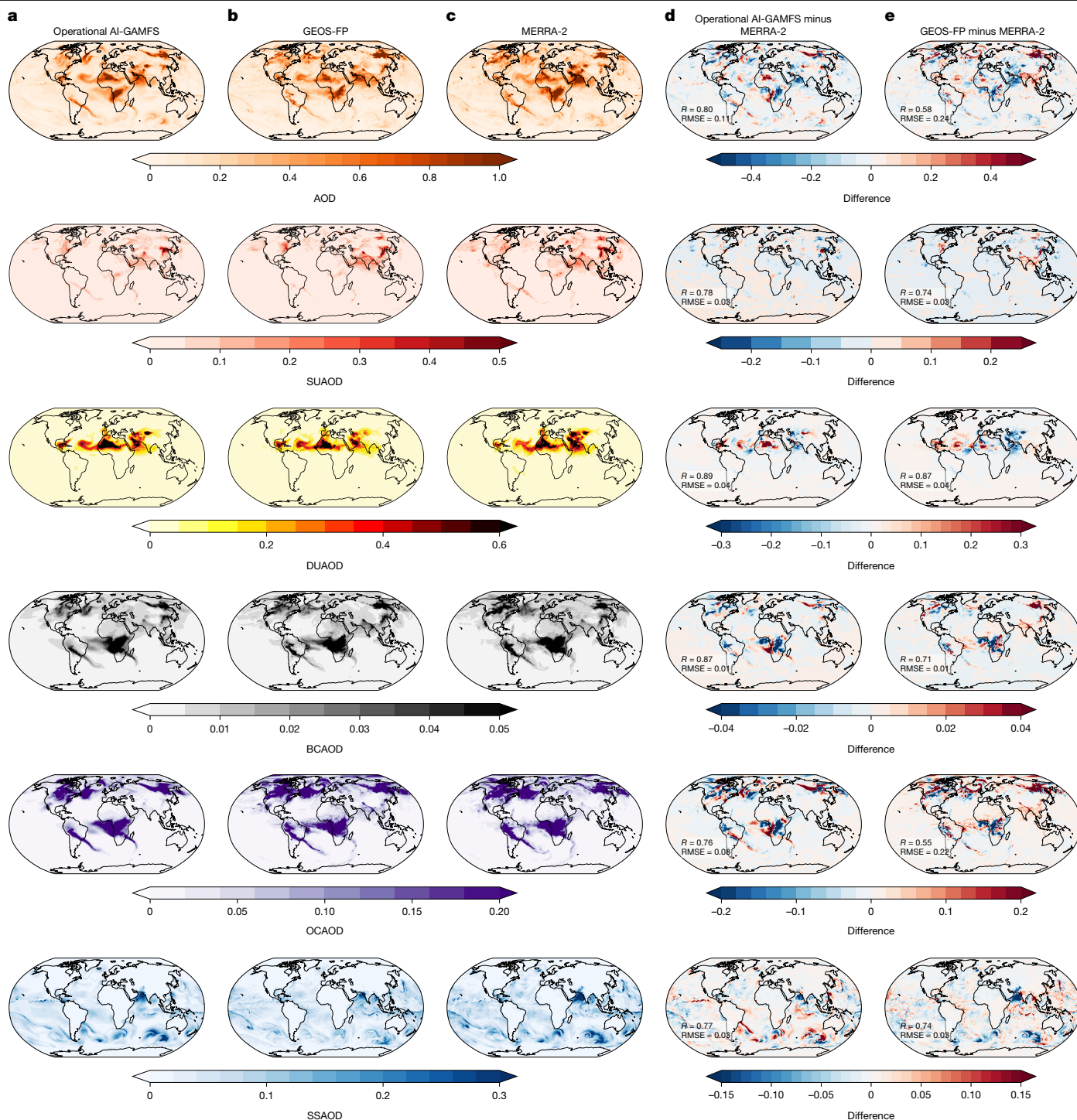


Fig. 4 | Case study of operational medium-range global aerosol optical component forecasts. a–c, The 3-day lead-time global forecasts for AOD and its 5 key components—SUAOD, DUAOD, BCAOD, OCAOD and SSAOD—from operational AI-GAMFS (a), GEOS-FP (b) and MERRA-2 (c), initialized at 22:30 UTC

on 26 July 2024. d, e, Forecasting errors of operational AI-GAMFS (d) and GEOS-FP (e) relative to MERRA-2 reanalysis data. Overall accuracy metrics (that is, spatial *R* and RMSE values) for operational AI-GAMFS and GEOS-FP are also indicated in the lower-left corner of each panel in d and e.

to enhance its aerosol simulations (Extended Data Fig. 6). However, forecast accuracy for wind speed declines beyond 2 days, which negatively impacts the forecast of SS aerosols.

Independent evaluation beyond MERRA-2 was conducted using global aerosol observations from July–August 2024. Using ground-based observations of AOD and AODc from AERONET, AOD from CARSONET, and BCSMC, OCSMC and SUSMC from the Interagency Monitoring of Protected Visual Environments (IMPROVE) network³³,

the Environmental Protection Agency Chemical Speciation Network (EPA-CSN)³³ and the China Atmospheric Watch Network (CAWNET)³⁴, we compared aerosol component forecast performance between operational AI-GAMFS and GEOS-FP. Globally, operational AI-GAMFS achieves mean RMSE values of 0.11–0.16 for AOD and 0.03–0.05 for DUAOD across forecast days 1–5 (Fig. 3c). It outperforms GEOS-FP for AOD, achieving lower RMSE in 36 out of 40 forecast steps against AERONET. Over China, as evaluated by 24 CARSONET sites, the model

shows RMSE values between 0.33 and 0.35 (Extended Data Fig. 3c). Despite these regionally elevated errors, it maintains an advantage over GEOS-FP, achieving lower RMSE in 20 out of 25 steps and at 58.3% of sites (Extended Data Fig. 3c,d). In contrast, the performance of global DUAOD forecasts is overall comparable to that of GEOS-FP during days 1–4, albeit weaker on day 5 (Fig. 3c). Furthermore, AOD observations from two independent CARSONET sites in dust-dominated regions provide additional evidence for the improved regional dust forecasting by operational AI-GAMFS in China (Supplementary Fig. 2b).

For surface aerosol components evaluated against the IMPROVE network across the USA, operational AI-GAMFS consistently shows lower RMSE in daily forecasts over the 5-day period compared with GEOS-FP. The RMSE ranges for BCSMC, OCSMC and SUSMC were 0.45–0.51 $\mu\text{g m}^{-3}$, 4.6–6.7 $\mu\text{g m}^{-3}$ and 1.0–1.3 $\mu\text{g m}^{-3}$, representing reductions of approximately 64.4–86.2%, 74.5–88.3% and 42.2–61.0%, respectively (Fig. 3c). Improvements in BCSMC and OCSMC are most pronounced across the western USA, influenced by frequent wildfires, whereas SUSMC improvements are clustered in the eastern USA, which is strongly influenced by anthropogenic emissions (Supplementary Figs. 3–5). Similar advantages were confirmed against the EPA-CSN network (Supplementary Figs. 6–8), indicating the overall superiority of operational AI-GAMFS in forecasting wildfire-related BC and OC, as well as anthropogenic sulfate. In China, CAWNET observations reveal that operational AI-GAMFS outperforms GEOS-FP in BCSMC forecasts across all 5 days, with a higher daily R at 52.5–75.0% of sites and a lower RMSE at 62.5–72.5% of sites (Supplementary Figs. 9 and 12). Comparable superiority is also observed for OCSMC and SUSMC (Supplementary Figs. 10–12).

Tracking pollutant-type events

A distinguishing feature of data-driven forecasting is its ability to rapidly track and segregate aerosol pollution types, closely mirroring real-world patterns, at a fraction of the computational cost and time required by traditional physics-based aerosol forecasting models. Figure 4 illustrates a case study with a 3-day lead time, highlighting the performance of operational AI-GAMFS in forecasting global AOD and its 5 key components: SUAOD, DUAOD, BCAOD, OCAOD and SSAOD. A more comprehensive evaluation is provided in Extended Data Figs. 7 and 8, which include maps of the spatiotemporal evolution of additional aerosol and meteorological variables.

In general, operational AI-GAMFS produces forecasts that align more closely with MERRA-2 than with GEOS-FP, effectively mitigating spatial smoothing artefacts commonly introduced by longer forecast lead times. In terms of AOD forecasting, it demonstrates superior R and notably lower RMSE values compared with those of GEOS-FP. This improvement is consistent across forecasts of various aerosol optical components and surface concentrations (Extended Data Fig. 7). Overall, the exceptional performance of operational AI-GAMFS is attributable largely to its precise forecasting of key meteorological variables (Extended Data Fig. 8).

Saharan dust and Central African wildfires are long-established sources of global aerosol loading, and their accurate forecasting remains a major challenge. We further evaluate the performance of operational AI-GAMFS in forecasting regional dust and BC at half-day intervals, as shown in Extended Data Figs. 9 and 10. Compared with GEOS-FP, operational AI-GAMFS markedly improves the simulation of Saharan dust and Central African wildfire aerosols, as evidenced by enhanced R and reduced RMSE values. Furthermore, it successfully captures the trans-Atlantic transport of dust and smoke aerosols, underscoring its robustness in forecasting long-range aerosol transport. Notably, it also captures the spatiotemporal evolution of smoke aerosols in South America, in addition to those in Africa (Extended Data Fig. 10).

Discussion

We introduce AI-GAMFS, a data-driven system for operational global aerosol forecasting. By extracting valuable insights from 42 years of coupled aerosol–meteorology reanalysis data, it demonstrates the potential to propel operational weather forecasting towards more complex forecasts in environmental meteorology. Our results show that operational AI-GAMFS outperforms several physics-based global and regional aerosol forecasting systems, delivering improved deterministic forecasts for AOD and multiple aerosol components. Unlike physics-based models, such as GEOS-FP and CAMS running on 6- or 12-hourly cycles, AI-GAMFS generates forecasts at 3-hourly intervals, thereby enhancing timeliness and better capturing the spatiotemporal variations of aerosols.

Although operational AI-GAMFS shows considerable potential in refining global aerosol forecasting, further enhancements are attainable through several strategic improvements. First, training strategies could be strengthened by incorporating multi-time-step rolling inputs²³ to improve temporal coherence and autoregressive performance. Future models should also integrate dynamic anthropogenic emission inventories²⁵ and time-varying background fields to better capture long-term aerosol trends and mitigate the impact of shifting anthropogenic activities. Second, the current training dataset of approximately 120,000 time steps is notably smaller than that of other data-driven weather models, such as Pangu-Weather (approximately 340,000 samples)²². This limited data volume probably contributes to the model's lower accuracy in forecasting key meteorological variables, including wind speed and temperature, compared with GEOS-FP, thereby affecting the accuracy of SS forecasts. Expanding the training dataset with higher temporal resolution records, where computationally feasible, will be critical to improving skill across aerosol components. In addition, owing to limitations in accessing GEOS-FP historical forecast data and the scarcity of surface aerosol component observations, accumulating longer time series of both observational and forecast data will be necessary to enable robust cross-seasonal evaluations of operational AI-GAMFS's stability and forecasting capabilities. Finally, a fundamental direction for future development lies in embedding physicochemical constraints and atmospheric dynamics as inductive biases into the model architecture—shifting from purely data-driven learning towards a hybrid physics–machine-learning framework that ensures forecasts remain inherently consistent with Earth system principles.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-026-10234-y>.

1. Prospero, J. M. et al. The atmospheric aerosol system: an overview. *Rev. Geophys.* **21**, 1607–1629 (1983).
2. Charlson, R. J. et al. Climate forcing by anthropogenic aerosols. *Science* **255**, 423–430 (1992).
3. Morcrette, J. J. et al. Aerosol analysis and forecast in the European centre for medium-range weather forecasts integrated forecast system: forward modeling. *J. Geophys. Res. Atmos.* **114**, D06206 (2009).
4. Kukkonen, J. et al. A review of operational, regional-scale, chemical weather forecasting models in Europe. *Atmos. Chem. Phys.* **12**, 1–87 (2012).
5. Peuch, V. H. et al. The Copernicus Atmosphere Monitoring Service from research to operations. *Bull. Am. Meteorol. Soc.* **103**, E2650–E2668 (2022).
6. Sofiev, M. et al. Construction of the SILAM Eulerian atmospheric dispersion model based on the advection algorithm of Michael Galperin. *Geosci. Model Dev.* **8**, 3497–3522 (2015).
7. Gong, S. L. & Zhang, X. Y. CUACE/Dust—an integrated system of observation and modeling systems for operational dust forecasting in Asia. *Atmos. Chem. Phys.* **8**, 2333–2340 (2008).
8. Tanaka, T. Y. et al. MASINGAR, a global tropospheric aerosol chemical transport model coupled with MRI/JMA98 GCM: model description. *Pap. Meteorol. Geophys.* **53**, 119–138 (2003).

9. Kim, M., Cho, J. H. & Ryoo, S. B. Development and assessment of ADAM3 ensemble prediction system. *Sci. Online Lett. Atmos.* **19**, 26–32 (2023).
10. Lucchesi, R. *File Specification for GEOS FP (Forward Processing)* GMAO Office Note No. 4 (Version 1.2) (NASA, 2018); http://gmao.gsfc.nasa.gov/pubs/office_notes.
11. Li, J. et al. Scattering and absorbing aerosols in the climate system. *Nat. Rev. Earth Environ.* **3**, 363–379 (2022).
12. Bellouin, N. et al. Bounding global aerosol radiative forcing of climate change. *Rev. Geophys.* **58**, 1–45 (2020).
13. Zhang, X. et al. Aerosol components derived from global AERONET measurements by GRASP: a new value-added aerosol component global dataset and its application. *Bull. Am. Meteorol. Soc.* **105**, E1822–E1848 (2024).
14. Dubovik, O. et al. Variability of absorption and optical properties of key aerosol types observed in worldwide locations. *J. Atmos. Sci.* **59**, 590–608 (2002).
15. Che, H. et al. Aerosol optical and radiative properties and their environmental effects in China: a review. *Earth Sci. Rev.* **248**, 104634 (2024).
16. Gui, K. et al. The significant contribution of small-sized and spherical aerosol particles to the decreasing trend in total aerosol optical depth over land from 2003 to 2018. *Engineering* **16**, 82–92 (2022).
17. Pöschl, U. Atmospheric aerosols: composition, transformation, climate and health effects. *Angew. Chem. Int. Ed.* **44**, 7520–7540 (2005).
18. Baklanov, A. & Zhang, Y. Advances in air quality modeling and forecasting. *Glob. Transit.* **2**, 261–270 (2020).
19. Hendricks, J., Righi, M. & Aquila, V. in *Atmospheric Physics. Research Topics in Aerospace* (ed. Schumann, U.) 561–576 (Springer, 2012); https://doi.org/10.1007/978-3-642-30183-4_34.
20. Sleeman, J. et al. Deep learning ensemble emulation for NASA's Goddard Earth Observing System (GEOS) composition forecast. *Artif. Intell. Earth Syst.* **4**, 1–12 (2025).
21. Lam, R. et al. Learning skillful medium-range global weather forecasting. *Science* **382**, 1416–1421 (2023).
22. Bi, K. et al. Accurate medium-range global weather forecasting with 3D neural networks. *Nature* **619**, 533–538 (2023).
23. Chen, L. et al. FuXi: a cascade machine learning forecasting system for 15-day global weather forecast. *npj Clim. Atmos. Sci.* **6**, 1–11 (2023).
24. Kochkov, D. et al. Neural general circulation models for weather and climate. *Nature* **632**, 1060–1066 (2024).
25. Bodnar, C. et al. A foundation model for the Earth system. *Nature* **641**, 1180–1187 (2025).
26. Xiong, W. et al. AI-GOMS: large AI-driven global ocean modeling system. Preprint at <https://arxiv.org/abs/2308.03152> (2023).
27. Cui, Y. et al. Forecasting the eddy ocean with a deep neural network. *Nat. Commun.* **16**, 2268 (2025).
28. Nowak, T. E., Augousti, A. T., Simmons, B. I. & Siebert, S. DustNet (v1): skilful neural network predictions of dust aerosols over the Saharan desert. *Geosci. Model Dev.* **18**, 3509–3532 (2025).
29. Cai, A. S., Fang, F., Peuch, V. & Alexe, M. Advancing operational PM_{2.5} forecasting with dual deep neural networks (D-DNet). Preprint at <https://arxiv.org/abs/2406.19154> (2024).
30. Randles, C. et al. The MERRA-2 Aerosol Reanalysis, 1980 onward. Part I: system description and data assimilation evaluation. *J. Clim.* **30**, 6823–6850 (2017).
31. Holben, B. N. et al. AERONET—a federated instrument network and data archive for aerosol characterization. *Remote Sens. Environ.* **66**, 1–16 (1998).
32. Che, H. et al. Instrument calibration and aerosol optical depth validation of the China aerosol remote sensing network. *J. Geophys. Res. Atmos.* **114**, D03206 (2009).
33. Hand, J. L., Schichtel, B. A., Pitchford, M., Malm, W. C. & Frank, N. H. Seasonal composition of remote and urban fine particulate matter in the United States. *J. Geophys. Res. Atmos.* **117**, D05209 (2012).
34. Zhang, X. Y. et al. Atmospheric aerosol compositions in China: spatial/temporal variability, chemical signature, regional haze distribution and comparisons with global aerosols. *Atmos. Chem. Phys.* **12**, 779–799 (2012).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons

Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2026

Methods

Datasets details

MERRA-2 reanalysis. MERRA-2, developed by NASA's Global Modeling and Assimilation Office (GMAO), is a comprehensive atmospheric reanalysis dataset that spans global atmospheric and climate conditions from 1980 to the present^{30,35}. By assimilating satellite, ground-based and additional observational data into the Goddard Earth Observing System, version 5 (GEOS-5) Earth system model, MERRA-2 provides high-precision meteorological parameters and multilayer atmospheric profiles. With its extensive temporal coverage, high spatial resolution (approximately 50 km) and robust consistency, MERRA-2 has become an indispensable tool for climate change research, air-quality monitoring and environmental policymaking. A defining innovation of MERRA-2 is its aerosol dataset, which integrates joint meteorological and aerosol data assimilation. To our knowledge, this marks the first time aerosol radiative effects have been incorporated directly into the atmospheric model^{30,35}, enhancing the fidelity of aerosol–meteorology interactions. MERRA-2 provides high-resolution data across multiple aerosol components—dust, sulfate, BC, OC and SS—with precise parameters such as spatial distribution, optical depth, concentration and radiative properties.

We used three subsets of the MERRA-2 time-averaged products— aerosol variables (tavg1_2d_aer_Nx), surface atmospheric variables (tavg1_2d_flux_Nx) and upper-air atmospheric variables (tavg3_3d_asm_Nv)—to train, test and evaluate AI-GAMFS, covering 44 years of data from 1980 to 2023. The dataset has spatial resolution of $0.5^\circ \times 0.625^\circ$ (361×576 latitude–longitude grid points). For each subset, only data from timestamps corresponding to the 3-hourly overlapping periods (01:30, 04:30, 07:30, ..., 22:30 UTC) were used. We focused on forecasting 12 aerosol variables: AOD, TSAOD, SUAOD, DUAOD, BCAOD, OCAOD, SSAOD, SUSMC, DUSMC, BCSMC, OCSMC and SSSMC. All aerosol optical variables are available at the wavelength of 550 nm. In addition, we forecasted 6 surface atmospheric variables and 4 upper-air atmospheric variables at 9 model levels (72, 68, 63, 60, 56, 53, 51, 48 and 45, corresponding to pressure levels of 985 hPa, 925 hPa, 850 hPa, 800 hPa, 700 hPa, 600 hPa, 525 hPa, 413 hPa and 288 hPa, respectively). Specifically, the six surface atmospheric variables are: surface specific humidity (QLML), surface air temperature (TLML), surface eastwards wind (ULML), surface northwards wind (VLML), sea-level pressure (SLP) and total precipitation (PRECTOT). The four upper-air atmospheric variables are: specific humidity (QV), air temperature (T), eastwards wind (U) and northwards wind (V). In total, we forecast and evaluated 54 variables. Detailed information on the MERRA-2 variables used in this study is provided in Supplementary Table 1.

GEOS-FP analyses and forecasts. GEOS-FP is a near-real-time analysis and forecasting system developed by the GMAO¹⁰. This system provides global meteorological and aerosol analyses (that is, assimilation fields) and generates 5-day (or 10-day) global forecasts, initialized daily at 00:00 UTC (12:00 UTC). It has grid resolution of approximately 25 km (latitude 0.25° , longitude 0.3125°). GEOS-FP uses the same model configuration as MERRA-2³⁰, including the simulation of dust, sulfate, BC, OC and SS via the Goddard Chemistry Aerosol Radiation and Transport model^{36,37}. In addition, GEOS-FP incorporates the assimilation of satellite-based bias-corrected AOD data³⁸.

We used three subsets from the GEOS-FP time-averaged analysis and forecast products (see also Supplementary Table 1), which are consistent with the nomenclature of the MERRA-2 data, to conduct a 5-day comparison experiment between AI-GAMFS historical deterministic and operational forecasts. These subsets contain 54 target variables that fully align with the inputs and outputs of the AI-GAMFS forecasts. To evaluate the forecast performance of AI-GAMFS relative to other global and regional aerosol forecasting models, we used the historical GEOS-FP analyses and MERRA-2 reanalysis data from 22:30 UTC each day in 2023 to drive AI-GAMFS and generate daily 5-day forecasts for the

entire year of 2023. In contrast, collecting historical GEOS-FP forecast data is more challenging because the GMAO archives only the most recent 2 weeks of forecast data. Consequently, we collected only the GEOS-FP analyses at 00:00 UTC and 5-day forecast data (initialized daily at 00:00 UTC) from July to August 2024 for the near-real-time operational comparison between AI-GAMFS and GEOS-FP. To drive AI-GAMFS and conduct the comparative analysis, we used bilinear interpolation to resample the GEOS-FP analysis and forecast data to match the spatial resolution of $0.5^\circ \times 0.625^\circ$.

CAMS aerosol forecasts. CAMS, developed by the European Centre for Medium-Range Weather Forecasts, is one of the most advanced global aerosol forecasting systems⁵. It provides twice-daily forecasts of global atmospheric composition, including 5-day forecasts of AOD and DUAOD. Using data assimilation techniques, CAMS integrates prior forecasts with current satellite observations to derive optimal initial conditions. It then applies a numerical atmospheric model based on physical and chemical principles to forecast the evolution of aerosols and other atmospheric compositions over the next 5 days^{5,39}. The spatial resolution of the CAMS aerosol forecast product at a single level is $0.4^\circ \times 0.4^\circ$, with 1-h temporal resolution.

In this study, CAMS served as the baseline for global AOD and DUAOD forecasts based on a physical model, facilitating comprehensive comparison with AI-GAMFS. We used the 5-day global AOD and DUAOD forecasts for the entire year of 2023, initialized daily at 00:00 UTC. To align with AI-GAMFS for comparison or analysis, we resampled the CAMS forecast data to match the spatial resolution of $0.5^\circ \times 0.625^\circ$ and the temporal resolution of 3 h, using time interpolation and bilinear interpolation.

Physical-based dust forecasts. In this study, we used the 2023 dust forecast products from five physical-based dust forecasting models developed by various institutions and deployed in the SDS-WAS Asian Regional Centre. These products include two global models: CAMS and FMI-SILAM⁶, with FMI-SILAM having 1-h temporal resolution and spatial resolution of $0.2^\circ \times 0.2^\circ$. In addition, we analysed 3 regional models: CMA-CUACE/Dust⁷, with 3-h temporal resolution and a spatial resolution of $0.5^\circ \times 0.5^\circ$; JMA-MASINGAR⁸, with 1-h temporal resolution and a spatial resolution of $0.5^\circ \times 0.5^\circ$; and KMA-ADAM3⁹, with 3-h temporal resolution and a spatial resolution of $0.5^\circ \times 0.5^\circ$ (Supplementary Table 1). The JMA-MASINGAR model provides forecasts with a 3-day lead time, and all other models provide forecasts with a 5-day lead time. Detailed descriptions of these models can be found in their respective technical documentation^{5–9}.

Owing to differences in initialization times and dust output variables across the models, we used DUAOD forecast outputs from CAMS, FMI-SILAM, CMA-CUACE/Dust, JMA-MASINGAR and KMA-ADAM3, and for DUSMC, we utilized outputs from all models except CAMS. Notably, except for JMA-MASINGAR, which initializes at 12:00 UTC, all other forecast products begin at 00:00 UTC. To facilitate comparison, we unified the spatial and temporal resolutions of all model outputs to match the AI-GAMFS spatial resolution of $0.5^\circ \times 0.625^\circ$ and 3-h temporal resolution.

AERONET and CARSNET measurements. AERONET is a global aerosol observation network that provides high-quality ground-based measurements of aerosol optical properties³¹. The network consists of numerous automated stations equipped with sun photometers to monitor AOD and other aerosol parameters in real time. AERONET data are widely regarded as the 'gold standard' in atmospheric aerosol observations, serving as high-precision references for climate studies, air-quality monitoring and satellite remote-sensing validation. In this study, we used instantaneous AOD and AODc data (Version 3.0, Level 2.0)⁴⁰ from all available AERONET sites worldwide during 2023 and July–August 2024. Owing to the lack of a direct method to derive

DUAOD, we used AODc as a proxy⁴¹ to evaluate the DUAOD forecasts. Furthermore, AERONET does not provide AOD or AODc measurements at 550 nm; therefore, we derived AODc at 550 nm from AODc at 500 nm using the Ångström exponent. We used the following quadratic polynomial interpolation method^{42,43} to convert AOD observations at 4 adjacent wavelengths (440 nm, 500 nm, 675 nm and 870 nm) into AOD values at 550 nm:

$$\ln(\tau_\lambda) = a_0 + a_1 \ln(\lambda) + a_2 [\ln(\lambda)]^2 \quad (1)$$

where a_0 , a_1 and a_2 represent fitting coefficients, and τ_λ denotes the AOD value at the respective wavelength λ .

To complement the sparse distribution of AERONET sites across China, we incorporated cloud-screened instantaneous AOD observations at 550 nm (Level 2.0) from CARSNET during 2023 and July–August 2024 into the regional evaluation. CARSNET, which is a ground-based aerosol monitoring network established by the China Meteorological Administration in 2002³², uses a systematic calibration protocol: field instruments are calibrated against CARSNET reference standards that are themselves regularly calibrated in coordination with the AERONET programme^{32,44}. Consequently, CARSNET provides AOD data with accuracy comparable to that of AERONET, showing an estimated uncertainty of 0.01–0.02 (ref. 32). To ensure the accuracy of the evaluation, we averaged the AERONET or CARSNET instantaneous observations within a half-hour window before and after the forecast lead time, which served as the reference truth.

In situ aerosol component measurements. To evaluate the forecast accuracy of operational AI-GAMFS for surface aerosol components against GEOS-FP, we collected in situ measurements of aerosol chemical components over the USA and China during July–August 2024. For the USA, daily data on BCSMC, OCSMC and SUSMC were obtained from the IMPROVE network³³, with additional daily OCSMC and SUSMC data sourced from the EPA-CSN network³³. All datasets were screened using available data-quality flags. For China, we used quality-controlled data from the CAWNET network³⁴, including hourly BCSMC and daily OCSMC and SUSMC measurements.

AI-GAMFS details

As illustrated in Fig. 1a, the AI-GAMFS architecture consists of three primary modules: cube embedding, a vision transformer and cube unembedding. The base model of AI-GAMFS is an autoregressive model that uses the spatial feature tensor at the previous time step (X^{t-n}) as input to forecast the spatial feature tensor at the next time step (X^t). Here $t-n$ and t represent the previous and upcoming time steps, respectively. The base model considers time steps of 3 h, 6 h, 9 h and 12 h. Using the output of the base model as input, AI-GAMFS can generate forecasts for different lead times. For detailed descriptions of the modelling process for each module and the sensitivity analysis of key hyperparameters, see Supplementary Note 1 and Supplementary Fig. 15.

Training strategy. We utilized the MERRA-2 reanalysis with 3-h temporal resolution to train the AI-GAMFS model. Data from 1980 to 2021 were used for training, data from 2022 served as the test set and data from 2023 were used for validation. All input variables, except for time features, were standardized before being processed by the embedding layer, and the output from the unembedding layer was unstandardized to generate the final forecasts. The model uses a rolling training approach, where pairs of samples from two consecutive time points (X^{t-n} and X^t) are fed iteratively into the model for training.

For the standardized samples, the mean absolute error was used as the loss function:

$$L_1 = \frac{1}{C \times H \times W} \sum_{c=1}^C \sum_{i=1}^H \sum_{j=1}^W |X_{c,i,j}^t - X_{c,i,j}^t| \quad (2)$$

where C , H and W denote the number of variables, the latitudinal grid points and the longitudinal grid points, respectively; c , i and j are the indices for variables, latitude and longitude coordinates, respectively; and $X_{c,i,j}^t$ and $\hat{X}_{c,i,j}^t$ represent the ‘ground truth’ (that is, MERRA-2) and the forecasted value at the specified forecasting time, respectively.

The AI-GAMFS framework was implemented on the PyTorch platform. Each model, corresponding to a specified lead time and containing approximately 1.2 billion parameters, was trained on a server equipped with 8 L40 GPUs for 80 epochs (approximately 10 days). We used the Adam optimizer with $\beta_1 = 0.9$ and $\beta_2 = 0.999$, and an initial learning rate of 3×10^{-4} , which was decayed using a cosine annealing schedule to 0.001 of its initial value. Training was conducted in 32-bit floating-point precision with a dropout rate of 0.15 to mitigate overfitting.

To evaluate the temporal robustness of AI-GAMFS against potential long-term aerosol trends within MERRA-2, we conducted a stratified cross-validation experiment. The training set (1980–2021) was partitioned into 7 contiguous 6-year subsets. The model was iteratively trained on six out of seven subsets and validated on the remaining withheld subset (Supplementary Fig. 13). Results from all seven validations were compared with the model’s performance on the 2022 test set, which was trained on data for the full 1980–2021 period. The overall consistent performance observed across all validations demonstrates that AI-GAMFS captures underlying evolution patterns rather than temporal artefacts, ensuring its reliability for extrapolative forecasting (Supplementary Fig. 14).

Forecasting strategy. Similar to physics-based forecasting models, we observed that forecast errors in deep-learning models accumulate and amplify as the number of rolling iterations increases. Inspired by the temporal aggregation method from Pangu-Weather²², we adopted a relay forecasting strategy that reduces the number of model iterations without compromising the forecast time resolution. Using the same modelling framework and configurations, we trained 4 pretrained AI-GAMFS models with lead times of 3 h, 6 h, 9 h and 12 h, referred to as the 3-h, 6-h, 9-h and 12-h models, respectively. For forecasts with specific lead times, we prioritize the 12-h model and combine it with shorter timescale models in a relay fashion (Fig. 1b and Extended Data Fig. 1). As an example, for a lead time of 54 h, the 12-h forecast model is first invoked 4 times, followed by a single invocation of the 6-h forecast model (Extended Data Fig. 1a). Although this strategy sacrifices some computational efficiency, it takes advantage of the high-speed capabilities of GPUs, enabling the model to produce a 5-day forecast in approximately 39 s on a single L40 GPU.

Evaluation experiment

To rigorously evaluate the forecasting capabilities of AI-GAMFS, we conducted a series of evaluation experiments, using MERRA-2 reanalysis and observational data as reference baselines.

AI-GAMFS relay forecast evaluation. We compared 4 AI-GAMFS model configurations on the 2022 test set, encompassing all 54 aerosol and meteorological variables. These configurations included: a 3-h single model, a 3-h and 6-h relay model, a 3-h, 6-h and 9-h relay model, and a 3-h, 6-h, 9-h and 12-h relay model. This evaluation provides insight into the optimal relay configurations for enhanced predictive performance.

AI-GAMFS versus regional dust forecasting models. We evaluated AI-GAMFS forecasts against five physics-based dust forecasting models across East Asia, using the 2023 validation dataset with MERRA-2 as the baseline. The models included in this comparison—CAM5, CMA-CUACE/Dust, FMI-SILAM, JMA-MASINGAR and KMA-ADAM3—are either specialized dust forecasting models or aerosol models with dust-specific outputs. The evaluation focused on two critical parameters: DUAOD and DUSMC, which allowed us to evaluate AI-GAMFS’s accuracy and reliability in forecasting dust storm events. We also used the full year

of AODc observations for 2023 from AERONET at the Beijing-CAMS site for independent evaluation.

AI-GAMFS versus CAMS in global AOD and DUAOD forecasts. We conducted a spatial comparison of AI-GAMFS and CAMS in terms of their 5-day AOD and DUAOD forecasts on a global scale in 2023, using MERRA-2 as the baseline. In addition, the forecasts from AI-GAMFS and CAMS were further evaluated against both global AERONET observations and CARSNET observations from China throughout 2023.

Operational performance of AI-GAMFS versus GEOS-FP. AI-GAMFS is designed for real-time operational forecasting and utilizes GEOS-FP real-time analyses to generate global 5-day aerosol–meteorology forecasts. To evaluate its operational forecasting capabilities for various aerosol components and meteorological variables, we analysed GEOS-FP forecast outputs for July and August 2024. A detailed comparative assessment of AI-GAMFS and GEOS-FP was performed using MERRA-2 as the reference baseline, focusing on all 54 target aerosol and meteorological variables. In addition, the aerosol component forecasts from AI-GAMFS and GEOS-FP were evaluated further against AERONET, CARSNET, IMPROVE, EPA-CSN and CAWNET observations.

Evaluation metrics. For the site-scale evaluation, using independent observations as the reference baseline, we used two metrics: RMSE and Pearson's *R*. For the spatial evaluation, using MERRA-2 as the reference baseline, we used two metrics: latitude-weighted RMSE and spatial *R*, defined as follows:

$$\text{RMSE}(c, t) = \sqrt{\frac{\sum_{i=1}^{N_{\text{lat}}} \sum_{j=1}^{N_{\text{lon}}} w_i (\hat{X}_{c,i,j}^t - X_{c,i,j}^t)^2}{N_{\text{lat}} \times N_{\text{lon}}}} \quad (3)$$

$$R(c, t) = \frac{\sum_{i=1}^{N_{\text{lat}}} \sum_{j=1}^{N_{\text{lon}}} (\hat{X}_{c,i,j}^t - \bar{X}_{c,i,j}^t) (X_{c,i,j}^t - \bar{X}_{c,i,j}^t)}{\sqrt{\sum_{i=1}^{N_{\text{lat}}} \sum_{j=1}^{N_{\text{lon}}} (\hat{X}_{c,i,j}^t - \bar{X}_{c,i,j}^t)^2} \times \sqrt{\sum_{i=1}^{N_{\text{lat}}} \sum_{j=1}^{N_{\text{lon}}} (X_{c,i,j}^t - \bar{X}_{c,i,j}^t)^2}} \quad (4)$$

where the latitude weight w_i is given by $w_i = N_{\text{lat}} \times \frac{\cos \phi_i}{\sum_{i=1}^{N_{\text{lat}}} \cos \phi_i}$, c represents the specified variable, ϕ_i refers to the latitudinal value, N_{lat} is the total number of latitudinal grids, and $\bar{X}_{c,i,j}^t$ and $\bar{X}_{c,i,j}^t$ correspond to the spatial averages over all grid points of the forecast field and the ground-truth field, respectively.

Data availability

MERRA-2 reanalysis are available at https://gmao.gsfc.nasa.gov/gmao-products/merra-2/data-access_merra-2/. GEOS-FP aerosol and meteorological analyses, as well as 5-day forecasts, are available from <https://portal.nccs.nasa.gov/datashare/gmao/geos-fp/>. The CAMS global aerosol forecasts are available from the Copernicus Atmosphere Data Store (<https://ads.atmosphere.copernicus.eu/datasets/cams-global-atmospheric-composition-forecasts>). Regional dust forecasting data from FMI-SILAM, CMA-CUACE/Dust, JMA-MASINGAR and KMA-ADAM3 are available from the Sand and Dust Storm Warning Advisory and Assessment System (SDS-WAS) Asian Regional Center (<http://www.asdf-bj.net/publish/forecasts/ens/ea/surface-concentration/index.html>). The global instantaneous aerosol observational data from AERONET sites are available at <https://aeronet.gsfc.nasa.gov/>. The instantaneous aerosol observational data from CARSNET and the daily surface aerosol component concentration data from CAWNET used in this study are available at <https://zenodo.org/records/18298799> (ref. 45). The surface daily aerosol component concentration measurements

from the IMPROVE network and EPA-CSN in the USA are available at <https://views.cira.colostate.edu/fed/QueryWizard/Default.aspx>. The AI-GAMFS has been operational on the China Science and Technology Cloud since 20 January 2025. Since then, historical global aerosol forecast data—updated 8 times daily, providing 5-day forecasts at 3-hourly intervals—have been freely available at <http://tapdata.org.cn> and <https://obs.cstcloud.cn/s/ai-gamfs/AI-GAMFS/>. In addition, the 5-day, 3-hourly forecasts of DUAOD and DUSMC from AI-GAMFS have also been available on the International Meteorological Early Warning Operational Support Platform hosted by the World Meteorological Centre Beijing (<http://ew4all.wmc-bj.net/EW4ALL/warning>). All plots were made using Matplotlib (<https://pypi.org/project/matplotlib/>) and the geographical maps were produced using Cartopy (<https://pypi.org/project/Cartopy/>). Source data are provided with this paper.

Code availability

The code for operational AI-GAMFS is available at <https://zenodo.org/records/18298799> (ref. 45).

35. Buchard, V. et al. The MERRA-2 aerosol reanalysis, 1980 onward. Part II: evaluation and case studies. *J. Clim.* **30**, 6851–6872 (2017).
36. Chin, M. et al. Tropospheric aerosol optical thickness from the GOCART model and comparisons with satellite and sun photometer measurements. *J. Atmos. Sci.* **59**, 461–483 (2002).
37. Colarco, P., Da Silva, A., Chin, M. & Diehl, T. Online simulations of global aerosol distributions in the NASA GEOS-4 model and comparisons to satellite and ground-based aerosol optical depth. *J. Geophys. Res. Atmos.* **115**, D14207 (2010).
38. Albayrak, A., Wei, J., Petrenko, M., Lynnes, C. & Levy, R. C. Global bias adjustment for MODIS aerosol optical thickness using neural network. *J. Appl. Remote Sens.* **7**, 073514 (2013).
39. Inness, A. et al. The CAMS reanalysis of atmospheric composition. *Atmos. Chem. Phys.* **19**, 3515–3556 (2019).
40. Sinyuk, A. et al. The AERONET Version 3 aerosol retrieval algorithm, associated uncertainties and comparisons to Version 2. *Atmos. Meas. Tech.* **13**, 3375–3411 (2020).
41. Pu, B. & Ginoux, P. How reliable are CMIP5 models in simulating dust optical depth? *Atmos. Chem. Phys.* **18**, 12491–12510 (2018).
42. Eck, T. F. et al. Wavelength dependence of the optical depth of biomass burning, urban, and desert dust aerosols. *J. Geophys. Res. Atmos.* **104**, 31333–31349 (1999).
43. Schuster, G. L., Dubovik, O. & Holben, B. N. Angstrom exponent and bimodal aerosol size distributions. *J. Geophys. Res. Atmos.* **111**, D07207 (2006).
44. Che, H. et al. Ground-based aerosol climatology of China: aerosol optical depths from the China Aerosol Remote Sensing Network (CARSNET) 2002–2013. *Atmos. Chem. Phys.* **15**, 7619–7652 (2015).
45. Gui, K. et al. AI-driven Global Aerosol–Meteorology Forecasting System (AI-GAMFS). *Zenodo* <https://doi.org/10.5281/zenodo.18298799> (2026).

Acknowledgements We thank NASA for providing the MERRA-2 reanalysis, GEOS-FP analysis and forecast data, and AERONET observational datasets; the European Centre for Medium-Range Weather Forecasts for supplying the CAMS aerosol forecast products; the SDS-WAS for the regional dust forecast products; and the IMPROVE and EPA-CSN networks for the surface aerosol component data. This research was supported by the National Natural Science Foundation of China (42588301, 42522509, 42090033, 42175153, 42375188 and 42275195), the National Key Research and Development Program of China (2023YFC3706305), the Youth Innovation Team of China Meteorological Administration (CMA2024QN13) and the Basic Research Fund of CAMS (2023Z021 and 2025XM002). JW's participation is supported by the Lichtenberger Family Chair professorship in the University of Iowa.

Author contributions K.G. and Xutao Zhang contributed equally to this work. H.C. and Xiaoye Zhang conceived of the study and led the overall scientific questions. K.G. and Xutao Zhang trained AI-GAMFS and analysed the data. H.C. and Xiaoye Zhang managed and oversaw the research project. K.G., Xutao Zhang, L.L., Y.Z., L.A., Y.M. and H.Z. prepared the data. O.D., B.H., J.W., P.G., E.S.L., C.T., H.W., Z.W., Y.W., X.H., K.D., X. Xia and X. Xu provided constructive suggestions to improve this study. K.G. and Xutao Zhang wrote the paper. All authors gave feedback and contributed to editing the paper.

Competing interests The authors declare no competing interests.

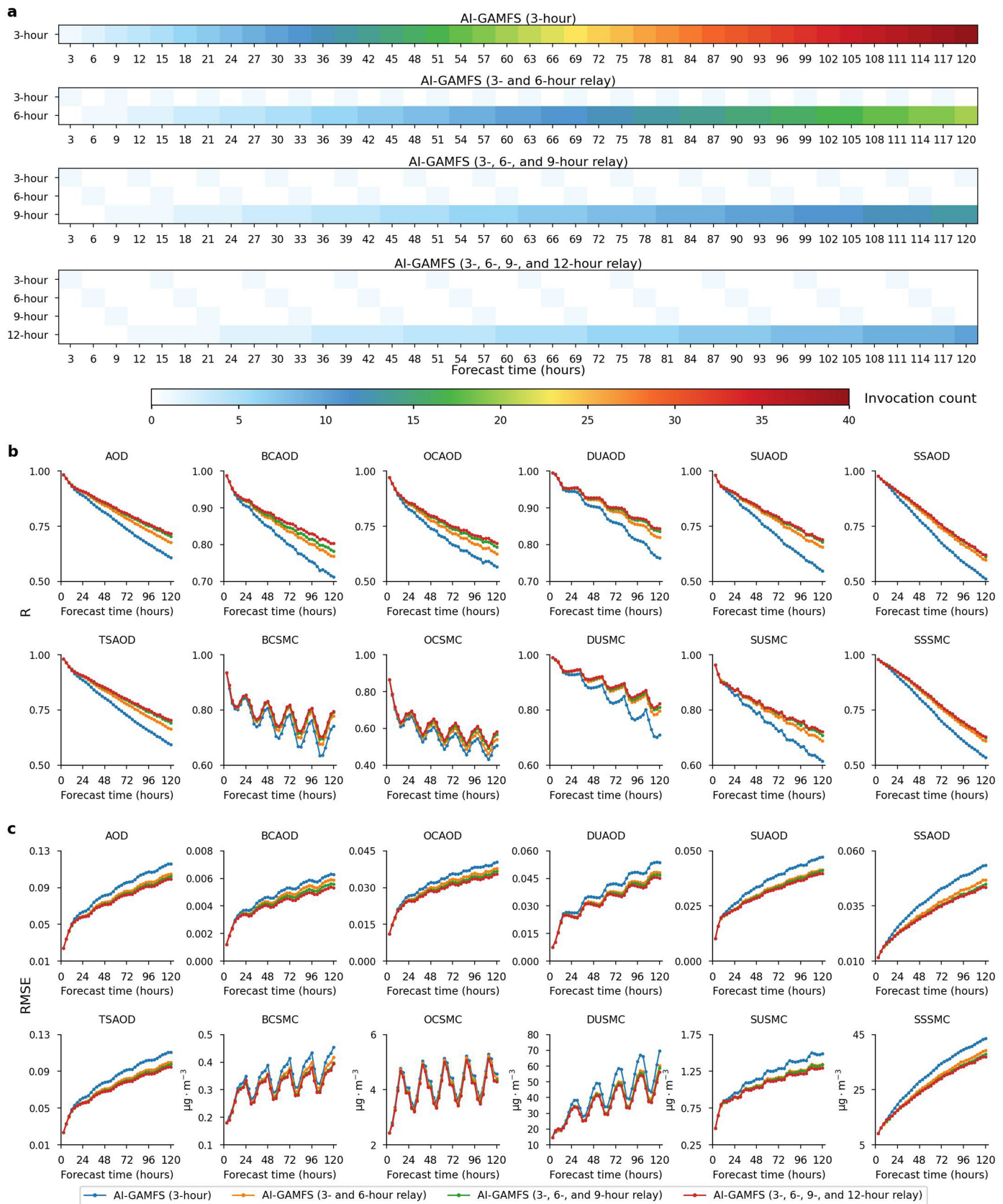
Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41586-026-10234-y>.

Correspondence and requests for materials should be addressed to Huizheng Che or Xiaoye Zhang.

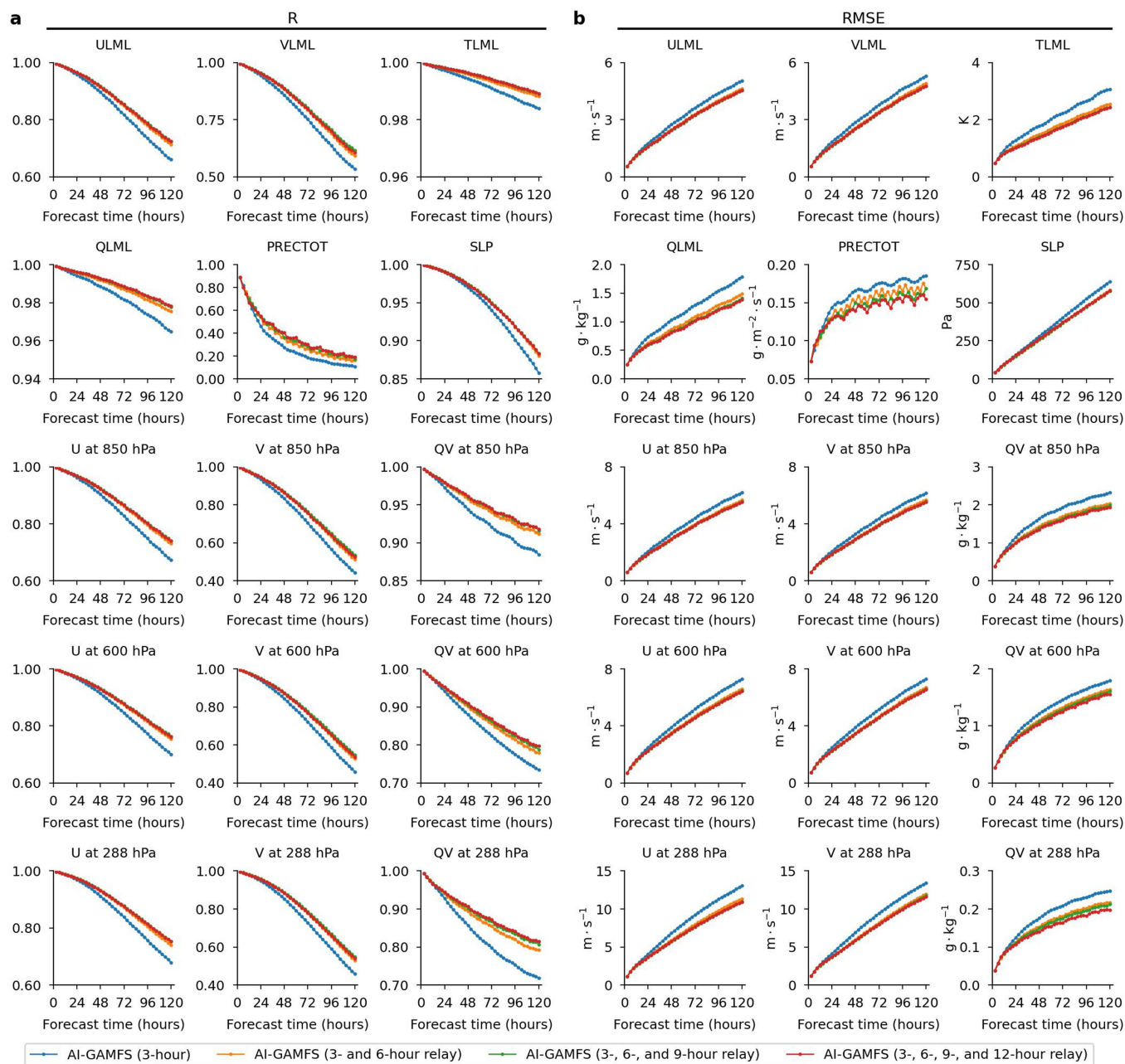
Peer review information *Nature* thanks Daniel Rothenberg and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permissions information is available at <http://www.nature.com/reprints>.



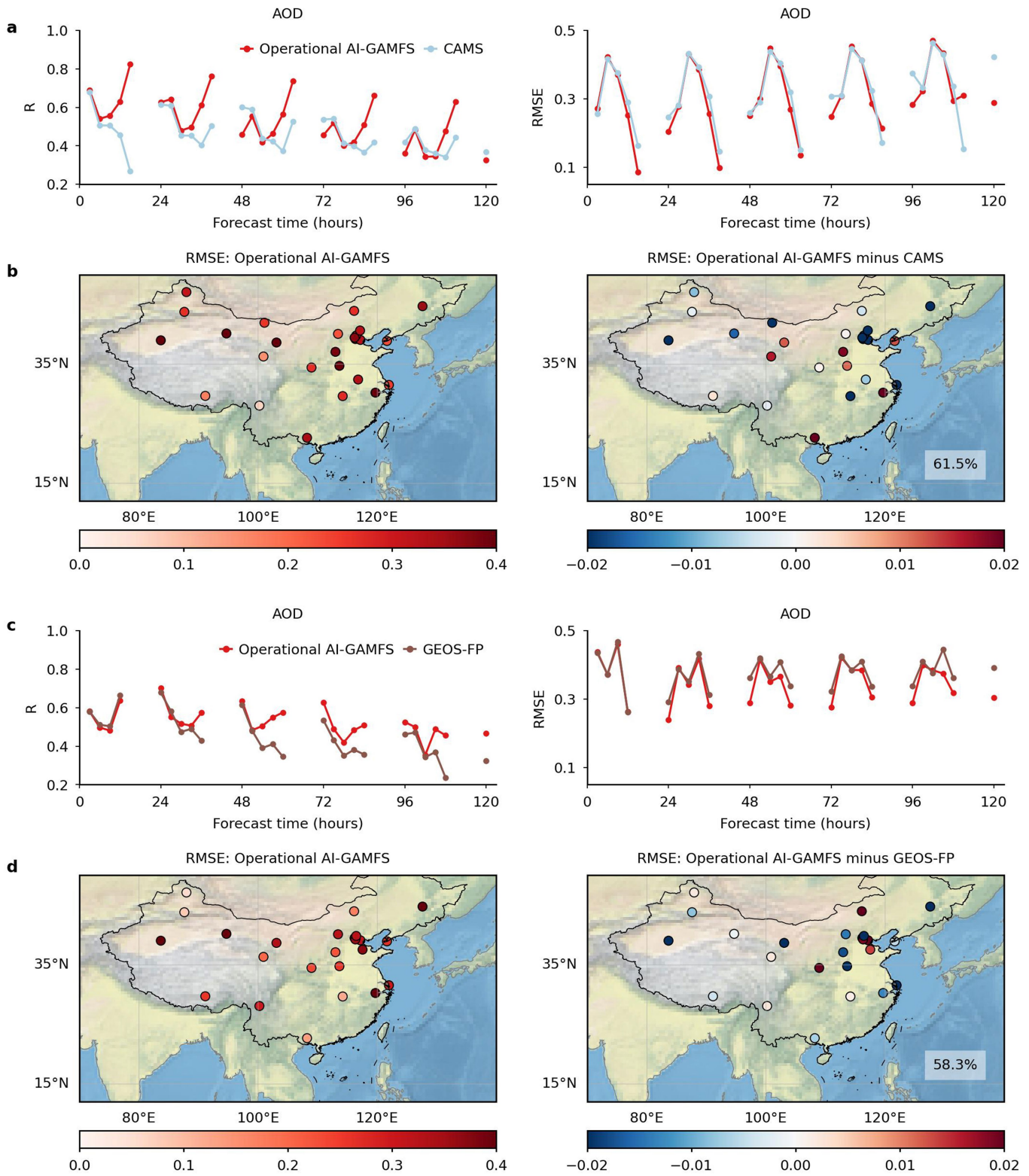
Extended Data Fig. 1 | Comparison of aerosol forecasting accuracy between single-model and multimodel relay forecasting approaches. a, Invocation count of the 3-, 6-, 9-, and 12-h models at different forecast lead times, using the 3-h single model, the 3- and 6-h relay model, the 3-, 6-, and 9-h relay model, and

the 3-, 6-, 9-, and 12-h relay model. **b, c,** Accuracy of global 5-day deterministic forecasts for 12 aerosol variables in the 2022 test set, showing spatial R (**b**) and latitude-weighted RMSE (**c**) over time, for the 3-h single model, the 3- and 6-h relay model, the 3-, 6-, and 9-h relay model, and the 3-, 6-, 9-, and 12-h relay model.



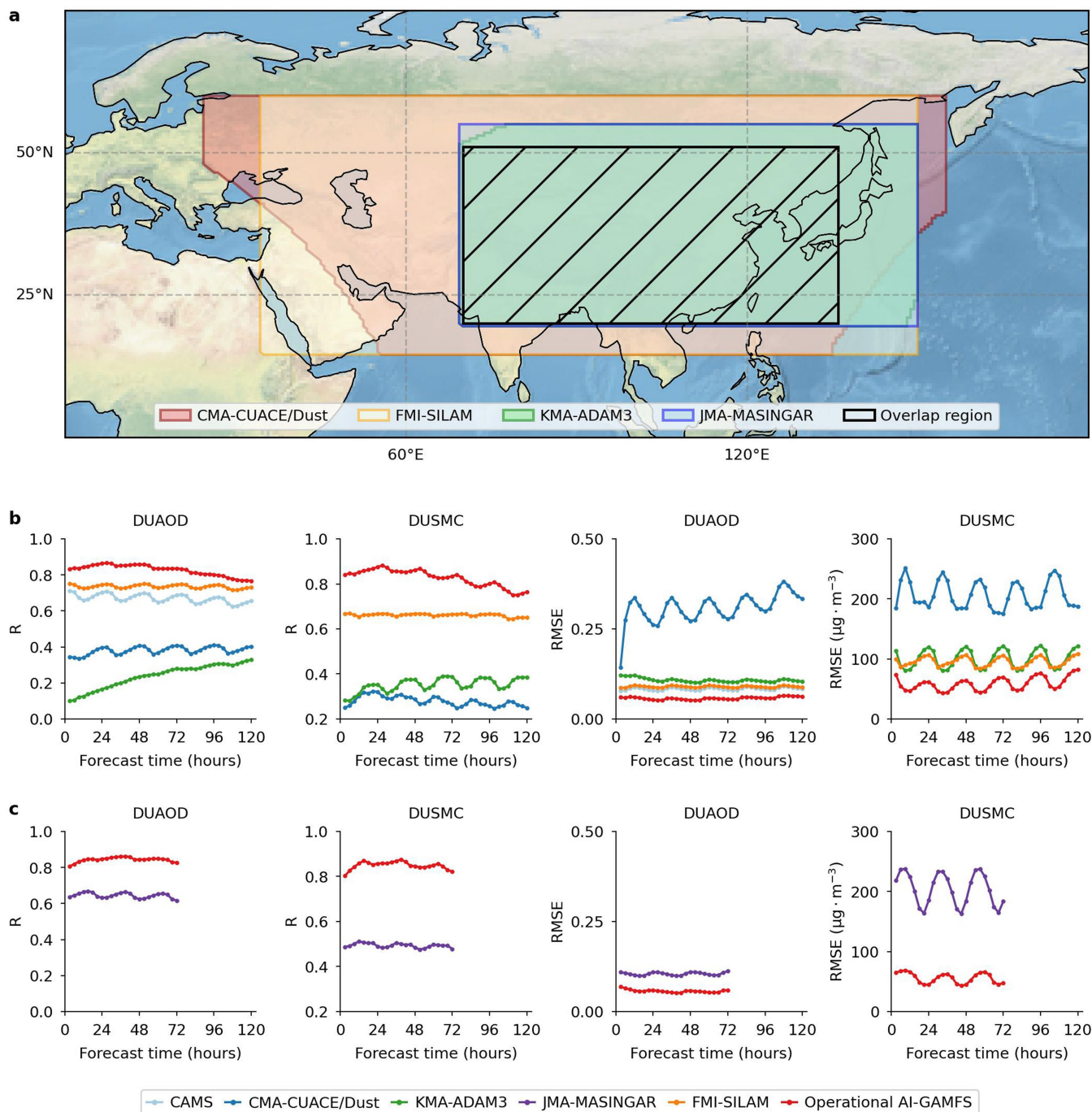
Extended Data Fig. 2 | Comparison of meteorological variable forecasting accuracy between single-model and multimodel relay forecasting approaches. a,b, Accuracy of global 5-day deterministic forecasts for six surface meteorological variables and nine selected upper-level meteorological variables in 2022, showing spatial R (**a**) and latitude-weighted RMSE (**b**) over time. The comparisons include the 3-h single-model, 3- and 6-h relay model,

3-, 6-, and 9-h relay model, and 3-, 6-, 9-, and 12-h relay model. The six surface meteorological variables comprise surface eastward wind (ULML), surface northward wind (VLML), surface air temperature (TLML), surface specific humidity (QLML), total precipitation (PRECTOT), and sea level pressure (SLP). The nine upper-level meteorological variables comprise eastward wind (U), northward wind (V), and specific humidity (QV) each at 850, 600, and 288 hPa.



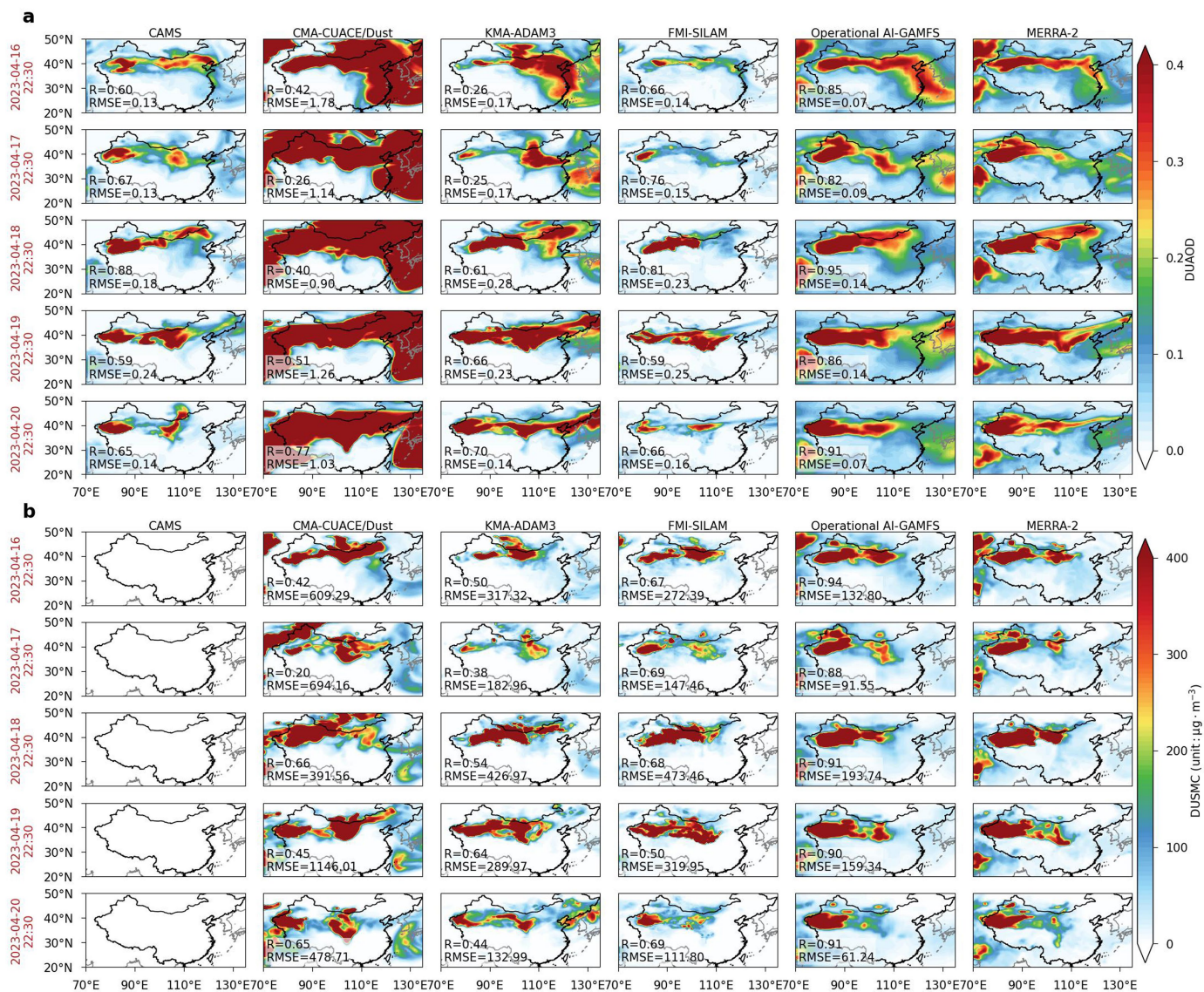
Extended Data Fig. 3 | Comparison of AOD forecasting accuracy among the operational AI-GAMFS, CAMS and GEOS-FP systems. a, R and RMSE of operational AI-GAMFS forecasts compared with CAMS. Forecasts are for 5 days at 3-h resolution evaluated against AOD observations from 26 CARSNET sites across China in 2023. Evaluation metrics (R and RMSE) for each step were calculated by aggregating all matched samples over the entire domain. **b,** Spatial distribution of the average RMSE for each step of the 5-day AOD forecast (a total of 30 steps) from operational AI-GAMFS, alongside the RMSE difference

between CAMS and operational AI-GAMFS. **c,** As in **a** but for operational AI-GAMFS forecasts compared with GEOS-FP, evaluated against 24 CARSNET sites during July–August 2024. **d,** Spatial distribution of the average RMSE for each step of the 5-day AOD forecast (a total of 25 steps) from operational AI-GAMFS, alongside the RMSE difference between GEOS-FP and operational AI-GAMFS. The percentages in the lower-right corners of the RMSE difference panels indicate the proportion of sites where operational AI-GAMFS showed lower RMSEs than those of CAMS or GEOS-FP.



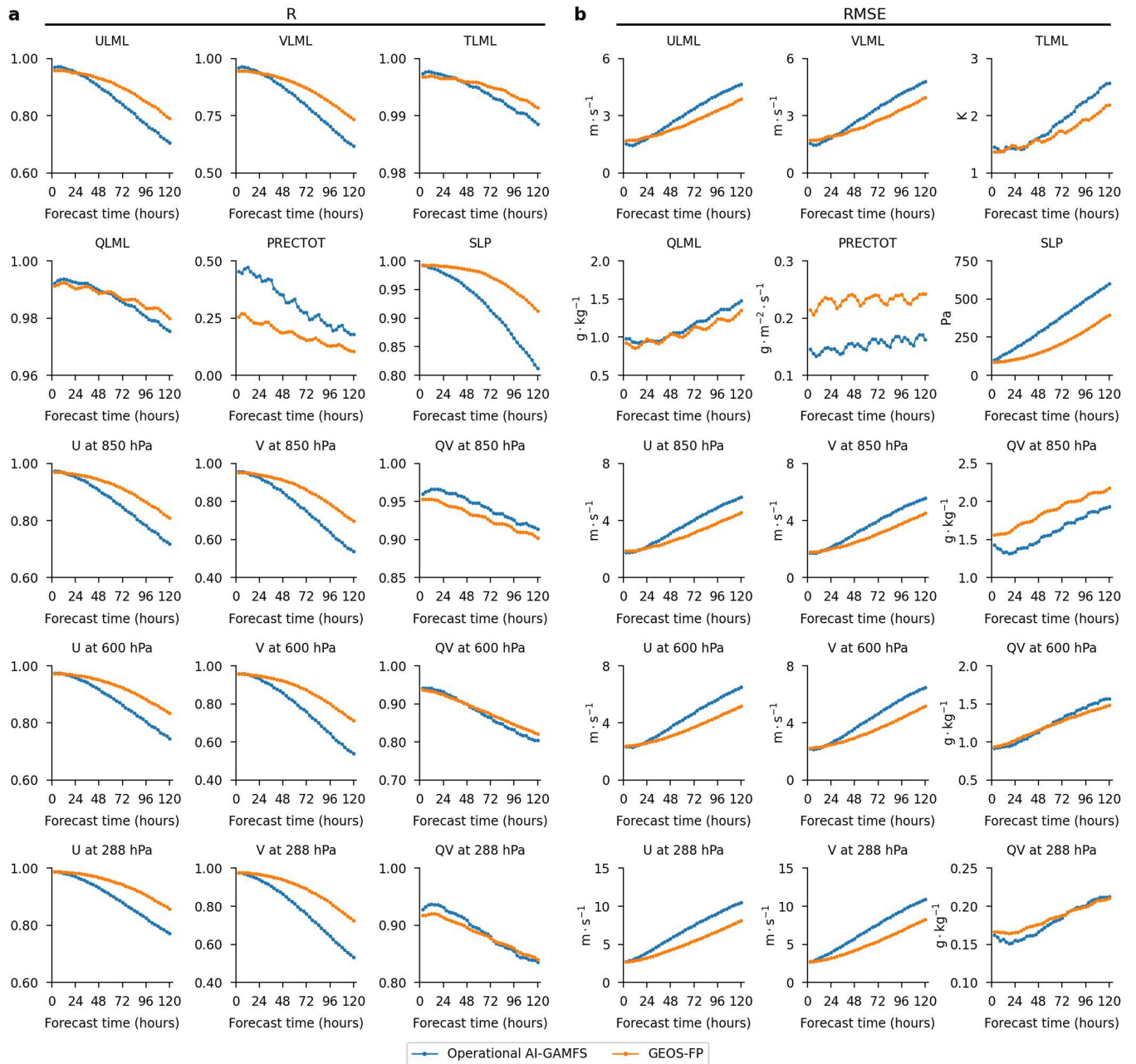
Extended Data Fig. 4 | Superior performance of operational AI-GAMFS in dust forecasting over East Asia throughout 2023. **a**, Spatial distribution of forecast products from four dust models deployed at the SDS-WAS Asian Regional Centre: CUACE/Dust from the China Meteorological Administration (CMA-CUACE/Dust), SILAM from the Finnish Meteorological Institute (FMI-SILAM), ADAM3 from the Korea Meteorological Agency (KMA-ADAM3), and MASINGAR from the Japan Meteorological Agency (JMA-MASINGAR). The area enclosed by the black border represents the overlap of the forecast

regions from all four models used in this study for comparative analysis and evaluation of the East Asia region. **b**, Comparison of 5-day forecast accuracy for DUAOD and DUSMC over East Asia using MERRA-2 as the reference baseline, evaluated by spatial R and latitude-weighted RMSE, with models including operational AI-GAMFS (initialized by GEOS-FP analyses), CAMS, CMA-CUACE, KMA-ADAM3, and FMI-SILAM. **c**, Comparison of 3-day forecast accuracy for DUAOD and DUSMC from operational AI-GAMFS and JMA-MASINGAR, evaluated against MERRA-2 as the reference baseline.



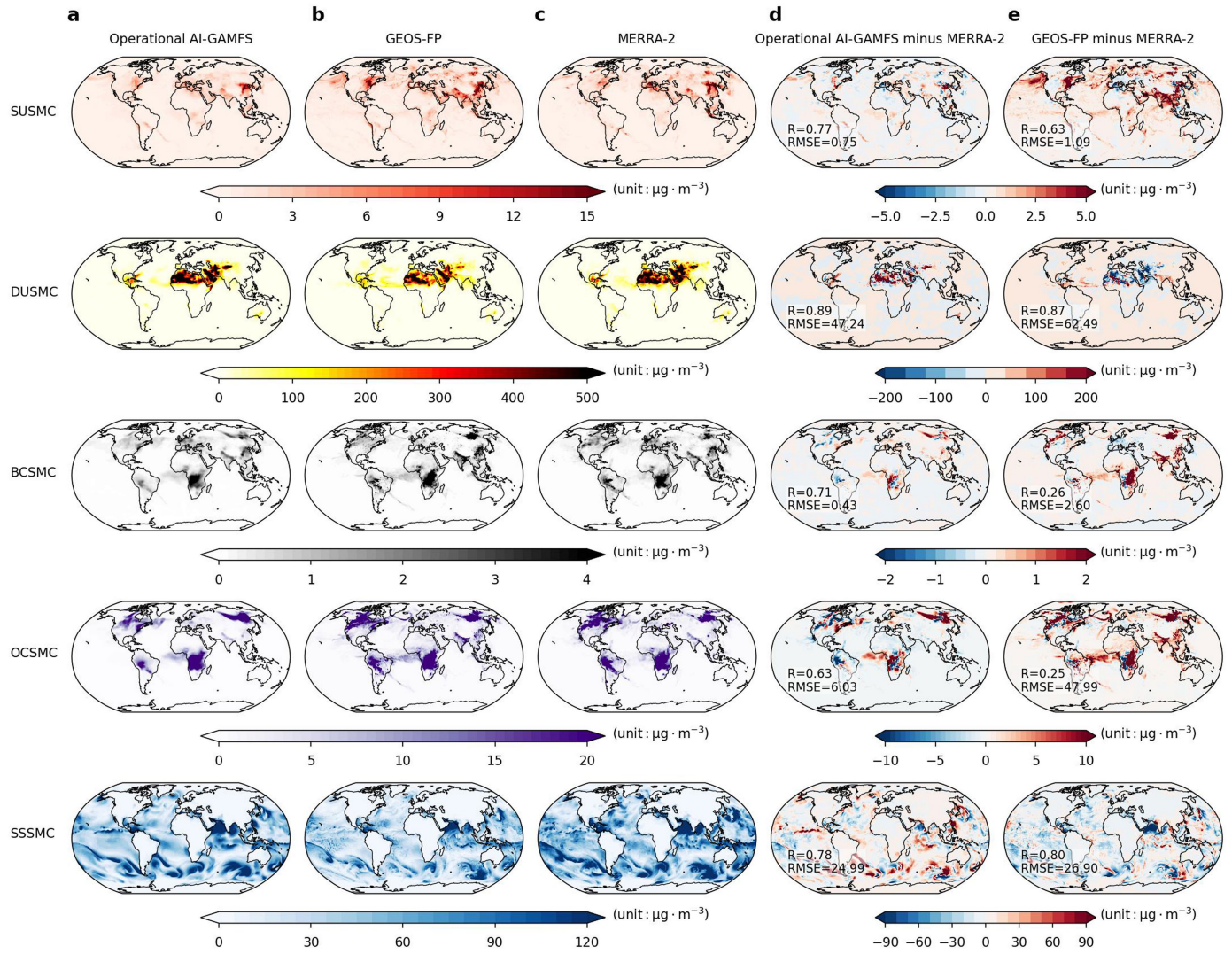
Extended Data Fig. 5 | Case study of a mega dust storm event in northern China during April 16–20, 2023. a, b. Forecasts of DUAOD (a) and DUSMC (b) during the event, produced by operational AI-GAMFS, CAMS, CMA-CUACE/Dust, KMA-ADAM3, and FMI-SILAM models, with a 5-day lead time at 24-h intervals, compared against MERRA-2 reference data. Operational AI-GAMFS was initialized at 22:30 UTC on April 15, 2023, while the other models were

initialized at 00:00 UTC on April 16, 2023. For comparison, the outputs of all models were interpolated to the same spatiotemporal resolution as AI-GAMFS. The overall accuracy metrics (i.e., spatial R and RMSE) for operational AI-GAMFS and the four dust models relative to MERRA-2 data are shown in the lower-left corner of each panel. Note that DUSMC is not included as a direct forecast variable for the CAMS model.



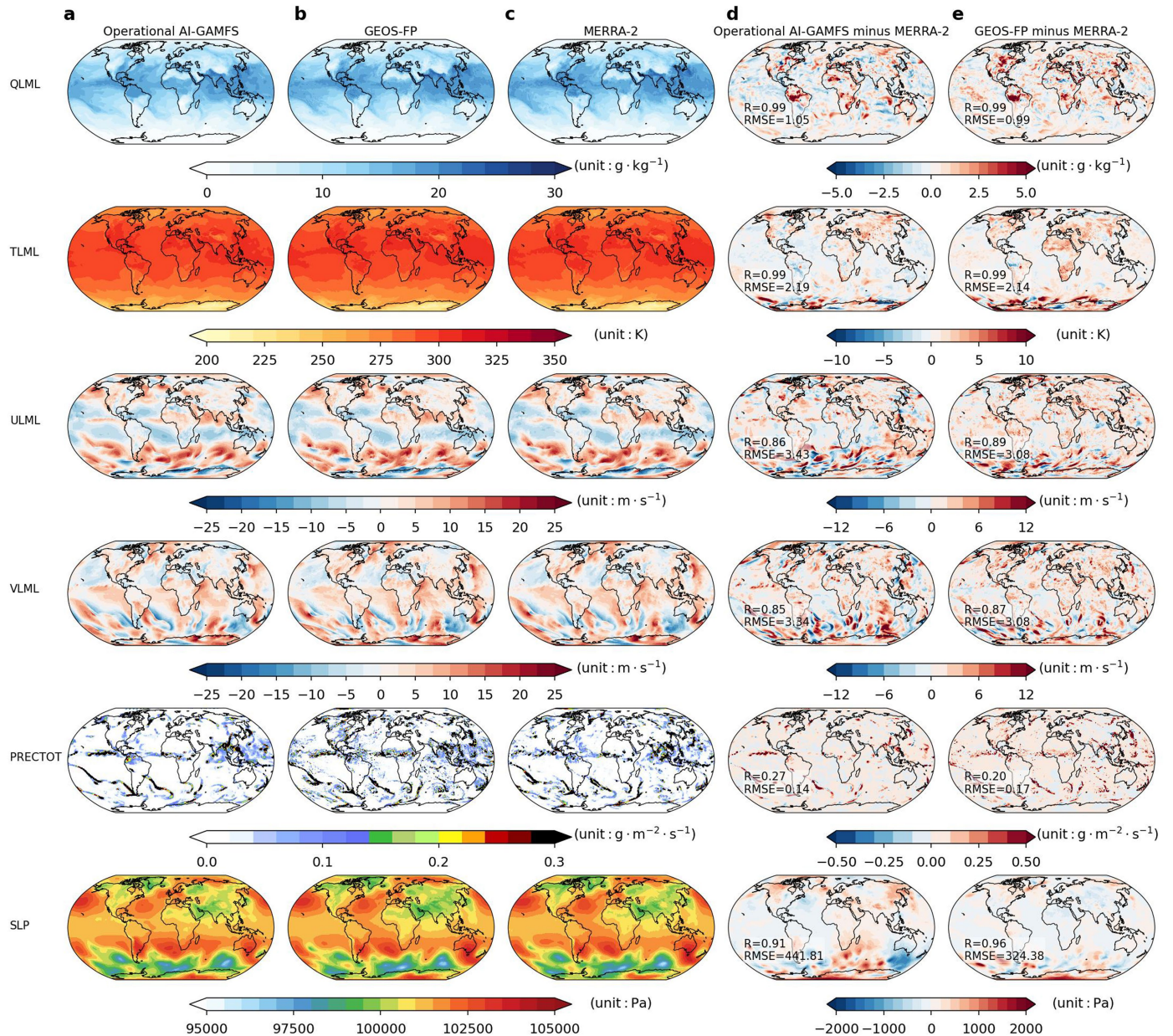
Extended Data Fig. 6 | Comparison of meteorological variable forecasting accuracy between operational AI-GAMFS and GEOS-FP. a,b, Comparison of spatial *R* (a) and latitude-weighted RMSE (b) between operational AI-GAMFS and GEOS-FP for six surface meteorological variables and nine selected

upper-level meteorological variables in 5-day global forecasts with 3-h temporal resolution. MERRA-2 data from July to August 2024 were used as the reference baseline.



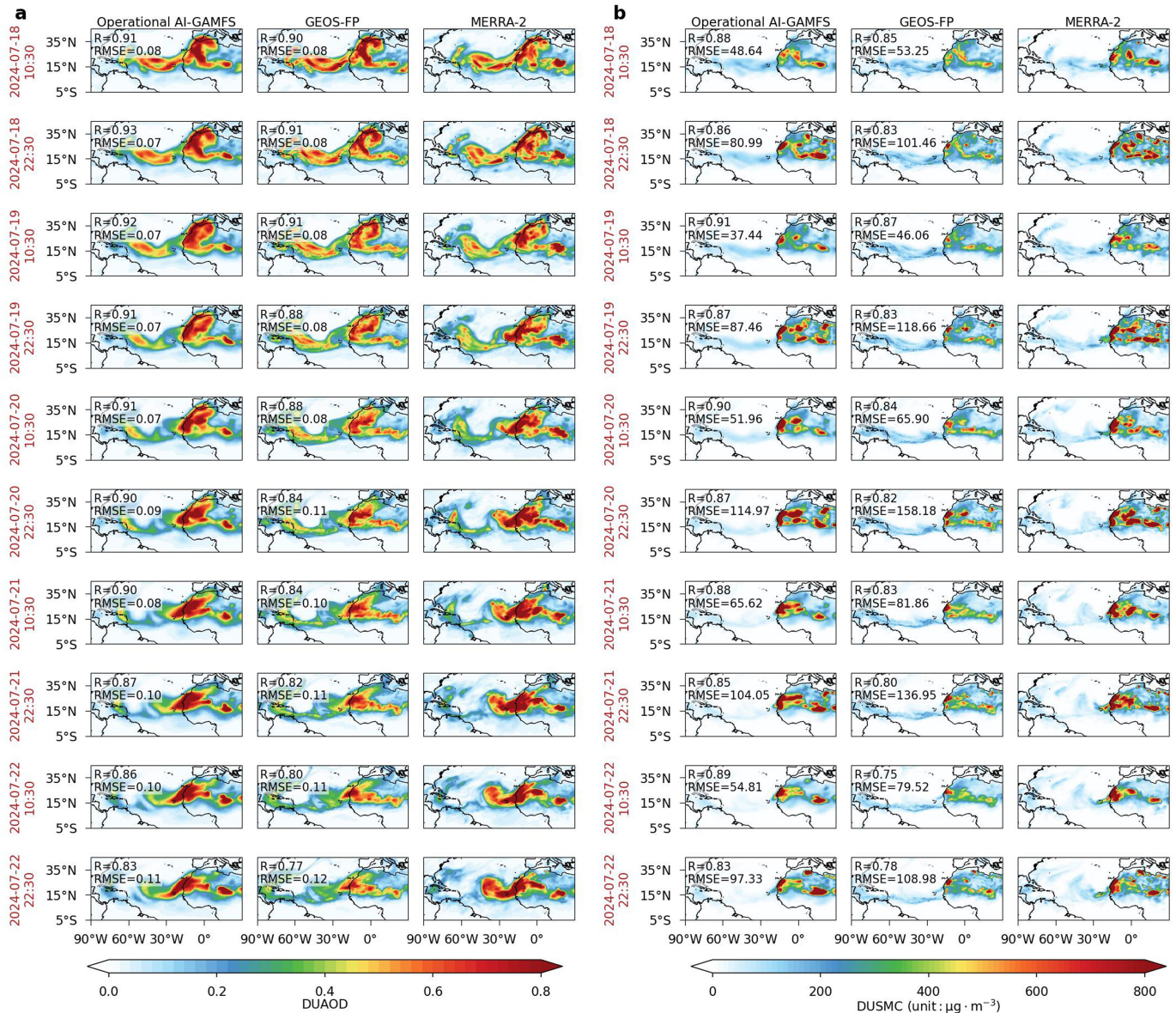
Extended Data Fig. 7 | Case study of global aerosol component surface concentrations forecasting. a, b, c. The 3-day lead-time global forecast for SUSMC, DUSMC, BCSMC, OCSMC, and SSSMC from operational AI-GAMFS (a), GEOS-FP (b), and MERRA-2 (c), initialized at 22:30 UTC on 26 July 2024.

d, e. Forecasting errors of operational AI-GAMFS (d) and GEOS-FP (e) relative to MERRA-2 reanalysis data. Overall accuracy metrics (i.e., spatial R and RMSE) for operational AI-GAMFS and GEOS-FP are shown in the lower-left corner of each panel in d and e.



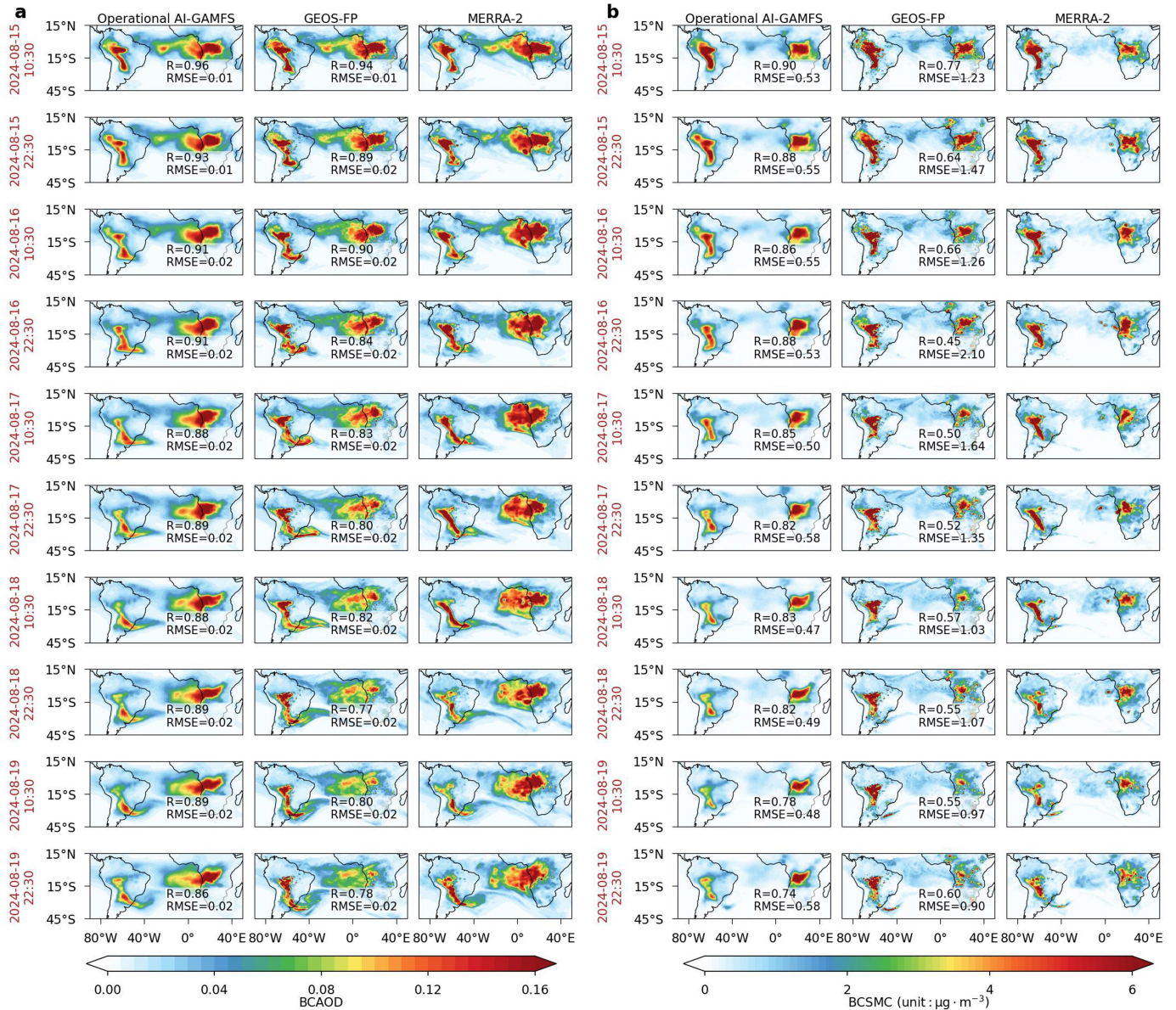
Extended Data Fig. 8 | Case study of global surface meteorological forecasts. **a, b, c**, The 3-day lead-time global forecast for QLML, TLML, ULML, VLML, PRECTOT, and SLP from operational AI-GAMFS (**a**), GEOS-FP (**b**), and MERRA-2 (**c**), initialized at 22:30 UTC on 26 July 2024. **d, e**, Forecasting errors of

operational AI-GAMFS (**d**) and GEOS-FP (**e**) relative to MERRA-2 reanalysis data. Overall accuracy metrics (i.e., spatial R and RMSE) for operational AI-GAMFS and GEOS-FP are shown in the lower-left corner of each panel in **d** and **e**.



Extended Data Fig. 9 | Case study of a trans-Atlantic African dust storm event from July 18 to 22, 2024. a, b, Forecasts of DUAOD (a) and DUSMC (b) during the event, produced by operational AI-GAMFS and GEOS-FP, with a 5-day lead time at 12-h intervals, compared with MERRA-2 reference data. Operational AI-GAMFS was initialized at 22:30 UTC on July 17, 2024, while GEOS-FP was

initialized at 00:00 UTC on July 18, 2024. For comparison, the outputs of GEOS-FP were interpolated to the same spatiotemporal resolution as AI-GAMFS. The overall accuracy metrics (i.e., spatial R and RMSE) for operational AI-GAMFS and GEOS-FP relative to MERRA-2 data are shown in the upper-left corner of each panel.



Extended Data Fig. 10 | Case study of simultaneous biomass burning events in Central Africa and Central South America from August 15 to 19, 2024. **a,b**, Forecasts of BCAOD (a) and BCSMC (b) during the event, produced by operational AI-GAMFS and GEOS-FP, with a 5-day lead time at 12-h intervals, compared with MERRA-2 reference data. Operational AI-GAMFS was initialized

at 22:30 UTC on August 14, 2024, while GEOS-FP was initialized at 00:00 UTC on August 15, 2024. For comparison, the outputs of GEOS-FP were interpolated to the same spatiotemporal resolution as AI-GAMFS. The overall accuracy metrics (i.e., spatial R and RMSE) for operational AI-GAMFS and GEOS-FP relative to MERRA-2 data are shown in the lower-right corner of each panel.