ARTICLE   OPEN

Check for updates

Improving air quality assessment using physics-inspired deep graph learning

Lianfa Li[1,2✉], Jinfeng Wang[1✉], Meredith Franklin[2,3], Qian Yin[1], Jiajie Wu[1], Gustau Camps-Valls[4], Zhiping Zhu[1], Chengyi Wang[5], Yong Ge[1] and Markus Reichstein[6]

Existing methods for fine-scale air quality assessment have significant gaps in their reliability. Purely data-driven methods lack any physically-based mechanisms to simulate the interactive process of air pollution, potentially leading to physically inconsistent or implausible results. Here, we report a hybrid multilevel graph neural network that encodes fluid physics to capture spatial and temporal dynamic characteristics of air pollutants. On a multi-air pollutant test in China, our method consistently improved extrapolation accuracy by an average of 11–22% compared to several baseline machine learning methods, and generated physically consistent spatiotemporal trends of air pollutants at fine spatial and temporal scales.

INTRODUCTION

Air pollution has major adverse effects on health and climate. Recent research[1] shows that in 2018 more than 8 million people died from acute and chronic respiratory diseases, lung cancer, stroke and heart disease related to fossil fuel pollution, indicating a much greater burden than suggested in previous studies[2]. Globally, air pollution plays a crucial but complex role in Earth's climate and ecosystems[3]. Similar to the demand for high-resolution climate model estimates, reliable and finely-resolved air pollution estimates are needed to accurately evaluate its effects on humans and ecosystems at regional and local scales, and to inform effective control strategies[4]. Air pollutant concentrations can vary sharply over short spatial distances due to unevenly distributed emissions sources, meteorological transport dynamics and complex secondary chemical processes[5], particularly in extreme air quality events[6].

Despite advances in mechanistic and statistical modeling, predicting fine-scale concentration gradients remains a challenge due to sparsely distributed measurement data, difficulty in characterizing complex atmospheric physiochemical processes, and modeling limitations[7]. For mechanism-based methods including dispersion, photochemical and plume models, issues such as missing and noisy input data (e.g., insufficient or outdated emission inventories), ill-posed boundary conditions, and complex accumulative parameterizations may lead to high uncertainties for many fine-scale local applications[5,8]. Statistics-based machine learning methods solely rely on the empirical statistical associations between air pollutant concentrations and explanatory factors to make predictions. However, the formation and transformation of air pollutants involves varying emission sources, atmospheric transport and physiochemical reactions within the planetary boundary layer, which are also affected by altitude, surrounding land characteristics and local geography[4,9]. These features cannot be well captured by classical statistical or machine learning techniques. Furthermore, while statistical methods have reported improved predictive performance, they may produce

physically inconsistent or implausible results[10] due to sampling bias, lack of physical constraints, and/or confounding effects arising from spatially stratified heterogeneous processes[11]. These issues may be amplified given that models are often tested by taking random samples across space and time for validation, which does not provide accurate generalizable performance statistics when applying the model to make predictions at unobserved spatial locations. To improve upon these limitations, machine learning has been combined with climate or atmospheric chemistry models to speed up computation or improve prediction accuracy, by replacing time-consuming mechanism components (e.g., chemical integrator)[12,13], memorizing the input-output relationship in mechanism models[14], correcting the bias[15] or representing subgrid process[16]. Furthermore, there has been a growing trend of integrating physical knowledge into deep learning techniques to model fluid fields, incorporating constraints such as partial differential equations (PDE)[17,18], yet apart from few limited applications[19,20] there continue to be gaps in effective strategies for embedding physics in deep learning to improve finely-resolved air quality assessments.

Here, we report a flexible deep graph hybrid modeling (DGM) approach (Fig. 1) incorporating fluid physics[21] in deep learning to capture spatiotemporal interactions of air pollutants, encoding their dynamic characteristics and transport at fine scales to improve spatiotemporal predictions. We are motivated by the applicability of deep graph learning to encode neighborhood information that represents the dynamics of air pollutants[9], and physical invariance to guide optimization[17]. For irregular non-Euclidean data with limited measurements, such as monitoring data of air pollutants, graph neural networks that are rooted on solid mathematical grounds of spectral graph theory[22] and locality[23], can be used to encode complex geometric structures and model their interactions. With graph operations such as Laplacian smoothing[24], graph convolution network (GCN) can message and aggregate neighborhood information to represent

¹State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China. ²Department of Preventive Medicine, University of Southern California, Los Angeles, CA, USA. ³Department of Statistical Sciences, University of Toronto, Toronto, Canada. ⁴Image Processing Laboratory (IPL), University of València, València, Spain. ⁵National Engineering Research Center for Geomatics, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China. ⁶Max Planck Institute for Biogeochemistry, Jena, Germany. ✉email: lspatial@gmail.com; wangjf@lreis.ac.cn
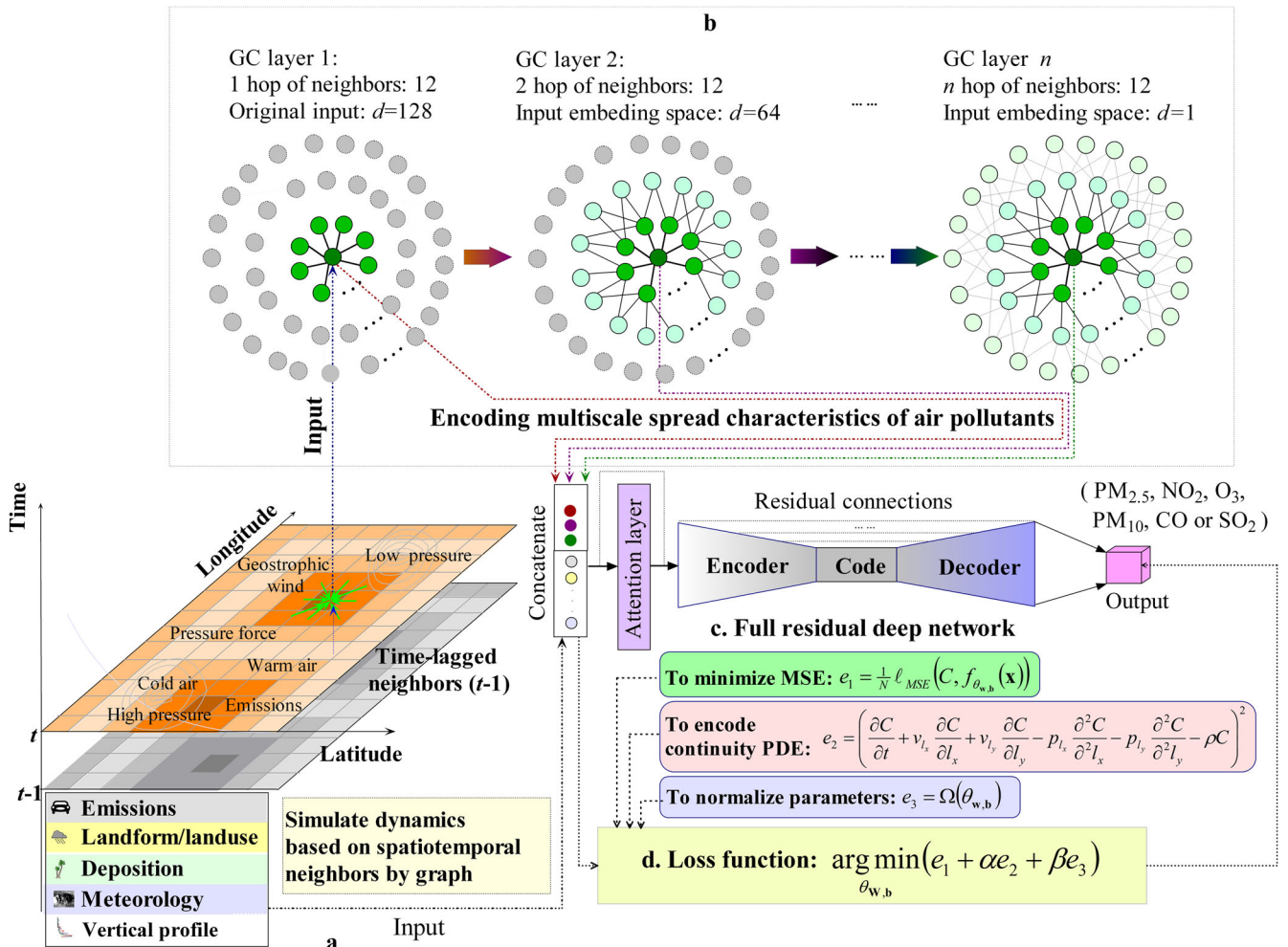
Fig. 1 Architecture of the deep graph network for air quality assessment. a The input data included atmospheric and surface grids, emissions or their proxies, meteorology, dry deposition, landuse and vertical profile. **b** The graph convolutions are constructed in multiple layers to simulate spatiotemporal dynamics of air pollutants, encoding their local spreading characteristics. **c** The GC multiscale outputs are concatenated with the input to enter full residual encoder-decoder to account for transformation and deposition. **d** The loss function included e_1 (mean square error (MSE) between observed and predicted values), e_2 (residual to encode PDE) and e_3 (normalization).

interactions between nodes, with similar applications in protein-protein[25] and disease-gene[26] interactions and drug discovery[27].

Our modeling architecture utilizes graph convolutions (GC) to simulate the dynamics of air pollutants. Additionally, by incorporating PDE residuals, the predictions are guided to conform to the principles of mass conservation and continuity in atmospheric fluid physics[28]: the former assumes that the total mass of the air pollutant of interest in the atmosphere is nearly constant in balance during the study period; the latter assumes that the air pollutant fluid of interest is continuous (it contains no gaps). Given the absence of vertical measurements of air pollutant concentrations in the atmosphere (only ground measurements were available), our approach focuses on simulating the horizontal ground advection and diffusion of air pollutants. This pertains to local or fine-scale dynamics along the north-south and east-west directions on the ground surface. We do not explicitly account for vertical convection in our methodology. Thus, our architecture is built upon a 2-D Eulerian system (Fig.1a), rather than a 3-D system[21]. In this system, the evolution of air pollutant concentrations at a target location of ground is modeled using partial differential equations for advection, diffusion, chemical transformation, emission and deposition. Our modeling architecture is meshfree, making it different from traditional numerical methods like finite element or finite difference methods. Unlike these methods, which rely on a predefined mesh or grid structure to discretize the problem domain, our approach offers greater flexibility and adaptability, particularly when dealing with complex geometries or irregularly distributed samples. By employing a meshfree approach, our architecture allows the neural network to directly approximate the dynamics or PDEs at irregularly distributed sampling points within the domain. This eliminates the constraints imposed by a predefined mesh, enabling more effective learning of the underlying physics[29].

Emissions

As an essential input, emission inventories are developed as a database listing the amount of air pollutants emitted over a specific time period (e.g., a year) by source. However, accurately quantifying emissions, especially in urban settings, poses significant challenges[30]; inventories tend to have a high degree of uncertainty and are often available at coarse spatiotemporal scales. In place of raw emission inventories, we used proxy variables including satellite-derived data (Moderate Resolution Imaging Spectroradiometer (MODIS) aerosol optical depth (AOD) for particulate matter, Ozone Monitoring Instrument (OMI) for nitrogen dioxide (NO_2), The Modern-Era Retrospective analysis for Research and Applications, version 2 - Global Modeling Initiative

(MERRA2-GMI), and traffic variables from OpenStreetMap). The satellite data have daily retrievals with global coverage. MERRA-2 GMI is driven by MERRA-2 meteorology to the Global Modeling Initiative's (GMI) stratosphere-troposphere chemical mechanism run at ~50 km horizontal resolution and output onto the native MERRA-2 grid. Compared with the emission inventories, the MERRA2-GMI provides background levels of air pollution at global to regional scales.

Driven by wind, air pollutants are transported across locations, cities, states, regions or even globally. Mechanical models like GMI and CMAQ are effective in simulating pollutant movement on a regional to global scale. However, they struggle to capture local and fine-scale variations in air pollutant gradients. This limitation may arise from inadequate fine-scale bottom-up emissions data, complex local terrain, the presence of transboundary pollutants under certain micro environments, high uncertainty in parameterizations, and challenging physical assumptions. Additionally, reanalysis methods like MERRA2-GMI operate at a coarse resolution due to limitations in data storage and the time required for long-term simulations. To overcome these challenges, we propose the integration of deep graph learning with prior, mechanism and/or empirical knowledge. This approach allows for the simulation of fine-scale air pollutant transport in a more time-efficient manner, while also improving generalization in prediction. By adopting physics-inspired deep graph learning, we can reduce the limitations of mechanical models and gain a more comprehensive understanding of the complex dynamics of air pollution at local and fine scales.

Graph convolutions for fine-scale dynamics of air pollutants

While our architecture is designed based on a 2-D Eulerian system due to the absence of vertical concentration measurements, it is important to note that atmospheric vertical processes, such as boundary layer venting, play a crucial role in controlling pollutant concentrations. This is particularly significant during winter when the planetary boundary layer height (PBLH) is very close to the ground[31]. Therefore, the input data in our horizontal transport model included PBLH and vertical meteorological variables (e.g., wind speed, specific humidity, and temperature), as well as other variables to account for the influence of vertical processes on ground-level pollutant concentrations, in a simplified way. According to the fundamental assumptions underlying atmospheric motions, the continuity equation[21] is used to simulate conservation of the air pollutant mass of interest on the ground. For a spatial location of interest, the continuity equation simulates the temporal evolution of the concentration of air pollutants, taking into account the pollutant sources, transport, sinks as well as meteorological, topological and geographical effects. Using Reynolds decomposition[32,33], advection and diffusion processes were summarized based on direction velocities, the advection partial derivatives, diffusion coefficients and the eddy diffusion terms (Eq. 2). Therefore, the air pollutant concentration changes over time (Supplementary Fig. 1) were decomposed into three balanced components, advection, diffusion, and the others.

For diffusion, the GCN provides an essential solution to simulate the second-order derivative on the discrete graph space[22,23], represented by the Laplacian convolution operator shown in Supplementary Fig. 2. Advection (Supplementary Fig. 1) can be decomposed along the directions of the connected nodes based on the differential formula.

A single fixed graph network is inappropriate for spatiotemporal estimation since such a network is inherently transductive and cannot naturally generalize to unseen data[34]. Here, we developed a hybrid semi-supervised deep graph model (Fig. 1), which uses local GC (Fig. 1b) to extract local spread characteristics of air pollutants from the multiscale nearest neighbors (cells in the regular grid system) of nodes, regardless of whether measured

concentrations were available in neighboring nodes. To simulate the spatial spread of air pollutants at a local scale in the graph space, we employed a Eulerian regular 2-D grid system with a spatial resolution of 1 km and daily temporal resolution. The retrieval of cell neighbors was based on this grid system where the proximity of cell samples to a target spatial location played a significant role, and cells closer to the target were given higher weights (inversely proportional to their distance) during the aggregation of graph convolutions. Additionally, we considered the lag influence of samples on the target spatiotemporal point. To enhance computational convenience, we simplified the construction of a local graph. We performed a search for the k nearest spatiotemporal cell samples to the target point and linked them to create the local topology graph, as depicted in Fig. 1a, b. This local graph aimed to simulate the spatiotemporal dynamics of air pollutants for the target location. By leveraging local graph networks, we obtained a flexible graph-level solution for spatiotemporal simulation. To model pollutant transport, we employed deep graph learning due to its flexibility in graph convolutions. This approach was also suitable for handling non-Euclidean and irregular spatiotemporal samples, which are common characteristics of air pollution measurement data.

Therefore, with neighboring grids, the dynamic spread of pollutants was able to be modeled using embedded learning techniques based on graph convolutions. The inputs to the model represented emissions or their proxies, meteorological driving factors, land-use data, spatially-varying covariates, and temporal indices at multiple scales (Supplementary Table 1 for the complete list). With neighboring grid cells, we constructed multiple interconnected local graph convolutions (as shown in Fig. 1b) to simulate the local spread of air pollutants (Fig. 1b). Mapped relationships between the continuity equation and the simulation of air pollution advection/diffusion on the graph space and feature space are presented in Supplementary Fig. 1. We defined the convolutional operators in multilevel GC layers to simulate the transport of air pollutants.

Coupling with residual learning for transformation and deposition

Air pollutants from a wide range of emission sources are transported by winds and deposited by gravitational sedimentation or precipitation. In the DGM architecture, we introduced MERRA2-GMI as inputs to represent background pollution levels and used GC to simulate local and fine scale transport. To account for vertical variation, chemical transformation and deposition, we coupled the GC with a full residual deep network by concatenating the GC output with the aforementioned inputs (Fig. 1c) in a hybrid systematic framework. In addition, the attention layers were added to weigh the influence of different parts of the input on the model's predictions[35]. In the hybrid systematic architecture, error information can be systematically back-propagated to the residual and GC layers during learning. Full residual deep learning has proved to be powerful for spatiotemporal regression of air pollutants[36] and imputation of missing satellite data[37]. The necessary covariates were also introduced: vertical meteorological factors (e.g., relative humidity, temperature, wind speed and PBLH) for vertical influence on ground pollutant concentration, OMI-NO$_2$ for O$_3$, NDVI for dry deposition, and precipitation for wet deposition.

In addition, to support the governing mass conservation, we encoded physical invariance in the total loss through the soft residual constraint of the continuity partial differential equation (PDE, Fig. 1d). Recent physics-informed neural networks (PINNs)[17,38] used the PDE residuals to improve physical relevant results. The deep learning process by gradient descent is also a simulation of the transport of the air pollutant of interest at a fine scale in the graph and feature space. Through this semi-supervised

learning, the goal of minimizing the target loss function was achieved, and the optimal solution was obtained, which ensured robust predictions while maintaining continuity of the PDE[17]. We classified our method as "hybrid" based on two aspects: model structure and learning loss function. In terms of model structure, we coupled two types of architectures, namely graph convolutional network (GCN) and full residual deep network (FRDN), to leverage their respective strengths and enhance overall learning efficiency. The GCN was employed to capture dynamic information from the neighborhood, while the FRDN was used to improve the efficiency of regression learning for local variables. Furthermore, we have incorporated a combination of traditional data loss terms and physics-based regularization loss terms into our loss function (Supplementary Note 1). The purpose of incorporating these physics-based regularization loss terms into the model is to make the predictions of the model as consistent as possible with the principles of physics. By doing so, we reduced inherent bias and significantly improved the model's generalization capabilities when making predictions in new scenarios.

RESULTS

Mainland China, a large region with heterogeneous topography, diverse meteorology and multiple emission sources, was selected as the testbed for the DGM method (Supplementary Note 2). We trained seven national hybrid models for six air pollutants including daily average carbon monoxide (CO), nitrogen dioxide (NO_2), fine particulate matter ($PM_{2.5}$), coarse particulate matter (PM_{10}) and sulfur dioxide (SO_2), as well as hourly maximum daily ozone (abbreviated as O_3M24) and 8 h average ozone (abbreviated as O_3A8). Daily gridded surfaces with a spatial resolution of 1×1 km^2 were also generated from 2015 to 2018. To assess the true generalization of the DGM method, three validations were

performed, including site-based independent testing, comparison with MERRA2-GMI pollutant variables and sensitivity analysis without MERRA2-GMI input of pollutants.

Trained air quality models are typically used to make predictions at new/unobserved geographic locations. Model validation based on test sets that are constructed by randomly sampling across space and time can result in inflated confidence in the model's out-of-sample prediction performance because it does not provide a genuine measure of the model's ability to generalize to new spatial locations. Instead, we used site-based independent testing, which uses complete time series data at untrained locations for testing. In addition, in order to have a thorough validation of the DGM, we reported the sensitivity analysis results without the MERRA2-GMI pollutant input.

Our site-based independent tests included 288 sites (about 18%) that were randomly selected from 1604 monitoring sites; all the data from these sites were used only for testing but not for training. From the remaining samples, we randomly selected 78% for training and 22% for regular testing across space and time (stratified by spatiotemporal factors, i.e., province and month). We chose the sample so that the proportion of the total samples for regular and site-based tests was similar to the proportion (~0.37) of the unique test samples in the bootstrap 0.632 rule[39] and 56 variables were used in training. In order to avoid the influence of random noise, we trained 100 models and summarized their average performance for each air pollutant, and then compared them with seven baseline machine learning or statistical regression models (for these methods and optimization of their hyperparameters, see Supplementary Table 2 and Materials and Methods). Sensitivity analysis was conducted to retrieve optimal weights (0.5 for α and 0.2 for β) for the regularizers. The results (Supplementary Tables 3, 4; Supplementary Fig. 3 for the boxplots of testing root mean square error (RMSE)) show that our hybrid
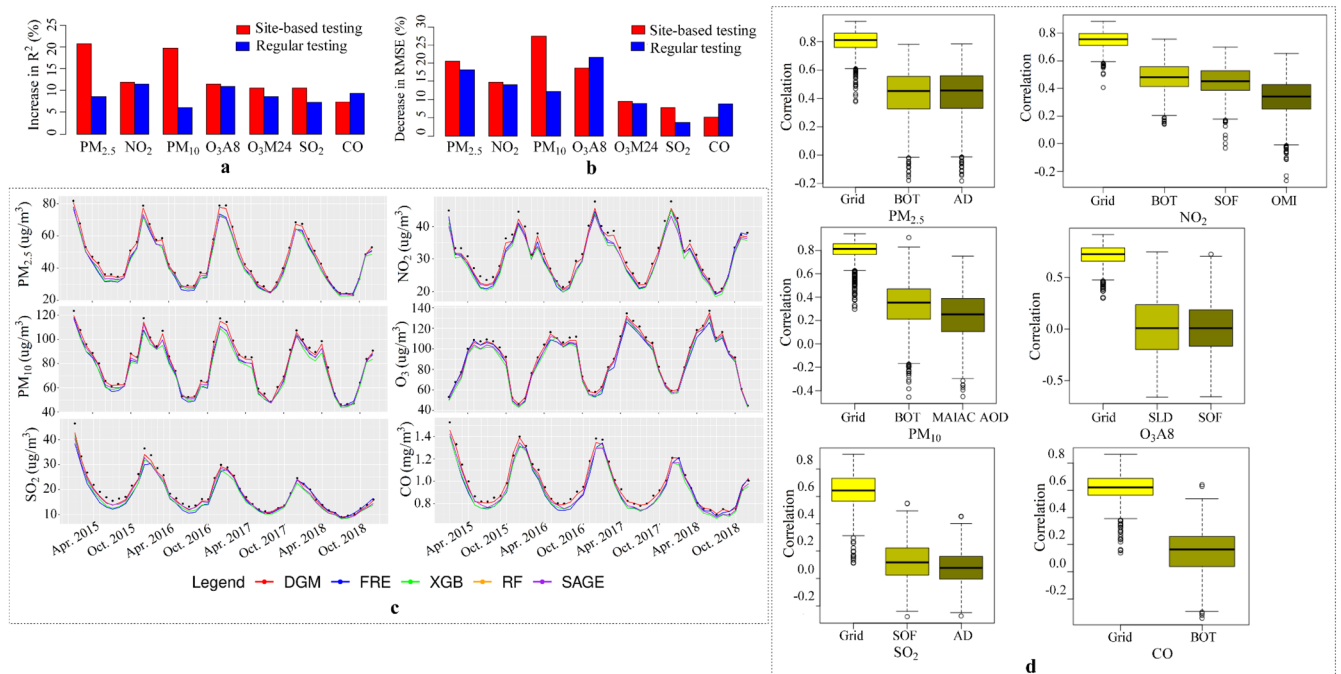
Fig. 2 Improvement in generalization in comparison with baseline machine learning methods and MERRA2-GMI data. a, b The increase in R^2 (**a**) and the decrease in RMSE (**b**) by the physics-aware deep learning method for site-based testing and regular testing. **c** Mean time series of observed and predicted values for all sites in the site-based testing dataset (black dots: observed values). **d** The boxplots between daily grid mean estimates and daily MERRA2-GMI variables (Grid: 1×1 km^2 grid mean estimates; BOT Bottom layer diagnostics, AD Aerosol diagnostics, SOF Daily satellite overpass fields, SLD Single-level diagnostics). In (**d**), the whiskers (solid horizontal lines) indicate the minimum and maximum correlations, excluding outliers; the box bounds the interquartile range from the 25th to 75th percentiles, with the 50^s percentile (median) marked (bold horizontal lines); the whiskers (dashed vertical lines) extend 1.5 times the interquartile range from the box, defining the range of typical correlations; data points (small circles) beyond the whiskers are outliers.
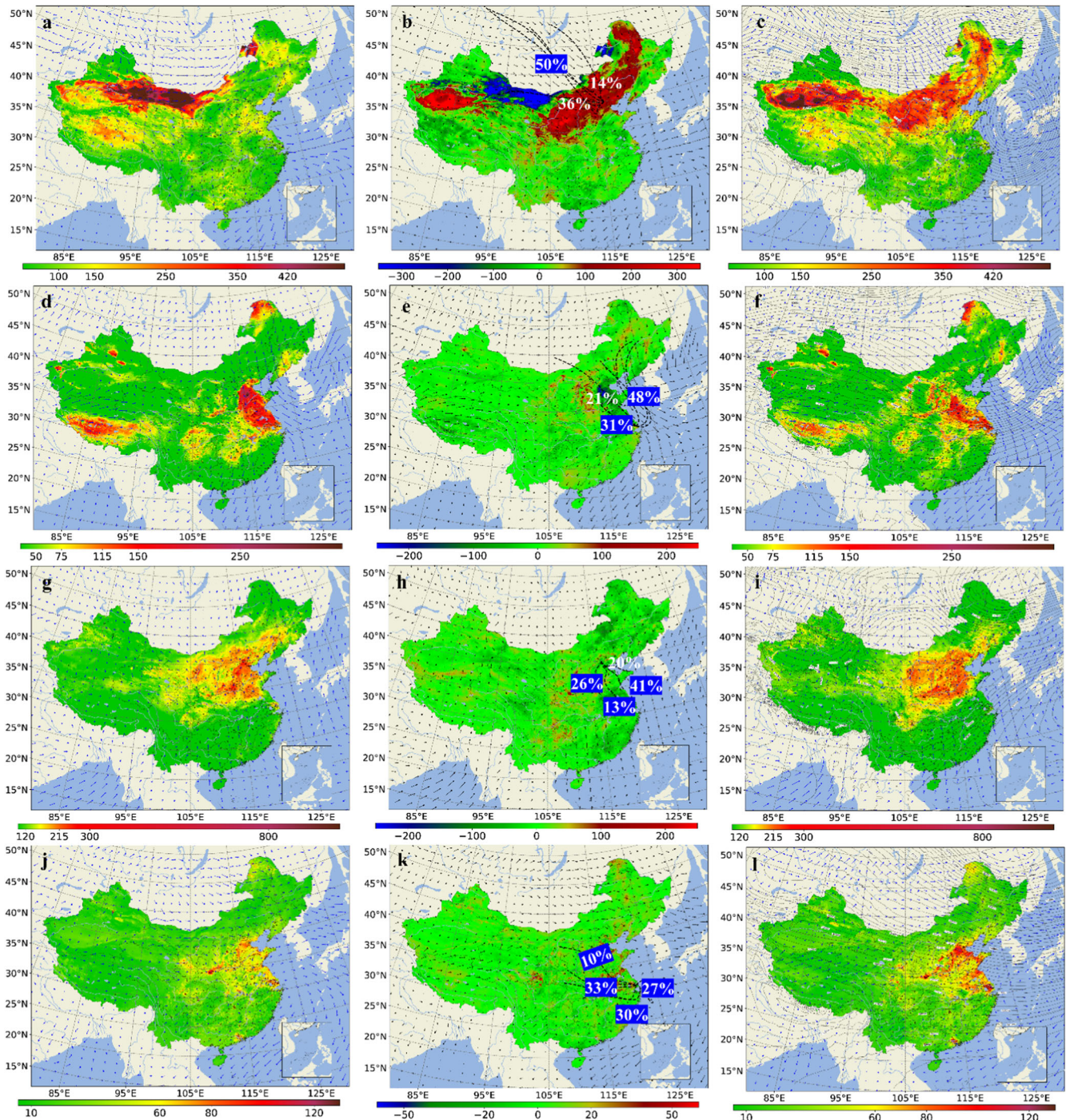
Fig. 3 Four representative events of air pollution presented by the predicted surface grids. a–c The predicted PM_{10} surfaces on a sandstorm event in North China (**c**) of April 15, 2015 and the day (**a**) before it, and the difference (**b**) between the two. **d–f** The predicted $PM_{2.5}$ surfaces on a haze event in East China (**f**) of December 23, 2016 and the day (**d**) before it, and the difference (**e**) between the two. **g–i** The predicted O_3A8 surfaces on a severe ozone event in Beijing (**i**) of July 1, 2017 and the day (**g**) before it, and the difference (**h**) between the two. **j–l** The predicted NO_2 surfaces on a haze event in Shanghai (**l**) of December 19, 2018 and the day (**j**) before it, and the difference (**k**) between the two. The blue arrow is mean ground wind vector (m/s) and the black isolines present GPH in meter at 500 hPa. The dash lines in (**b, e, h, k**) show the HYSPLIT clusters of backward air trajectories at 500 m with the percentage.

method consistently performed better than all seven other machine learning and statistical regression models. For $PM_{2.5}$ and PM_{10}, our method had the highest site-based test mean R2 (0.85–0.87), an improvement of 13–85% over the other methods. Compared to $PM_{2.5}$, we found a lower site-based test mean R2 for NO_2 and O_3A8 (0.66–0.78), but overall our hybrid method showed an improvement of 6–61% over the other methods. For SO_2,

random forest had a slightly better testing R2 (0.74 vs. 0.73) than our method but had lower site-based test R2 (0.58 vs. 0.63) and higher RMSE (13.11 vs. 12.37 µg/m3). Although the training R2 and RMSE of the full residual network[36] were similar to our hybrid DGM (0.84–0.93 vs. 0.83–0.93), the test R2 and RMSE, especially independent test R2 and RMSE, showed poorer performance. Compared with the regular network, the hybrid method improved

test R2 by 6–11%, and site-based independent test R2 by 7–21% (Fig. 2a), and it decreased test RMSE of the monthly mean by 4–18% and site-based test R2 by 5–21% (Fig. 2b). Our sensitivity analysis showed that GCN made a considerable contribution to the improvement in prediction (by 6–19% for testing and sited-based testing R2) and the PDE residual made additional 1–4% improvement in test and sited-based test R2. Furthermore, we evaluated the extent of mass conservation by estimating the e_2 component of the total RMSE, based on the ratio of e_2 to the overall loss. As shown in Supplementary Table 5, the converged small values of the e_2 RMSE (ranging from 1.62E-5 to 0.134, approaching 0) indicate a high level of mass conservation with minimal bias. The models with GCN simulation and constrained by the PDE residual performed the best, with a steady learning process, and better generalization and extrapolation. In site-based independent tests, the samples from selected sites also exhibit fairly consistent distributions between observed and predicted values (Supplementary Fig. 4). In the site-based tests, on average, our DGM method consistently improved explained variance (R2) by 11–22%, and reduced RMSE by 12–35% compared to baseline statistical learning methods (Fig. 2a; Supplementary Table 6).

We also conducted a comparison between our method and the baseline machine learning methods by analyzing observed and predicted time series. For each independent test site, we calculated the root mean square error (RMSE) for each machine learning method. The results indicate that our method effectively captures temporal trends, as evidenced by the violin plots of the site-based RMSE in Supplementary Fig. 5. Additionally, when comparing the total monthly means of predicted and observed values (Fig. 2c) and the total RMSE (Supplementary Table 7) for all these sites, our method considerably outperforms other machine learning methods. It produced more stable time series with significantly smaller deviations (much less RMSE) from the observed values.

Correlations of our DGM point and $1 \times 1 \, km^2$ grid estimates with ground measurements were considerably stronger than those of MERRA2 GMI reanalysis data (Supplementary Table 8 and Fig. 2d: 0.76–0.92/0.66–0.83 vs. 0.005–0.53). For example, the MERRA2-GMI $PM_{2.5}$ and NO_2 had a correlation of about 0.41–0.53 and the other MERRA2-GMI pollutant variables had a poor correlation (0.005–0.36). This comparison suggests that the coarse-scale reanalysis MERRA2-GMI data could be quite biased against ground measurements. With finer spatial resolution, our estimates improved correlation coefficients significantly. We also reported the seasonal and yearly performance metrics for the site-based independent test (Supplementary Tables 9, 10). In comparison to a recent study that utilized the Weather Research and Forecasting model with Chemistry (WRF-Chem) at high spatial resolution for the same region and year (2015)[40], our results demonstrate significantly stronger Pearson's correlation and lower mean bias (as shown in Supplementary Table 10). Specifically, our method exhibited correlations ranging from 0.77 to 0.92 for $PM_{2.5}$, PM_{10}, O_3, and NO_2, whereas the previous study showed correlations of only 0.29 to 0.51 for the same pollutants. Additionally, when considering the correlation computed from the mean temporal evolution averaged across all monitoring stations, our method achieved an impressive correlation of 0.99, surpassing the range of 0.71 to 0.90 reported in the previous study. It is worth noting that our approach demonstrated comparable or superior performance to several other investigations[41–43] employing WRF-Chem for O_3, NO_2, and/or PM_{10} in different regions, although making direct comparisons may not be entirely appropriate due to potential variations in methodology and data sources.

On the other hand, using MERRA2-GMI or other reanalysis pollutant data in our DGM provided the information on regional and global scale emissions and transport of air pollutants. For example, through the SHAP (SHapley Additive exPlanations), we summarize the contributions of the emission-related components

to the $PM_{2.5}$ predictions in mainland China and its four representative regions (Supplementary Note 3; Supplementary Fig. 6). Nevertheless, sensitivity analyses (Supplementary Tables 3, 4) showed a small (0–4%) contribution to the independent test R2 by the MERRA2-GMI pollutant data, illustrating the robust performance of our approach for fine-scale air quality assessment even without the MERRA2-GMI pollutant input.

Interpretation of improved air quality assessment

Compared to purely data-driven machine learning methods, our approach offers a simulation-based solution for capturing the advection and diffusion of air pollutants. We achieve this by constructing local graphs to simulate spatiotemporal dynamics and incorporating the continuity equation into the training cost function. As a result, our method effectively captured the fine-scale transport of pollutants and significantly reduced estimation bias. For PM_{10}, we generated nationwide concentration grids with similar or finer spatial resolution (1 km vs. 1–10 km) and higher generalization accuracy (site-based R^2: 0.85 vs. CV or out-of-station R^2: 0.82–0.83) compared to[44–47]. For the recent studies of $PM_{2.5}$ in China[47–51], we had higher generalization (site-based R^2: 0.87 vs. CV or out-of-station R^2: 0.64–0.85) at the finer spatial scale (1 km vs. 1–50 km). For O_3A8 and NO_2, our method had similar generalization (site-based R^2: 0.72–0.78 vs. out-of-station or out-of-city R^2 0.71–0.80) at 1 km spatial resolution compared to recent studies[52,53]. Although direct comparison is inappropriate, compared to the existing studies, our physics-inspired approach demonstrates improved accuracy in capturing the temporal evolution of fine-scale air pollutants. By incorporating fluid physics principles and encouraging mass conservation, our model yields predictions that more effectively reflect the dynamics of these pollutants. Additionally, our model produces daily spatial surfaces of air pollutants that exhibit continuous variations without abrupt changes or discontinuities, in contrast to the outputs of tree-based machine learning algorithms. This demonstrates our model's ability to accurately simulate air continuity. Based on the $1 \times 1 \, km^2$ grid estimates of air pollutants, we investigate the evolutions of their spatial variability and interpret them in terms of regional/ global scale transport and fine-scale dynamics of air pollutants and meteorological characteristics. For a daily grid of $1 \times 1 \, km^2$, we extracted a heatmap of neighborhood features from the output of the final graph convolution layer, representing the spatial agglomeration, and visualized them using the contours.

For inert air pollutants including $PM_{2.5}$ and PM_{10}, multiscale pollutant transport has important influence upon their spatiotemporal variability. On April 15, 2015, Beijing experienced the biggest sandstorm with the highest PM_{10} levels seen since 2002[54] (Fig. 3a–c). Meteorologically, the Rossby waves and intense frontal activities in the spring resulted in strong wind gusts that injected large amounts of material into the lower and middle troposphere. Dust uplifted from the arid and semi-arid regions was transported to the downwind regions of North China[55]. Backward trajectories (http://ready.arl.noaa.gov/HYSPLIT.php) of the winds showed two main sources of dust, i.e. Northwest China and Southern Mongolia for the sandstorm incidence in the Beijing city (Fig. 3b). The spatiotemporal evolution of PM_{10} (Supplementary Fig. 7) shows that the dust transport pattern was well captured by the extracted graphical convolution heatmap, and the estimated PM_{10} exactly matched the trends in the surface wind and geopotential height (GPH) fields. The spatiotemporal evolution of $PM_{2.5}$ (Supplementary Fig. 8) during this sandstorm was different from that of PM_{10}, because the latter was mainly from the long-distance transport of dust and aerosols, while the former was more from local transport and sources, including power plants, industries, and automobiles, etc. At the local or fine spatial scale, our estimated $PM_{2.5}$ and PM_{10}, respectively, matched well with the ground measurements on the sandstorm day, as shown in the enlarged local maps and indicated
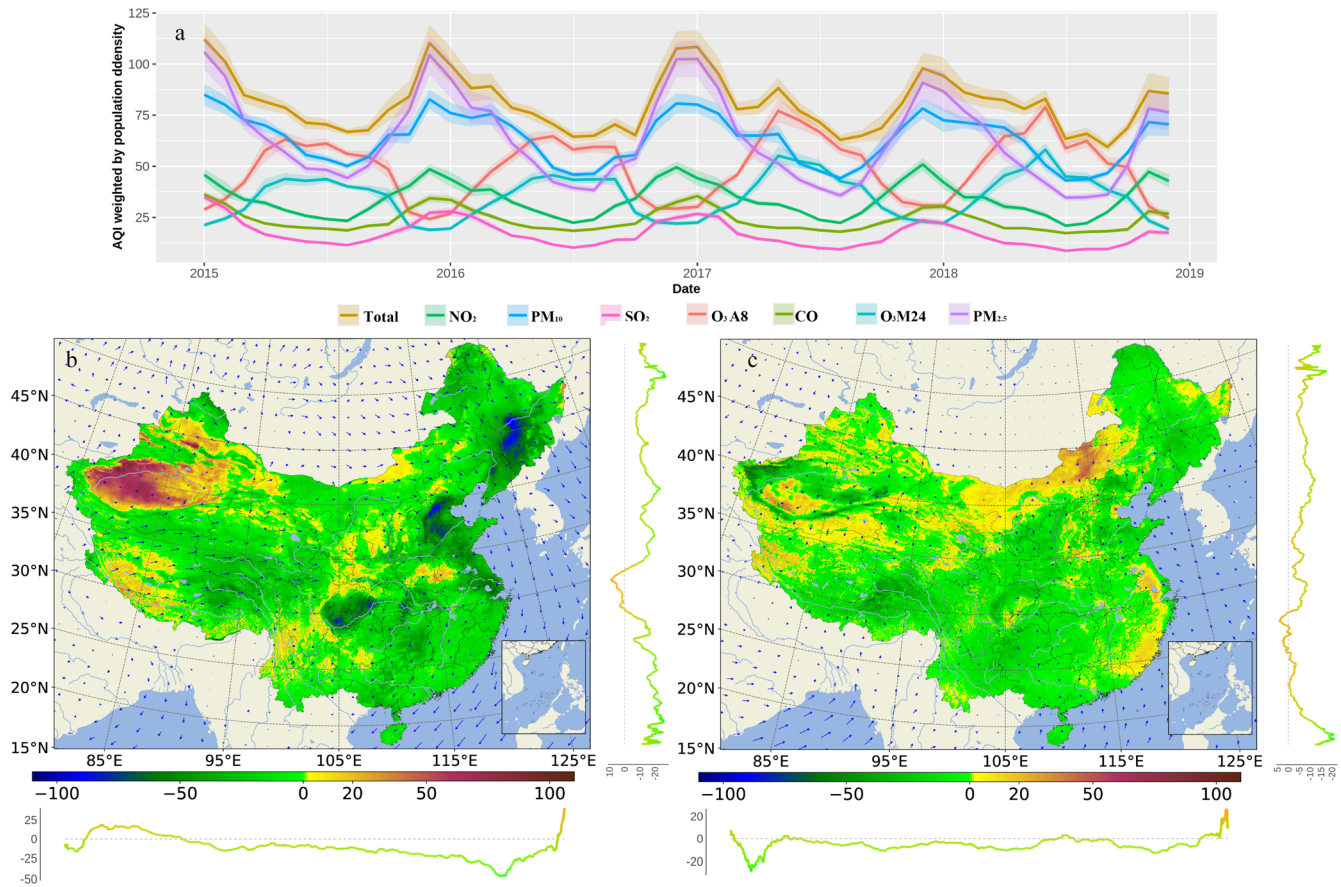
Fig. 4 National population-weighted AQI changes. a Monthly trend for total AQI and each air pollutant AQI with the shade of 95% confidence intervals. **b, c** Total air quality index (AQI) changes in winter between 2015 and 2018 (**b**), and in summer between 2015 and 2018 (**c**) in mainland China (mean AQI change by latitude, below in **b, c**; mean AOI change by longitude on the right in **b, c**).

by the high correlations in the independent point estimates and the grid estimates (0.80–0.87), that were much higher than the reanalysis estimates. In addition, compared to the measured data, our estimated concentrations had similar normal distributions, and the 2D histogram of $PM_{2.5}$ vs. PM_{10} of the measurement data was consistent with that of the finely resolved estimated data (Supplementary Fig. 9).

Similar to PM_{10}, interregional transport may play an important role in $PM_{2.5}$ pollution, although the latter is partially from local emission sources. For the haze incident in East China in late December 2016 (Fig. 3d–f), the updraft in front of the 500 hPa high-altitude trough was conducive to the regional transport of pollutants, and in the northwest wind field before high air pressure, the majority of $PM_{2.5}$ in the haze incidence was composed of upstream pollutants being transported to the downwind areas of East China such as Jiangsu Province and Shanghai. The backward trajectory analysis showed two main sources of haze in Shandong Province and the Beijing-Tianjin-Hebei region (Fig. 3e). In the spatiotemporal evolution of $PM_{2.5}$ (Supplementary Fig. 10), the graph convolution features were shown to capture regional and local meteorological transport of $PM_{2.5}$ well.

As a reactive pollutant, O_3 is mainly formed by photochemical reactions of volatile organic compounds (VOCs), CO and nitrogen oxides under sunlight, especially the UV spectrum. An intensive and persistent regional ozone pollution episode occurred in eastern China from 25 June to 5 July 2017, mostly located in the Beijing-Tianjin-Hebei and the surrounding area (BTHS) (Fig. 3g–i). A study[56] showed that the prevailing northwesterly winds in the mid-high troposphere hindered the northward movement of the

Western Pacific Subtropical High, thereby inhibiting the transport of water vapor from low latitudes and providing favorable meteorological conditions (low relative humidity, less cloud cover, strong solar radiation and temperature inversion) for local formation of O_3; in addition, warm advection due to southerly winds prevailing in the lower troposphere also brought pollutants to BTHS, as shown in the spatiotemporal evolution of ozone during this period (Supplementary Fig. 11). Backward trajectory analysis showed considerable contributions from southern regions. Compared with other reactive pollutants such as CO and NO_2, O_3 had much higher spatial autocorrelation[56] as seen in the grid estimates, which showed a more gradual spatial gradient pattern at fine spatial scales and the estimates were in high agreement with the ground measurements (correlation: 0.81–0.92).

As a highly reactive gas and precursor of surface and tropospheric O_3, NO_2 has a relatively short atmospheric lifetime compared to PM. In winter, when solar radiation, relative humidity and precipitation are low, the cool and dry conditions of the troposphere slow down photochemical reactions, allowing pollutants to accumulate, and the lifetime of NO_2 is longer than in spring and autumn. For the 2018 winter haze in Shanghai (Fig. 3j–l), in addition to local emission sources (power plants, biomass burning, and fossil fuel combustion etc.), the analysis also revealed partial contributions from regional transport of pollution in nearby cities in North, Central, and East China. The dominant NO_2 source regions usually correspond to areas with high population, heavy traffic and large industry scale. Given high traffic leading to high NO_2, local hot spots in urban areas show up around heavy traffic routes, traffic intersections and transportation

hubs[57,58]. The spatiotemporal evolution (Supplementary Fig. 12) of NO_2 during this episode presented such hot spot patterns on fine scales, with high correlation (0.81–0.88) with measurements.

Further investigation of other days and pollutants (SO_2 and CO) also revealed the critical role of graph convolutions and PDEs in simulating advection and dispersion. For inert background air pollutants ($PM_{2.5}$ and PM_{10}), their concentration profiles are dominated by advection and diffusion, with regional and local transport contributing much more to their spatiotemporal variability than other pollutants. The results showed that graph convolutions for simulating local pollutant transport achieved the largest improvement (14–15% in R2) for $PM_{2.5}$ and PM_{10}, which was not seen using pure deep learning. Furthermore, MERRA2-GMI inputs had little effect on the $PM_{2.5}$ and PM_{10} estimates as they involved fewer chemical transformations compared with reactive pollutants. On the other hand, for reactive air pollutants involving complex physiochemical reactions that are subject to meteorological conditions, although the contribution of global and regional pollutant transport may be limited, by simulating fine-scale pollutant transport, graph convolutions still achieved significant improvements (R^2: 7–13% for NO_2 and O_3; 4–6% for CO and SO_2). The MERRA2-GMI input provided the relevant background-level data on precursor chemicals and vertical meteorology for reactive air pollutants, that helped improve their estimates, albeit limited.

DISCUSSION

By simulating spatiotemporal dynamics of air mass, our deep graph learning method achieved better generalization with physical interpretability than purely mechanistic or machine learning methods. The fine-scale (1×1 km^2) gridded estimates accurately represented the spatiotemporal distribution of all air pollutants and enabled more reliable and finer statistics of national air pollutant trends than previously reported. A literature review shows that to date few studies have reported high-resolution predicted surfaces for all criteria pollutants and total air quality index (AQI, for the quantification definition, refer to Supplementary Note 4). This study fills this gap and provides a fine-scale grid of AQI for China, providing important information over a large geographic area that does not have monitoring stations. Nationwide, the population-weighted means (Supplementary Fig. 13) were much higher than the raw means for NO_2 (by about 9 $\mu g/m^3$) and $PM_{2.5}$ (by about 13.5 $\mu g/m^3$), indicating that they were mainly distributed in densely populated areas. The yearly summary (Supplementary Fig. 14 and Supplementary Table 11) shows the there was a national decline from 2015 to 2018 for the population-weighted $PM_{2.5}$ (by 11 $\mu g/m^3$), PM_{10} (by 11 $\mu g/m^3$), SO_2 (by 6 $\mu g/m^3$) and CO (by 0.12 mg/m^3) but a national rise in O_3 (7 $\mu g/m^3$ for the 8 h daytime average). The national trends of air pollutants were generally consistent with ground measurements and reanalysis data, but since most monitoring sites were located in urban areas, the limited measurement data mostly overestimated the national averages by about 6–52%, and the reanalysis data also substantially underestimate the national averages due to coarse spatial scales. In addition, the daily national air quality index (AQI) results (Supplementary Figs. 15, 16) provided a fine representation of the spatiotemporal distribution of each air pollutant, and the aggregated trend was consistent with measurement-based AQI[59] (Pearson's correlation: 0.76–0.92). The highest concentrations in summer were PM_{10} for the Xinjiang desert area and O_3 for most of the other areas; while in winter it was $PM_{2.5}$ for most of the eastern areas of China and PM_{10} for the other areas (Supplementary Figs. 17, 18).

Compared with the national point-based AQI (https://www.iqair.cn/cn/china), the statistics of the predicted grids more accurately show the spatiotemporal distribution patterns of air pollutants. From 2015 to 2018, the population-weighted daily AQI (defined in[60]) declined by about 6 (by about 3 for the naïve AQI mean; Fig. 4a for monthly mean; Supplementary Table 11 and Supplementary Fig. 19). Compared with other air pollutants, $PM_{2.5}$ and SO_2 decreased by about 18–30%, and PM_{10} and CO decreased by about 12–13%. The mean value of NO_2 dropped slightly by about 4% (by about 3% for its population-weighted mean). The national daily population-weighted 8 h average O_3 rose by about 8%. The results show that there were higher $PM_{2.5}$ concentrations partially in the eastern areas of China in winter, and the northwestern region of Xinjiang had higher PM_{10} concentrations than the other regions. For mainland China, the air quality in summer is much better than in winter (AQI: 35–41 vs. 47–58). The grids (Fig. 4b) of AQI change show that from January 2015 to December 2018, the monthly AQI in winter declined by about 11 in AQI and the $PM_{2.5}$ and SO_2 had the maximum decline (by about 13–15 in AQI); from July 2015 to July 2018, the monthly AQI in summer declined by about 5 and $PM_{2.5}$ and SO_2 had the maximum decline (by about 10) (Fig. 4c). In China, the primary pollutants in winter included $PM_{2.5}$ and PM_{10}, while the primary pollutant in summer was O_3. Spatially, concentrations of $PM_{2.5}$, PM_{10} and CO were higher in the Beijing-Tianjin-Tangshan and Xinjiang areas than in the Yangtze River Delta and the Pearl River Delta. From 2015 to 2018, the Beijing-Tianjin-Tangshan area, Jilin Province and Sichuan Basin experienced the largest declines in AQI.

Our results show that from 2015 to 2018, air pollution in mainland China dropped significantly, which is consistent with the clean air action in China[61]. However, as one criteria air pollutant, O_3 has presented a national upward trend (by about 8%) from 2015 to 2018, which is consistent with reports[62]. Climate change can result in the increase of tropospheric O_3 since increased air temperature results in O_3 formation. Data on daily air temperature statistics (Supplementary Fig. 20) support this finding. As a greenhouse gas, O_3 can raise the global temperature and has adverse effects on health. For China, while maintaining lower emissions policies for other criteria air pollutants, it is time to take a cooperative strategy to efficiently control emissions of volatile organic compounds and other pollutants that contribute to the formation of O_3.

Our extensive evaluation effectively reveals the power of a physics-inspired deep graph learning approach for finely resolved air quality assessments. First, multilevel graph convolutions can be constructed to simulate the spatiotemporal dynamics of air pollutants to capture pollutant transport on a fine scale. Graph convolution is a tool to model local interactions and movements by its Laplace differential operator as the gradient divergence[22]. Graph-theoretical methods have brought about a paradigm shift in the exploration of complex systems[63], including atmospheric chemical mechanisms. In their study, Silva et al.[64] utilized species-reaction graphs to uncover similarities among different chemical reaction systems, unveiling novel characteristics that transcend existing domain knowledge and thereby advance our understanding of atmospheric chemical mechanisms within intricate geosystems. Diverging from the species-reaction graph approach, this research introduced a pioneering methodology by utilizing local graphs as a roadmap to precisely simulate the transport of air pollutants originating from immediate and geographically intricate surroundings at a fine scale. Through efficient embedding learning, our approach significantly enhanced the prediction of product outcomes resulting from the advection and diffusion of air masses, as well as atmospheric chemical reactions. These intricate processes are profoundly influenced by a multitude of factors, including the surrounding atmosphere, surface characteristics, and geographical environment. Here, local graph convolutions are used to overcome the limitation of convolution and recurrent neural networks to handle the ubiquitous irregularity and sparsity of air quality monitoring data. They facilitate the utilization of substantial quantities of unlabeled or unmeasured

data in a semi-supervised approach. Whether inert or reactive pollutants, pollutant transport plays an important role in finely resolved air quality assessment, although we found that it may contribute more to inert regional air pollutants such as $PM_{2.5}$ and PM_{10}. Second, in order to make the final predictions tractable, we encoded the constraint of the continuity partial differential equation as a loss residual so as to strengthen physical invariance. Different from the simulation in the graph space and hard constraint of mass conservation in ref. [20], the PDE residual constraint acted as a soft constraint that encouraged the final predictions to conform to the mass conservation for the ground air pollutants along the timeline. Both ways of encoding physical concepts resulted in parameters that were optimized to better represent physical processes than the model without any physical constraints. Third, using a smaller spatial resolution, we were able to simulate spatiotemporal dynamics on a finer spatial scale, which may be useful for reactive air pollutants with rapid distance decay gradients from the emission sources, such as NO_2, SO_2 and CO. Using more graph convolutional layers, we could extend the simulation to a longer range of pollutant transport, which favors regional pollutants with slower decay gradients such as $PM_{2.5}$, PM_{10}, and O_3. To account for the transformation and deposition of air pollutants, the full residual and attention layers were systematically concatenated with the graph convolution layers to reduce potential over-smoothing in GCs and strengthen the contribution of local and non-traffic factors to the final prediction[65].

Essentially, the physics-inspired hybrid deep graph network is built on a flexible and unified architecture of air pollutant emissions, pollutant transport, and deposition, and is designed for spatiotemporal prediction. Machine learning has played a crucial role in improving air quality assessments over the past few decades. Its effectiveness and computational efficiency have made it a popular choice for enhancing accuracy and scalability in this field[66,67]. In previous efforts, Keller[12] and Kelp[13] used machine learning to replace time-consuming components in the physical models to accelerate emulating of the chemistry components; Sturm[14] designed the mass- and energy-conserving framework for using machine learning; Ivatt[15] used gradient-boosted regression trees to reduce the bias in physical model outputs. However, the lack of physical mechanisms in most machine learning methods may lead to solutions that seriously violate physical principles such as conservation laws, with potential bias and low interpretability[68]. For example, tree-based machine learners (e.g., random forest, XGBoost and extremely randomized trees) have been widely used alone[46,50,69] or in physical models[12,15] with reported high accuracy, but the binary discretization in these algorithms inherently violate the continuity law of air pollutants, which can lead to discontinuity bias[36]. In contrast to these existing methods[70,71], based on deep neural networks as a general nonlinear approximator of any continuous functions[72], the DGM directly simulated dynamics of air mass by graph convolution and encoded the fluid physical laws of mass conservation and continuity by PDE residuals, which is a creative combinational strategy for computing efficiency and generalization. Our approach outperformed simulations conducted using mechanism models like MERRA2-GMI or high-resolution WRF-Chem, highlighting the significance of integrating mechanism models with advanced deep learning techniques for substantial improvements in spatiotemporal predictions of air pollutants. This underscores the importance of leveraging the strengths of both methodologies to achieve more accurate and reliable results in predicting air quality dynamics.

In physical models, the advection velocity and diffusion coefficients are affected by meteorological factors, gravitational settling, viscosity of specific air pollutant, terrain, soil and land cover, which are hard to obtain due to missing or ill-posed boundary data and complicated atmospheric process[8,21], and require to be specified manually. The proxy coefficients in the

DGM overcame this as they were adaptively learned through end-to-end automatic differential learning. When the concentration measurements alone do not provide sufficient information to guarantee a single solution for these coefficients, we could provide additional data on them to enforce model training. However, it is impossible to acquire these coefficient measurements in practice and solving inverse problems with hidden physics is prohibitively expensive due to high uncertainty in simulation formulas[17]. Similar to the simulation by hidden fluid mechanics[18], we achieved a high quantitative agreement between observed and predicted proxy coefficients by relying solely on the information contained in the observed concentration data.

By simulating spatiotemporal dynamics of air mass and encouraging the adherence to the physical laws of mass conservation and continuity for all samples, including prediction points and those beyond the observation dataset, the trained model exhibits robust accuracy in spatiotemporal extrapolation (low generalization error), less affected by the values outside the training samples in predictions. Further, in this flexible architecture, satellite variables and reanalysis data of air pollutants provided clues for pollution background; alternatively, with reanalysis or inversed meteorology, the data from historical and future emission inventories can be used to predict future concentrations of criteria pollutants under certain carbon emissions and climate change scenarios. The MERRA2-GMI or other reanalysis data provide vertical background pollution levels that can be used to model regional-scale pollutant transport, or as precursor proxies for reactive air pollutants, but these coarse-scale data, if missing, have limited effect on generalization of DGM.

METHODS
Dataset
The observed data consisted of concentrations of six air pollutants (daily average of CO, NO_2, $PM_{2.5}$, PM_{10}, and SO_2; hourly maximum of daily O_3 and 8 h average of O_3) collected from monitoring stations of the China Environmental Monitoring Station. From January 1, 2015 to December 31, 2018. A total of 1,913,012 samples were obtained from 1604 monitoring stations covering mainland China. We had 56 spatial and temporal input variables including 2 spatial coordinates, elevation, day of year, ground aerosol coefficient, Aura Ozone Monitoring Instrument (OMI) NO_2, planetary boundary layer height (PBLH), 4 high-resolution meteorological variables, normalized difference vegetation index (NDVI), an indicator variable of workday, roadway length within a 10 km buffer around each monitoring site, 26 component pollutant variables and 10 reanalysis vertical meteorological variables. For more details, please see Supplementary Note 1 and Supplementary Table 1.

Modeling architecture
The architecture is constructed based on a Eulerian system where the change of air pollutant concentration at a specific spatial location of interest (Fig. 1a) is modeled using the following simplified partial differential equations (PDE):

$$\frac{\partial C}{\partial t} = -\nabla(VC) + p\nabla^2 C + R + E - F \quad (1)$$

where C is air pollutant concentration, ∂t represents the time derivative, $\partial C/\partial t$ represents concentration change in each target location over time, $\nabla(VC)$ represents advection, V represents velocity, p is the coefficient for the diffusion term ($\nabla^2 C$), R represents chemical transformation, E stands for emission, and F represents deposition.

To model over a surface, Eq. 1 uses horizontal advection and diffusion (l_x and l_y directions); vertical convection (l_z direction) is not explicitly considered, but proxy variables are used to account

for variation of vertical convection. Advection and diffusion can be simplified[32,33]:

$$\frac{\partial C}{\partial t} = -v_{l_x}\frac{\partial C}{\partial l_x} - v_{l_y}\frac{\partial C}{\partial l_y} + p_{l_x}\frac{\partial^2 C}{\partial^2 l_x} + p_{l_y}\frac{\partial^2 C}{\partial^2 l_y} + R + E - F \quad (2)$$

where v_{l_x} ($v_{l_y\,ss}$) is the Reynolds velocity of the air pollutant of the $l_x(l_y)$ direction of the spatial location of interest, $\partial C/\partial l_x$ ($\partial C/\partial l_y$) denotes the advection partial derivative, and p_{l_x} (p_{l_y}) is the diffusion coefficient (simplified as the sum of eddy and molecular diffusion coefficients), $\partial^2 C/\partial^2 l_x$ and $\partial^2 C/\partial^2 l_y$ represent the eddy diffusion terms. The air pollutant concentration changes over time (Supplementary Fig. 1) are decomposed into three components, advection, diffusion and the others.

For the diffusion process, the GCN provides an essential solution to simulate the second-order derivative on the discrete graph space[22,23], represented by the Laplacian convolution operator shown in Supplementary Fig. 2. Advection can be decomposed along the directions of the connected nodes based on the differential formula and the conservation of mass. We obtain the total convolutional operation to simulate the pollutant traffic on the graph space (Supplementary Fig. 1d):

$$\frac{\partial C}{\partial t} = -\overline{V}LC - pLC + \rho C = -(\overline{V}L + pD^{-\frac{1}{2}}AD^{-\frac{1}{2}})C + \rho C \quad (3)$$

where D is the degree matrix, L is the Laplacian matrix, A is the adjacency matrix to quantify spatial relationship, p is the diffusion coefficient, $\overline{V}$ is the *learnable* velocity matrix based on the Laplacian matrix, A, and ρ represents a proportion factor to account for the concentration changes from chemical transformation, emissions and deposition.

We constructed the semi-supervised multilevel local graph modeling architecture (Fig. 1). For each spatiotemporal node of interest, multilevel local graph convolutions were constructed in sequence based on the k nearest neighbors[73] to capture neighborhood information to encode multilevel dynamic processes of air mass. We characterized the space and time dimensions using a discrete data model to simulate the variation of spatiotemporal fields. Specifically, we used k-NN to find the nearest spatiotemporal samples based on the spatiotemporal distance to construct graph convolutions for each target node. The temporal and spatial dimensions had different units (coordinate space: meter; time: day) and needed to be fused to construct the deep graph network using k-NN. To fuse them, the standard deviation was used to normalize the coordinates (x and y) and time (day) to make them unitless. The reciprocal of the distance was also used as a weight in aggregating neighborhood information. The multilevel neighbors were retrieved based on the search depths to capture distance-varying neighborhood influence on a target node. For each graph convolution, we trained a set of functions that learned to aggregate a spread feature from a node's local neighborhood. Our GCN is an inductive framework[34], different from the transductive fixed graph, and it can be easily generalized to the same form of features or invisible nodes. In our application we used four graph convolutions where each node had k ($k = 12$ was an optimal solution of grid search) neighbors. For graph convolutions, the vector size was (in order) 128, 64, 32, and 1.

The output of each graph convolution was merged with the original input features as the input to the downstream components, a full residual deep network with the attention layers (Fig. 1a–c). The full residual encoder-decoder[36] was fused with the neighborhood features to account for local sources and sinks, and transformation. In order to reduce or avoid gradient vanishing in training, residual connections were introduced from each encoding layer to its corresponding decoding layer. This fusion also alleviated the potential over-smoothing issue in GC[74]. The attention layers were added with the residual layers after the input layer to allow the model to focus and place more "attention" (weights) on the important parts of the input features[35,75]. For the residual network topology, the number of nodes in each encoding or coding layer was $512 \to 320 \to 256 \to 128 \to 96 \to 64 \to 32 \to 16$; the number of nodes in each decoding layer was symmetric with its corresponding encoding layer. The final output was inversely normalized and exponentially transformed to represent the air pollutant concentrations in their original units. Graph convolutions extracted the spread feature from the input data of the multilevel neighbors for a node of interest. Irregularly distributed sparse or limited measured data at sample locations was used as the dependent variable to drive training of the entire model.

In addition, in order to make the final concentration predictions to maintain the continuity partial differential equation (Eq. 1), we introduce a residual term, e_2 to encode such physical invariance of PDE in the loss function:

$$L(\theta_{\mathbf{W,b}}) = e_1 + \alpha e_2 + \beta e_3 \quad (4)$$

where e_1 is the mean square error (MSE) function for the observed values and the predicted values, e_2 is the residual errors for the continuity equation (defined according to Eq. 1), e_3 is the normalization for the parameters, and α and β are the scaled weights for e_2 and e_3, respectively. The weight (α or β) can vary from 0 to 1, where a larger weight indicates under-fitting, while a smaller weight implies over-fitting. In order to determine the optimal weights, a sensitivity analysis was performed by testing a range of values between 0 and 1, with an interval of 0.1 for α and β. See Fig. 1d.

For each of the seven air pollutants considered, the system model was constructed and trained, using the logarithmic and normalized transformation of the pollutant concentrations as the dependent variable. In training, the minibatch size was 2048, and the initial training rate was 0.001 and adjusted in optimization, and the Adam optimizer[76] was used. The number of epochs was set at 200. In total, we sampled the dataset and trained the model 100 times and the final test performance was summarized.

Graph convolutions

As the core component in our method, multilevel graph convolutions were used to encode and represent local spread of air pollutants. Based on Eqs. 1–3 and Supplementary Fig. 1, the convolutional operator in the kth convolution layer is defined to simulate the spatiotemporal dynamics of air pollutant concentration:

$$\frac{\partial \mathbf{x}}{\partial t} = \frac{\mathbf{x}'_i - \mathbf{x}_i}{\Delta t} = -\mathbf{v}^k \sum_{j \in N(i)} \overline{A}_{ij}^k (\mathbf{x}_j - \mathbf{x}_i) + \mathbf{p}^k \sum_{j \in N(i)} (\mathbf{x}_j - \mathbf{x}_i) + \rho^k \mathbf{x}_i$$

$$(5)$$

where $\mathbf{x}_i$ (or $\mathbf{h}_i$) are the features (or the output) of the kth layer, representing the concentration of a specific air pollutant, C_i in the target location, i, $\mathbf{x}'_i$ (or $\mathbf{h}'_i$) is the updated status after an evolution timestep of mini-batch gradient descent, Δt, $\mathbf{v}^k$ is the velocity matrix, $\mathbf{p}^k$ can be interpreted as the proxy diffusion matrix, $\overline{A}_{ij}^k$ is the neighborhood weight matrix, and $N(i)$ is the set of spatial or spatiotemporal neighbors for i.

In order to strengthen representation learning, $\overline{A}_{ij}^k$ is simplified as the product of a learnable velocity (v^k) and the weight, $w_{d_{ij}}^k$ (the reciprocal distance from a neighbor to the target node), and two additional learnable matrixes, $\mathbf{W}_c^k$ and $\mathbf{W}_u^k$ are introduced for the current input and the gradient-derived change, respectively. Empirical thresholds can be set up for the learning parameters such as $\mathbf{v}^k$, $\mathbf{p}^k$ and ρ^k. Thus, we define the update of the kth graph

convolution after an evolution timestep (Δt) as:

$$\mathbf{x}'_i = \mathbf{W}^k_c \mathbf{x}_i - \mathbf{W}^k_u \left(-\mathbf{v}^k \sum_{j \in N(i)} w_{d^k_{ij}}(\mathbf{x}_j - \mathbf{x}_i) + \mathbf{p}^k \sum_{j \in N(i)} (\mathbf{x}_j - \mathbf{x}_i) + \rho^k \mathbf{x}_i \right)$$

(6)

The learning process involved the process of graph convolutions to simulate the spatiotemporal dynamics of air pollutants, as shown in Supplementary Fig. 1. Through the evolutional training driven by the measured concentration and covariate data, the model learned optimal parameters to inverse the target estimates. Accordingly, the minibatch forward propagation algorithm of Multilevel local Graph convolutions for Air pollutant spread was developed and shown as follows (Supplementary Algorithm 1).

Although GCs can efficiently extract spread features of air pollution, multiple graph convolutions may result in over-smoothing due to Laplacian smoothing. Thus, the full residual deep layers with multiple optimal attention layers was used as the downstream components for the output of neighborhood features to enhance the influence of non-traffic factors from emission, transformation and deposition on the final predictions. Here, the original input features and the aggregate feature were concatenated as the input for the full residual deep network. The residual error of continuity equation (e_2) was introduced to the loss function. For details, please see Supplementary Note 2 and Supplementary Algorithm 2.

Training, testing and sensitivity analysis

In total, approximately 18% (288) of monitoring sites were randomly selected for site-based testing, and all samples from these sites were used for independent testing only and not for training. Then, from the remaining samples, we randomly selected 78% for training and 22% for regular testing stratified by spatiotemporal factors (province and month). We chose the samples so that the proportion of the total samples for regular and site-based tests was similar to the proportion (~ 0.37) of the unique test samples in the bootstrap 0.632 rule[39]. In semi-supervised training, all 56 variables were used for model learning; all the training, regular testing and predicting samples were used to construct the graph convolutions to simulate fluid dynamics, and encouraged to satisfy the continuity equations in PDE residuals; only the training samples were used to minimize the MSE loss between observations and predictions. Theoretically, it is possible to utilize site-based testing samples in model training, encouraging them to satisfy the continuity equations. However, in practice, we opted not to include these site-based samples in our model training to avoid testing bias and ensure a fair comparison. To fairly compare our method with other machine learning methods, the same training, regular testing, and site-based testing samples were used for each method, and the hyperparameters of each model were optimally set. In addition, in order to avoid the influence of random noise, we trained 100 models and summarized their average performance for each air pollutant and each method.

We undertook a rigorous comparison of our method against several machine learning techniques commonly employed in similar domains. The methods included GraphSAGE (local graph)[34], full residual deep network[36], random forest[77], XGBoost[78], non-linear generalized additive model (GAM)[79], and spatial statistical models such as ordinary kriging and regression kriging[80]. To ensure a fair comparison, we used the same set of covariates for all the machine learning and statistical regression methods, with the exception of ordinary kriging. For ordinary kriging, it relies solely on constructing the variogram of spatial dependency from neighbors and does not utilize any covariates for predictions. In order to identify the most suitable

hyperparameters, we conducted a grid search for each method. The hyperparameter search was based on suggested value ranges specific to each method. Supplementary Table 2 provides a succinct overview of each method, including its default value or value range for the hyperparameters, as well as the search step. After obtaining the optimal hyperparameter values for each method, we proceeded to compare their performance using appropriate evaluation metrics including total and time series metrics. This approach allowed us to gain insights into the strengths and weaknesses of each technique and assess how our method measures up against these established alternatives.

For the sensitivity analysis of missing MERRA2-GMI pollutants, we also used the same training, regular testing and site-based testing samples. But we removed 28 variables of MERRA2-GMI pollutants (shown in the gray background in Supplementary Table 1) and kept the vertical meteorological variables and other variables. We developed our hybrid deep graph learning framework using PyTorch Geometric (Version 2.0), a specialized software tool known for its efficiency[81]. Model training, testing, and prediction were carried out in a server environment equipped with 128GB memory, 16 Intel(R) Xeon(R) E5–2620 CPUs, and three NVIDIA 1080Ti GPUs (12GB memory). On average, training and testing a national model of an air pollutant took approximately 3–4 days (72–96 h) with the support of a GPU. To expedite the process, we utilized parallel computing, allowing us to complete the training and testing of three national models simultaneously. For daily national predictions of each air pollutant spanning from 2015 to 2018, it took approximately 10 days. Again, parallel computing was employed to accelerate the national predictions of air pollutants.

REFERENCES

1. Vohra, K. et al. Global mortality from outdoor fine particle pollution generated by fossil fuel combustion: Results from GEOS-Chem. *Environ. Res.* **195**, 110754 (2021).
2. Cohen, A. J. et al. Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: An analysis of data from the Global Burden of Diseases Study 2015. *Lancet* **389**, 1907–1918 (2017).
3. Tao, Z. *Air Pollution and Greenhouse Gases.* (Springer, 2014).
4. Li, X. D., Jin, L. & Kan, H. D. Air pollution: A global problem needs local fixes. *Nature* **570**, 437–439 (2019).
5. Apte, J. S. et al. High-resolution air pollution mapping with google street view cars: Exploiting big data. *Environ. Sci. Technol.* **51**, 6999–7008 (2017).
6. Crawforda, B. et al. Mapping pollution exposure and chemistry during an extreme air quality event (the 2018 Kīlauea eruption) using a low-cost sensor network. *PNAS* **118**, e2025540118 (2021).

7. Zannetti, P. *Air Pollution Modeling: Theories, Computational Methods and Available Software*. (Springer Science & Business Media, 2013).

8. EPA. *CMAQ: The Community Multiscale Air Quality Modeling System*, https://www.epa.gov/cmaq (2021).

9. Sorbjan, Z. in *AIR QUALITY MODELING - Theories, Methodologies, Computational Techniques, and Available Databases and Software* (The EnviroComp Institute (http://www.envirocomp.org/) and the Air & Waste Management Association (http://www.awma.org/), 2003).

10. Camps-Valls, G., Tuia, D., Zhu, X. X. & Reichstein, M. *Deep Learning for the Earth Sciences: A Comprehensive Approach to Remote Sensing, Climate Science, and Geosciences*. (John Wiley & Sons Ltd, 2021).

11. Wang, J. F., Zhang, T. L. & Fu, B. J. A measure of spatial stratified heterogeneity. *Ecol. Indic.* **67**, 250–256 (2016).

12. Keller, C. A. & Evans, M. J. Application of random forest regression to the calculation of gas-phase chemistry within the GEOS-Chem chemistry model v10. *Geosci. Model Dev.* **12**, 1209–1225 (2019).

13. Kelp, M. M., Jacob, D. J., Kutz, J. N., Marshall, J. D. & Tessum, C. W. Toward stable, general machine-learned models of the atmospheric chemical system. *J. Geophys. Res.-Atmos.* **125** (2020).

14. Sturm, P. O. & Wexler, A. S. A mass- and energy-conserving framework for using machine learning to speed computations: a photochemistry example. *Geosci. Model Dev.* **13**, 4435–4442 (2020).

15. Ivatt, P. D. & Evans, M. J. Improving the prediction of an atmospheric chemistry transport model using gradient-boosted regression trees. *Atmos. Chem. Phys.* **20**, 8063–8082 (2020).

16. Rasp, S., Pritchard, M. S. & Gentine, P. Deep learning to represent subgrid processes in climate models. *Proc. Natl Acad. Sci. Usa.* **115**, 9684–9689 (2018).

17. Karniadakis, G. E. et al. Physics-informed machine learning. *Nat. Rev. Phys.* **3**, 422–440 (2021).

18. Raissi, M., Yazdani, A. & Karniadakis, G. E. Hidden fluid mechanics: Learning velocity and pressure fields from flow visualizations. *Science* **367**, 1026 (2020).

19. Hahnel, P., Marecek, J., Monteil, J. & O'Donncha, F. Using deep learning to extend the range of air pollution monitoring and forecasting. *J. Comput. Phys.* **408**, https://doi.org/10.1016/j.jcp.2020.109278 (2020).

20. Sturm, P. O. & Wexler, A. S. Conservation laws in a neural network architecture: enforcing the atom balance of a Julia-based photochemical model (v0. 2.0). *Geosci. Model Dev.* **15**, 3417–3431 (2022).

21. Jacobson, Z. M. *Fundamentals of Atmospheric Modeling, 2nd Edition*. (Cambridge University Press, 2005).

22. Spielman, D. Spectral Graph Theory. *Combinatorial Scientific Computing*, 495-524 (Chapman and Hall/CRC Press, 2012).

23. Bruna, J., Zaremba, W., Szlam, A. & LeCun Y. Spectral networks and deep locally connected networks on graphs. (arXiv, 2013).

24. Li, Q., Han, Z. & Wu, X.-M. Deeper insights into graph convolutional networks for semi-supervised learning. *Thirty-Second AAAI Conference on Artificial Intelligence*, 3538-3545 (2018).

25. Fout, M. A. *Protein Interface Prediction Using Graph Convolutional Network*, Master thesis, Colorado State University, (2017).

26. Han, P. et al. GCN-MF: Disease-gene association identification by graph convolutional networks and matrix factorization. *KDD'19: Proceedings of the 25th Acm Sigkdd International Conferencce on Knowledge Discovery and Data Mining*, 705–713 (2019).

27. Stokes, J. M. et al. A deep learning approach to antibiotic discovery. *Cell* **180**, 688–702.e613 (2020).

28. Andrews, D. *An Introduction to Atmospheric Physics*. (Cambridge University, 2010).

29. Cuomo, S. et al. Scientific machine learning through physics–informed neural networks: where we are and what's next. *J. Sci. Comput.* **92**, 88 (2022).

30. Li, M. et al. Anthropogenic emission inventories in China: A review. *Natl Sci. Rev.* **4**, 834–866 (2017).

31. Sujatha, P., Mahalakshmi, D., Ramiz, A., Rao, P. & Naidu, C. Ventilation coefficient and boundary layer height impact on urban air quality. *Cogent Environ. Sci.* **2**, 1125284 (2016).

32. Pedlosky, J. *Geophysical Fluid Dynamics*. 10–13 (Springer, 1987).

33. Ulfah, S., Awalludin, S. A. & Wahidin. Advection-diffusion model for the simulation of air pollution distribution from a point source emission. *1st International Conference of Education on Sciences, Technology, Engineering, and Mathematics (Ice-Stem)* **948** (2018).

34. Hamilton, W. L., Ying, R. & Leskovec, J. Inductive representation learning on large graphs. *Advances in Neural Information Processing Systems 30 (NIPS 2017)* **30** (2017).

35. Vaswani, A. et al. Attention is all you need. (arXiv preprint arXiv:1706.03762., 2017).

36. Li, L., Fang, Y., Wu, J., Wang, J. & Ge, Y. Encoder-decoder full residual deep networks for robust regression prediction and spatiotemporal estimation. *IEEE Trans. Neural Netw. Learn. Syst.* **32**, 4217–4230 (2021).

37. Li, L. F. et al. Spatiotemporal imputation of MAIAC AOD using deep learning with downscaling. *Remote Sens. Environ.* **237**, 11584 (2020).

38. Raissi, M., Perdikaris, P. & Karniadakis, G. E. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Computational Phys.* **378**, 686–707 (2019).

39. Berrar, D. & Dubitzky, W. in *Encyclopedia of Systems Biology* (eds W. Dubitzky, O. Wolkenhauer, K. H. Cho, & H. Yokota) (Springer, 2013).

40. Sicard, P. et al. High spatial resolution WRF-Chem model over Asia: Physics and chemistry evaluation. *Atmos. Environ.* **244**, 118004 (2021).

41. Adedeji, A., Dagar, L., Petra, M. & De Silva, L. Sensitivity of WRF-Chem model resolution in simulating tropospheric ozone in Southeast Asiain. In *IOP Conference Series: Earth and Environmental Science*, Vol. 489, 12030 (IOP Publishing, 2020).

42. Guo, W.-K. et al. Establishment of a high-resolution anthropogenic emission inventory and its evaluation using the WRF-Chem model for Lanzhou. *Environ. Sci. (China)* **42**, 634–642 (2021).

43. Žabkar, R. et al. Evaluation of the high resolution WRF-Chem (v3. 4.1) air quality forecast and its comparison with statistical ozone predictions. *Geoscientific Model Dev.* **8**, 2119–2137 (2015).

44. Chen, B. et al. Estimation of Atmospheric PM_{10} Concentration in China Using an Interpretable Deep Learning Model and Top-of-the-Atmosphere Reflectance Data From China's New Generation Geostationary Meteorological Satellite, FY-4A. *J. Geophys. Res.: Atmospheres* **127**, e2021JD036393 (2022).

45. Wei, J. et al. The ChinaHighPM_{10} dataset: generation, validation, and spatio-temporal variations from 2015 to 2019 across China. *Environ. Int.* **146**, 106290 (2021).

46. Chen, G. et al. Spatiotemporal patterns of PM_{10} concentrations over China during 2005–2016: A satellite-based estimation using the random forests approach. *Environ. Pollut.* **242**, 605–613 (2018).

47. Wang, Y., Yuan, Q., Li, T., Tan, S. & Zhang, L. Full-coverage spatiotemporal mapping of ambient $PM_{2.5}$ and PM_{10} over China from Sentinel-5P and assimilated datasets: Considering the precursors and chemical compositions. *Sci. Total Environ.* **793**, 148535 (2021).

48. He, Q., Gao, K., Zhang, L., Song, Y. & Zhang, M. Satellite-derived 1-km estimates and long-term trends of $PM_{2.5}$ concentrations in China from 2000 to 2018. *Environ. Int.* **156**, 106726 (2021).

49. Ma, Z., Hu, X., Huang, L., Bi, J. & Liu, Y. Estimating ground-level $PM_{2.5}$ in China using satellite remote sensing. *Environ. Sci. Technol.* **48**, 7436–7444 (2014).

50. Wei, J. et al. Estimating 1-km-resolution $PM_{2.5}$ concentrations across China using the space-time random forest approach. *Remote Sens. Environ.* **231**, 111221 (2019).

51. Wei, J. et al. Reconstructing 1-km-resolution high-quality $PM_{2.5}$ data records from 2000 to 2018 in China: spatiotemporal variations and policy implications. *Remote Sens. Environ.* **252**, 112136 (2021).

52. Wei, J. et al. Full-coverage mapping and spatiotemporal variations of ground-level ozone (O_3) pollution from 2013 to 2020 across China. *Remote Sens. Environ.* **270**, 112775 (2022).

53. Wei, J. et al. Ground-level NO_2 surveillance from space across China for high resolution using interpretable spatiotemporally weighted artificial intelligence. *Environ. Sci. Technol.* **56**, 9988–9998 (2022).

54. Zheng, S. & Singh, R. P. Aerosol and meteorological parameters associated with the intense dust event of 15 April 2015 over Beijing, China. *Remote Sens.* **10**, 957 (2018).

55. Creamean, J. M. et al. Dust and biological aerosols from the Sahara and Asia influence precipitation in the western U.S. *Science* **339**, 1572–1578 (2013).

56. Mao, J. et al. Meteorological mechanism for a large-scale persistent severe ozone pollution event over eastern China in 2017. *J. Environ. Sci. (China)* **92**, 187–199 (2020).

57. Beckwith, M., Bates, E., Gillah, A. & Carslaw, N. NO_2 hotspots: Are we measuring in the right places? *Atmos. Environ.: X* **2**, 100025 (2019).

58. WHO, R. O. f. E. C. in *Air Quality Guidelines for Europe* (WHO Regional Publications, 2000).

59. World Air Quality Index. *Air Pollution in China: Real-time Air Quality Index Visual Map*, https://aqicn.org/map/china/ (2021).

60. Ministry of Environmental Protection of China. *Technical Regulation on Ambient Air Quality Index*. (China Environmental Science Press, 2012).

61. Zhang, Q. & Geng, G. N. Impact of clean air action on $PM_{2.5}$ pollution in China. *Sci. China-Earth Sci.* **62**, 1845–1846 (2019).

62. Lu, X. et al. Severe surface ozone pollution in China: A global perspective. *Environ. Sci. Technol. Lett.* **5**, 487–494 (2018).

63. Estrada, E. *The Structure of Complex Networks: Theory and Applications*. (Oxford University Press, 2016).

64. Silva, S. J., Burrows, S. M., Evans, M. J. & Halappanavar, M. A graph theoretical intercomparison of atmospheric chemical mechanisms. *Geophysical Research Letters* **48**, https://doi.org/10.1029/2020GL090481 (2021).

65. Zhou, J. et al. Graph neural networks: A review of methods and applications. *AI Open* **1**, 57–81 (2020).

66. Liu, X., Lu, D., Zhang, A., Liu, Q. & Jiang, G. Data-driven machine learning in environmental pollution: Gains and problems. *Environ. Sci. Technol.* **56**, 2124–2133 (2022).

67. Zhang, B. et al. Deep learning for air pollutant concentration prediction: A review. *Atmos. Environ.* **290**, 119347 (2022).

68. Liao, Q. et al. Deep learning for air quality forecasts: A review. *Curr. Pollut. Rep.* **6**, 399–409 (2020).

69. Hu, X. et al. Estimating $PM_{2.5}$ concentrations in the conterminous United States using the random forest approach. *Environ. Sci. Technol.* **51**, 6936–6944 (2017).

70. Delworth, T. L. et al. Simulated climate and climate change in the GFDL $CM_{2.5}$ high-resolution coupled climate model. *J. Clim.* **25**, 2755–2781 (2012).

71. Grell, G. A. et al. Fully coupled "online" chemistry within the WRF model. *Atmos. Environ.* **39**, 6957–6975 (2005).

72. Hornik, K. Approximation capabilities of multilayer feedforward networks. *Neural Netw.* **4**, 251–257 (1991).

73. Altman, N. S. An introduction to kernel and nearest-neighbor nonparametric regression. *Am. Stat.* **46**, 175–185 (1992).

74. Yang, C., Wang, R., Yao, S., Liu, S. & Abdelzaher, T. Revisiting oversmoothing in deep GCNs. arXiv preprint:2003.13663 (arXiv 2020).

75. Li, Y., Zeng, J. B., Shan, S. G. & Chen, X. L. Occlusion aware facial expression recognition using CNN with attention mechanism. *IEEE Trans. Image Process.* **28**, 2439–2450 (2019).

76. Kingma, D. P. & Ba, J. Adam: A method for stochastic optimization. (arXiv, 2014).

77. Liaw, A. & Wiener, M. Classification and regression by randomForest. *R. News* **2**, 18–22 (2002).

78. Chen, T. et al. XGBoost: extreme gradient boosting. *R. package version 0. 4-2* **1**, 1–4 (2015).

79. Wood, S. *Generalized Additive Models: An Introduction with R*. (Chapman and Hall/CRC 2006).

80. Hengl, T., Heuvelink, G. B. & Stein, A. *Comparison of Kriging with External Drift and Regression Kriging*. (ITC Enschede The Netherlands, 2003).

81. Fey, M. & Lenssen, J. E. Fast graph representation learning with PyTorch Geometric. *arXiv preprint:1903.02428* (arXiv 2019).

ACKNOWLEDGEMENTS

AUTHOR CONTRIBUTIONS

L.L., and J.F.W. designed research; L.L. performed research and analyzed data; J.J.W. and Z.Z. conducted data preparation and preprocessing; Q.Y., C.W., and Y.G. provided technical support and assistance; J.F.W., M.F., G.C.-V., and M.R. provided critical intellectual contributions throughout the study; and L.L. and J.F.W. wrote the paper.

COMPETING INTERESTS

The authors declare no competing interest.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41612-023-00475-3.

Correspondence and requests for materials should be addressed to Lianfa Li or Jinfeng Wang.

Reprints and permission information is available at http://www.nature.com/reprints

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