

<https://doi.org/10.1038/s44407-026-00052-x>

# TransNet: a transport-informed graph neural network for forecasting PM<sub>2.5</sub> concentrations across South Korea

Check for updates

Rijul Dimri, Yunsoo Choi , Deveshwar Singh, Jincheol Park &amp; Nima Khorshidian

Accurate forecasting of PM<sub>2.5</sub> concentrations is essential for air quality management and public health protection. Previous methods for correcting biases in chemical transport model simulations are computationally expensive, limiting their operational applications. We introduce the Transport-Informed Graph Neural Network (TransNet), a machine learning model that forecasts PM<sub>2.5</sub> concentrations for +72-h by learning coupled Advection-Diffusion-Reaction (ADR) operators, enabling efficient operational deployment. Evaluated over 170 stations across South Korea, TransNet outperforms the bias-correction model Adaptive Graph Attention Network (AGATNet) at short-term lead times (+1 h to +48 h), achieving initial MAE of 2.60 µg/m<sup>3</sup> compared to AGATNet's 4.27 µg/m<sup>3</sup>. However, AGATNet demonstrates greater stability at extended lead times (>+48 h) and better captures extreme events. Operator analysis reveals that coupled advection-diffusion systems are essential for capturing transport dynamics, with error increment by 153.5% at +1 h, while isolated advection or diffusion produced 72-73% error increment. TransNet represents a physics-informed framework with improved capability for operational PM<sub>2.5</sub> forecasting.

South Korea has emerged as a global technological powerhouse over the past several decades, accompanied by rapid urbanization and industrial growth. Such advances have introduced significant environmental trade-offs, particularly in relation to urban air quality. As of 2025, more than 82% of the population resides in urban areas<sup>1</sup>, and the combined effects of economic development, intensive industrial activity, and high population density have contributed to persistent air pollution across metropolitan areas<sup>2</sup>. Foremost among the current air quality concerns, fine particulate matter (PM<sub>2.5</sub>) has been a major issue despite rigorous regulatory interventions. Recent observations indicate encouraging trends, with surface and satellite measurements from 2015 to 2023 revealing measurable declines in PM<sub>2.5</sub> concentrations across South Korea<sup>3,4</sup>, reflecting the effectiveness of policy interventions and evolving urban systems. Nevertheless, PM<sub>2.5</sub> remains a critical indicator of air pollution that significantly impairs visibility and contributes to regional haze formation<sup>5,6</sup>. Its impact also extends beyond national boundaries, intensifying transboundary pollution across East Asia<sup>7,8</sup>. The associated public health risks are considerable. Chronic exposure to PM<sub>2.5</sub> is strongly associated with increased incidence of both respiratory and non-respiratory illnesses<sup>9,10</sup>, heightened risk of cardiovascular diseases<sup>11,12</sup>, and elevated premature mortality within the population<sup>9,12,13</sup>.

Traditional PM<sub>2.5</sub> forecasting methods generally fall into two categories: numerical and statistical approaches. Numerical approaches, such as

chemical transport models (CTMs), explicitly simulate atmospheric processes by solving governing physical and chemical equations for the advection, diffusion, transformation, and deposition of air pollutants<sup>13</sup>. While these models provide detailed mechanistic insights into pollutant behavior<sup>14,15</sup>, their high computational cost often limits their utility for real-time forecasting<sup>16</sup>. On the other hand, data-driven statistical methods leverage historical data to identify time series patterns in air pollutant concentrations<sup>17</sup>. Most commonly used methods include regression techniques such as Multiple Linear Regression (MLR)<sup>18</sup> and Support Vector Regression (SVR)<sup>19</sup>. Although these are computationally efficient, their underlying assumptions such as linear relationships between variables and temporal stationarity limit their ability to capture the non-linear, dynamic behaviors of air pollutants<sup>20,21</sup>. Consequently, they often fail to generalize under rapidly changing meteorology and evolving emissions patterns. The computational constraints of numerical models and the narrow applicability of conventional statistical methods have motivated the development of more advanced modeling techniques capable of real-time, non-linear forecasting.

In response to the limitations of numerical and statistical methods, deep learning (DL) approaches have emerged as promising alternatives to accurately forecast air quality at significantly reduced computational costs<sup>21-28</sup>. Recently, Graph Neural Networks (GNNs) have gained particular

attention in air quality applications due to their ability to capture complex, non-linear spatial dependencies across locations<sup>29–35</sup>. For example, air quality monitoring stations can be assigned as nodes, and their unique connections, represented by an adjacency matrix, can be constructed to represent environmental factors such as geographic distance, prevailing wind direction, meteorological similarity, or pollution patterns<sup>30,33</sup>. This capability is particularly useful when monitoring stations are unevenly distributed across complex urban settings, where conventional spatial interpolation becomes less reliable. However, despite their architectural strengths, GNN-based models often experience degradation in forecast accuracy at mid-range lead times. For example, previous studies have shown that the performance of GNN architectures tends to decline beyond 6 to 12 h<sup>25,29–31,33</sup>. One major limitation stems from the use of static adjacency matrices, which are typically constructed based on fixed geographic distances or long-term spatial correlations<sup>35,36</sup>. Such static structures cannot effectively represent temporally evolving meteorological processes, such as changes in wind direction and speed, which strongly influence pollutant transport. Some studies have introduced wind-informed graph construction methods that encode directional relationships between stations using wind vectors and relative positioning of station pairs<sup>33,53</sup>. However, such wind-informed dynamic adjacency approaches may lack the temporal stability to sustain forecast accuracy beyond 6–12 h, contributing to the observed performance degradation at mid-range lead times<sup>30,33,37</sup>. Recent physics-guided methods have attempted to address this limitation by incorporating transport processes directly into the learning framework<sup>25,46,53</sup>. For example, AirPHyNet integrates diffusion and advection as differential equation networks to capture pollutant dispersion dynamics, demonstrating improvements over purely data-driven methods with forecasts extending to 72 h. Despite these advances, fundamental challenges persist. Physics-guided approaches often rely on simplified transport representations by omitting reaction processes or encoding transport information through input features rather than learnable physical operators, limiting their ability to capture directional transport dynamics.

Recent progress in Physics-Informed Neural Networks (PINNs) has enabled the integration of physical laws into data-driven models<sup>22,27,38–42,46,53</sup>. However, their application in air quality forecasting remains constrained by several limitations. Most existing approaches either rely on output from CTMs for supervised training<sup>30,43,44</sup> or embed governing equations into the loss function to enforce physical consistency. While these strategies introduce physical inductive biases<sup>38</sup>, they often treat transport processes as fixed constraints rather than learnable components. This rigid formulation can restrict the model's capacity to represent complex, non-linear pollutant dynamics, especially in transport-dominated regimes, due to optimization difficulties and sensitivity to loss formulation<sup>45</sup>. To address this, operator learning provides a more flexible and principled alternative. For example, operators such as Fourier Neural Operators (FNO)<sup>46</sup> have demonstrated success in learning function-to-function mappings for partial differential equations, enabling resolution-invariant and transferable predictions. In the context of air quality forecasting, transport mechanisms such as advection, diffusion, and reaction<sup>47</sup> can be explicitly modeled as learnable operators<sup>42</sup>. These operators can be parameterized and trained directly from data to approximate the underlying transport dynamics, enabling neural networks to simulate pollutant evolution in a physically grounded and computationally efficient manner.

Motivated by these opportunities and building upon prior approaches, this study proposes Transport-informed Graph Neural Network (TransNet), a novel physics-informed framework designed to enhance temporal stability while eliminating reliance on CMAQ output by integrating atmospheric transport processes with data-driven feature engineering. TransNet models the full advection–diffusion–reaction system, discretized over graph structures where monitoring stations serve as nodes and their spatiotemporal interactions define edges. The framework implements three coupled physical processes: (1) advection, through learnable directional edge weights that represent wind-driven  $PM_{2.5}$  transport while preserving mass conservation; (2) diffusion, via adaptive graph Laplacian operators

with feature-wise coefficients that control spatial smoothing; and (3) reaction, modeled using pointwise multilayer perceptrons (MLPs) that integrate meteorological variables with the historical pollutant features. These components are integrated using an operator splitting strategy, which maintains numerical stability while allowing for sequential optimization of each physical process. To evaluate the performance of TransNet, experiments were conducted on  $PM_{2.5}$  forecasting using data spanning 2018–2021, with the model trained on 2018–2019 data, validated on 2020 data, and tested on the 2021 dataset. Results were compared with AGATNet<sup>30</sup> (an Adaptive Graph Attention Network), a data-driven bias-correction model that leverages the Community Multiscale Air Quality (CMAQ) model to refine forecasts. Unlike data-driven bias-correction approaches that require CTM outputs, TransNet maintains physical relevance through explicit learnable transport operators while generating better forecasts (till day 2) directly from observational and meteorological data over in-situ monitoring sites.

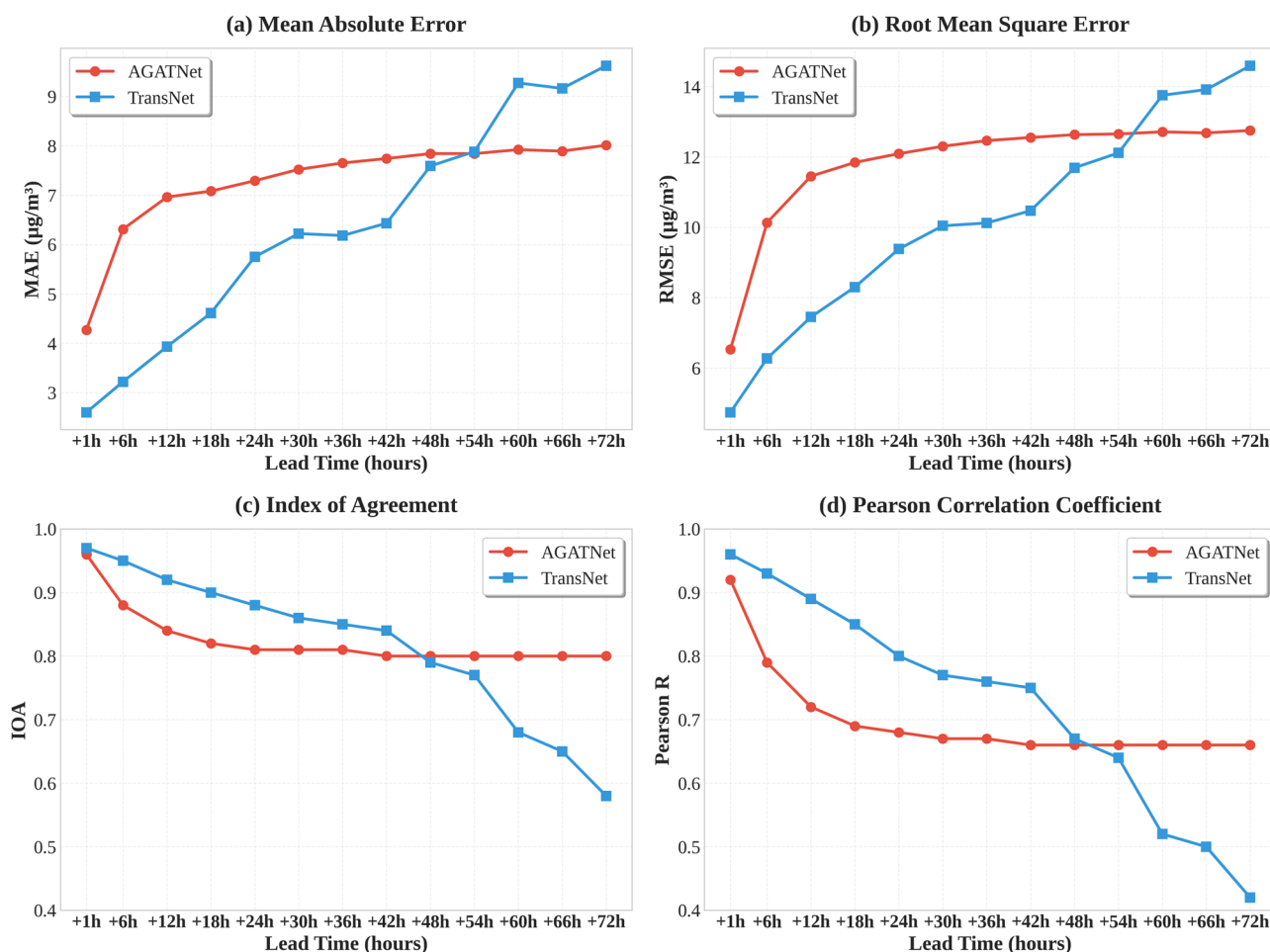
## Results

### Comparative performance: data-driven vs. physics-informed models for forecasting $PM_{2.5}$ concentrations

To evaluate TransNet's performance in forecasting surface  $PM_{2.5}$  concentrations across South Korea, its predictions were benchmarked against those from AGATNet, a recent data-driven bias-correction model (see Supplementary Note 4 for AGATNet model description and Supplementary Table 1 and Supplementary Table 2 for the input variables and comparison with CMAQ, respectively). This involved comparing the models' forecasts to ground-truth observations at 170 air quality monitoring stations for the year 2021.

Across 13 lead times from +1 h to +72 h (Fig. 1), a clear trade-off was found in performance. TransNet demonstrated higher forecast accuracy at initial lead times, while AGATNet showed greater stability at extended lead times within the forecast period. During the +1–12 h window, TransNet showed substantially lower prediction biases, with a Mean Absolute Error (MAE) lower by 39% at +1 h lead time, 49% at +6 h, and 44% at +12 h approximately, compared to AGATNet. Corresponding Root Mean Square Error (RMSE) was approximately lower by 27%, 38%, and 35% for the same lead times. TransNet's predictions also showed higher agreement with the station measurements, with Pearson's correlation coefficients (R) and Index of Agreement (IOA) of 0.96 and 0.97 at +1 h and 0.89 and 0.92 at +12 h, respectively; AGATNet achieved R and IOA of 0.92 and 0.96 at +1 h and 0.72 and 0.84 at +12 h. TransNet's advantage in having lower forecast biases was maintained for subsequent lead times, with MAE values that were lower by 35% at +18 h, 21% at +24 h, and 17% at +42 h approximately. This performance was attributed to TransNet's operator-based formulation that explicitly decomposes  $PM_{2.5}$  evolution into advection, diffusion, and reaction operators, consistent with the structure of the governing transport equation. This operator splitting strategy maintains numerical stability and physical consistency through two mechanisms: the small integration step size ( $h = 0.01$ ) limits the magnitude of state updates per operator, and physics-based constraints including mass conservation in advection and bounded diffusivity coefficients prevent unphysical amplification of errors. The resulting controlled error propagation yields gradual forecast degradation across lead times, particularly at short-to-medium horizons (+1 h to +48 h) where atmospheric transport dominates  $PM_{2.5}$  evolution. In contrast, AGATNet's architecture jointly learns spatial and temporal dependencies through coupled graph attention and temporal convolution layers without explicit physical constraints on intermediate representations. While this integrated design exhibits sharper initial accuracy decline, AGATNet's access to CTM-derived chemical information provides superior stability over extended forecast horizons where chemistry-dominated processes become increasingly important.

AGATNet demonstrated a sharper initial decline in forecast accuracy but achieved greater stability at extended lead times. Following a pattern common to GNN-based data-driven models, AGATNet's accuracy declined after the +6–12 h window, with the error propagation stabilizing after +18 h. However, while TransNet's R declined from 0.67 to 0.42 from



**Fig. 1 | Comparative performance of TransNet and AGATNet across 13 forecast lead times for PM<sub>2.5</sub> concentration predictions.** Performance metrics comparing AGATNet and TransNet forecasts of surface PM<sub>2.5</sub> concentrations at 170 monitoring stations across South Korea for the year 2021. **a** shows Mean Absolute Error

(MAE;  $\mu\text{g}/\text{m}^3$ ), panel **(b)** shows Root Mean Square Error (RMSE;  $\mu\text{g}/\text{m}^3$ ), **(c)** shows Index of Agreement (IOA), and **(d)** shows Pearson correlation coefficient (R). Red circles connected by solid lines represent AGATNet performance, while blue squares connected by solid lines represent TransNet performance.

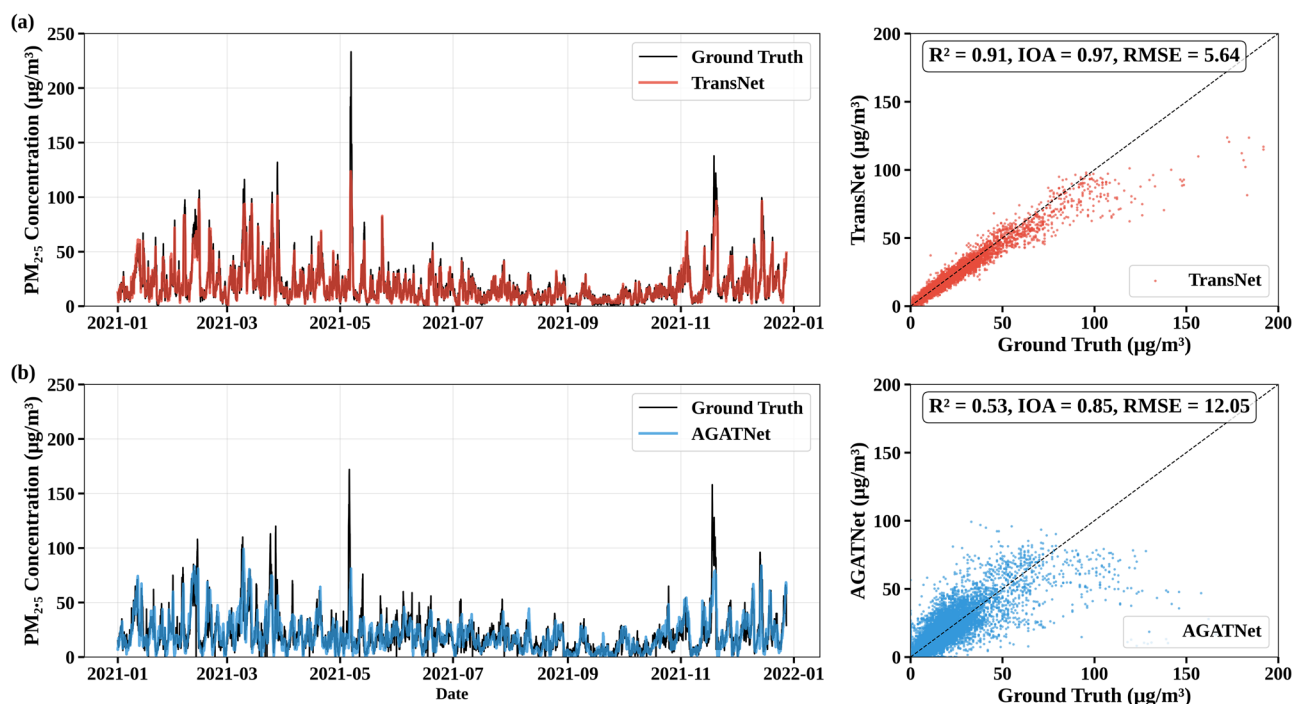
the +48 h to +72 h lead times, AGATNet’s R remained stable at 0.66. This can be attributed to its training on CMAQ outputs, which are inherently less noisy than station observations.

Additionally, the models’ +1 h forecast performance was compared at a station in Seoul to further investigate their behavior during severe pollution events (Fig. 2). AGATNet’s forecasts, benefiting from access to detailed chemical speciation in its CTM input, more accurately captured the peak concentrations during extreme pollution events exceeding 120  $\mu\text{g}/\text{m}^3$ . However, TransNet’s forecasts demonstrated a better overall fit to the station observations, achieving a higher coefficient of determination ( $R^2$ ) of 0.93 and a lower RMSE of 5.64  $\mu\text{g}/\text{m}^3$ , compared to AGATNet’s corresponding values of 0.88, and 7.14  $\mu\text{g}/\text{m}^3$ .

Detection analysis (Supplementary Note 5) at South Korea’s Very Unhealthy threshold ( $\geq 76 \mu\text{g}/\text{m}^3$ , 1.2% of observations; see Supplementary Figure 11) reveals distinct capabilities (see Supplementary Note 4). At +1 h, AGATNet achieves 84.5% recall (see Supplementary Fig. 12), detecting 16,079 of 17,921 severe events through CMAQ chemical speciation that captures secondary formation and transboundary transport mechanisms. TransNet achieves 53.9% recall, detecting 9655 events, representing 6424 fewer detections. However, AGATNet generates 6077 false alarms versus TransNet’s 619, with precision of 72.6% versus TransNet’s 94.0%. At extended horizons, AGATNet maintains detection (15.4% at +12 h, 9.1% at +72 h) while TransNet’s recall approaches zero beyond +12 h. Critically, AGATNet misses 75–90% of severe events at extended horizons, reflecting inherited CTM limitations: CMAQ systematically under-predicts extreme

episodes due to emission uncertainties and incomplete secondary aerosol schemes. Both approaches demonstrate fundamental limitations during the rarest episodes ( $> 100 \mu\text{g}/\text{m}^3$ , 0.4% of observations).

The spatial distribution of forecast error was analyzed to identify geographic patterns in model performance across South Korea’s 170 monitoring stations (Fig. 3), classified into urban ( $n = 35$ ), suburban ( $n = 88$ ), and coastal ( $n = 47$ ) categories (Supplementary Figs. 13–15). TransNet exhibited consistently lower forecast errors at most stations during shorter lead times, while AGATNet demonstrated superior performance at extended horizons. At +1 h, TransNet achieved lower RMSE at 166 of 170 stations (mean RMSE: 4.58  $\mu\text{g}/\text{m}^3$  vs 6.47  $\mu\text{g}/\text{m}^3$ ), with categorical analysis revealing the strongest performance advantage at coastal locations (mean RMSE difference: 2.12  $\mu\text{g}/\text{m}^3$ ), where TransNet outperformed at all 47 stations, followed by comparable advantages at urban (1.82  $\mu\text{g}/\text{m}^3$  difference, 34 of 35 stations) and suburban (1.81  $\mu\text{g}/\text{m}^3$  difference, 85 of 88 stations) locations. This spatial advantage extended to 168 stations at +6 h (mean RMSE: 6.02 vs 9.98  $\mu\text{g}/\text{m}^3$ ) and persisted through +24 h at 153 stations (9.07 vs 11.88  $\mu\text{g}/\text{m}^3$ ), with TransNet maintaining high outperformance rates of 85.7%, 87.5%, and 97.9% at urban, suburban, and coastal stations, respectively. The crossover in model performance began at +48 h, where TransNet’s advantage narrowed substantially (mean RMSE: 11.42 vs 12.41  $\mu\text{g}/\text{m}^3$  across 131 stations), with outperformance rates declining to 85.7%, 79.5%, and 66.0% at urban, suburban, and coastal stations, respectively. By +72 h, AGATNet demonstrated lower errors at 167 of 170 stations (mean RMSE: 12.54 vs 14.37  $\mu\text{g}/\text{m}^3$ ), with



**Fig. 2 | Comparison of +1 h forecast performance between TransNet and AGATNet at a Seoul monitoring station throughout 2021.** **a** Shows TransNet performance with a time series plot (left) displaying ground truth as a black solid line and TransNet forecasts as a red solid line, alongside a scatter plot (right) showing predicted versus observed concentrations as red dots with a black dashed one-to-one

reference line ( $R^2 = 0.91$ ,  $IOA = 0.97$ ,  $RMSE = 5.64 \mu\text{g}/\text{m}^3$ ). **b** Shows AGATNet performance with a time series plot (left) displaying ground truth as a black solid line and AGATNet forecasts as a blue solid line, alongside a scatter plot (right) showing predicted versus observed concentrations as blue dots with a black dashed one-to-one reference line ( $R^2 = 0.53$ ,  $IOA = 0.85$ ,  $RMSE = 12.05 \mu\text{g}/\text{m}^3$ ).

TransNet outperformance declining to 0% at urban stations, 1.1% at suburban stations, and 4.3% at coastal stations. The standard deviation of RMSE differences across stations remained relatively consistent throughout most forecast horizons ( $0.75\text{--}1.94 \mu\text{g}/\text{m}^3$ ). The gradual erosion of TransNet's advantage, from 168 stations at +6 h to 153 at +24 h and 131 at +48 h, suggests that the model's learned transport mechanisms become increasingly limited at extended forecast horizons. TransNet's wind-informed graph construction establishes connections when wind direction aligns with inter-station vectors (see Eq. 3), effectively capturing direct advective transport pathways that dominate pollutant evolution at short lead times. However, without explicit topographic information, coastal meteorological constraints, and regional circulation patterns as additional inputs, the model cannot adequately represent the spatial complexity introduced by these atmospheric processes. Incorporating these atmospheric processes as geometric and meteorological constraints in the transport operators could potentially reduce model error.

To further assess TransNet's year-to-year robustness, its performance was evaluated on an independent dataset for 2020 and 2021 (Fig. 4). Despite interannual variability in meteorological conditions and observed  $\text{PM}_{2.5}$  concentrations, TransNet maintained consistent skill across both years. IOA remained above 0.85 for the first 36 forecast hours in 2020 and above 0.80 for the first 30 h in 2021, reflecting a reliable alignment with the observations. RMSE remained below  $10 \mu\text{g}/\text{m}^3$  for the first 48 h in 2020 and for up to 36 h in 2021, showing stable error characteristics that are essential for reliable short-range forecasting. R were also notably high for earlier lead times, achieving 0.96 at +1 h in 2020 and 2021, and remained above 0.70 for up to +48 h in 2020. These findings highlight TransNet's year-to-year generalization capacity and suggest its applicability as a reliable, standalone air quality forecasting model for operational deployment.

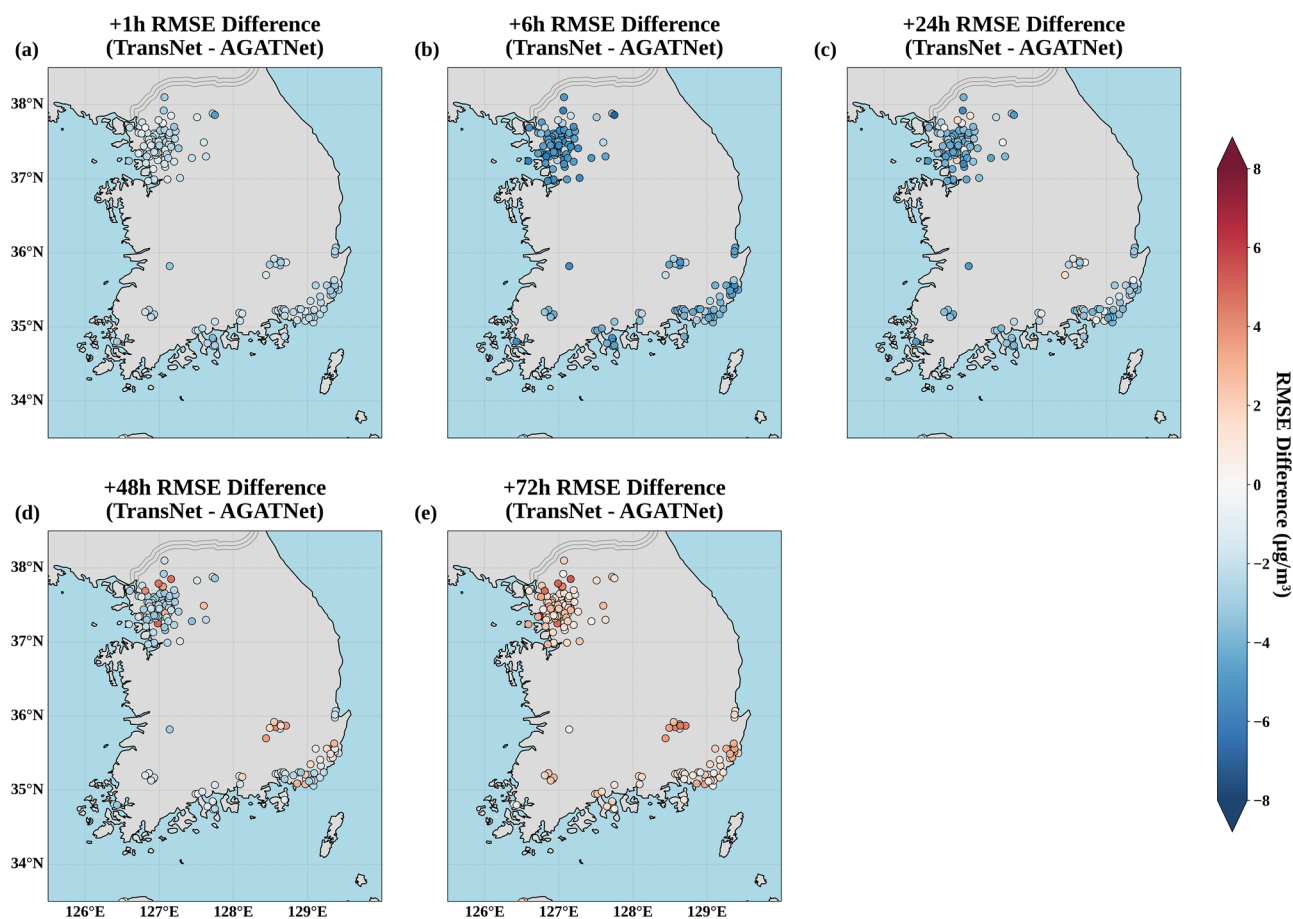
### Evaluation of physics-informed transport pathways

To validate the physical basis of TransNet's transport mechanisms, learned transport connections were compared against Weather Research and

Forecasting (WRF) model wind fields in the Busan metropolitan area (Fig. 5). TransNet's graph construction incorporates wind-driven transport physics through Eq. 3, where edge weights are computed by projecting wind velocity onto inter-station vectors. Effective forecasting in GNNs depends on establishing a meaningful "message passing" structure between nodes. The results demonstrate that TransNet's transport-informed design overcomes the limitations of traditional static distance-based connections by learning physically plausible transport pathways.

The analysis reveals strong correspondence between wind fields and the connections learned by TransNet. Across the Busan domain, where wind vectors exhibit complex patterns including onshore and offshore flows, TransNet identifies transport pathways that are generally aligned with the prevailing wind directions. For example, in coastal areas exhibiting a predominantly eastward flow, TransNet correctly establishes connections between monitoring stations along similar directional corridors. However, the results also reveal a more sophisticated behavior, as some learned pathways deviate from the instantaneous wind patterns. This occurs because TransNet's directional module integrates wind information from the entire preceding 24-h period. By establishing connections based on these historical wind patterns rather than instantaneous wind conditions or geographic proximity alone, TransNet learns to prioritize dominant and recurring transport corridors. This ensures that information propagates along pathways that reflect persistent atmospheric phenomena. The strong overall alignment between the wind fields and the model's connections, combined with these expected deviations, validates that TransNet's design effectively captures both immediate and time-averaged spatial dependencies, maintaining physical consistency in South Korea's complex atmospheric environment.

To demonstrate robustness across diverse meteorological conditions and geographic contexts, the transport pathway analysis was extended to 18 representative periods distributed throughout 2021, spanning winter monsoons, spring transitions, summer convection, and autumn circulation patterns (see Supplementary Note 5). The analysis encompasses three



**Fig. 3 | Spatial distribution of RMSE differences (TransNet - AGATNet) at 170 stations across South Korea.** **a** Shows +1 h lead time, panel **(b)** shows +6 h, panel **(c)** shows +24 h, panel **(d)** shows +48 h, and panel **(e)** shows +72 h. Each station is represented by a filled circle, with colors indicating RMSE difference

in  $\mu\text{g}/\text{m}^3$  according to the vertical color bar. Blue circles indicate lower RMSE for TransNet (negative difference), while red circles indicate lower RMSE for AGATNet (positive difference). White circles indicate negligible difference. The color scale ranges from  $-8 \mu\text{g}/\text{m}^3$  (dark blue) to  $+8 \mu\text{g}/\text{m}^3$  (dark red).

geographically distinct regions: Busan (coastal metropolitan area), Seoul (inland urban center), and Daegu (southeastern basin). Across all analyzed snapshots, learned pathways maintained physical consistency with WRF-forecasted wind fields, confirming that the wind-informed graph construction produces realistic transport structures across varying topographic and atmospheric regimes (Supplementary Figs. 16–69).

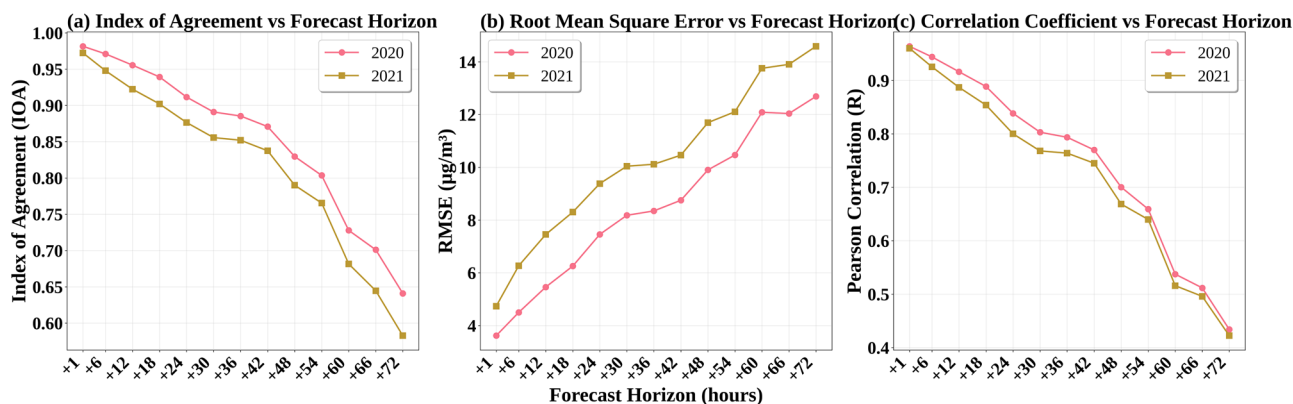
### Operator analysis

To evaluate the individual contributions of TransNet’s operators, analysis across six forecast horizons (+1 h, +6 h, +12 h, +24 h, +48 h, and +72 h) was conducted, systematically isolating advection, diffusion, and reaction components to quantify their relative importance. Figure 6 presents the percentage increase in MAE when specific operators are isolated, with the baseline representing the full TransNet model performance.

During short-term forecasts (+1–12 h), with the reaction-only configuration without state, (no influence of  $R_{\theta_1}$ ; and  $U^{\text{state}(+\frac{2}{3})}$  and  $U^{\text{state}(+\frac{1}{3})}$  is replaced with  $U^{\text{state}(+0)}$  in Eq. (15)) showing the most severe impact, increasing error by 153.5% at +1 h, 244.0% at +6 h, and 194.9% at +12 h. This demonstrates that the advection-diffusion system is essential for capturing the relevant dynamics governing pollutant transport during the early forecast period. When isolated, both advection-only and diffusion-only configurations shows similar error patterns, with increases of 72.3% and 73.1% respectively at +1 h, 111.9% and 111.8% at +6 h, and 82.4% and 82.6% at +12 h. The comparable performance degradation when isolating advection or diffusion operators aligns with established findings that operator splitting between transport processes introduces systematic errors<sup>47,48</sup>, indicating that neither alone can adequately represent complete

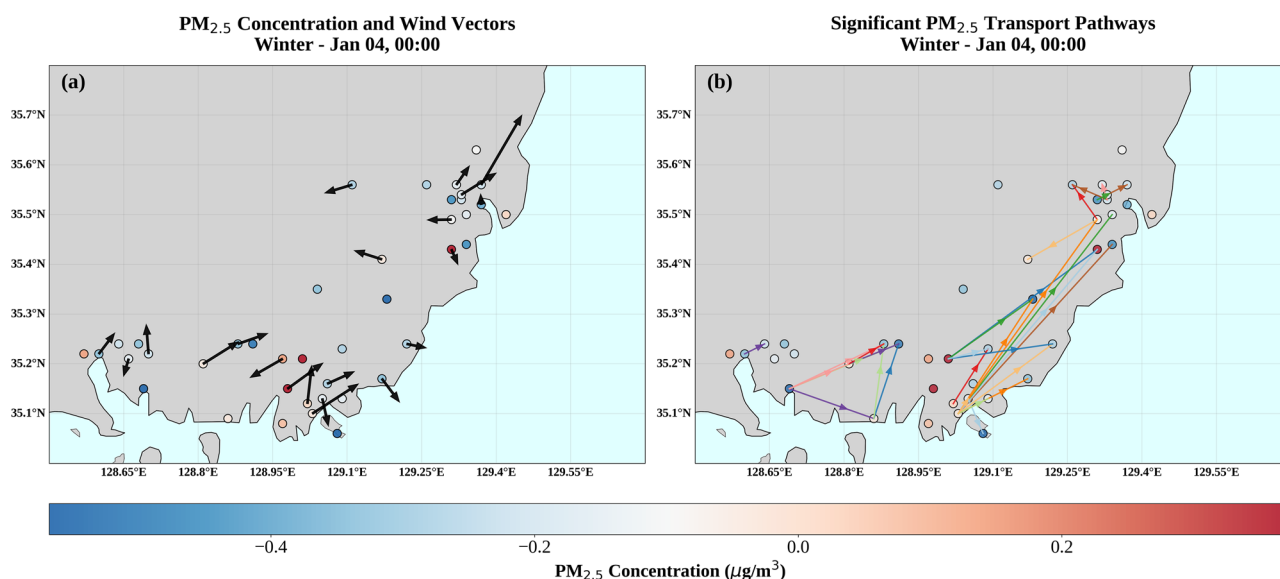
atmospheric transport system. Their combined action determines pollutant evolution along with the reaction system. The reaction-only configuration with state (influence of  $R_{\theta_1}$ ; and  $U^{\text{state}(+\frac{2}{3})}$  is replaced with  $U^{\text{state}(+0)}$  in Eq. (15)) shows more moderate performance degradation (13.5% at +1 h, 31.7% at +6 h, 37.7% at +12 h), indicating that while the reaction operator can capture temporal patterns through state integration, it cannot replace the spatial transport mechanisms provided by advection-diffusion coupling. The substantial difference between reaction-only with state (13.5% increase) versus without state (153.5% increase) at +1 h emphasizes the critical importance of the transport-evolved pollutant state in maintaining physical consistency.

At medium-term horizons (24–48 h), the relative importance of individual operators converges, with advection and diffusion showing comparable contributions (49.6% vs 49.7% at +24 h, 17.4% vs 38.0% at +48 h), while the impact of removing meteorological inputs becomes minimal (0.7% at +24 h, effectively 0% at +48 h). The temporal evolution of operator importance across forecast horizons aligns with the physical understanding of atmospheric pollutant transport mechanisms. At short-term horizons (+1–12 h), spatial transport processes dominate pollutant evolution, as evidenced by severe performance degradation when the advection-diffusion system is excluded. At medium-term horizons (24–48 h), operator contributions converge with advection and diffusion showing nearly identical error increases (49.7% and 49.6% at +24 h), while at extended horizons (+72 h), all operators stabilize at modest levels (11.5–29.8%), indicating that long-term forecast skill becomes increasingly dependent on the integrated operator framework rather than individual transport mechanisms. This temporal progression reflects the limitations of



**Fig. 4 | TransNet’s forecasting performance at 170 stations across South Korea for 2020 and 2021.** a Shows Index of Agreement (IOA) versus forecast horizon, panel (b) shows Root Mean Square Error (RMSE;  $\mu\text{g}/\text{m}^3$ ) versus forecast horizon, and panel (c) shows Pearson correlation coefficient (R) versus forecast horizon. Pink circles connected by solid lines represent 2020 performance, while yellow-gold squares connected by solid lines represent 2021 performance. All panels display metrics across 13 lead times: +1 h, +6 h, +12 h, +18 h, +24 h, +30 h, +36 h, +42 h,

+48 h, +54 h, +60 h, +66 h, and +72 h. TransNet maintains consistent performance across both years, with IOA remaining above 0.85 for the first 36 h in 2020 and above 0.80 for the first 30 h in 2021. RMSE stays below  $10 \mu\text{g}/\text{m}^3$  for the first 48 h in 2020 and for up to 36 h in 2021. Pearson correlation coefficient achieves 0.96 at +1 h for both years and remains above 0.70 through +48 h in 2020, demonstrating robust year-to-year generalization.



**Fig. 5 | Validation of TransNet’s learned transport pathways against WRF-forecasted wind fields in Busan metropolitan area.** Comparison of WRF-forecasted wind fields and TransNet’s learned transport pathways in Busan metropolitan area on 4<sup>th</sup> January 00:00 (12:00 am). a displays wind fields as black arrows overlaid on normalized  $\text{PM}_{2.5}$  concentrations shown as filled circles colored according to the horizontal color bar, ranging from  $-0.4$  (dark blue, below-average

concentrations) through 0.0 (white, average) to  $+0.2$  (dark red, above-average concentrations). Panel (b) shows the corresponding transport pathways learned by TransNet, where arrows represent the direction and magnitude of connections originating from each station. Different arrow colors (red, orange, yellow, blue) are used to visually distinguish between different transport pathways for clarity.

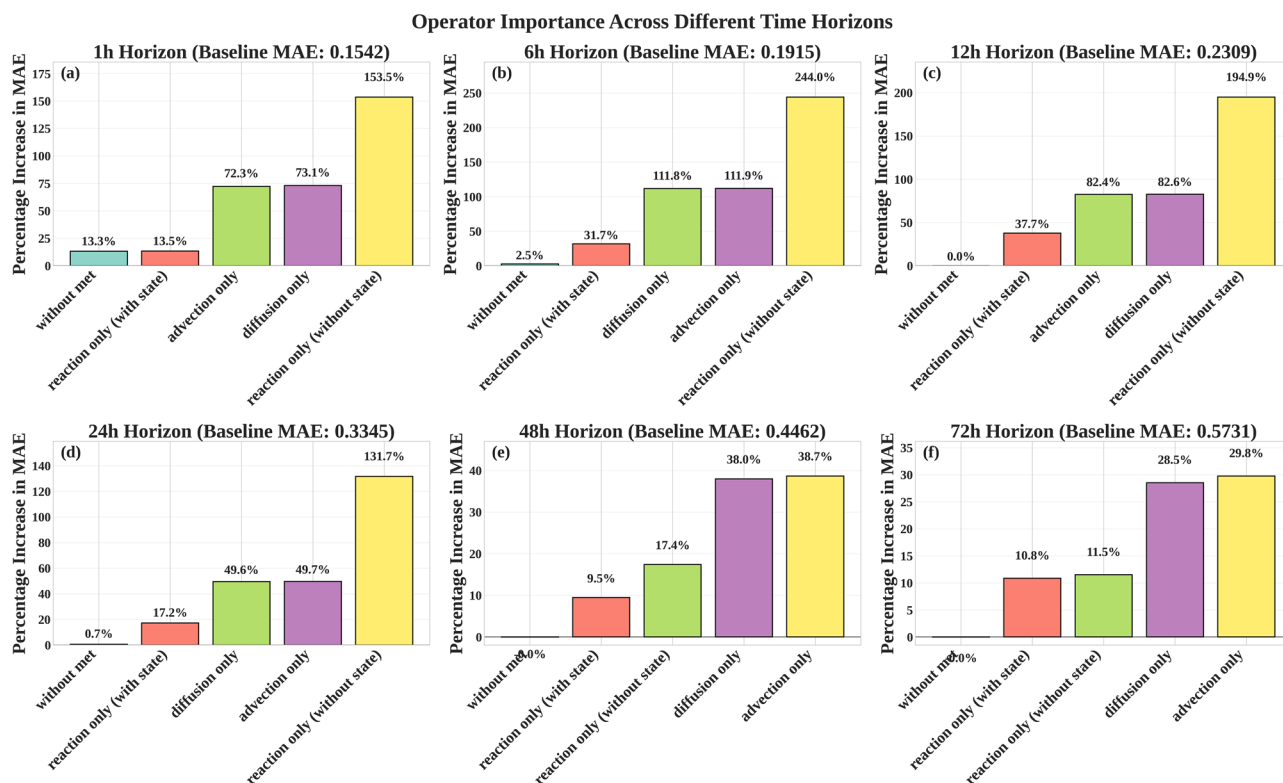
our simplified transport formulation, where reaction processes are realized through neural networks parameterization rather than explicit physical parameterization of turbulent mixing and chemical transformation processes that dominate atmospheric evolution at longer time scales.

### Discussion

This study demonstrates the standalone capability of TransNet, a transport-informed graph neural network that integrates advection, diffusion, and reaction operators into a unified operator-learning framework for  $\text{PM}_{2.5}$  concentrations forecasting. Unlike hybrid methods that rely on CTMs output for generating training inputs for bias corrections, TransNet leverages discretized advection and diffusion equations to learn the transport dynamics directly from the observational data. This independence not only reduces computational costs but also removes the reliance on external

model infrastructures, thereby enhancing its suitability for real-time operational deployment.

A key strength of TransNet lies in its ability to maintain forecast accuracy through +48 h, significantly extending the reliable horizon beyond the typical 6–12-h limitation observed in existing GNN-based approaches. The use of operator splitting ensures that advection, diffusion, and reaction processes are learned as coupled but distinct physical mechanisms, preserving mass balance and constraining error propagation in the short-to-medium range. However, despite these advantages, performance degradation becomes evident beyond +48 h. This limitation reflects the transition from transport-dominated to chemistry-dominated regimes. TransNet’s explicit transport operators effectively control error propagation when these processes dominate, but the graph-based ODE approximation lacks chemical kinetics, secondary aerosol formation, and detailed turbulent mixing



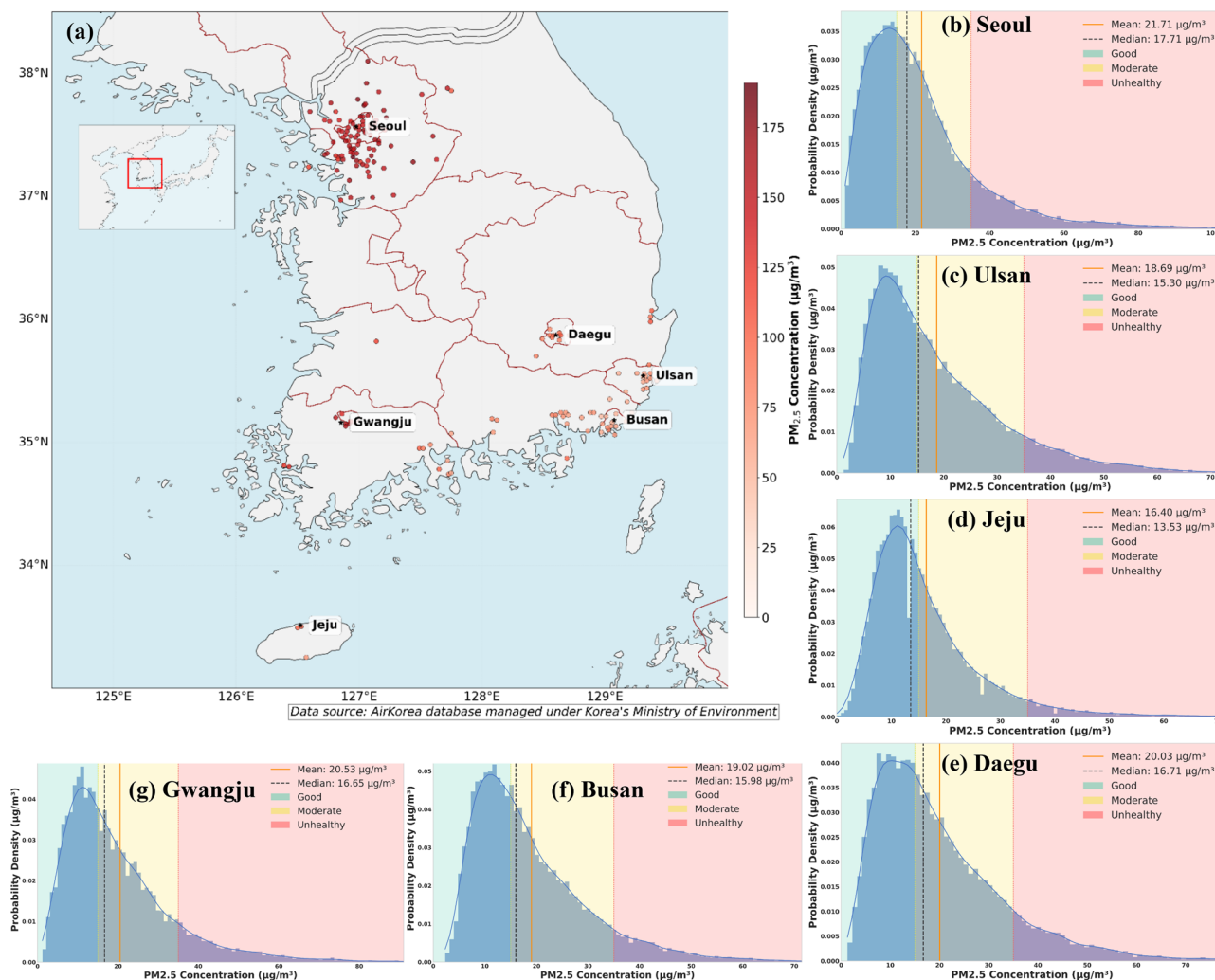
**Fig. 6 | Relative importance of TransNet’s physical operators evaluated through ablation analysis across six forecast horizons.** Percentage increase in MAE when isolating individual TransNet operators across forecast horizons. Each panel shows performance degradation when specific operators are removed, with baseline MAE values indicated in panel titles. Panel (a) shows +1 h horizon (baseline MAE: 0.1542), panel (b) shows +6 h (baseline MAE: 0.1915), panel (c) shows +12 h

(baseline MAE: 0.2309), panel (d) shows +24 h (baseline MAE: 0.3345), panel (e) shows +48 h (baseline MAE: 0.4462), and panel (f) shows +72 h (baseline MAE: 0.5731). Five configurations are tested: without meteorological inputs (yellow bars), reaction only with state (red bars), advection only (green bars), diffusion only (purple bars), and reaction only without state (pale yellow bars). Percentage values above each bar indicate the MAE increase relative to the full model.

that become critical at extended horizons. In contrast, AGATNet’s superior long-range performance stems from: (1) information-rich Hankel matrix embedding providing simultaneous access to 96 h (past (24) + future (72)) of simulated chemical speciation and meteorological forecasts at each forecast time horizon, and (2) architectural stability mechanisms including residual connections and adaptive graph attention that compensate for its absence of explicit transport operators.

While TransNet is directly benchmarked against AGATNet, it is important to note that AGATNet itself demonstrated superior performance over multiple established baselines including Multi-Layer Perceptron, Long Short-Term Memory networks, and PM<sub>2.5</sub>-GNN in comprehensive evaluations across South Korea<sup>30</sup>. AGATNet achieved up to 51.56% RMSE reduction compared to raw CMAQ outputs and demonstrated higher hit rates for severe pollution episodes compared to these data-driven approaches. Since TransNet outperforms AGATNet at short-to-medium lead times while eliminating CTM dependency, it transitively advances beyond these established machine learning baselines. The fundamental distinction lies in TransNet’s operator learning paradigm, which explicitly parameterizes transport physics through learnable advection, diffusion, and reaction operators rather than treating atmospheric processes as emergent features of attention mechanisms and temporal convolutions. This approach aligns with recent advances in neural operators for PDEs, which have demonstrated that learning solution operators improves generalization and physical consistency compared to point-wise mappings<sup>40</sup>. The trade-off manifests at extended horizons, where AGATNet’s access to CMAQ chemical speciation enables better extreme event capture, while TransNet’s simplified reaction parameterization limits performance beyond +48 h. However, for operational early warning systems prioritizing short-term threshold detection, TransNet provides a computationally efficient alternative without sacrificing accuracy.

These findings suggest clear future directions. Incorporating turbulence-resolving operators or stochastic parameterizations could extend temporal stability by better capturing unresolved mixing processes. Similarly, operator-learning modules that embed simplified chemical transformation pathways or emission variability could further enhance the reaction component, providing more realistic long-term pollutant evolution. Importantly, to the best of our knowledge, this study is the first to apply an operator-splitting framework based on advection, diffusion, and reaction operators within a GNN specifically for PM<sub>2.5</sub> forecasting, and the first to combine operator learning with VMD to capture the multi-scale temporal variability of pollutant signals. This dual innovation differentiates TransNet from earlier approaches such as VMD-based deep learning hybrids, which lacked explicit physical operator integration, and operator-inspired GNNs like ADR-GNN, which were not designed for air quality applications. These results indicate complementary operational capabilities. AGATNet’s CTM-derived chemical speciation enables substantially higher severe event detection (84.5% recall at +1 h, maintained at 15.4% through +12 h), making it preferable for applications requiring maximum detection coverage where computational infrastructure is available. TransNet’s physics-informed operators provide better precision (94–100%), computational independence, and superior performance across typical conditions (>96% of observations, MAE: 2.60 vs 4.27 μg/m<sup>3</sup> at +1 h). These attributes make TransNet well-suited for routine operational forecasting and rapid deployment scenarios where warning reliability, computational efficiency, and performance across normal concentration ranges are prioritized. Operational suitability depends on specific application requirements, severe event-focused systems benefit from AGATNet’s detection coverage, while routine forecasting systems benefit from TransNet’s precision and computational independence. Future development should explore integrating simplified chemical operators into TransNet’s framework to improve



**Fig. 7 | Spatial distribution of monitoring stations across South Korea and PM<sub>2.5</sub> concentration probability distributions for six major cities during 2021.** **a** Shows the distribution of 170 monitoring stations across South Korea (33.1°N to 38.6°N, 126.0°E to 129.5°E) with PM<sub>2.5</sub> concentrations displayed as filled circles during the maximum pollution event on March 5th, 2019. Circle colors correspond to the vertical color bar ranging from 0 (white) to 175+ µg/m<sup>3</sup> (dark red). Major cities labeled include Seoul, Gwangju, Daegu, Ulsan, Busan, and Jeju. Gray contour lines indicate provincial boundaries, and an inset map (upper left) shows the regional

context with a red box indicating the study domain. Panels (b–g) show probability density distributions of hourly PM<sub>2.5</sub> concentrations for 2021 for Seoul, Ulsan, Jeju, Daegu, Busan, and Gwangju, respectively. Each distribution is color-coded by air quality categories: blue regions indicate “Good” conditions (0–15 µg/m<sup>3</sup>), yellow regions indicate “Moderate” conditions (16–35 µg/m<sup>3</sup>), and pink regions indicate “Unhealthy” conditions (≥ 36 µg/m<sup>3</sup>). Red solid vertical lines mark the mean concentration, and red dashed vertical lines mark the median concentration, with values displayed in the legend of each panel.

extreme event detection while preserving precision advantages and computational independence.

## Method

### Study domain and data sources

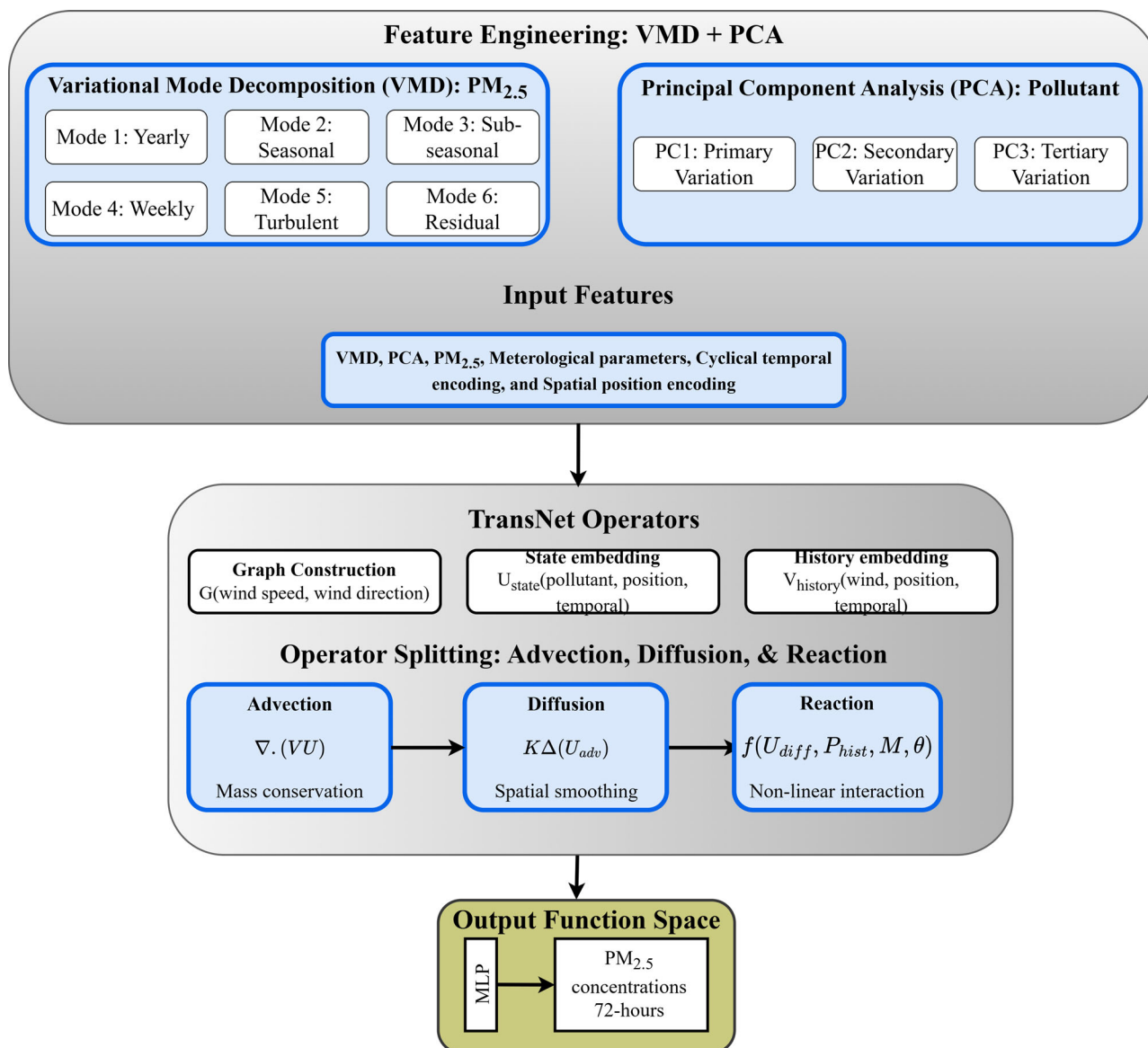
The study domain encompasses 170 air quality monitoring stations across South Korea (see Fig. 7), distributed across major metropolitan regions, including Seoul (25 stations), Gyeonggi province (65 stations), Gyeongnam province (32 stations), and Gyeongbuk province (23 stations). This network provides comprehensive coverage of the country’s most populated areas, spanning from 33.1°N to 38.6°N latitude and 126.0°E to 129.5°E longitude (Fig. 7a). Figure 7b–g further detail the PM<sub>2.5</sub> concentration distributions for six major cities. This domain is particularly challenging for air quality forecasting due to complex topography and frequent transboundary pollution events. During an extreme event on March 5, 2019, for example, PM<sub>2.5</sub> concentrations were recorded as high as 190.94 µg/m<sup>3</sup>, with foreign contributions accounting for 87% of the total<sup>7</sup>.

The dataset for this study spans the years 2018 to 2021. Hourly surface concentrations of six air pollutants, including PM<sub>2.5</sub>, coarse particulate

matter (PM<sub>10</sub>), sulfur dioxide (SO<sub>2</sub>), carbon monoxide (CO), ozone (O<sub>3</sub>), and nitrogen dioxide (NO<sub>2</sub>), were obtained from the AirKorea database, maintained by South Korea’s Ministry of Environment<sup>49</sup>. Corresponding meteorological variables were obtained from the WRF model version 3.8 at a 27 km spatial PBL, TEMP2, Q2, RN, and RC resolution, following the same modeling configuration established in the previous study<sup>13,30</sup>. Seven (7) WRF variables were used which include planetary boundary layer (PBL) height, temperature at 2 m (TEMP2), water vapor mixing ratio at 2 m (Q2), wind speed at 10 m (WSPD10), wind direction at 10 m (WDIR10), non-convective precipitation (RN), and convective precipitation (RC). For each station, the meteorological fields were extracted and matched to the nearest coordinate as done in the previous study<sup>30,43</sup>.

### Data pre-processing and feature engineering

Data quality is ensured by selecting 170 stations from the original network, excluding any station with more than 10% of missing data during the observation period. Remaining gaps in the time-series data were imputed using linear interpolation for short-term gaps and cubic interpolation for longer gaps to preserve temporal continuity.



**Fig. 8 | TransNet architecture integrating feature engineering, transport operators, and operator splitting for 72-h PM<sub>2.5</sub> forecasting.** TransNet framework showing the complete modeling pipeline from feature engineering to forecast generation. The top section shows feature engineering through Variational Mode Decomposition (VMD) applied to PM<sub>2.5</sub> time series, producing six intrinsic mode functions (Mode 1: Yearly, Mode 2: Seasonal, Mode 3: Sub-seasonal, Mode 4: Weekly, Mode 5: Turbulent, Mode 6: Residual), and Principal Component Analysis (PCA) applied to pollutant concentrations (6 pollutants coming from in-situ stations), producing three principal components (PC1: Primary Variation, PC2: Secondary Variation, PC3: Tertiary Variation). These processed features combine with meteorological parameters, cyclical temporal encoding, and spatial position

encoding to form the input feature set shown in the blue box. The middle section displays TransNet operators including graph construction based on wind speed and wind direction, state embedding containing pollutant, position, and temporal information, and history embedding containing wind, position, and temporal information. The core operator splitting framework implements three sequential physical processes shown in blue boxes: advection ( $\nabla \cdot (VU)$ ) for mass conservation, diffusion ( $K\Delta(U^{adv})$ ) for spatial smoothing, and reaction ( $f(U^{diff}, P_{hist}, M, \theta)$ ) for non-linear interactions. The bottom section shows the output function space where a multilayer perceptron (MLP) projects the evolved features to generate 72-h PM<sub>2.5</sub> concentration forecasts.

A set of 43 features was constructed from the raw data (Fig. 8). To address the multi-scale and non-stationary characteristics of the PM<sub>2.5</sub> time-series, Variational Model Decomposition (VMD) was applied, decomposing the signal into six intrinsic mode functions (IMFs) that represent different frequency components (see Supplementary Note 1). To reduce the dimensionality and multicollinearity of the six observed pollutants, a Principal Component Analysis (PCA) was performed, retraining three components that capture approximately 85% of the total variance (see Supplementary Note 2). The final feature set comprises: pollutant features (10 variables, including VMD modes and PCA components; see Supplementary Figure 1 and Supplementary Fig. 2); wind components (U,

V, U<sub>cos</sub>, V<sub>sin</sub>); temporal encodings (8 sinusoidal features for hour, month, weekday, and day-of-year; see Supplementary Note 3); positional embeddings (16 features; see Supplementary Note 3); and WRF-forecasted meteorological variables (planetary boundary layer height, 2 m temperature, 2 m specific humidity, convective and non-convective precipitations). All features were normalized using Standard scaling and the data was structured into 24-h sliding windows to capture diurnal patterns and transport cycles. This pre-processing ensures that TransNet receives physically relevant input features which capture the different cycles/frequencies of the PM<sub>2.5</sub> pollutants which are essential to the physically relevant processes.

### TransNet model architecture

TransNet is built upon the Advection-Diffusion-Reaction Graph Neural Network (ADR-GNN) architecture proposed by Eliasof et al.<sup>24</sup>, extended for operational PM<sub>2.5</sub> forecasting across South Korea’s 170 monitoring stations. Our implementation incorporates domain-specific feature engineering, wind-informed graph construction, and transport physics tailored for air quality applications. The architecture processes multi-dimensional spatio-temporal features through four main components: (1) feature-specific embedding networks that transform heterogeneous input data into compatible latent representations, (2) graph construction based on real-time meteorological conditions, (3) transport physics operators implementing advection-diffusion-reaction processes, and (4) a forecast network that maps evolved features to 72-h forecasts of PM<sub>2.5</sub> concentrations (see Fig. 8).

The model processes 43-dimensional spatiotemporal features through specialized embedding networks. The pollutant state vector comprises ten PM<sub>2.5</sub>-related features, including six modal concentrations from VMD (PM<sub>2.5</sub><sup>(0)</sup> through PM<sub>2.5</sub><sup>(5)</sup>), three PCA-reduced pollutant representations, and total PM<sub>2.5</sub> concentrations. Wind history is encoded using four-dimensional wind vectors (zonal, meridional, cosine and sine components) across 24 temporal steps. Eight cyclical temporal features encode diurnal, monthly, weekly, and annual patterns across the 24-h input window, while 16-dimensional positional encodings provide geographic context. These features are processed by specialized embedding networks tailored to their temporal characteristics; wind and other meteorological features are flattened across the 24-h input window, resulting in 96 and 120 dimensions, respectively, while temporal and positional features maintain their full temporal resolution. These two embeddings capture different aspects of the atmospheric conditions. Pollutant state embedding represents the instantaneous atmospheric states by combining current PM<sub>2.5</sub> concentrations with their spatiotemporal context. In contrast, the wind history encoding integrates transport conditions across the full input sequence.

TransNet constructs a graph that adapts to meteorological conditions through a physics-based edge weighting scheme. The module first recovers interpretable wind parameters from the decomposed wind components<sup>50</sup>:

$$wind\ speed = \sqrt{U^2 + V^2} \tag{1}$$

$$wind\ direction = 270^\circ - atan2(V, U) \times \frac{180^\circ}{\pi} \tag{2}$$

The wind direction conversion follows standard atmospheric modeling practices using the 270° offset variant of the meteorological conversion to ensure compatibility with the WRF model output<sup>51</sup>. Edge weights are then computed by projecting wind velocity onto inter-station vectors, scaled by inverse distance:

$$w_{ij}(t) = ReLU\left(\frac{wind\ speed_i(t) \cdot \cos(\theta_{ij} - wind\ direction_i(t))}{d_{ij}}\right) \tag{3}$$

where  $\theta_{ij}$  represents the geographical bearing from station  $i$  to station  $j$ ,  $wind\ direction_i(t)$  is the meteorological wind direction at the source station, and  $d_{ij}$  is the Euclidean distance between stations. This formulation ensures that edge weights are maximized when wind direction aligns with the inter-station vector, effectively modeling wind-driven pollutant transport while using a Rectified Linear Unit (ReLU) activation that inherently handles opposing wind directions.

### TransNet physics implementation

The transport physics module implements three learnable operators that approximate the fundamental processes governing atmospheric pollutant

evolution through graph-based discretization.

$$\frac{\partial U}{\partial t} = \nabla \cdot (VU) + K\Delta U + f(U, X, \theta_r) \tag{4}$$

where  $\nabla \cdot (VU)$  represents advection,  $K\Delta U$  represents diffusion, and  $f(U, X, \theta_r)$  represents reaction. The framework processes VMD-decomposed PM<sub>2.5</sub> modes and PCA-reduced pollutant features alongside wind parameters, meteorological parameters, temporal encoding, and spatial position encoding to create physics-informed embeddings that drive the transport operators.

The framework employs two specialized embeddings that capture distinct physical aspects of atmospheric transport. The state embedding  $U^{state}$  combines pollutant concentrations (VMD modes, PCA components, and raw PM<sub>2.5</sub> concentrations), spatial position encoding, and temporal encoding to represent the instantaneous atmospheric pollutant distribution. The history embedding  $V^{history}$  encodes 24-h wind evolution (U-wind, V-wind, and wind direction components), spatial encodings, and temporal encodings to represent the velocity field that drives atmospheric transport. The graph construction then leverages the current wind speed and direction to establish a time-varying connectivity that reflects the prevailing atmospheric transport pathways.

The graph advection operator discretizes the continuous advection equation onto the monitoring station network, where each station represents a node and wind-driven transport pathways define dynamic edges. For each station  $i$ , the operator computes the net advective flux by balancing inbound transport from neighboring stations against outbound transport to connected neighbors:

$$\nabla_i(VU) = \sum_{j \in \mathcal{N}_i} V_{j \rightarrow i}^{history} \odot U_j^{state} - U_i^{state} \odot \sum_{j \in \mathcal{N}_i} V_{i \rightarrow j}^{history} \tag{5}$$

$$\nabla_i(VU) = \sum_{j \in \mathcal{N}_i} V_{j \rightarrow i}^{history} \odot U_j^{state} - U_i^{state}, \text{ where } \sum_{j \in \mathcal{N}_i} V_{i \rightarrow j}^{history} = 1 \tag{6}$$

where  $V_{j \rightarrow i}^{history}$  represents the learned directional weight quantifying transport from station  $j$  to station  $i$  based on the previous 24-h wind,  $\mathcal{N}_i$  represents the neighbors of station  $i$ , and  $\odot$  denotes element-wise multiplication. The inbound term  $\sum_{j \in \mathcal{N}_i} V_{j \rightarrow i}^{history} \odot U_j^{state}$  aggregates pollutant mass arriving from upwind stations, weighted by learned transport efficiencies. The outbound term  $U_i^{state} \odot \sum_{j \in \mathcal{N}_i} V_{i \rightarrow j}^{history}$  represents pollutant mass leaving the station  $i$  toward downwind neighbors. The directional weights  $V_{i \rightarrow j}$  and  $V_{j \rightarrow i}$  are derived from  $V^{history}$  through a learnable asymmetric mapping that ensures mass conservation while capturing wind-driven transport preferences. This asymmetry ( $V_{i \rightarrow j}$  not equal to  $V_{j \rightarrow i}$ ) is crucial for representing realistic atmospheric transport, where prevailing winds create preferential directional flows between station pairs. The constraint  $\sum_{j \in \mathcal{N}_i} V_{i \rightarrow j} = 1$  ensures that each station can transfer at most its total pollutant mass to neighboring locations, maintaining physical consistency with mass conservation principles. The graph advection operator is now defined as:

$$U_i^{state(+\frac{1}{3})} = U_i^{state(0)} + h \nabla_i(V^{history} \cdot U^{state}) \\ = U_i^{state(0)} + h(\sum_{j \in \mathcal{N}_i} V_{j \rightarrow i}^{history(0)} \odot U_j^{state} - U_i^{state}) \tag{7}$$

where  $h$  represents the temporal step size and the superscript  $+\frac{1}{3}$  indicates the intermediate state following the operator splitting approach.

The directional weights  $V_{i \rightarrow j}$  and  $V_{j \rightarrow i}$  are computed from the historical wind features ( $V^{history}$ ) through a sequence of learnable transformations (see Supplementary Fig. 70). For each edge connection stations  $i$  and  $j$ , the algorithm first compute directional affinity scores by processing the historical wind features of both endpoints through the learned

transformations:

$$Z_{ij} = A_3 \cdot (\text{ReLU}(V_i^{\text{history}} \cdot A_1 + V_j^{\text{history}} \cdot A_2)) \quad (8)$$

$$Z_{ji} = A_3 \cdot (\text{ReLU}(V_j^{\text{history}} \cdot A_1 + V_i^{\text{history}} \cdot A_2)) \quad (9)$$

where  $A_1$  and  $A_2$  are linear transformation that map the wind features from feature dimension to hidden dimension with layer normalization to maintain the stability of the model, and  $A_3$  projects the learnable hidden representation to a scalar value. The transformation matrices  $A_1$  and  $A_2$  learn to extract relevant wind patterns from each station's 24-h wind data, while their different parameterizations allow the model to distinguish between source and target characteristics. The computation of  $Z_{ij}$  swaps the roles of stations  $i$  and  $j$  in the transformation, creating inherently different representations for each direction based on which station serves as the transport origin versus destination. The directional weights are then derived by computing opposing difference terms processed through a scaling transformation as:

$$V_{i \rightarrow j} = \text{ReLU}(A_4(Z_{ij} - Z_{ji})) \quad (10)$$

$$V_{j \rightarrow i} = \text{ReLU}(A_4(-Z_{ij} + Z_{ji})) \quad (11)$$

where  $A_4$  is the learnable scaling factor. Since  $(Z_{ij} - Z_{ji})$  and  $(-Z_{ij} + Z_{ji})$  have opposite signs, the ReLU activation ensures that for each edge pair, only one direction will be non-zero while the other becomes zero, as ReLU output ranges from 0 to infinity. The learnable scaling matrix ( $A_4$ ) allows the model to calibrate the magnitude of directional preferences learned from the wind patterns. This mechanism creates the essential asymmetric directionality required for realistic transport, where wind patterns establish preferential flow directions between station pairs rather than symmetric bidirectional exchange. Softmax normalization is applied per source node to enforce mass conservation. This normalization ensures that  $\sum_{j \in \mathcal{N}_i} V_{i \rightarrow j} = 1$ , maintaining physical consistency with mass conservation principles.

The graph diffusion operator models spatial smoothing of pollutant concentrations using the normalized graph Laplacian derived from wind-based connectivity. Unlike advection which uses learned directional weights, diffusion operates on the same wind-derived graph structure but processes it through the normalized Laplacian ( $\hat{L}$ ) to achieve symmetric spatial smoothing. This transformation converts the directional wind-based connectivity into a symmetric diffusion matrix that captures spatial diffusivity between neighboring stations. The operator employs learned feature-wise diffusion coefficients  $K = [\kappa_1, \kappa_2, \dots, \kappa_c]$  where each  $\kappa_i \geq 0$  independently controls diffusion strength for feature channel  $i$ . These coefficients are learned parameters constrained through:

$$K = \text{hardtanh}(\theta_d, 0, 1) \quad (12)$$

ensuring non-negative values that prevent unphysical anti-diffusive behavior. The diffusion computation implements the discretized diffusion equation:

$$U^{\text{state}(+\frac{2}{3})} = \text{mat}((I + hK \otimes \hat{L})^{-1} \text{vec}(U^{\text{state}(+\frac{1}{3})})) \quad (13)$$

This is solved as an implicit linear system for numerical stability. For each batch  $b$  and feature channel  $c$ , the graph diffusion operator is defined by the equation:

$$M = I + h \cdot \kappa_c \cdot \hat{L} + \varepsilon I \quad (14)$$

$$U^{\text{state}(+\frac{2}{3})}[b, c] = M^{-1} U^{\text{state}(+\frac{1}{3})}[b, c]$$

where  $I$  is the identity matrix,  $h = 0.01$  is the temporal step size,  $\kappa_c$  is the learned diffusion coefficient for channel  $c$ , and  $\varepsilon = 1e-6$  provides numerical

regularization to ensure matrix invertibility. The system is solved independently for each batch  $b$  and feature channel  $c$ , with robust fallback mechanisms including least squares (`torch.linalg.lstsq`) and pseudoinverse solutions (`torch.linalg.pinv`) for ill-conditioned cases.

The reaction operator represents pointwise non-linear transformations that modify pollutant concentrations locally at each station, capturing processes that operate independently of spatial transport. In the context of  $\text{PM}_{2.5}$  forecasting without explicit emission or chemistry data, the reaction operator learns to capture local temporal patterns, meteorological influences, and non-linear state dependencies from the available observational data. The reaction operator processes four distinct input components through dedicated neural networks. The current pollutant state  $U^{\text{state}(+\frac{2}{3})}$  from the diffusion step provides the baseline concentrations for transformation. The 24-h pollutant history  $U^{\text{history}}$  captures temporal persistence and cyclical patterns in pollutant evolution. The meteorological variables include planetary boundary layer height, surface temperature, specific humidity, and convective and non-convective precipitation rates that influence local processes. The position and temporal encodings provide spatial and temporal context for location-specific behaviors. The reaction function combines additive and multiplicative transformations to capture complex non-linear interactions<sup>52,53</sup>:

$$f(U^{\text{state}(+\frac{2}{3})}, U^{\text{history}}, \text{Met}; \theta_r) = \text{ReLU}(R_{\theta_1}(U^{\text{state}(+\frac{2}{3})}) + \tanh(R_{\theta_2}(U^{\text{state}(+\frac{2}{3})})) \odot U^{\text{state}(+\frac{2}{3})} + R_{\theta_3}(U^{\text{history}}) + R_{\theta_4}(\text{Met})) \quad (15)$$

where  $R_{\theta_1}$ ,  $R_{\theta_2}$ ,  $R_{\theta_3}$ , and  $R_{\theta_4}$  represent specialized MLPs with learnable parameters  $\theta_r = \{R_{\theta_1}, R_{\theta_2}, R_{\theta_3}, R_{\theta_4}\}$ , and  $\odot$  denotes element-wise multiplication (see Supplementary Figure 71). The  $R_{\theta_1}$  network processes the current state through linear transformations with layer normalization, learning direct additive corrections. The  $R_{\theta_2}$  network computes multiplicative gating weights  $\tanh(R_{\theta_2}(U^{\text{state}(+\frac{2}{3})}))$  that modulate the current state, enabling adaptive feature scaling. The  $R_{\theta_3}$  network processes the pollutant history to capture temporal dependencies, while  $R_{\theta_4}$  integrates meteorological influences on local pollutant transformation processes. The reaction operator updates the pollutant state through forward Euler integrations:

$$U^{\text{state}(+1)} = U^{\text{state}(+\frac{2}{3})} + h \cdot f(U^{\text{state}(+\frac{2}{3})}, U^{\text{history}}, \text{Met}; \theta_r) \quad (16)$$

where  $h$  is temporal step size ( $h = 0.01$ ) controlling the magnitude of reaction-driven changes. The additive formulation preserves the base pollutant state while incorporating learned modifications, ensuring numerical stability throughout the transformation process.

### Model training, hyperparameters, and evaluation

TransNet was implemented in PyTorch and used distributed data parallel processing across 4 NVIDIA V100 GPUs to accelerate training. The temporal data spanning 2018–2021 was partitioned into training (2018–2019; 17,520 hourly samples across two annual cycles covering all four seasons, around 50% of the dataset), validation (2020; 8760 samples, around 25% of the dataset), and test (2021; 8760 samples, around 25% of the dataset) sets, following a 2:1:1 split ratio. This chronological split prevents data leakage and ensures a robust evaluation on unseen future data. The model was trained using the Adamax optimizer<sup>54</sup> with an initial learning rate of 0.01, a weight decay of 0.0005, and a batch size of 32. To ensure stable convergence and prevent overfitting, a series of standard techniques was employed. The learning rate was adaptively reduced using ReduceLROnPlateau<sup>55</sup> scheduling (patience=3 epochs, reduction factor=0.2, minimum learning rate= $1 \times 10^{-6}$ ). Gradient clipping was applied with a maximum norm of 1.0 to stabilize training dynamics. Early stopping was implemented with a patience of 5 epochs and a minimum delta of  $1 \times 10^{-4}$  to prevent overfitting.

The negative IOA was used as the loss function, with its negative value minimized during optimization (so minimum for IOA is  $-1$ ) to maximize prediction-observation agreement. The model was trained for a maximum of 100 epochs, with checkpoints automatically saved every 100 batches to ensure training continuity.

The model's forecasting accuracy was evaluated on the year 2021 test dataset across all 170 monitoring stations. The performance was assessed at forecast lead times ranging from 1 to 72 h using four complementary metrics: MAE, RMSE, IOA, and  $R^2$ . MAE provides an interpretable error magnitude the original unit  $\mu\text{g}/\text{m}^3$ ; RMSE emphasizes the impact of larger prediction errors; IOA offers normalized performance assessment independent of concentration ranges; and  $R^2$  quantifies the proportion of variance in the observed  $\text{PM}_{2.5}$  concentrations explained by model predictions. The metrics are defined as follows:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |P_i - O_i| \quad (17)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (P_i - O_i)^2} \quad (18)$$

$$\text{IOA} = 1 - \frac{\sum_{i=1}^n (O_i - P_i)^2}{\sum_{i=1}^n (|P_i - \bar{O}| + |O_i - \bar{O}|)^2} \quad (19)$$

$$R^2 = \left( \frac{\sum_{i=1}^n (O_i - \bar{O})(P_i - \bar{P})}{\sqrt{\sum_{i=1}^n (O_i - \bar{O})^2 \sum_{i=1}^n (P_i - \bar{P})^2}} \right)^2 \quad (20)$$

where  $P_i$  and  $O_i$  represent predicted and observed  $\text{PM}_{2.5}$  concentrations at time step  $i$ ,  $\bar{P}$  and  $\bar{O}$  denote their respective means, and  $n$  is the total number of predictions. An  $R^2$  closer to 1 indicates that the model explains a greater proportion of the variance in the observed  $\text{PM}_{2.5}$  concentrations.

### Data availability

The datasets generated and analyzed during the current study are available in the <https://doi.org/10.5281/zenodo.10963116>.

### Code availability

The underlying code for this study is available in TransNet repository and can be accessed via this link <https://github.com/rijul01/TransNet>.

Received: 31 October 2025; Accepted: 10 January 2026;

Published online: 26 February 2026

### References

- South Korea Population. *Worldometer* <https://www.worldometers.info/world-population/south-korea-population/> (2025).
- Choi, H. & Myong, J.-P. Association between air pollution in the 2015 winter in South Korea and population size, car emissions, industrial activity, and fossil-fuel power plants: an ecological study. *Ann. Occup. Environ. Med.* **30**, 60 (2018).
- Oak, Y. J. et al. Air quality trends and regimes in South Korea inferred from 2015–2023 surface and satellite observations. *Atmospheric Chem. Phys.* **25**, 3233–3252 (2025).
- Chun, Y. & Kim, K. Temporal changes in the urban system in South Korea. *Front. Sustain. Cities* **4**, 1013465 (2022).
- Lin, Y. et al. Formation, radiative forcing, and climatic effects of severe regional haze. *Atmospheric Chem. Phys.* **22**, 4951–4967 (2022).
- Hao, H., Wang, K., Wu, G., Liu, J. & Li, J.  $\text{PM}_{2.5}$  concentrations based on near-surface visibility in the Northern Hemisphere from 1959 to 2022. *Earth Syst. Sci. Data* **16**, 4051–4076 (2024).
- Lee, D. et al. Analysis of a severe  $\text{PM}_{2.5}$  episode in the Seoul metropolitan area in South Korea from 27 February to 7 March 2019: focused on estimation of domestic and foreign contribution. *Atmosphere* **10**, 756 (2019).
- Park, M.-J. et al. Inter-annual changes in transboundary air quality from KORUS-AQ 2016 to SIJAQ 2022 campaign periods and assessment of emission reduction strategies in Northeast Asia. *Environ. Pollut.* **363**, 125114 (2024).
- Kim, N. R. & Lee, H. J. Ambient  $\text{PM}_{2.5}$  exposure and rapid population aging: a double threat to public health in the Republic of Korea. *Environ. Res.* **252**, 119032 (2024).
- Feng, Y. et al. Long-term exposure to ambient  $\text{PM}_{2.5}$ , particulate constituents and hospital admissions from non-respiratory infection. *Nat. Commun.* **15**, 1518 (2024).
- Sagheer, U. et al. Environmental pollution and cardiovascular disease. *JACC Adv.* **3**, 100805 (2024).
- Kosanpipat, B. et al. Impact of  $\text{PM}_{2.5}$  exposure on mortality and tumor recurrence in resectable non-small cell lung carcinoma. *Sci. Rep.* **14**, 24660 (2024).
- Park, J., Jung, J., Choi, Y., Mousavinezhad, S. & Pouyaei, A. The sensitivities of ozone and  $\text{PM}_{2.5}$  concentrations to the satellite-derived leaf area index over East Asia and its neighboring seas in the WRF-CMAQ modeling system. *Environ. Pollut.* **306**, 119419 (2022).
- Park, J. et al. Satellite-based, top-down approach for the adjustment of aerosol precursor emissions over East Asia: the TROPOspheric Monitoring Instrument (TROPOMI)  $\text{NO}_2$  product and the Geostationary Environment Monitoring Spectrometer (GEMS) aerosol optical depth (AOD) data fusion product and its proxy. *Atmosp. Meas. Tech.* **16**, 3039–3057 (2023).
- Pouyaei, A. et al. Investigating the long-range transport of particulate matter in East Asia: introducing a new Lagrangian diagnostic tool. *Atmos. Environ.* **278**, 119096 (2022).
- Zhang, Y., Bocquet, M., Mallet, V., Seigneur, C. & Baklanov, A. Real-time air quality forecasting, part II: state of the science, current research needs, and future prospects. *Atmos. Environ.* **60**, 656–676 (2012).
- Madsen, H. *Time Series Analysis* (Chapman and Hall/CRC, New York, 2007). <https://doi.org/10.1201/9781420059687>.
- Ng, K. Y. & Awang, N. Multiple linear regression and regression with time series error models in forecasting  $\text{PM}_{10}$  concentrations in Peninsular Malaysia. *Environ. Monit. Assess.* **190**, 63 (2018).
- Yang, W., Deng, M., Xu, F. & Wang, H. Prediction of hourly  $\text{PM}_{2.5}$  using a space-time support vector regression model. *Atmos. Environ.* **181**, 12–19 (2018).
- Qiu, M., Zigler, C. & Selin, N. E. Statistical and machine learning methods for evaluating trends in air quality under changing meteorological conditions. *Atmos. Chem. Phys.* **22**, 10551–10566 (2022).
- Abuouelezz, W. et al. Exploring  $\text{PM}_{2.5}$  and  $\text{PM}_{10}$  ML forecasting models: a comparative study in the UAE. *Sci. Rep.* **15**, 9797 (2025).
- Bedi, S., Tiwari, K., Kota, S. H. & Krishnan, N. A. A neural operator for forecasting carbon monoxide evolution in cities. *Npj Clean Air* **1**, 2 (2025).
- Zeng, T. et al. A hybrid optimization prediction model for  $\text{PM}_{2.5}$  based on VMD and deep learning. *Atmospheric Pollut. Res.* **15**, 102152 (2024).
- Tian, J. et al. Air quality prediction with physics-guided dual neural ODEs in open systems. <https://doi.org/10.48550/arXiv.2410.19892> (2025).
- Hettige, K. H. et al. AirPhyNet: Harnessing Physics-Guided Neural Networks for Air Quality Prediction. <https://doi.org/10.48550/arXiv.2402.03784> (2024).
- Chen, Q. et al. An adaptive adjacency matrix-based graph convolutional recurrent network for air quality prediction. *Sci. Rep.* **14**, 4408 (2024).
- Benson, V. et al. Atmospheric transport modeling of  $\text{CO}_2$  with neural networks. *J. Adv. Model. Earth Syst.* **17**, e2024MS004655 (2025).
- Chen, H. et al. Visibility forecast in Jiangsu province based on the GCN-GRU model. *Sci. Rep.* **14**, 12599 (2024).

29. Teng, M. et al. 72-hour real-time forecasting of ambient PM<sub>2.5</sub> by hybrid graph deep neural network with aggregated neighborhood spatiotemporal information. *Environ. Int.* **176**, 107971 (2023).
30. Dimri, R., Choi, Y., Salman, A. K., Park, J. & Singh, D. AGATNet: an adaptive graph attention network for bias correction of CMAQ-forecasted PM<sub>2.5</sub> concentrations over South Korea. *J. Geophys. Res. Mach. Learn. Comput.* **1**, e2024JH000244 (2024).
31. Wang, X. et al. Air quality forecasting using a spatiotemporal hybrid deep learning model based on VMD–GAT–BiLSTM. *Sci. Rep.* **14**, 17841 (2024).
32. Su, I.-F., Chung, Y.-C., Lee, C. & Huang, P.-M. Effective PM<sub>2.5</sub> concentration forecasting based on multiple spatial–temporal GNN for areas without monitoring stations. *Expert Syst. Appl.* **234**, 121074 (2023).
33. Wang, S. et al. PM<sub>2.5</sub>-GNN: A Domain Knowledge Enhanced Graph Neural Network For PM<sub>2.5</sub> Forecasting. in *Proceedings of the 28th International Conference on Advances in Geographic Information Systems* 163–166 (ACM, Seattle WA USA, 2020). <https://doi.org/10.1145/3397536.3422208>.
34. Zhou, X., Wang, J., Wang, J. & Guan, Q. Predicting air quality using a multi-scale spatiotemporal graph attention network. *Inf. Sci.* **680**, 121072 (2024).
35. Kim, D.-Y., Jin, D.-Y. & Suk, H.-I. Spatiotemporal graph neural networks for predicting mid-to-long-term PM<sub>2.5</sub> concentrations. *J. Clean. Prod.* **425**, 138880 (2023).
36. Zhao, G., He, H., Huang, Y. & Ren, J. Near-surface PM<sub>2.5</sub> prediction combining the complex network characterization and graph convolution neural network. *Neural Comput. Appl.* **33**, 17081–17101 (2021).
37. Teng, M. et al. A new hybrid deep neural network for multiple sites PM<sub>2.5</sub> forecasting. *Journal of Cleaner Production* **473**, 143542 (2024).
38. Krishnapriyan, A. S., Gholami, A., Zhe, S., Kirby, R. M. & Mahoney, M. W. Characterizing possible failure modes in physics-informed neural networks. *Adv. Neural Inf. Process. Syst.* **34**, 26548–26560 (2021).
39. 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).
40. Kovachki, N. et al. Neural Operator: Learning Maps Between Function Spaces. <https://doi.org/10.5555/3648699.3648788> (2024).
41. Eliasof, M., Haber, E. & Treister, E. P. D. E.-G. C. N.: Novel Architectures for Graph Neural Networks Motivated by Partial Differential Equations. <https://doi.org/10.48550/arXiv.2108.01938> (2021).
42. Eliasof, M., Haber, E. & Treister, E. Feature Transportation Improves Graph Neural Networks. <https://doi.org/10.48550/arXiv.2307.16092> (2023).
43. Singh, D. et al. Deep-BCSI: A deep learning-based framework for bias correction and spatial imputation of PM<sub>2.5</sub> concentrations in South Korea. *Atmospheric Res.* **301**, 107283 (2024).
44. Sayeed, A., Eslami, E., Lops, Y. & Choi, Y. CMAQ-CNN: a new-generation of post-processing techniques for chemical transport models using deep neural networks. *Atmos. Environ.* **273**, 118961 (2022).
45. Wang, S., Teng, Y. & Perdikaris, P. Understanding and mitigating gradient flow pathologies in physics-informed neural networks. *SIAM J. Sci. Comput.* **43**, A3055–A3081 (2021).
46. Li, Z. et al. Fourier Neural Operator for Parametric Partial Differential Equations. <https://doi.org/10.48550/arXiv.2010.08895> (2021).
47. Lanser, D. & Verwer, J. G. Analysis of operator splitting for advection–diffusion–reaction problems from air pollution modelling. *J. Comput. Appl. Math.* **111**, 201–216 (1999).
48. Sportisse, B. An analysis of operator splitting techniques in the stiff case. *J. Comput. Phys.* **161**, 140–168 (2000).
49. Air Korea: Check final measurement data. [https://www.airkorea.or.kr/web/last\\_amb\\_hour\\_data?pMENU\\_NO=123](https://www.airkorea.or.kr/web/last_amb_hour_data?pMENU_NO=123).
50. Wind Direction Quick Reference | Earth Observing Laboratory. <https://www.eol.ucar.edu/content/wind-direction-quick-reference>.
51. Practical Meteorology An Algebra-based Survey of Atmospheric Science - version 1.02b. isbn 978-0-88865-283-6. [https://www.eoas.ubc.ca/books/Practical\\_Meteorology/](https://www.eoas.ubc.ca/books/Practical_Meteorology/).
52. Jayakumar, S. M. et al. Multiplicative Interactions and Where to Find Them. In *International conference on learning representations* (2020).
53. Ben-Shaul, I., Galanti, T. & Dekel, S. Exploring the Approximation Capabilities of Multiplicative Neural Networks for Smooth Functions. <https://doi.org/10.48550/arXiv.2301.04605> (2023).
54. Adamax — PyTorch 2.9 documentation. <https://docs.pytorch.org/docs/stable/generated/torch.optim.Adamax.html>.
55. ReduceLRonPlateau—PyTorch 2.9 documentation. [https://docs.pytorch.org/docs/stable/generated/torch.optim.lr\\_scheduler.ReduceLRonPlateau.html](https://docs.pytorch.org/docs/stable/generated/torch.optim.lr_scheduler.ReduceLRonPlateau.html).

## Acknowledgements

This work was partially supported by the high priority of the University of Houston. We acknowledge the use of the Carya Cluster and the advanced support from the Research Computing Data Core at the University of Houston to carry out the research presented here.

## Author contributions

R.D. conceived and designed the study, developed the computational codes, conducted all computational experiments, performed formal analysis, visualized the data, and wrote the original draft. Y.C. contributed to manuscript review and editing. D.S. contributed to preprocessing code and manuscript review. J.P. contributed to manuscript review, editing, and provided supervision for result interpretation. N.K. contributed to visualization and manuscript review. All authors discussed the results, reviewed the manuscript, and approved the final version.

## 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/s44407-026-00052-x>.

**Correspondence** and requests for materials should be addressed to Yunsoo Choi.

**Reprints and permissions 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.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, 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 changes were made. 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/4.0/>.

© The Author(s) 2026