npj | climate and atmospheric science

Review article

Published in partnership with CECCR at King Abdulaziz University

https://doi.org/10.1038/s41612-024-00716-z

Dynamics of $PM_{2.5}$ and network activity during extreme pollution events

Nail F. Bashan, Weiyu Li & Qi R. Wang ✉

In an era where air pollution poses a significant threat to both the environment and public health, we present a network-based approach to unravel the dynamics of extreme pollution events. Leveraging data from 741 monitoring stations in the contiguous United States, we have created dynamic networks using time-lagged correlations of hourly particulate matter ($PM_{2.5}$) data. The established spatial correlation networks reveal significant $PM_{2.5}$ anomalies during the 2020 and 2021 wildfire seasons, demonstrating the approach's sensitivity to detecting regional pollution phenomena. The methodology also provides insights into smoke transport and network response, highlighting the persistence of air quality issues beyond visible smoke periods. Additionally, we explored meteorological variables' impacts on network connectivity. This study enhances understanding of spatiotemporal pollution patterns, positioning spatial correlation networks as valuable tools for environmental monitoring and public health surveillance.

Air pollution remains a critical global health risk[1,2]. In the United States, over 30% of the population resides in areas with hazardous air pollution levels and this figure is expected to rise significantly due to the growing impacts of extreme pollution events on yearly trends (Fig. 1a). This, in turn, contributes to an estimated annual death toll of 85,000–200,000[3,4].

Particulate matter with an aerodynamic diameter of 2.5 μm or less ($PM_{2.5}$) severely compromises respiratory health[5–7], costs trillions of US dollars in healthcare[8–10], and exacerbates social inequalities[11–15]. The need to understand the spatial distribution and scales of $PM_{2.5}$ becomes even more urgent during extreme air pollution events, when atmospheric factors can significantly extend the reach of these pollutants, exposing distant communities to hazardous concentrations. Current air quality assessments tend to underestimate these particles' varied toxicity and disparate health impacts across different regions[7,16,17]. Therefore, a deeper comprehension of the spatial distribution of $PM_{2.5}$ at both local and regional levels is essential. Such knowledge is not only critical for accurately evaluating the health risks associated with air pollution but also for establishing effective risk communication mechanisms to mitigate these health burdens[18,19].

Despite its importance and urgency, the accurate estimation of $PM_{2.5}$ distribution and scale in the U.S. remains a challenge. The research-grade monitoring stations, maintained according to the federal equivalent methods (FEM) and federal reference methods (FRM), are sparse and unevenly distributed due to high costs (Fig. 1b, c). In fact, only 21% of the 3100 U.S. counties are equipped with FRM/FEM $PM_{2.5}$ monitors. Many counties also only have a single monitor, insufficient to accurately represent $PM_{2.5}$ levels across wider areas[20–22]. Alternative data sets, from low-cost sensors or air quality models using satellite remote sensing and meteorological data, frequently yield lower-quality data and inaccurate estimates[23–30]. During extreme pollution events, relying solely on sparsely distributed sensors or employing error-prone models for monitoring system changes can result in inaccurate assessments of the scale and impact[31].

Our study employs the complex network approach to analyze air quality data as an alternative solution under the pollution events when toxicity assessment is out of the limits for conventional particulate matter measurement methods. Using $PM_{2.5}$ time series correlations between FRM/FEM monitors across the contiguous United States for 2019–2021, we construct *spatial correlation networks* and updated the network structure daily[32]. We then examine the dynamic changes in the network structure with the progress of extreme air pollution events to discern system-wide and local impacts. In previous studies, network-based approaches have proven effective in capturing critical transitions in environmental events[33,34] and diffusion of pollution particles[35–38]. Here we undertake a comprehensive nationwide examination of the multiple extreme pollution scenarios, significantly expanding beyond traditional point-based analyses. The network-based approach improves our understanding of regional air pollution impacts. It also can provide a more integrative risk communication strategy aimed at mitigating the adverse effects of air pollution on public health.

Results

Network conceptualization

We conceptualize air quality monitors as individual nodes in a network and construct undirected, unweighted networks using hourly $PM_{2.5}$ data

Department of Civil and Environmental Engineering, Northeastern University, 360 Huntington Ave., Boston, 02115 MA, USA.
✉e-mail: q.wang@northeastern.edu

Fig. 1 | Network conceptualization. a Number of monitoring stations exceeding the EPA's yearly $PM_{2.5}$ standard (12 µg/m^3) by the U.S. climate regions. **b** Distances (in km) between adjacent monitoring stations within each climate region. **c** Spatial distribution of FRM/FEM sensors (blue circles) across the climate regions[53]. **d–g** Conceptualization of spatial correlation networks responses to extreme pollution events. **d** Illustration of an urban area in normal conditions. **e** Urban area during a wildfire, displaying deteriorated air quality. Under typical conditions, the network shows localized correlations (**f**), while during extreme pollution, regional correlations emerge (**g**), with the network's complexity measured by the total degree ($2L$).

gathered from FRM/FEM monitors across the contiguous United States for the years 2019, 2020, and 2021. In line with the methodology developed in ref. 32, links between two nodes are established if hourly time-lagged cross-correlation between them exceeds a critical threshold c_{ij}, as shown in Fig. 1d–g. A detailed definition and mathematical exposition of the methodology are presented in the "Methods" section. This conceptual network is termed a *spatial correlation network* of $PM_{2.5}$.

Figure 1d and f illustrate a scenario with no large-scale pollution events, where air quality monitors (A–E) measure local air quality trends. In such conditions, air quality readings tend to differ more across locations, leading to a loosely connected network—evidenced by the fewer inter-monitor links in the network diagram. This is due to the low correlations between individual monitor measurements, suggesting independent local air quality trends rather than a synchronized regional phenomenon. The resulting low total degree ($2L = 4$) signifies the monitors' operational independence, with each responding to potentially unique, localized events.

In contrast to the air quality variations captured under normal conditions, Fig. 1e presents a starkly different situation characterized by large-scale pollution stemming from wildfires adjacent to an urban area. This emergency is mirrored in the network behavior of air quality monitors A through E, which now exhibit a dense web of connections indicative of more and higher correlation between the readings of the monitors. This high-level connectivity, quantified as a total degree $2L = 20$, represents a substantial increase in the sum of all connections compared to the previous scenario. Such extensive connectivity suggests a homogenized distribution of pollutants across the region, with the monitors collectively detecting a uniform environmental disturbance over local differences.

US-Wide impact analysis

In our investigation of nationwide $PM_{2.5}$ levels and network dynamics (Fig. 2a), we first established a baseline for understanding seasonal patterns and the structural response of the monitoring network during extreme pollution events. The baseline for daily median $PM_{2.5}$ levels was

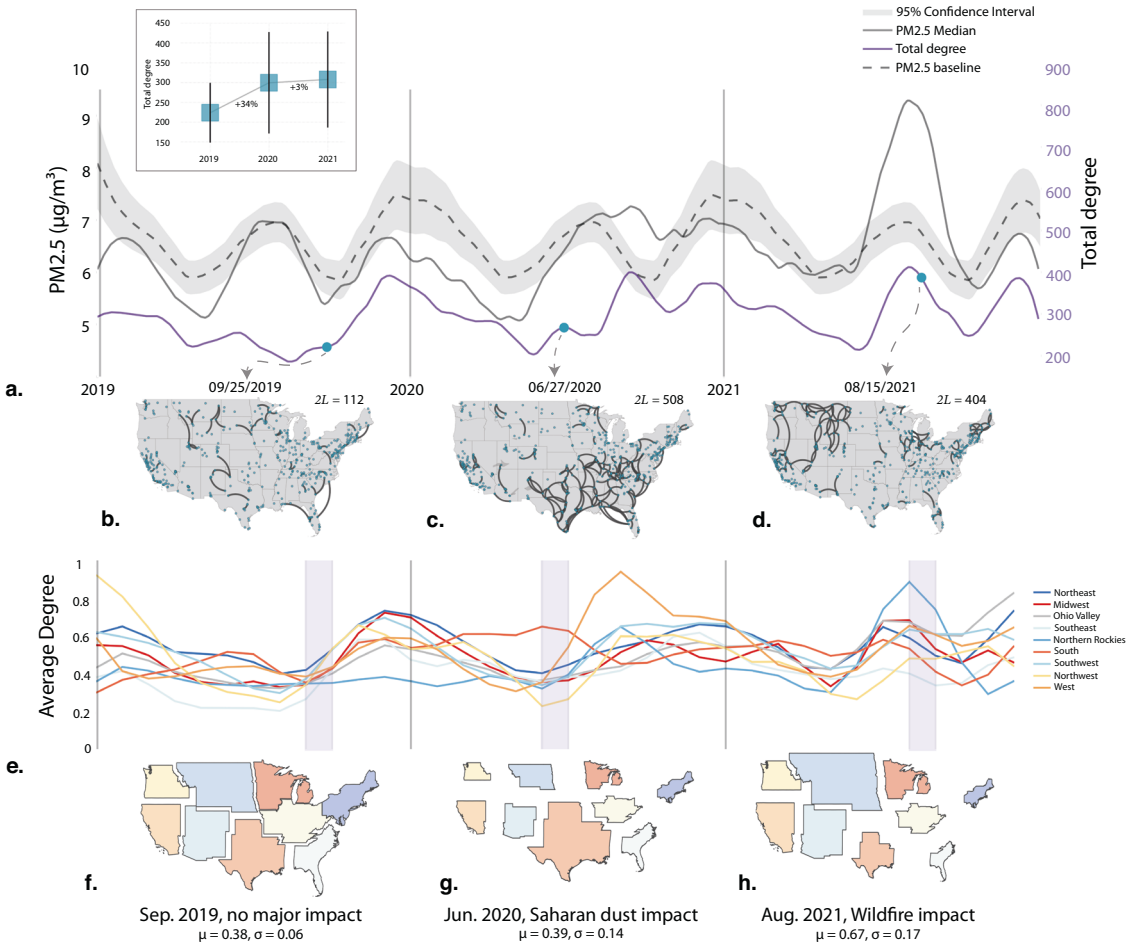

Fig. 2 | US-Wide and climate region impact analysis. a Time series of daily median $PM_{2.5}$ values in the US (black line), network total degree (purple line), and baseline levels of $PM_{2.5}$ (dashed line). The inset shows percentage changes in the yearly average total degree. **b–d** Daily networks on representative days illustrating different pollution scenarios: September 25, 2019 (**b**) under normal conditions with local correlations, June 27, 2020 (**c**), and August 15, 2021 (**d**) during extreme pollution events with regional correlations. **e** Monthly average network degree by US climate region over 3 years. **f–h** Rescaled U.S. climate regions for September 2019, June 2020, and August 2021 according to the average network degrees. National mean (μ) and standard deviation (σ) are provided.

set using data from years identified with the lowest impact from wildfires, namely 2013, 2014, and 2016.

The year 2019 served as a control period for comparison purposes because $PM_{2.5}$ levels were predominantly aligned with expected seasonal fluctuations and remained close to the baseline. Summer months typically exhibit natural $PM_{2.5}$ peaks due to the photochemical generation of secondary $PM_{2.5}$ particles from the reactions between volatile organic compounds (VOC_s) and nitrogen oxides (NO_x)[39]. However, in 2019, despite the usual summer increases, $PM_{2.5}$ measurements and network connectivity indicated no significant regional pollution events. Exceptions occurred in June 2019, when smoke from Alberta's wildfires reached the Mid-West and South USA, prompting several cities to issue air quality alerts.

Compared to the control period, the years 2020 and 2021 exhibited pronounced anomalies due to increased extreme pollution events. $PM_{2.5}$ concentrations during these summer seasons surged, with increases of 23% and 35%, respectively, when compared to the baseline. These events resulted in an increased uniformity of pollutant distribution, as depicted by the heightened correlation patterns among the air quality monitors. Compared to the same periods in 2019, we observed a substantial rise in the summer total degree averages up to 34% in 2020 and 38% in 2021, as illustrated in the inset of Fig. 2a. This drastic change underscores the extensive reach of major pollution events, affecting air quality across vast distances well beyond their immediate hotspots.

The observations from the yearly analysis are validated and signified by the case studies in 2019, 2020, and 2021. Figure 2b illustrates a snapshot of the spatial correlation network for $PM_{2.5}$ under normal conditions on September 25, 2019, across the contiguous United States. The connections between nodes are sparse, and the total degree of the network ($2L$) is low at 112 (average degree of $\mu = 0.23$). In comparison, the total degree of the networks more than doubled during the pollution event caused by Saharan dust originating in Africa Fig. 2c ($2L = 508$, $\mu = 1.02$) and wildfire events originating from the northwestern USA Fig. 2d, ($2L = 404$, $\mu = 0.81$), indicating the scale of national impacts.

Climate regions and spatial correlation networks of $PM_{2.5}$

We next delve deep into the spatial correlation networks broken down in different climate regions to uncover the detailed ways in which our network-based approach could help understand the spatial dispersion of air pollutants. The network-based approach allows us to investigate the inter-regional influences and the potential for widespread air quality impacts due to transboundary pollution transport[40].

Regional climatic conditions profoundly influence air quality, and thus $PM_{2.5}$ distribution shows distinct regional patterns across the country

(Fig. 2e). We compare the average degree values over various periods and scenarios to discern these patterns. In September 2019, a period characterized by the absence of significant pollution events, all regions displayed similar average degree values, with the highest at 0.47 in the Northeast and the lowest at 0.28 in the Southwest (Fig. 2f).

Contrastingly, in June 2020, $PM_{2.5}$ concentrations surpassed the EPA's air quality standards at nearly 40% of the monitoring stations in the southern US due to the massive dust plume traversed the Caribbean Basin[41,42]. This event led to the Southern US demonstrating the highest average degree in the network, peaking at 0.78, while other regions were less affected, maintaining an average degree of $\mu = 0.39$ (Fig. 2g) close to the national average of $\mu = 0.38$ in 2019. In August 2021, the highest average degree was observed over the Northern Rockies, directly linked to the wildfires originating from California. During this time, an increase in the average degree was noted across the entire nation (Fig. 2h) ($\mu = 0.67$), highlighting the extensive reach of pollutants beyond their initial hotspots. This pattern of increased connectivity is primarily attributed to the long-range transport of pollutants from wildfire events in Canada and California. The synchronization of monthly peaks in network connectivity with the occurrence of extreme pollution events emphasizes the widespread environmental impact of such events. Furthermore, a comparative analysis of different regions shows that while some maintained a relatively stable trend, others experienced significant fluctuations between different years, corresponding to the localized impacts they encountered.

Smoke coverage and pollutant homogeneity

Building on the established framework, we then test the efficacy of our connectivity index in explaining smoke-covered days. This specific measure is chosen because of its substantial health implications documented by prior research[16,43]. Smoke exposure is not only a marker of air quality degradation but is also associated with various adverse health outcomes, making it critical for public health surveillance and response.

To assess the explanatory power of our index, we analyze its performance over the 4 months characterized by wildfire impact in the year 2021. Figure 3a shows the spatial patterns of $PM_{2.5}$, and Fig. 3b presents the spatial distribution of smoke-covered days across the United States during this time. The spatial patterns of impacted regions and severities misalign in these two panels, suggesting that while smoke from wildfires can create visually dense pollution areas, it does not always correspond to the ground $PM_{2.5}$ measurements across regions. On the contrary, Fig. 3c, which illustrates the total degree over 120 days, shows a spatial distribution that has a greater alignment with smoke-covered days (Fig. 3b) than the median $PM_{2.5}$ concentrations (Fig. 3a). The similarity suggests that the network connectivity, derived from our spatial correlation network, is a better proxy for the spread and impact of wildfire smoke over large geographical scales.

Statistical analysis reveals the same insight. Figure 3d presents a scatter plot showing the relationship between median $PM_{2.5}$ concentrations from June to September 2021 and the number of days covered by smoke for different climate regions across the United States. The plot reveals a low coefficient ($R^2 = 0.04$), suggesting a weak relationship between the median $PM_{2.5}$ levels and the number of smoke-covered days across the regions. This implies that the presence of smoke in the atmosphere, indicative of pollution from wildfires, does not necessarily correlate strongly with higher ground-level $PM_{2.5}$ concentrations. The low R^2 value indicates that other factors may influence $PM_{2.5}$ levels. On the contrary, we observe a much more substantial positive correlation ($R^2 = 0.74$), suggesting a strong relationship between the network's total degree and the incidence of smoke-covered days across the regions Fig. 3e. The consistency between these two analyses emphasizes that our network approach better captures the true spatial footprint of smoke coverage. It thus can potentially be a more effective tool in understanding air quality and health risk assessments associated with wildfire smoke exposure.

Daily network dynamics and feature importance

Beyond capturing the national and regional impacts of air pollution events, we test our network-based approach's ability to capture the dynamics of air pollution events with high temporal resolution. We thus delve further into the 2020 fire season and examine the spatial correlation network on a day-to-day basis. The dynamics of our daily networks reveal a time lag between network connectivity and the transport of smoke. Figure 4a–c presents a series of maps displaying the spatial correlation network over the United States on three different days in late August 2020. On August 21, 2020, Fig. 4a, the network is highly interconnected in the western regions, again corresponding to areas directly affected by wildfire smoke. As we progress to August 25 (Fig. 4b), the network shows increasing connectivity across the entire United States caused by the spreading of the smoke to receptor regions. By August 27 (Fig. 4c), there is a notable reduction in visible smoke

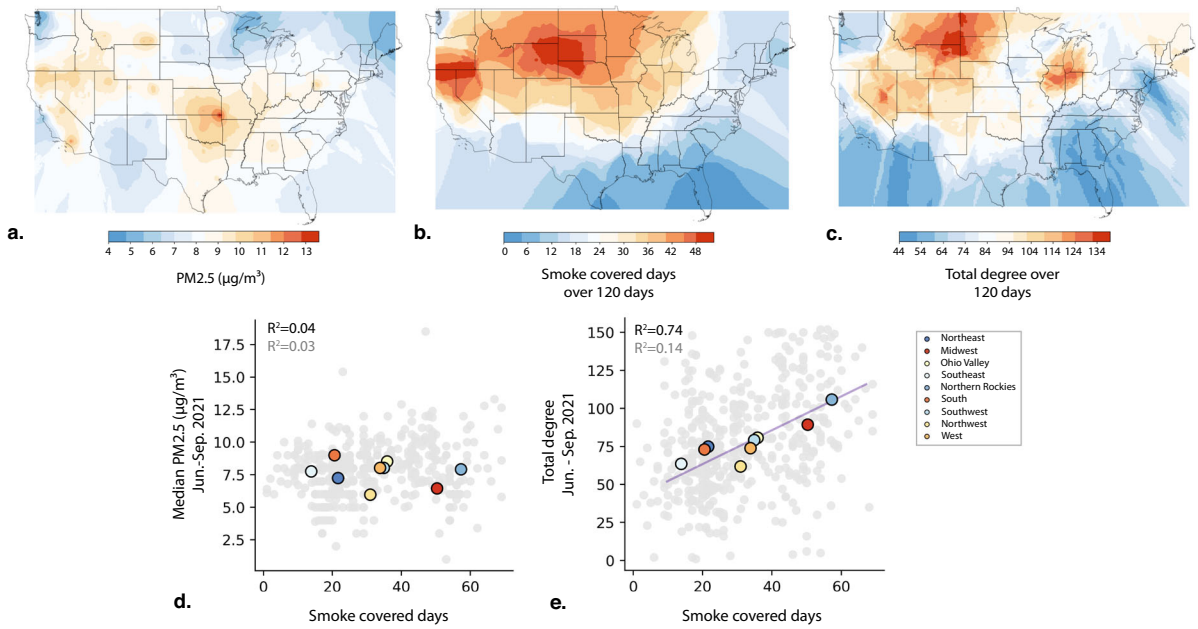

Fig. 3 | Smoke coverage and connectivity. a–c Spatial distribution of interpolated seasonal values: Interpolated median $PM_{2.5}$ concentrations (**a**), smoke-covered days (**b**), and total degrees (**c**) during the 2021 wildfire season (June–September). **d** Correlation plot between seasonal $PM_{2.5}$ and smoke-covered days for individual U.S. sensors (gray points) and mean values for climate regions (colored circles). **e** Seasonal total degree and smoke-covered days.

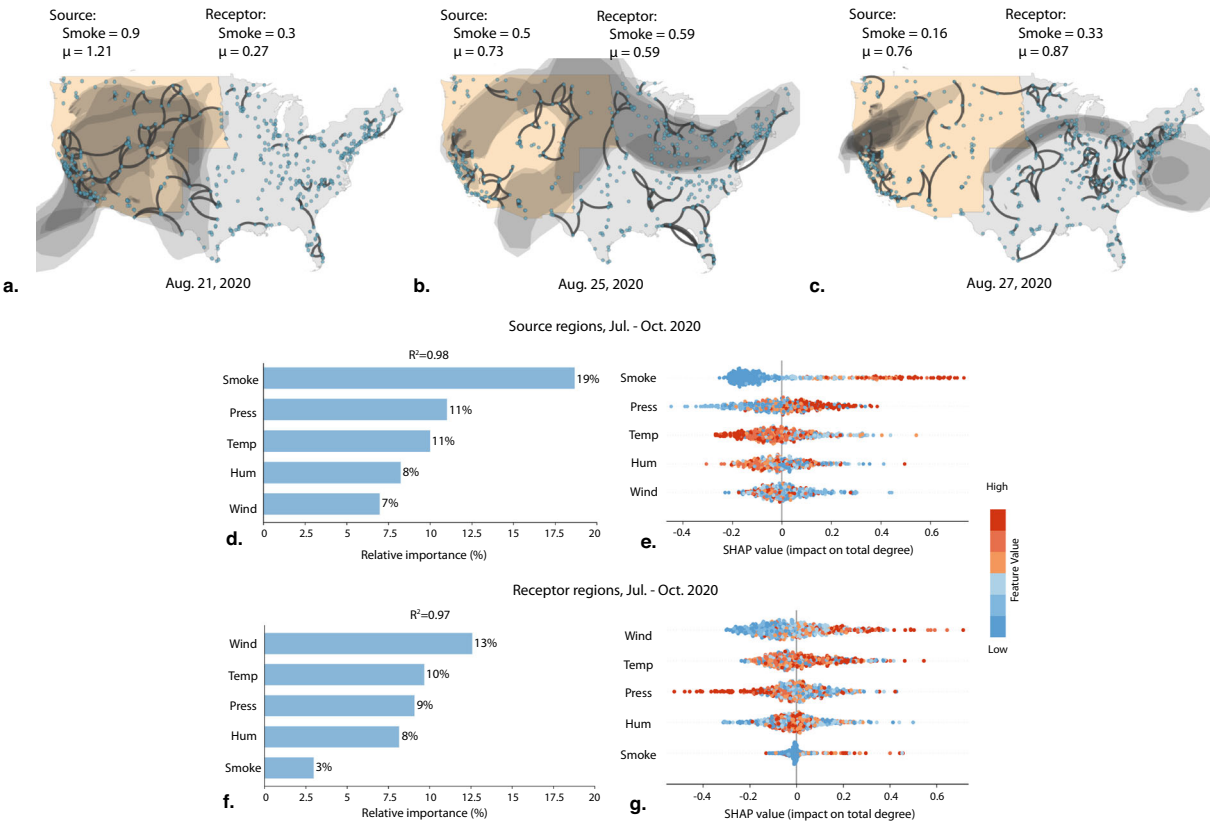

Fig. 4 | Daily network dynamics and feature importance. a–c Network connectivity and smoke transport time lag analysis. On Aug. 21, 2020 (**a**), smoke plumes spread to the western U.S. They reached the eastern U.S. 4 days later on August 25, 2020 (**b**), contributing to increased regional connectivity in the eastern part. The connectivity remained high even though smoke plumes were not visible based on satellite images on August 27, 2020 (**c**). **d–g** Assessment of the impacts of meteorological variables wind speed (knots), temperature (F°), humidity (%), atmospheric pressure (millibars), and smoke coverage on the network's average degree. The bar charts in **d** show relative importance, calculated as the ratio of absolute SHAP values' means to the average degree, for source regions. The scatter plots of SHAP values in **e** display the impact of each factor on network connectivity. **f** and **g** display the same values in the receptor regions.

coverage; however, the network remains densely connected. The observed phenomenon is due to the limitation of satellite imagery[44,45]: although providing a comprehensive view of its columnar extent, it falls short in capturing the smoke during cloudy days and pinpointing the smoke's vertical distribution—whether hovering near the ground or being suspended at higher altitudes. Even when visual indications of smoke diminish, the persistent connectivity suggests that ground-level $PM_{2.5}$ impact remained high, signaling that the air quality issues are not solely confined to the periods of visible smoke but may persist as the particulate matter settles closer to the earth's surface. Our network-based index adeptly captures the impact of pollution even when conventional data may be incomplete or absent.

Given the usefulness of the connectivity from our spatial correlation model, our last effort is to gain deeper insights into the factors influencing network connectivity by employing LightGBM (light gradient-boosting machine) for modeling daily average degrees with meteorological variables in different regions[46]. In source regions, our observations highlight that smoke coverage holds the highest relative importance for average degree changes (Fig. 4d, e), while other meteorological variables exhibited indeterminate trends. In receptor regions, smoke coverage showed the least relative importance, aligning with the time lag behavior (Supplementary Fig. 3). Despite the absence of a clear trend in receptor regions, wind appears to be the most influential variable (Fig. 4f, g). This suggests that when receptor regions experience high wind speeds during the transportation of aerosols from extreme pollution events, ground homogeneity increases, resulting in elevated network connectivity. This analysis indicates that network connectivity can be influenced by multiple factors, including meteorological conditions (Supplementary Fig. 2), long-range transport of aerosols, and their resultant effects on ground concentrations. Therefore, a careful examination of these causative factors is essential to comprehensively understand the underlying mechanisms driving these observed patterns in network connectivity.

Discussion

Our network model for air pollution analysis faces limitations, notably in data dependency. The network's effectiveness relies on the high quality and consistent availability of $PM_{2.5}$ measurements from FRM/FEM stations, which may have uneven distribution, potentially causing data blind spots[29]. While low-cost air sensors present an opportunity to augment data density, their accuracy requires thorough validation. Furthermore, the non-real-time nature of the validated FRM/FEM data limits the model's ability for immediate analysis. Our methodology elucidates the spatial and temporal dynamics of $PM_{2.5}$ through network analysis, though it does not differentiate types of emissions, which limits our ability to identify variations in particulate size and composition, such as primary versus secondary particulate formations. Our approach does not model the detailed chemical transformations of $PM_{2.5}$, yet recognizing the influence of atmospheric chemistry on network correlations is crucial. Future studies should integrate chemical transport models and low-cost sensors to refine our understanding of $PM_{2.5}$ dynamics and improve real-time data accuracy (see Supplementary Figs. 4 and 5).

The large-scale spatial correlation network developed in this study represents a significant theoretical advancement in our understanding of air quality dynamics. Instead of relying on computationally expensive atmospheric models or error-prone estimation models, this study explained pollutant homogeneity and associated risk factors by delving into the

correlation mechanisms inherent in physical systems. By conceptualizing air quality monitors as a network of interconnected nodes, we provide a framework that captures both the independence and interdependence of local and regional air quality events. This network transcends the traditional point-based analysis, offering a holistic view that reflects the complexity of air pollution as a multifaceted phenomenon. The theoretical implications extend beyond mere data aggregation; they redefine our understanding of air quality patterns as emergent properties of a complex system, where localized events can have ripple effects across vast geographical scales. This shift encourages a re-evaluation of how air pollution is modeled, moving towards more integrative and system-oriented approaches.

Our approach excels in capturing and characterizing large-scale pollution events. By analyzing daily variations in network connectivity, we have successfully identified significant anomalies in $PM_{2.5}$ concentrations on both national and regional levels, as was evident during the wildfire and Saharan dust events. These findings underscore the approach's sensitivity to detecting synchronized phenomena and its ability to differentiate between typical environmental conditions and periods of heightened pollution. On the scale of climate regions, the network model effectively identifies areas of pollutant homogeneity, allowing for a nuanced understanding of the spatial extent of extreme pollution events. The model's capacity to trace the evolution of such events over time further highlights its potential as a critical tool for environmental monitoring.

The practical applications of our spatial correlation network are particularly compelling in the context of public health. The ability to define smoke-covered days, a critical marker of air quality degradation, is invaluable for public health surveillance and response[16,40,44]. Our network's high temporal resolution can potentially provide early warnings for smoke dispersion, enabling health authorities to issue timely advisories and take preemptive action to protect vulnerable populations. From a policy-making perspective, the insights gleaned from our network analysis could inform the development of air quality standards and pollution control measures. By elucidating the transboundary nature of air pollutants, our methodology can drive the creation of more collaborative and effective environmental policies that reflect the interconnectedness of ecosystems and transcend political boundaries. This network-based approach, therefore, has the potential to transform air quality management and public health policy by providing a more responsive and accurate assessment of pollution-related risks.

Methods

Data

In this study, we analyze $PM_{2.5}$ FRM/FEM measurements obtained from 741 monitoring stations located across the United States. To maintain consistency in our observations over the years, we excluded monitoring stations that underwent equipment upgrades between 2019 and 2021. After removing these monitors and focusing our study on the contiguous United States, we had 496 monitoring stations as our network nodes. Data covers the time period from January 1, 2019 to December 31, 2021 and local $PM_{2.5}$ conditions are reported hourly in micrograms per cubic meter ($\mu g/m^3$). This multi-year data allows us to compare network dynamics across various time periods characterized by different extreme pollution event scenarios. The year 2020 was particularly notable as the year with the most devastating wildfires in California's history, leading to nearly 10,000 wildfires consuming over 4.2 million acres[47]. The air quality during this time was further worsened by the largest recorded Saharan dust event impacting the Southern USA. The following year, 2021, became the second most severe wildfire season, with over 2.5 million acres affected[42,48].

In addition, we utilized smoke data sourced from The National Oceanic and Atmospheric Administration, focusing only on smoke plumes categorized as 'heavy' and 'medium' in our calculations. The meteorological variables used to analyze feature importance, as discussed in the section "Daily network dynamics and feature importance", are obtained from the EPA's pre-generated data files, which are available along with the FRM/FEM measurements. Wind speed (knots), temperature (F°), humidity (%), and pressure (millibars) are selected as the independent variables as they are considered pivotal for the vertical and horizontal dispersion of aerosols[49,50].

Network definition

We conceptualize air quality monitors as network nodes and construct undirected, unweighted networks using hourly $PM_{2.5}$ data. Links are established based on time-lagged cross-correlation calculations[32,35]. These networks are spatially constrained, where each node can only form connections with its neighboring nodes. We employ the Voronoi diagram to ensure system-wide connectivity, defining two nodes as neighbors if they share a border[51]. Over a period of three years, each day's network begins with N disconnected nodes. Links are then formed if the cross-correlation between two neighboring nodes i and j exceeds a critical bonding threshold c_{ij}.

Cross-correlation is calculated by Eq. (1):

$$\hat{C}_{ij}^{(\tau)} = \frac{\left\langle gX_i(t).gX_j(t+\tau)) \right\rangle}{\sqrt{\left\langle [gX_i(t)]^2 \right\rangle} \cdot \sqrt{\left\langle [gX_j(t+\tau)]^2 \right\rangle}} \quad (1)$$

where τ is the time-lag defined in the interval of $-\tau_{\max} < \tau < \tau_{\max}$, X_i represents $PM_{2.5}$ readings of the monitor i at time t, and fluctuation series is given as $gX_i(t) = X_i(t) - \langle X_i \rangle$ with respect to its mean value over T periods $\langle X_i \rangle = \frac{1}{T}\sum_{t=1}^{T} X_i(t)$.

Selecting an optimal time lag has been challenging due to the variety of sources impacting the physical and chemical processes involved in the formation, transport, and transformation of $PM_{2.5}$. Considering that $PM_{2.5}$'s atmospheric residence time is 3–5 days, previous studies have applied time lags of up to 5 days[35]. However, factors like deposition, suspension, and secondary aerosol formation progression can also influence the time lag selection. To ensure robust correlation calculations based on a sufficient data set, we adopted a maximum time lag of 9 h, guaranteeing the inclusion of at least 15 hourly measurements in our analysis. Selecting a time lag shorter than 12 h allows us to capture correlations resulting from external impacts rather than routine physical changes such as periodic boundary layer depth changes and anthropogenic emissions. Additionally, given an average distance of 137 km between neighboring monitors, a shorter time lag would inadequately capture vertical pollutant transportation under moderate to low wind speeds.

Previous studies have employed various methods to identify the threshold for the cross-correlations, including shuffling the data and computing the average of the absolute values of correlations[37], calculating the summation of the mean and standard deviation of a rolling window correlation matrix[38], and setting a global threshold[36]. However, as illustrated in Supplementary Fig. 1, the sensitivity to a global threshold can vary depending on the locations of the monitors. Due to similar local dynamics, the threshold values between two neighboring urban monitoring sites are higher than those between neighboring urban and rural sites. Furthermore, distance plays a crucial role, with monitoring sites in closer proximity exhibiting higher thresholds. Consequently, a universally applied threshold may inaccurately suggest that these nearby monitors are consistently under significant pollutant impact. Therefore, we propose the need for assigning pair-specific thresholds. We set 2912 individual bonding thresholds c_{ij} for each possible interaction by the summation of the mean and standard deviation of their three-year connectivity calculations. This threshold assists in distinguishing between normative conditions and extreme events in our analysis. Assigning unique i, j pairs also enables the detection of regional impacts, even in areas with differing geographical characteristics or shared background concentration.

We measured the overall impact of extreme pollution events on the system by calculating the total degree of our network ($2L$). This is determined by summing all edges formed between monitoring stations, which result from correlated time series. To compare this impact across different climate regions, we employed the average degree, considering that each

climate region has a distinct number of monitoring stations. The average degree (μ) is determined by dividing the total degree by the number of monitoring stations within a climate region.

LightGBM calculations and feature Importance

We employed the LightGBM (light gradient-boosting machine) machine learning algorithm along with Shapley additive explanation (SHAP) feature importance assessment to analyze the impact of various meteorological variables on network connectivity. LightGBM uses gradient boosting decision tree (GBDT) techniques incorporating gradient-based one-side sampling (GOSS) and exclusive feature bundling (EFB). In the process, decision trees are trained sequentially, and LightGBM significantly accelerates the training process while achieving nearly the same accuracy in a shorter time span[46]. We utilized region-specific smoke coverage, temperature, humidity, wind speed, and pressure as explanatory variables and trained a model for an average degree in both receptor and source monitors. Source and receptor regions are divided according to the presence of wildfires during our analysis period (July to October 2020), and smoke coverage is calculated as the percentage of monitors under the smoke captured by the satellite. The trained models with high R^2 values were then employed for feature importance analysis. SHAP interaction values, a game-theoretic approach for interpretability of tree-based models, were used to explain the output of the machine-learning model[52]. In the analysis, the conjunction of a high feature value (in red) and a positive SHAP value implies a significant and positive impact (Fig. 4d–g). The mixture of red and blue dots signifies an indeterminate impact of meteorological variables on average degree. Subsequently, we measured the relative importance of each factor by calculating the ratio between the means of their absolute SHAP values and the average degree for each region.

Data availability

$PM_{2.5}$ FRM/FEM measurements along with meteorological variables are available at https://aqs.epa.gov/aqsweb/airdata/download_files.html and smoke data sourced from The National Oceanic and Atmospheric Administration is available at https://www.ospo.noaa.gov/Products/land/hms.html.

Code availability

Code to reproduce all results in the paper is available at https://github.com/nnbashan/pm2.5_network_dynamics/tree/main.

Received: 5 January 2024; Accepted: 8 July 2024;
Published online: 22 July 2024

References

1. Shaddick, G., Thomas, M. L., Mudu, P., Ruggeri, G. & Gumy, S. Half the world's population are exposed to increasing air pollution. *npj Clim. Atmos. Sci.* **3**, 23 (2020).
2. Rentschler, J. & Leonova, N. Global air pollution exposure and poverty. *Nat. Commun.* **14**, 4432 (2023).
3. Tessum, C. W. et al. PM2.5 polluters disproportionately and systemically affect people of color in the United States. *Sci. Adv.* **7**, eabf4491 (2021).
4. American Lung Association. *State of the Air* https://www.lung.org/research/sota/key-findings (2023)
5. Woodruff, T. J., Parker, J. D. & Schoendorf, K. C. Fine particulate matter (PM2.5) air pollution and selected causes of postneonatal infant mortality in California. *Environ. Health Perspect.* **114**, 786–790 (2006).
6. Apte, J. S., Marshall, J. D., Cohen, A. J. & Brauer, M. Addressing global mortality from ambient $PM_{2.5}$. *Environ. Sci. Technol.* **49**, 8057–8066 (2015).
7. Burnett, R. et al. Global estimates of mortality associated with long-term exposure to outdoor fine particulate matter. *Proc. Natl Acad. Sci. USA* **115**, 9592–9597 (2018).
8. Yin, H. et al. Population ageing and deaths attributable to ambient $PM_{2.5}$ pollution: a global analysis of economic cost. *Lancet Planet. Health* **5**, e356–e367 (2021).
9. Xie, Y., Dai, H., Dong, H., Hanaoka, T. & Masui, T. Economic impacts from $PM_{2.5}$ pollution-related health effects in China: a provincial-level analysis. *Environ. Sci. Technol.* **50**, 4836–4843 (2016).
10. Yang, S., Fang, D. & Chen, B. Human health impact and economic effect for $PM_{2.5}$ exposure in typical cities. *Appl. Energy* **249**, 316–325 (2019).
11. Jbaily, A. et al. Air pollution exposure disparities across US population and income groups. *Nature* **601**, 228–233 (2022).
12. Deguen, S. & Zmirou-Navier, D. Social inequalities resulting from health risks related to ambient air quality—a European review. *Eur. J. Public Health* **20**, 27–35 (2010).
13. Liu, J. et al. Disparities in air pollution exposure in the United States by race/ethnicity and income, 1990–2010. *Environ. Health Perspect.* **129**, 127005 (2021).
14. Kioumourtzoglou, M.-A., Schwartz, J., James, P., Dominici, F. & Zanobetti, A. PM2.5 and mortality in 207 US cities: modification by temperature and city characteristics. *Epidemiology* **27**, 221 (2016).
15. Tessum, C. W. et al. Inequity in consumption of goods and services adds to racial–ethnic disparities in air pollution exposure. *Proc. Natl Acad. Sci. USA* **116**, 6001–6006 (2019).
16. Aguilera, R., Corringham, T., Gershunov, A. & Benmarhnia, T. Wildfire smoke impacts respiratory health more than fine particles from other sources: observational evidence from Southern California. *Nat. Commun.* **12**, 1493 (2021).
17. Kim, Y. H. et al. Mutagenicity and lung toxicity of smoldering vs. flaming emissions from various biomass fuels: implications for health effects from wildland fires. *Environ. Health Perspect.* **126**, 017011 (2018).
18. Schwartz, J., Dockery, D. W. & Neas, L. M. Is daily mortality associated specifically with fine particles? *J. Air Waste Manag. Assoc.* **46**, 927–939 (1996).
19. Childs, M. L. et al. Daily local-level estimates of ambient wildfire smoke $PM_{2.5}$ for the contiguous US. *Environ. Sci. Technol.* **56**, 13607–13621 (2022).
20. Li, J. et al. Integrating low-cost air quality sensor networks with fixed and satellite monitoring systems to study ground-level $PM_{2.5}$. *Atmos. Environ.* **223**, 117293 (2020).
21. Sullivan, D. M. & Krupnick, A. Using satellite data to fill the gaps in the US air pollution monitoring network. *Resour. Future Work. Pap.* 18–21 (2018).
22. Tang, M., Wu, X., Agrawal, P., Pongpaichet, S. & Jain, R. Integration of diverse data sources for spatial PM2.5 data interpolation. *IEEE Trans. Multimed.* **19**, 408–417 (2017).
23. Gao, M., Cao, J. & Seto, E. A distributed network of low-cost continuous reading sensors to measure spatiotemporal variations of $PM_{2.5}$ in Xi'an, China. *Environ. Pollut.* **199**, 56–65 (2015).
24. Li, Y. et al. From air quality sensors to sensor networks: Things we need to learn. *Sens. Actuators B Chem.* **351**, 130958 (2022).
25. Barkjohn, K. K., Gantt, B. & Clements, A. L. Development and application of a United States-wide correction for $PM_{2.5}$ data collected with the PurpleAir sensor. *Atmos. Meas. Tech.* **14**, 4617–4637 (2021).
26. Han, J., Liu, X., Chen, D. & Jiang, M. Influence of relative humidity on real-time measurements of particulate matter concentration via light scattering. *J. Aerosol Sci.* **139**, 105462 (2020).
27. Ma, Z. et al. A review of statistical methods used for developing large-scale and long-term $PM_{2.5}$ models from satellite data. *Remote Sens. Environ.* **269**, 112827 (2022).
28. Li, J. et al. Estimation of ambient $PM_{2.5}$ in Iraq and Kuwait from 2001 to 2018 using machine learning and remote sensing. *Environ. Int.* **151**, 106445 (2021).

29. Gupta, P. et al. Impact of California fires on local and regional air quality: the role of a low-cost sensor network and satellite observations. *GeoHealth* **2**, 172–181 (2018).
30. Lin, C. Q. et al. High-resolution satellite remote sensing of provincial $PM_{2.5}$ trends in China from 2001 to 2015. *Atmos. Environ.* **180**, 110–116 (2018).
31. Hua, Z., Sun, W., Yang, G. & Du, Q. A full-coverage daily average $PM_{2.5}$ retrieval method with two-stage IVW fused MODIS c6 AOD and two-stage GAM model. *Remote Sens.* **11**, 1558 (2019).
32. Gozolchiani, A., Yamasaki, K., Gazit, O. & Havlin, S. Pattern of climate network blinking links follows El Niño events. *Europhys. Lett.* **83**, 28005 (2008).
33. Mondal, S., K. Mishra, A., Leung, R. & Cook, B. Global droughts connected by linkages between drought hubs. *Nat. Commun.* **14**, 144 (2023).
34. Boers, N. et al. Complex networks reveal global pattern of extreme-rainfall teleconnections. *Nature* **566**, 373–377 (2019).
35. Vlachogiannis, D. M., Xu, Y., Jin, L. & González, M. C. Correlation networks of air particulate matter ($PM_{2.5}$): a comparative study. *Appl. Netw. Sci.* **6**, 32 (2021).
36. Jin, Q., Fang, X., Wen, B. & Shan, A. Spatio-temporal variations of PM2.5 emission in China from 2005 to 2014. *Chemosphere* **183**, 429–436 (2017).
37. Zhang, Y., Chen, D., Fan, J., Havlin, S. & Chen, X. Correlation and scaling behaviors of fine particulate matter ($PM_{2.5}$) concentration in China. *Europhys. Lett.* **122**, 58003 (2018).
38. Du, R. et al. Percolation analysis of urban air quality: a case in China. *Phys. A: Stat. Mech. Appl.* **541**, 123312 (2020).
39. Parkhurst, W. J., Tanner, R. L., Weatherford, F. P., Valente, R. J. & Meagher, J. F. Historic $PM_{2.5}/PM_{10}$ concentrations in the southeastern United States—potential implications of the revised particulate matter standard. *J. Air Waste Manag. Assoc.* **49**, 1060–1067 (1999).
40. Burke, M. et al. The contribution of wildfire to $PM_{2.5}$ trends in the USA. *Nature* **622**, 761–766 (2023).
41. Pu, B. & Jin, Q. A record-breaking trans-Atlantic African dust plume associated with atmospheric circulation extremes in June 2020. *Bull. Am. Meteorol. Soc.* **102**, E1340–E1356 (2021).
42. Yu, H. et al. Observation and modeling of the historic "Godzilla" African dust intrusion into the Caribbean basin and the southern US in June 2020. *Atmos. Chem. Phys.* **21**, 12359–12383 (2021).
43. Li, X., Jin, L. & Kan, H. Air pollution: a global problem needs local fixes. *Nature* **570**, 437–439 (2019).
44. Reisen, F., Duran, S. M., Flannigan, M., Elliott, C. & Rideout, K. Wildfire smoke and public health risk. *Int. J. Wildland Fire* **24**, 1029–1044 (2015).
45. Li, C., Li, J., Dubovik, O., Zeng, Z.-C. & Yung, Y. L. Impact of aerosol vertical distribution on aerosol optical depth retrieval from passive satellite sensors. *Remote Sens.* **12**, 1524 (2020).
46. Ke, G. et al. LightGBM: a highly efficient gradient boosting decision tree. *Adv. Neural Inf. Process Syst.* **30**, 3146–3154 (2017).
47. NOAA National Centers for Environmental Information. *Monthly Wildfires Report for Annual 2020* (National Centers for Environmental Information, accessed 18 June 2024); https://www.ncei.noaa.gov/access/monitoring/monthly-report/fire/202013.
48. Keeley, J. E. & Syphard, A. D. Large California wildfires: 2020 fires in historical context. *Fire Ecol.* **17**, 22 (2021).
49. Kumar, M., Tiwari, S., Murari, V., Singh, A. K. & Banerjee, T. Wintertime characteristics of aerosols at middle indo-gangetic plain: impacts of regional meteorology and long range transport. *Atmos. Environ.* **104**, 162–175 (2015).
50. Li, J. et al. Effects of different stagnant meteorological conditions on aerosol chemistry and regional transport changes in Beijing, China. *Atmos. Environ.* **258**, 118483 (2021).
51. Barthélemy, M. Spatial networks. *Phys. Rep.* **499**, 1–101 (2011).
52. Lundberg, S. M. et al. From local explanations to global understanding with explainable AI for trees. *Nat. Mach. Intell.* **2**, 56–67 (2020).
53. Karl, T. & Koss, W. J. *Regional and National Monthly, Seasonal, And Annual Temperature Weighted by Area, 1895–1983* (National Climatic Data Center, 1984).

Acknowledgements

This material is based upon work supported by the National Science Foundation (NSF) under Grant No. 2125326 and Northeastern University iSUPER Impact Engine. The authors are grateful for the support of NSF and Northeastern University. Any opinions, findings, conclusions, or recommendations expressed in the paper are those of the authors and do not necessarily reflect the views of the funding agencies.

Author contributions

N.F.B., Q.R.W., and W.L. developed the study concept and plan. N.F.B. performed the analyses and wrote the first draft of the manuscript. N.F.B. and W.L. created the visualizations, Q.R.W. provided feedback on the analysis, visualizations, and manuscript. All authors participated in reviewing and editing the final version of the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41612-024-00716-z.

Correspondence and requests for materials should be addressed to Qi R. Wang.