ARTICLE

Check for updates

Ultra-high-resolution mapping of ambient fine particulate matter to estimate human exposure in Beijing

Yongyue Wang[1,5], Qiwei Li[1,5], Zhenyu Luo[1], Junchao Zhao[1], Zhaofeng Lv[1], Qiuju Deng[2], Jing Liu[2], Majid Ezzati [3], Jill Baumgartner[4], Huan Liu [1✉] & Kebin He [1✉]

With the decreasing regional-transported levels, the health risk assessment derived from fine particulate matter ($PM_{2.5}$) has become insufficient to reflect the contribution of local source heterogeneity to the exposure differences. Here, we combined the both ultra-high-resolution $PM_{2.5}$ concentration with population distribution to provide the personal daily $PM_{2.5}$ internal dose considering the indoor/outdoor exposure difference. A 30-m $PM_{2.5}$ assimilating method was developed fusing multiple auxiliary predictors, achieving higher accuracy ($R^2 = 0.78$–0.82) than the chemical transport model outputs without any post-simulation data-oriented enhancement ($R^2 = 0.31$–0.64). Weekly difference was identified from hourly mobile signaling data in 30-m resolution population distribution. The population-weighted ambient $PM_{2.5}$ concentrations range among districts but fail to reflect exposure differences. Derived from the indoor/outdoor ratio, the average indoor $PM_{2.5}$ concentration was $26.5\,\mu g/m^3$. The internal dose based on the assimilated indoor/outdoor $PM_{2.5}$ concentration shows high exposure diversity among sub-groups, and the attributed mortality increased by 24.0% than the coarser unassimilated model.

[1] State Environmental Protection Key Laboratory of Sources and Control of Air Pollution Complex, State Key Joint Laboratory of Environmental Simulation and Pollution Control, School of Environment, Tsinghua University, Beijing 100084, China. [2] Centre for Clinical and Epidemiologic Research, Beijing An Zhen Hospital, Capital Medical University, Beijing Institute of Heart, Lung and Blood Vessel Diseases, Beijing 100029, China. [3] School of Public Health, Imperial College London, London SW72AZ, UK. [4] School of Population and Global Health, McGill University, Montréal, QC H3A0G4, Canada. [5] These authors contributed equally: Yongyue Wang, Qiwei Li. ✉email: liu_env@tsinghua.edu.cn; hekb@tsinghua.edu.cn

Anthropogenic air pollution stemming from rapid economic development and industrialization is a significant global public health issue[1–3], causing approximately 7 million premature deaths worldwide each year[4]. Fine particulate matter (particles that are 2.5 microns or less in aerodynamic equivalent diameter, $PM_{2.5}$), as a major air pollutant, has been proved to be linked to many diseases[5–7]. Several exposure assessment models were developed to estimate the premature mortality attributed to $PM_{2.5}$ exposure, linking the exposure effects directly to the ambient pollution level[8,9]. The ambient $PM_{2.5}$ levels was once sufficed to assess the significant difference in $PM_{2.5}$ exposure between cities in the earlier years[10–14], and the contribution of spatial heterogeneity from local sources to the exposure differences was negligible at such high ambient levels[15–18]. However, with the ambient $PM_{2.5}$ level declining sharply[19–22], the contribution of local $PM_{2.5}$ sources has grown to 50–80%[23–25]. The contribution of spatial heterogeneity from local sources to population exposure has become increasingly pronounced. $PM_{2.5}$ levels in some micro-environments can still be high[26–28], up to hundreds of micrograms per cubic meter[29,30]. Solely adopting ambient $PM_{2.5}$ levels is no longer enough to evaluate the exposure differences caused by local sources.

A more accurate way to assess $PM_{2.5}$ exposure is by measuring the personal internal dose. It reflects the cumulative interaction between human physiological activities and environmental pollution[31–33]. Direct approaches involving portable monitoring suits can effectively obtain the personal real-time exposure[34,35], but the cost highly limits its application in extensive sampling. Indirect approaches, estimated from the observed or simulated $PM_{2.5}$ distribution, are more suitable for obtaining adequate samples at the city level[36–38]. However, the monitoring value is only representative in the adjacent area[39], and the chemical transport models (CTMs) can usually reach a resolution of at most 1 km[40–42], making it all challenging to identify exposure differences among micro-environments. The development of ultra-high-resolution concentration field can effectively improve the population coverage and granularity together with the high-resolution population activity data.

To derive an ultra-high-resolution $PM_{2.5}$ distribution with better accuracy, auxiliary data sets have been fused with the observation data in recent studies, such as satellite measurements, CTM outputs, and meteorological variables[43,44]. However, the model resolution can sometimes still be limited, and some ultra-high-resolution variables (30-m land use or elevation) had to be resampled to a coarser level (i.e., 1 km) for match to avoid additional uncertainties[45,46]. Until recently, the availability of a vast amount of ultra-high-resolution data[47] like aerosol optical depth (AOD) data from Sentinel-2A satellite and top of atmosphere (TOA) data from Landsat-8 satellite[48,49] has made it no longer an obstacle for air quality applications. The linear land use regression (LUR) model has been used to fuse these auxiliary variables[50–53]. However, as the relationship between $PM_{2.5}$ and auxiliary variables is usually nonlinear and complex[54], the machine learning methods were introduced, achieving better performance and considerable extrapolation robustness[55–57]. The incorporation of both multiple ultra-high resolution data and machine learning algorithms is more likely to gain better performance than taking only a single one of them.
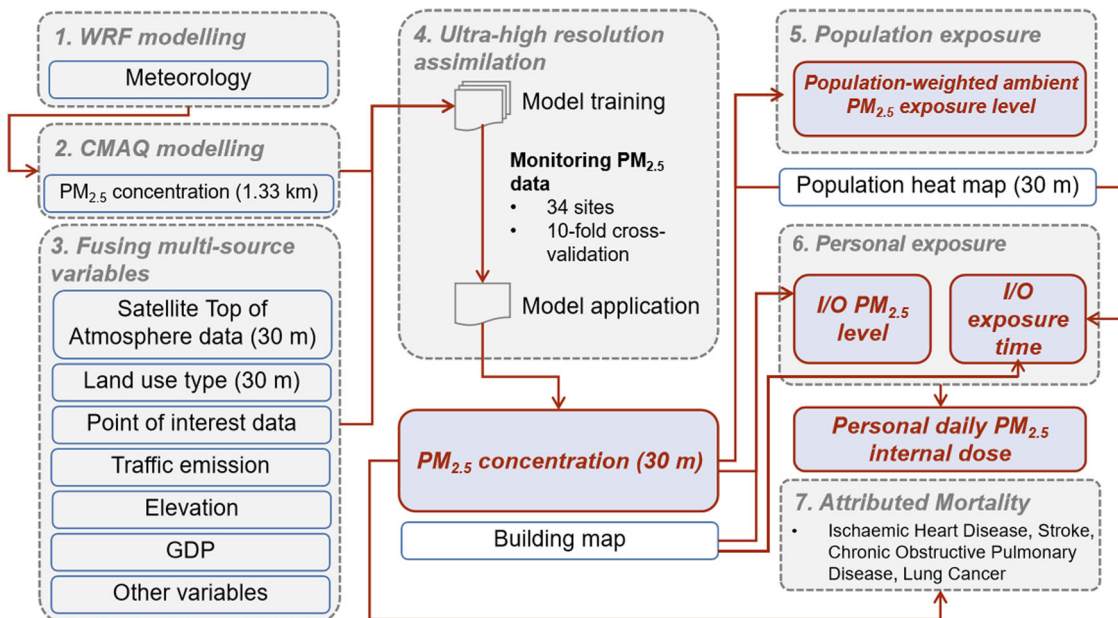
Here, the overall research process of this study is briefly shown in Fig. 1, including performing the CTM model, ultra-high resolution assimilation, population exposure, personal indoor exposure, and health risk assessment. By fusing ultra-high-resolution multi-source auxiliary data, we simulated the street-level distribution of $PM_{2.5}$, and combine it with the ultra-high-resolution population activity data to calculate the personal internal dose based on considering the indoor/outdoor (I/O) exposure differences. The

mortality burden assessment was performed to show the difference between the city-level ambient $PM_{2.5}$ distribution derived from 30-m and 1-km resolution result without any further consideration of the I/O exposure patterns, for the existing exposure-response relationship was only based on the ambient $PM_{2.5}$ level. The potential of this method for exposure inequality identification and health risk assessment was discussed.

Results and discussions

Ultra-high resolution $PM_{2.5}$ mapping in Beijing. Assimilated by other auxiliary variables, the ultra-high resolution $PM_{2.5}$ concentration distribution was mapped throughout Beijing based on the WRF-CMAQ simulation output (annual average in Fig. 2 and seasonal result in Supplementary Fig. 3). With the increase of the resolution and the monitoring-data-based assimilation, the 30-m mapping achieved higher accuracy ($R^2 = 0.78–0.82$) to the national monitoring data, compared with the WRF-CMAQ output ($R^2 = 0.31–0.64$), resulting in the annual average increased from 30.9 μg/m^3 to 34.3 μg/m^3. In the $PM_{2.5}$ concentration probability density distribution curve, the peaks at 10–30 μg/m^3 and 55–70 μg/m^3 intervals moved to the 20–40 μg/m^3 and 55–70 μg/m^3, respectively. It indicated that the overestimation in the southeastern and underestimation in the northwestern were both greatly amended. Area (i) showed a 30.84% increase, and area (iv) showed a 16.92% decrease. The areas (ii) and (iii) with moderate concentration variations but with local hotspots effectively identified could not be achieved by the WRF-CMAQ model on its own. Furthermore, the original concentration probability density distribution curve from WRF-CMAQ was quite zigzag, while became smoother after improving the spatial resolution, for the changes between the adjacent coarse grids are more finely transited by the smaller grids. It also showed that while maintaining the overall spatial distribution trend, the description of far more detailed information on a finer scale was effectively achieved after conducting the comparison result of two-dimensional discrete Fourier transform and high pass filtering on the $PM_{2.5}$ concentration distribution picture (Supplementary Method; Supplementary Discussion 2; Supplementary Fig. 4).

As the output of the traditional method and one of the auxiliary variables of the new method, the WRF-CMAQ model owns deemed acceptable performance (detailed in Supplementary Table 1) judged by the simulation performance criteria[58]. Compared with the LUR model, the RF model had better performance with higher R^2 and lower RMSE (Supplementary Fig. 5) and was thus selected for further downscaling and assimilation. The feature importance of the auxiliary variables in the RF model is shown in Supplementary Fig. 6. The $PM_{2.5}$ from WRF-CMAQ and planetary boundary layer height (PBLH) always played vital roles, followed by the meteorological variables like surface temperature (SFCTMP), wind speed at the altitude of 10 m (WSPD10), and relative humidity (RH). The high ranking of PBLH shows that the concentration of $PM_{2.5}$ in Beijing had not fallen low enough to decouple the influence of unfavorable meteorological factors, thus owning a high possibility of vertical aggregation to cause heavy pollution. Apart from that, the high contribution of RH and low contribution of WSPD10 in winter indicated that low diffusivity in the horizontal direction and high RH always triggered the secondary formation of $PM_{2.5}$, which is consistent with previous studies[59,60]. The SFCTMP contributed the highest for spring and summer, mainly caused by the high radiant fluxes, possibly leading to the intensified formation of secondary pollution[61]. Contributions of other auxiliary variables such as TOA, land use type, population, and road emission varied in different seasons. The feature importance of TOA is relatively lower since it only has a monthly resolution compared to the

Fig. 1 The overall research process of this study. The red part refers to the output dataset from this study, and the gray part refers to the input variables or the processing progress.
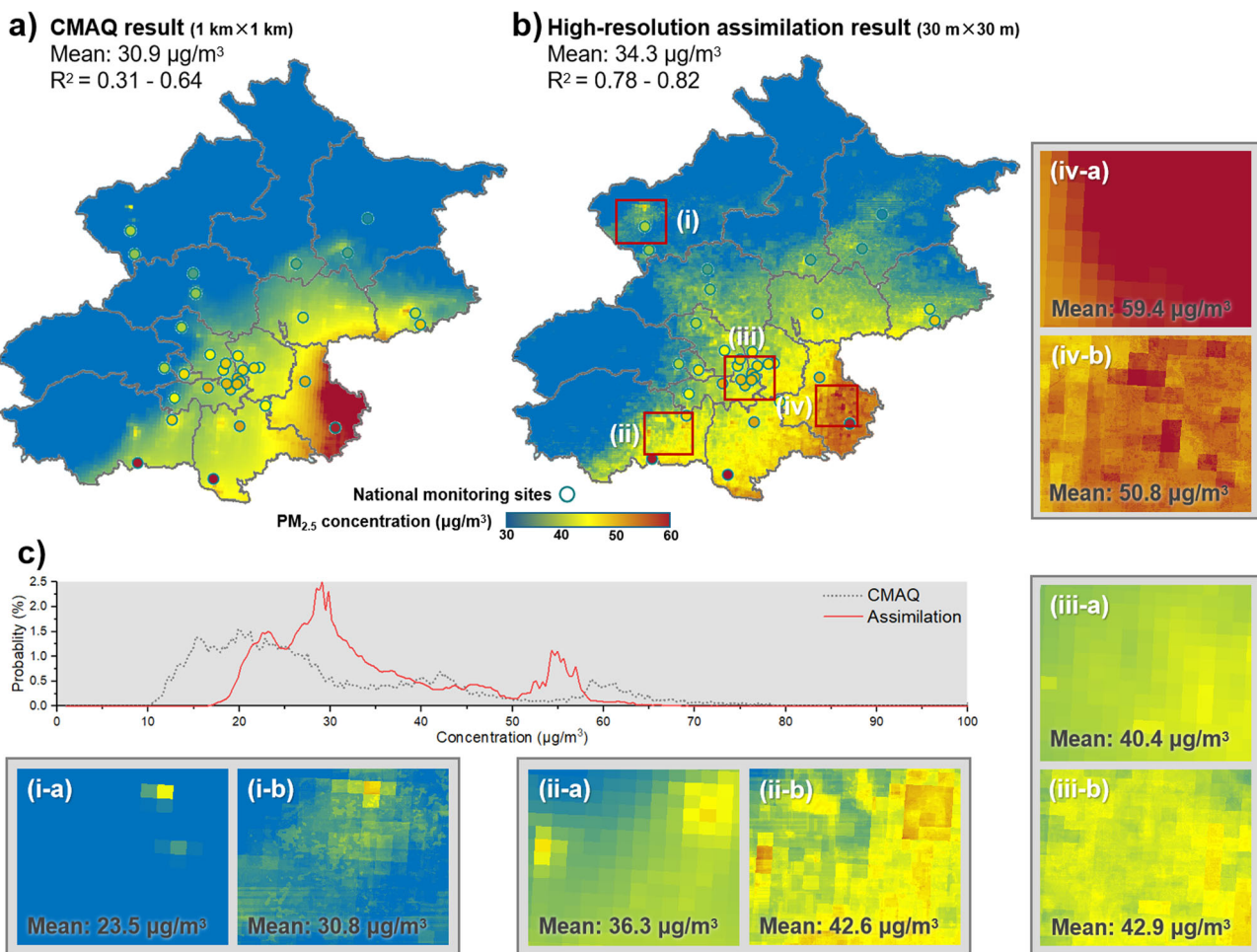
Fig. 2 Comparison of simulated PM$_{2.5}$ concentration before and after assimilation. a, b Model performance and mapping comparison among Beijing from WRF-CMAQ model, our method, and monitoring data, with 4 extracted local areas (i) – (iv) for detailed difference. **c** Probability distribution comparison of PM$_{2.5}$ concentration intervals.
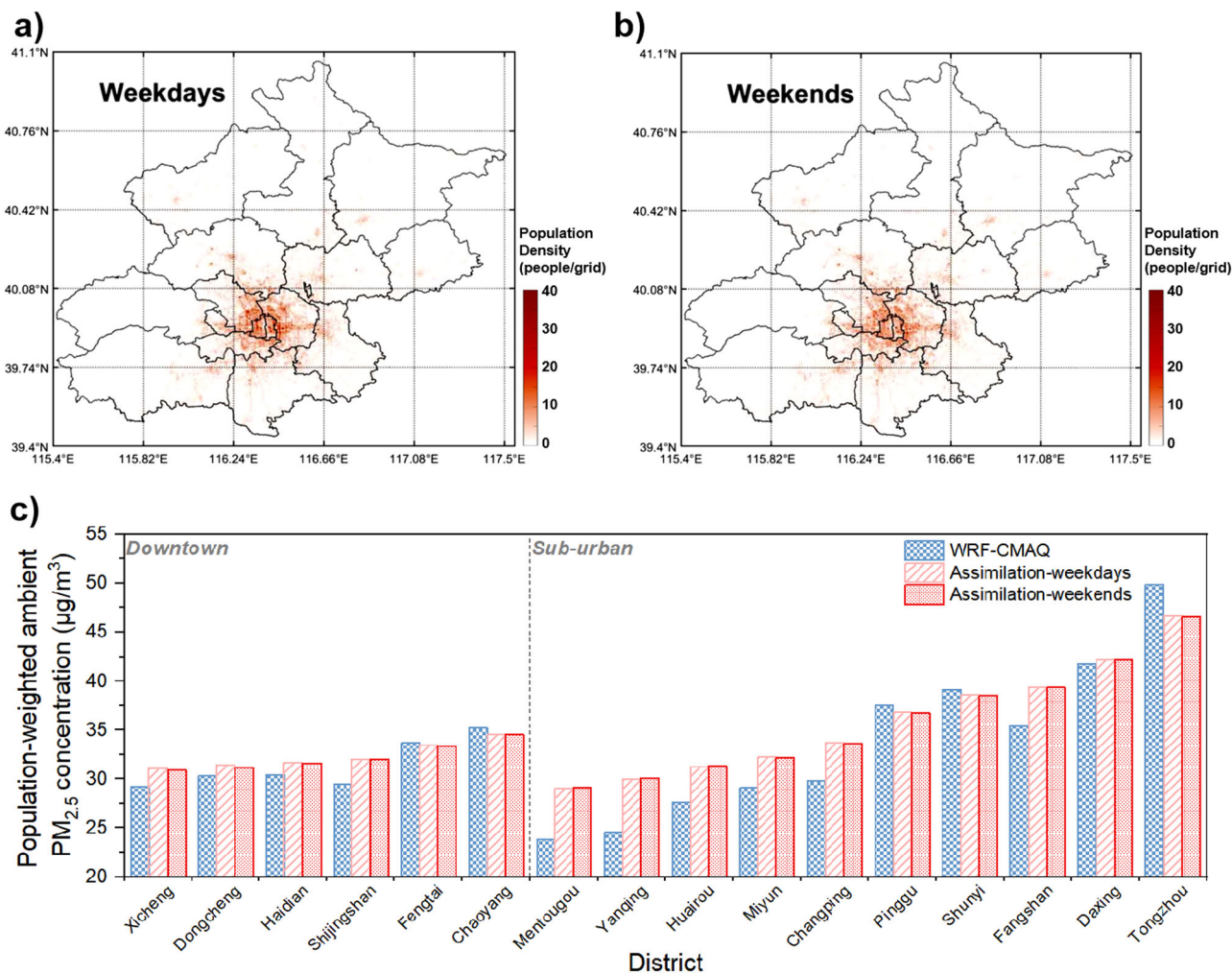
Fig. 3 Population distribution and PM$_{2.5}$ exposure pattern in Beijing. a, **b** population distribution heat map during weekdays and weekends. **c** Population-weighted ambient PM$_{2.5}$ concentration of 16 Beijing districts during weekdays and weekends.

hourly-resolution 1 km-PM$_{2.5}$ concentration and meteorological variables. The original raw TOA data has a higher temporal resolution of 10–15 days. The feature importance of TOA is relatively low due to its lower temporal resolution compared to the hourly-resolution 1 km-PM$_{2.5}$ concentration and meteorological variables. Although the original raw TOA data should be able to support higher temporal resolution (10–15 days), it is prone to abnormal values caused by cloud cover or high reflectance underlying surfaces. Therefore, to balance between temporal resolution and data accuracy, we use the median of valid data in each grid within a month as the representative TOA value. However, as the study period extends, the feature importance of TOA is expected to increase due to its close relationship with PM$_{2.5}$ concentration[62–67].

Spatial-temporal PM$_{2.5}$ exposure level for the population. Population distribution and PM$_{2.5}$ exposure pattern of 16 Beijing districts during weekdays and weekends are shown in Fig. 3. The population density was higher in urban area during weekdays than in weekends, as indicated in sub-figure (a, b) in Fig. 3. The population-weighted ambient PM$_{2.5}$ concentration of the whole Beijing area was 34.6 μg/m³ during weekdays and 34.5 μg/m³ during weekends, which increased by about 15% compared with the unweighted averaged ambient PM$_{2.5}$ level. Adopting the ultra-high resolution PM$_{2.5}$ concentration field produced a significant difference in the average concentration of population exposure

than the WRF-CMAQ model. The main difference of population-weighted ambient PM$_{2.5}$ concentration between the WRF-CMAQ and assimilation result is attributed to the amendment of estimation bias of the WRF-CMAQ result. For the northwestern district with low PM$_{2.5}$ levels, like Shijingshan and Haidian in the urban area and Mentougou, Yanqing, Miyun, Huairou, Changping and Fangshan in sub-urban area, the assimilation lifted the estimated results. In contrast for Tongzhou and Pinggu, the estimated level was lowered. For the other districts in central and southern Beijing, the estimated PM$_{2.5}$ level was moderate and changed little from the WRF-CMAQ to the assimilated result.

The population-weighted PM$_{2.5}$ concentration shows almost no significant difference between weekdays and weekends, indicating a relatively similar overall population distribution pattern. Nonetheless, the exposure surroundings of the population still exist significant difference during weekdays and weekends. Figure 4 shows an example of the 30-m resolution Population distribution difference between weekdays and weekends. Taking the Beijing China World Trade Centre adjacent as an example, the population were gathered more in office buildings, streets and subway stations during weekdays, while more in parks, residence areas and railway stations during weekends. Such a difference in exposure pattern could not be reflected using merely the population-weighted ambient PM$_{2.5}$ concentration, and the consideration of exposure micro-environment should be further taken into account.
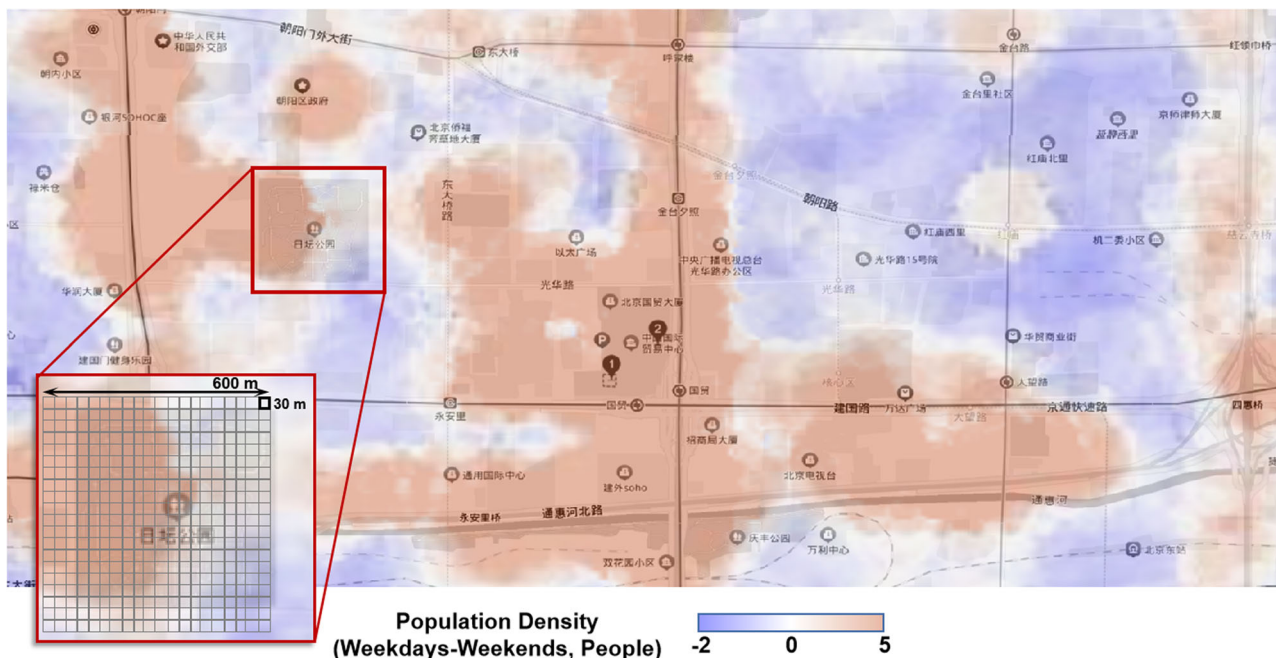
Fig. 4 Population distribution difference between weekdays and weekends using 30-m population heat map and building map. Taking the Beijing China World Trade Centre adjacent as an example.

Personal PM$_{2.5}$ internal dose assessment. The average indoor PM$_{2.5}$ concentration in Beijing was 26.5 μg/m^3, ranging from 10.5 μg/m^3 to 39.5 μg/m^3. The PM$_{2.5}$ exposure level considering the I/O difference within the Fifth Ring Road is shown in sub-figure (a) in Fig. 5, 26.9 μg/m^3 indoor and 41.6 μg/m^3 outdoor on average. As the indoor PM$_{2.5}$ level was estimated by applying the I/O ratio to the ambient level, its trend stayed consistent with the ambient PM$_{2.5}$ map, high in the southeast and low in the northwest. The indoor PM$_{2.5}$ level within the Third Ring Road was moderate and showed no significant difference, but outside the Third Ring Road, the indoor PM$_{2.5}$ level showed an increase in the southeast while a decrease in the north and west. Compared with the result of 57.6 μg/m^3 from Yang et al.'s study[68] in 2013–2014 and 38.6 μg/m^3 from Zuo et al.'s study[69] in 2017, the indoor PM$_{2.5}$ concentration in this study had dropped a lot as a consequence of air quality improvement. However, it was still over-high compared to indoor PM$_{2.5}$ in other developed countries. For instance, the indoor PM$_{2.5}$ level was mostly lower than 10 μg/m^3 in England[70]. The indoor environment is one of the main exposure surroundings for most Chinese people, so there is still a great urgency to ease the indoor health burden in our country.

The difference between population distribution and indoor concentration distribution indicated the existence of exposure inequality in the population. On weekdays, people were populated in buildings within the Fourth Ring Road, much higher than the population density in areas outside the Fourth Ring Road. On weekends, the population density in the urban area decreased, and the population spread wider in the sub-urban area. High indoor exposure concentration means that there might be a high exposure risk to the individual health. However, thanks to the low population density in such areas, they would not have a sufficient impact on the population. Conversely, areas with moderate pollution levels owned lower individual exposure risk, but because of the high population density, its impact on the health of the population is more significant. In order to control and reduce exposure risk, the priority of individual and population exposure risk should be further considered during the decision progress. Generally, individual exposure risk evidence was used to formulate the exposure-response relationship and further for formulating the environmental criteria, while the exposure risk for the whole population was taken to formulate the environmental standard. In our study, the exposure risk of the whole population depends mainly on the population distribution. The mobility of the people means that the exposure risk of the whole population was also spatially and temporally dynamic, suggesting the theoretical possibility of dynamic environmental standards can be formulated.

The spatial distribution of age-standardized personal PM$_{2.5}$ internal dose is shown in sub-figure (b) in Fig. 5, developed by the I/O PM$_{2.5}$ exposure level in sub-figure (a). People within the Fifth Ring Road suffered 24.6 μg/h PM$_{2.5}$ from ambient exposure, 22.6 μg/h from residential indoor exposure, and 16.0 μg/h from public indoor exposure on average. The distinguishing of indoor and outdoor exposure allowed us to calculate possibly more accurate personal daily PM$_{2.5}$ internal dose combined with the 30-m population distribution data. The age- and gender-standardized daily PM$_{2.5}$ internal dose was 568.2 μg/d for a single person based on the ambient WRF-CMAQ result, while 594.5 μg/d based on the ambient assimilation result, which increased by about 5%. However, after considering the I/O difference in the exposure level and population distribution, the age-standardized daily PM$_{2.5}$ internal dose was 512.9 μg/d, which decreased significantly by 14%. More specifically, people of different age or gender groups were exposed to different PM$_{2.5}$ doses, as shown in sub-figure (c) in Fig. 5. Males inhaled more PM$_{2.5}$ dose than females because of larger inhalation volume[71], about 120–150 μg/d above 10-year-old people, while 20 μg/d among children. Estimated with the assimilation result considering the I/O exposure difference, the internal dose of children under 10 was about 250 μg/d. The internal dose of 10 to 64-year-old people was highly similar and the highest during the lifetime, up to 450–600 μg/d, and then began to drop significantly after 65 years old, to about 350–450 μg/d.

Further sensitivity analysis was performed in Supplementary Table 4. As internal doses are divided into indoor and outdoor
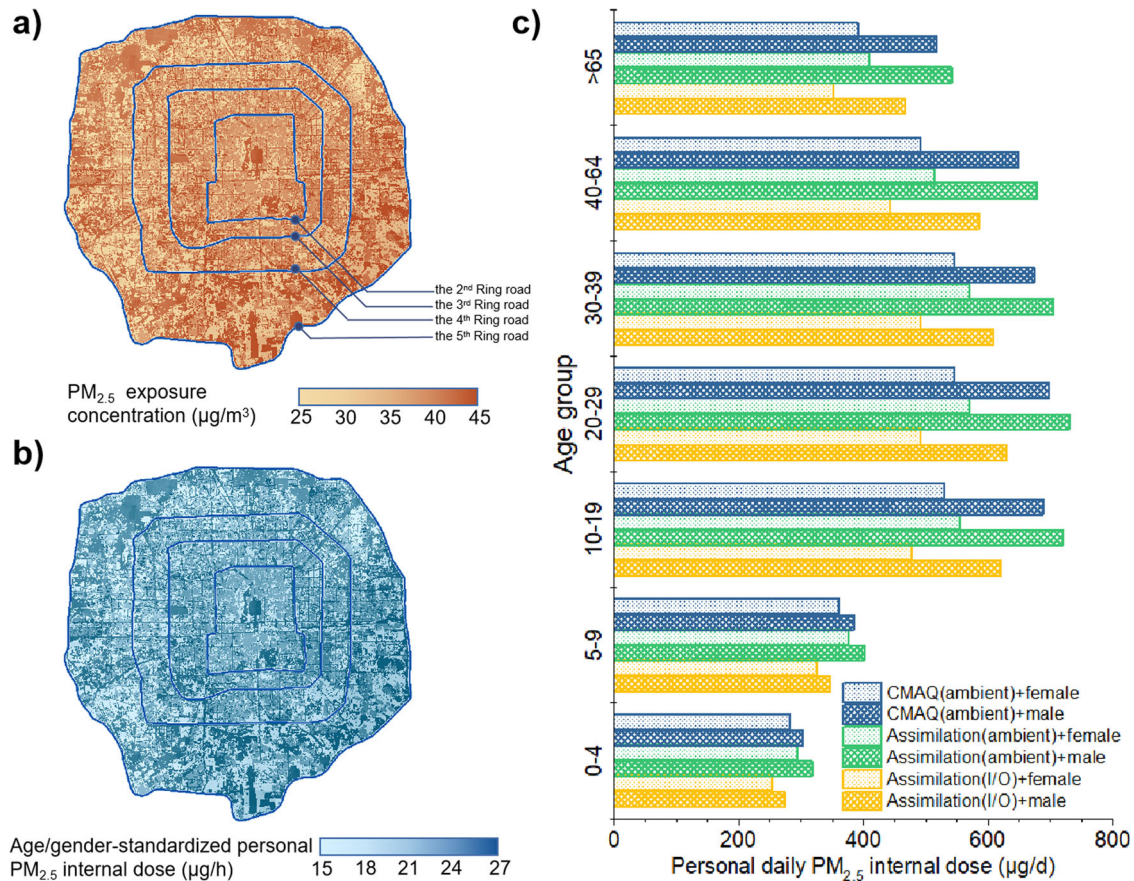
Fig. 5 PM$_{2.5}$ exposure pattern of the population within the Fifth Ring Road area in Beijing. a PM$_{2.5}$ exposure concentration distribution considering land use type and PM$_{2.5}$ I/O ratio; (**b**) age- and gender-standardized personal hourly PM$_{2.5}$ internal dose distribution; (**c**) comparison of age- and gender-specific average personal daily PM$_{2.5}$ internal dose basing on merely ambient WRF-CMAQ result, ambient assimilation result, and assimilation result considering the I/O exposure difference.

parts, the sensitivity of total internal dose to the I/O ratio variation demonstrates a sublinear relationship. Within the normal long-term fluctuations, the I/O ratio of residential and public building area changes by 0.2 independently, or fluctuate by ±0.1 simultaneously, the variation in internal dose both does not exceed ±10%. Fluctuations in the population distribution ratio between residential and public buildings do not have a significant impact on the internal dose (within ±2% under ±0.1 ratio changes). All these sensitivity analyses indicated the adoption of the PM$_{2.5}$ I/O ratio was reasonable with low estimating bias.

Potentials for mortality burden assessment. The substantial improvement of ambient PM$_{2.5}$ resolution will also largely impact the attributed mortality assessment results. By applying the same GEMM model to the annual WRF-CMAQ result and the ambient high-resolution assimilation result, the estimated annual mortality of IHD, stroke, COPD, and LC disease attributed to PM$_{2.5}$ is shown in Fig. 6. The total mortality (with 95% CI) of the 4 specific health endpoints estimated by WRF-CMAQ result was 20540 (16908-24086) people, while for high-resolution assimilation result was 25462 (20901-29881) people. Stroke was the health endpoint with the highest risk of PM$_{2.5}$ exposure, contributing to almost half of the total attributed mortality, while the lowest was for LC. Overall, the method improved the level of PM$_{2.5}$ by 25.9%, causing the mortality estimation to increase by 24.0%, indicating that the accuracy and resolution improvement of developing ambient PM$_{2.5}$ level may also lead to slightly higher mortality

results. It was worth noticing that there is currently no epidemiological evidence to evaluate the pros and cons of using a coarser or higher resolution PM$_{2.5}$ distribution in the mortality burden assessment, not to mention the trade-off between environmental concentration and internal dose. However, in this study, we calculated a series of indicators, including the averaged value, population-weighted averaged concentration, personal internal dose, and mortality estimation within the same framework. The evaluation results showed significant differences between the coarser- and higher-resolution result. This shows that the choice of the modelling resolution might have potential impact on health assessment[72].

In the previous cohort studies, many methods were adopted to derive the PM$_{2.5}$ field because the pollution data were usually not recorded together with the occurrence of health endpoints. The simplest method was directly using the one-hand data, such as the monitoring data from the nearest in-situ stations[73] or the spatial interpolation method[74]. Some studies also use remote satellite data to derive PM$_{2.5}$ field[75]. However, the spatial resolution of the above method was too coarse, usually 10 km level. Many studies used the CTMs to derive PM$_{2.5}$ field at the 1 km resolution[76,77], while still too coarse for human activities. To date, some street-level models have been adopted to derive finer PM$_{2.5}$ distribution in order to match with the scale of the activity pattern. For example, the LUR model to derive PM$_{2.5}$ at low concentration levels (<30 μg/m^3)[78] might face limitations in high-level regions like Beijing. Moreover, some multi-scale integrated model systems were also adopted[79], integrating the large and
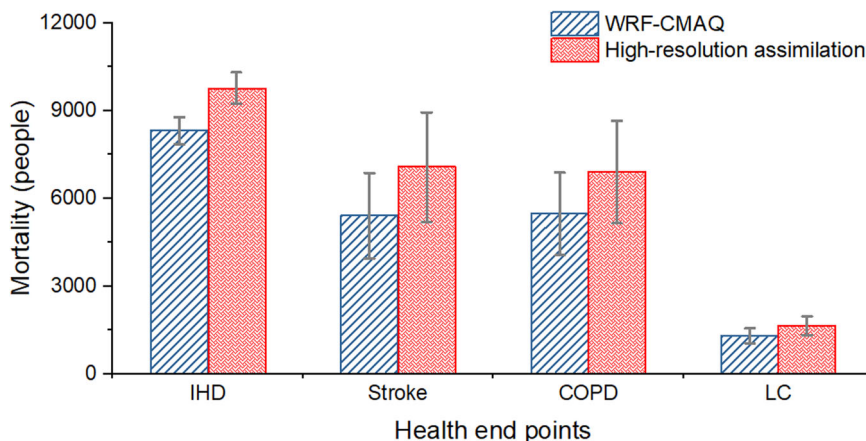
Fig. 6 Annual mortality of four certain health endpoints attributed to PM$_{2.5}$ in Beijing, 2019. Annual mortality estimation (with 95% confidential intervals) of ischaemic heart disease, stroke, chronic obstructive pulmonary disease, and lung cancer disease attributable to PM$_{2.5}$ in Beijing based on WRF-CMAQ and high-resolution assimilation results.

medium-scale CTMs with small scale simulation model under the AirGIS framework. Though with good performance and accuracy, the such kind integrated model relies highly on complex city-level databases and costing enormous computation resources, limiting its application in other cities. Our study provides a 30-m resolution model with good performance ($R^2 = 0.78$–0.82) while moderate complexity. It is likely to bring new understanding to the exposure characteristics, making it potential to be adopted in more accurate exposure assessments in the future.

Uncertainties and perspectives. Still, there are several uncertainties and limitations in this study. The primary source of uncertainty comes from the WRF-CMAQ air quality modeling and RF regression, both of which introduced uncertainties to a different degree, though with acceptable intervals. Furthermore, assumptions were made that the available heat maps accurately represented actual activity patterns of the Beijing population, leading to the underestimation of non-permanent residents in this study. Since various age groups (e.g., elderly individuals) exhibit significantly different activity patterns, they spend varying amounts of time in different areas of buildings. Given that our population activity data cannot be stratified by age, there exists an additional level of uncertainty regarding the exposure of specific population sub-groups. During the estimation of indoor concentration based on the outdoor PM$_{2.5}$, only the I/O ratio without air cleaner was adopted, given the potential complexity of real-life situations and data limitation. Our framework may still underestimate indoor PM$_{2.5}$ concentrations in buildings with complex indoor sources. Studies conducted in environments such as sports arenas and large shopping malls have revealed significant variations in PM$_{2.5}$ I/O ratios, sometimes exceeding 1[80]. Indoor air cleaner plays an essential part in human indoor exposure, there are also available PM$_{2.5}$ I/O ratio with air cleaner working[80], but currently, there remains no datasets capable for further consideration. Another limitation of the I/O ratio approach is its inability to account for the impact of meteorological factors on I/O PM$_{2.5}$ transport. Researches on office buildings has indicated that factors such as humidity and wind speed can influence the I/O ratio of PM$_{2.5}$. For example, an increase in humidity from less than 40% to 90% can lead to a decrease in the I/O ratio by up to 0.2, and changing of wind speed from less than 1 m/s to over 6 m/s can result in an increase in the I/O ratio of approximately 0.15[81]. Researches also suggested that wind direction can also be a significant influencing factor. However, quantifying the influence of meteorological factors on the I/O ratio is challenging,

particularly in densely populated urban areas like Beijing. Street canyon effects, where meteorological conditions within street canyons can significantly differ from those in the urban canopy layer, further complicate the description of short-term I/O ratio variations[82]. Some studies have constructed complex simulation models based on the ventilation characteristics of buildings together with consideration of indoor sources[83]. However, due to data availability constraints, applying these models within the scope of our research remains considerably challenging. Also, this study only provided daily personal indoor PM$_{2.5}$ internal doses in downtown area, for the I/O activity pattern may exist high uncertainty in the sub-urban area with sparse building density.

With further improvements in future air quality, the regional transmission of PM$_{2.5}$ will weaken, while the contribution of local emission sources to PM$_{2.5}$ will increase, leading to an intensification of the spatiotemporal heterogeneity of PM$_{2.5}$ distribution and population exposure. Therefore, it will be necessary to integrate population exposure assessment methods more closely with the effects of local sources. Currently, researches in this area are still limited by technical methods and data availability. Therefore, some prospects have been put forward: (1) The 30-m resolution PM$_{2.5}$ concentration distribution will provide basic data for environmental health assessments; (2) Such method can be applied to multi-year assessment to explore the changing trends in the spatial distribution heterogeneity of PM$_{2.5}$ at the urban scale; (3) This study has proposed a technical framework based on exposure-dose relationship, and future research can further refine in all aspects of this assessment method, e.g. population activity, exposure patterns and I/O PM$_{2.5}$ exposure differences; (4) China has begun to formulate environmental air quality baselines, and understanding the exposure-response relationship based on PM$_{2.5}$ internal dose will provide the most reasonable scientific basis for the baseline formulation.

Methods
Study domain and CTM model configuration. The CTMs used in this study were: The Weather Research and Forecasting (WRF) model (version 3.8.1), and the Community Multi-Scale Air Quality (CMAQ) model (version 5.2), which were developed by U.S. National Centre for Atmospheric Research (NCAR) and the U.S. Environmental Protection Agency (EPA), respectively. To simulate PM$_{2.5}$, the WRF–CMAQ system was applied during January, April, July, and October in 2019, with three days of spin-up time for each run, representing corresponding seasons, respectively[84]. To further evaluate the representativeness of the
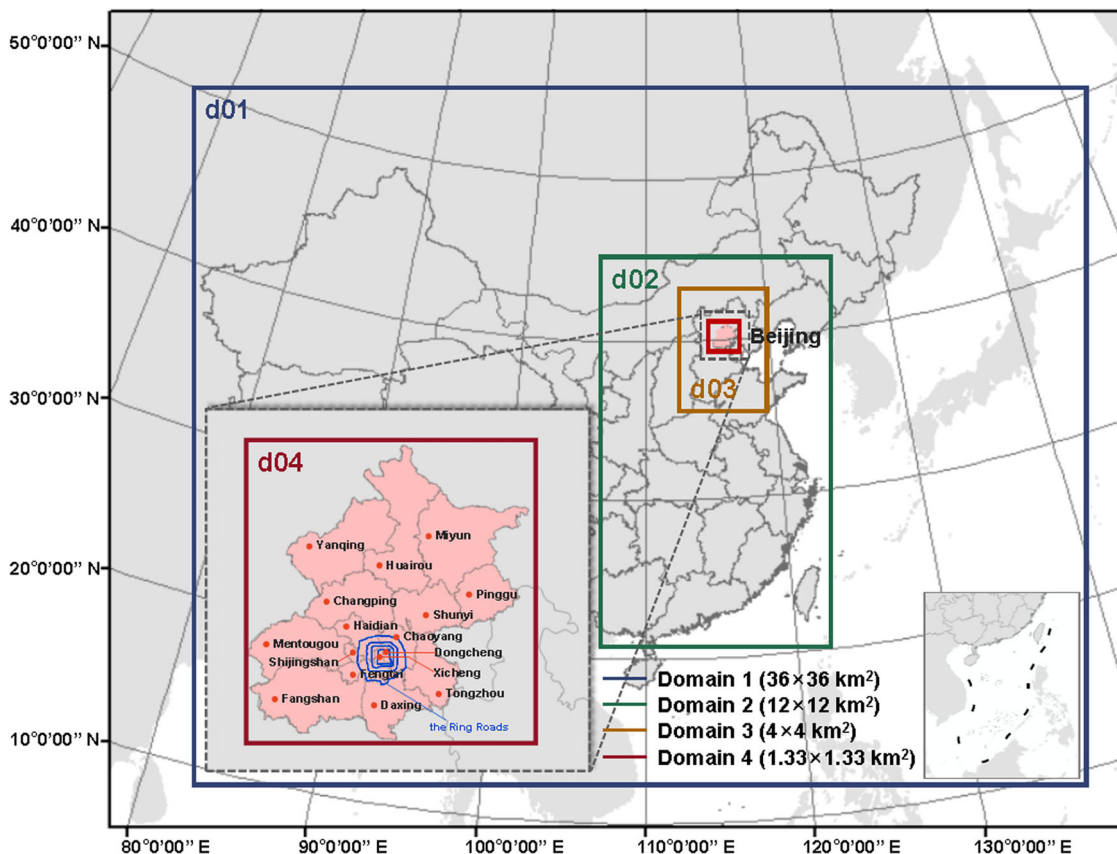
Fig. 7 Study domains of 4-level CMAQ modeling and distribution of Beijing districts and the Ring Roads.

specific months for the corresponding seasons, the similarity of the average value for specific months and corresponding seasons of $PM_{2.5}$ and meteorology conditions in Beijing, 2019 is shown in Supplementary Fig. 1. It exhibits good consistency for the vast majority of the time and monitoring locations, indicating that it is reasonable to use the pollution level of a specific month to represent the whole corresponding season. Beijing, one of the megacities in China, has been taken as the study area. As shown in Fig. 7, the largest modeling domain (d01) covered the area of East Asia and West Pacific, with a temporal resolution of $36 \times 36 \text{ km}^2$. The land-based anthropogenic emissions inventory for mainland China was from the Multi-resolution Emission Inventory for China (MEIC) data at a resolution of $0.25° \times 0.25°$ for the base year of 2015 (MEIC, http://www.meicmodel.org/, last access: 25 October 2018). Base emission of Beijing was gridded as the local emission sources input of the CMAQ model. Here, the local sources refer to the outdoor emission sources within the corresponding modeling domain, with no consideration given to indoor emission sources.

The first guess field and boundary conditions for WRF were generated from the 6 h NCEP FNL Operational Model Global Tropospheric Analyses dataset. The four-dimensional data assimilation (FDDA) was enabled using the NCEP ADP global surface and upper air observational weather data (http://rda.ucar.edu, last access: 25 October 2018). These datasets include all the necessary meteorological parameters required by the WRF model. CMAQ was initialized using the profile file output by the ICON module as the first guess field and boundary conditions, and was pre-run for a period of three days prior to the start date to develop a sufficiently precise monitoring filed. WRF and CMAQ used 32 vertical layers up to 100 hPa, and the lowest layer had a thickness of approximately 37 m. The modeling field and the

boundary condition were transferred into a smaller modeling domain with finer resolution (d02, $12 \times 12 \text{ km}^2$) and started another repeated calculation, same as the 3rd (d03, $4 \times 4 \text{ km}^2$) and 4th domain (d04, $1.3 \times 1.3 \text{ km}^2$). Domain 1 covers a larger region in East Asia than the entire country of China. Therefore, the boundary conditions for China were directly derived from the initial meteorological fields provided by WRF. The boundary conditions of CMAQ and the emission inventory follows the same logic as that of WRF in Domain 1. As for the nested grids within the inner layers, the boundary conditions of WRF, the emission and CMAQ for Domain 2 were all derived from the corresponding grid in its parent domain. The same procedure was applied for Domain 3 and 4. In the 4th domain, the Single-Layer Urban Canopy Model (SLUCM)[85] was coupled with the Noah land-surface model to improve meteorological predictions in the urban area[82]. It assumes the geometric structure of the city is an infinitely long street canyon, and considers the shadowing effect, radiation capture effect and surface reflection effect of the buildings in the street canyon, and specifies the wind profile index, which includes more than 20 parameters such as building height, road width, anthropogenic heat, urban area ratio, and surface albedo.

Ultra-high-resolution assimilation approach. Since a large proportion of $PM_{2.5}$ in Beijing comes from regional transportation and secondary generation, the level of $PM_{2.5}$ at the county scale is relatively similar[86–88]. The difference of $PM_{2.5}$ in the ultra-fine scale is mainly caused by the spatial-temporal distribution of local emission sources. This difference can be reflected through land use information when there is no high-precision data on sources and sinks. Thus, for assimilation, the

auxiliary data set combined 9 major type variables as listed in Supplementary Table 2, containing monitoring data from the national monitoring network, 30-m land use type and satellite data, 1 km-$PM_{2.5}$ level and meteorology derived from CTM results, as well as point of interests (POI), building location, population distribution, traffic emission, and other variables. Ambient air pollution measurements have been conducted routinely by the China Environmental Monitoring Centre (CEME) and Beijing Environmental Monitoring Centre (BJMEMC) since 2013. Hourly $PM_{2.5}$ concentrations were available from 34 of the 35 sites in Beijing (http://113.108.142.147:20035/emcpublish/ and http://zx.bjmemc.com.cn/) from 1 November 2018 to 1 November 2019. The satellite data taken in this study is the top of atmosphere reflectance (TOA) of band 2 of Landsat 8 with 30-m spatial and 16 day-temporal resolution. In order to obtain the complete TOA image covering the whole Beijing area and avoid the influence of extreme values produced by measurement anomalies, the median of the validate TOA value within each 30×30 m^2 grid was chosen and stitched as a whole map, providing a relative spatial trend of monthly concentration distribution at the ultra-high resolution. We realized that the temporal resolution of the satellite TOA data might be low for $PM_{2.5}$ representativeness, but the inclusion of the hourly WRF/CMAQ simulation result, to some extent, had solved the problem, as suggested earlier in the literature[89–91]. The satellite data provide ultra-high-resolution distribution trend of $PM_{2.5}$, and the WRF/CMAQ simulation result provide its temporal changing characteristics. Further information supporting the adoption of satellite data for $PM_{2.5}$ simulation can be found in Supplementary Discussion 1. POI numbers, road line length, and emission intensity were generated from a 1000 m-buffering area around the central 30-m grid to make it more effective in models. All data with a resolution coarser than 30 m were further divided into finer meshes with the same properties.

The LUR and RF models were trained and tested to compare the regression performance of the multi-variable dataset. A supervised forward stepwise linear regression was used to develop the LUR model to maximize the adjusted R^2 value. Due to the total iteration steps were limited by the number of variables, the model was considered to have approximately converged when increasing gradient of R^2 is not greater than 1%. The RF model was performed using an optimized integrated-tree model, which could be approximately adopted as an RF model. The 10-fold cross-validation and the leave-one-out cross-validation (LOOCV) were taken to evaluate model performance. For 10-fold validation, the training set was divided into 10 random subgroups. One of the subgroups was excluded as a validation set, and the model was recalculated on the remaining nine subgroups. The R^2, root mean squared error (RMSE) values, and the standard deviation of predicted values from the monitoring data in both the LOOCV and the 10-fold cross-validations were compared to show the validation results. The trained model was applied to a monthly auxiliary dataset, which includes meteorological variables and $PM_{2.5}$ distribution output from WRF-CMAQ averaged at hourly resolution, and other variables with monthly or annual temporal resolution.

Statistical analyses were performed in MATLAB R2021a. ESRI ArcGIS 10.3 was used for geospatial extraction of the auxiliary predictors, and final visualizations of the assimilated $PM_{2.5}$ concentrations by mapping.

Activity-adjusted population spatial distribution. The population is exposed to different pollution levels according to their activity patterns in various micro-environment during the daytime. In order to capture the location changes of the population

more precisely, hourly mobile signaling data from Baidu Smart Eye was adopted. Baidu Smart Eye is a commercial geographic intelligence data platform launched by Baidu Maps with the advancement of technology of cell-phone signal system, describing the population density in real-time. The relative crowdedness in each grid was derived by calculating the population proportion of the grid to the whole Beijing area. An assumption was made that the relative crowdedness of the weekdays and weekends remain similar in other periods of the year, which means, the distribution of the population was assumed unchanged for weekdays and weekends respectively. The population distribution was calculated by the relative crowdedness multiplied by the total population from Worldpop (available from https://www.worldpop.org/geodata/). To date, the resolution of the raw mobile signal data reaches only hundreds of meters[92], which is even larger than most building scales and is not sufficient to distinguish the I/O exposure differences. Thus, we have to incorporate an inverse distance weighting (IDW) method to interpolate the population distribution to a 30-m resolution in the sparse data area, so that the grid size can be comparable to the building scale. Similarly, due to the lack of supportive data, this study did not further investigate the spatial heterogeneity characteristics of population structure, although there should be differences in gender and age structure among different social places.

Indoor $PM_{2.5}$ concentration and I/O ratio. The rough resolution of the CMAQ model averages the environmental attributes within 1.33 km. Thus, it is unable to further reflect the differences in indoor and outdoor $PM_{2.5}$ concentration and population distribution at finer scales. Nevertheless, with the newly-derived 30-m resolution $PM_{2.5}$ concentration map, we are able to derive the estimated indoor $PM_{2.5}$ concentration together with the land use type. In the literature, it has been suggested that the indoor $PM_{2.5}$ transported from ambient surroundings can be described by the I/O ratio, infiltration factor, and penetration factor, etc. However, obtaining the infiltration and penetration factors requires a relatively complex model framework, which is difficult to achieve with a comprehensive input dataset covering the entire Beijing area. Additionally, while the I/O ratio is simplified compared to the other mentioned two factors, only a few studies provide representative I/O ratios of the certain kind of buildings[93] rather than the comprehensive geographic distributions. Thus, limited by data availability, only a unified I/O ratio representing a certain type of buildings was selected from the literature after a review, accounting for the study period and purpose, ambient sampling size, $PM_{2.5}$ level, and comparability with this study, shown in Supplementary Table 3. Typically, the I/O ratio of 0.7 in public building area and 0.9 in residential building area were taken in this study[80]. It is worth noticing that the I/O ratio of $PM_{2.5}$ was influenced by many factors, thus there may exists fluctuations, for example, complexity of indoor sources, air cleaner, etc. Several sensitivity analyses were carried out under different scenarios. The sensitivity of changing in the applied $PM_{2.5}$ I/O ratio to the personal daily $PM_{2.5}$ internal dose was listed in Supplementary Table 4, including sensitivity analysis of: (a) distinguishing $PM_{2.5}$ I/O ratios in residential and public buildings, and (b) population distribution ratio between residential and public buildings, to personal daily $PM_{2.5}$ internal dose.

Population-weighted ambient $PM_{2.5}$ concentration. Population-weighted $PM_{2.5}$ concentration is widely used to characterize the collective exposure concentration of a population

for long-term assessment[94,95]. Based on the 30-m PM$_{2.5}$ concentration and the ultra-high spatial-temporal resolution population distribution, the population-weighted ambient PM$_{2.5}$ concentration of each Beijing district was calculated, shown as the following equation:

$$C_{pop} = \frac{\sum_i C_i \times P_i}{\sum_i P_i} \qquad (1)$$

where:

i is a single grid in this district;

C_{pop} is the annual population-weighted ambient PM$_{2.5}$ concentration in a certain district (μg/m³);

C_i is the annual-averaged ambient PM$_{2.5}$ concentration in grid i (μg/m³);

P_i is the daily averaged population in grid i.

Personal indoor PM$_{2.5}$ internal dose calculation. The quantity of PM$_{2.5}$ inhaled by a person into the lung within a specific time (the internal dose) derived from the PM$_{2.5}$ concentration among the external exposure surroundings (the exposure level) can reflect the PM$_{2.5}$ exposure extent of a person. The higher the personal PM$_{2.5}$ internal dose, the greater the exposure risk a person is likely to suffer. Previous studies have highlighted various factors that significantly impact exposure assessments, such as commuting ways[96] and age period[97] that lead to exposure differences. However, due to limited access to fundamental data, we could only use I/O population distribution to characterize the exposure differences, failed to conduct more precise evaluations based on commuting patterns. The 30-m resolution was sufficiently fine in the urban area to classify the I/O exposure difference with high building density. However, in the sub-urban area, 30-m resolution would introduce high uncertainty of the I/O classification due to the small building scale and low building density. Thus, an inhalation model was adopted to develop the personal PM$_{2.5}$ internal dose within the Fifth Ring Road area in Beijing. The age-standardized internal dose of PM$_{2.5}$ was calculated as the following equation:

$$D_i = \sum_{j,k} \overline{P_{j,k}} \times IR_{j,k} \times \overline{C_i} \times t \qquad (2)$$

where:

D_i is the personal internal dose of indoor PM$_{2.5}$ in grid i (μg/d);

i is a single 30-m grid;

j stands for psychological gender, male or female;

k stands for age groups divided into 21 groups (5 years as a group, from 0-4, 5-9 to 95-99 and 100+) from the China Statistical Yearbook[98];

$\overline{P_{j,k}}$ is the average proportion of gender j and age group k, adopted from the China Statistical Yearbook[98];

$IR_{j,k}$ is the inhalation rate of gender j and age group k taken from EPA's Exposure Factors Handbook[71] (m³/h);

C_i is the estimated PM$_{2.5}$ exposure level in grid i (μg/m³). For comparison, the ambient CMAQ result, ambient assimilation result, and the assimilation result considering the I/O exposure difference were all adopted individually.

t is the daily exposure time (a total 24 h). The I/O ratio of the daily average population within the Fifth Ring Road area reflected the possibility of a person being indoors or outdoors; thus was suitable to be taken as the I/O exposure time ratio of the population. The population within a single 30-m grid was identified as indoor if the land use type was building, otherwise as outdoor (as shown in Fig. 2).

For age- and gender-specific personal daily PM$_{2.5}$ internal dose, the proportion of age and gender was set as 100%, and the inhalation rate was set as the typical value of the age and gender group individually. Detailed information on the population proportion of certain age and gender groups and its mean inhalation rate is listed in Supplementary Table 5.

Mortality estimation of specific health endpoints. In order to estimate long-term mortality attributable to PM$_{2.5}$ exposure, the epidemiological hazard index (HI) and hazard risk (HR) have been widely used in epidemiological studies. The equation is as follows:

$$HI = PAF \times m \times EP \qquad (3)$$

where:

HI the health impact of a specific disease during the assessment period, here annual PM$_{2.5}$-exposure-attributable mortality specifically;

m is the age-and-gender standardized cross-sectional mortality (or the so called baseline mortality in the literature) rate of the 25+ population from the GBD study 2019 (available from http://www.healthdata.org/results/data-visualizations);

EP is the exposed population;

PAF (population attributable fraction), calculated by (HR-1)/HR, refers to the potential reduction in morbidity or mortality when the entire population is exposed to the baseline concentration.

The Global Exposure Mortality Model (GEMM) function[95] was widely taken to calculate HR value for estimating long-term PM$_{2.5}$ exposure-attributed disease burden for the 25+ population. Ischaemic heart disease (IHD), cerebrovascular disease (Stroke), chronic obstructive pulmonary disease (COPD), and lung cancer (LC) are among the most important causes of death covered by HR functions. The International Classification of Diseases 10th Revision (ICD-10) codes of the 4 certain health end points stays consistent with the GBD 2019 Cause-ICD Codes Map (https://ghdx.health data.org/record/ihme-data/gbd-2019-cause-icd-code-mappings). The GEMM developed the exposure-response curve from ambient PM$_{2.5}$ level as follows:

$$HR(z) = \exp\left\{ \frac{\theta \log(\frac{z}{\alpha} + 1)}{1 + \exp\{-\frac{z-\mu}{\nu}\}} \right\} \qquad (4)$$

where:

z is the exceedance ambient PM$_{2.5}$ exposure level over the counterfactual concentration (C_0, the threshold concentration below which no additional health impacts are considered);

HR(z) is the hazard ratio of the 25+ population exposed under the exceedance PM$_{2.5}$ exposure level z of PM$_{2.5}$, derived from ambient PM$_{2.5}$ level;

α, γ and δ are parameters used to describe the shape of hazard risk curves of different health outcomes[95].

Details of the parameterization of the GEMM model taken in this study were summarized in Supplementary Table 6. The resolution of the HR, baseline mortality rate and the exposure population was all at the city level, and was all age-and-gender-normalized, here with no specificity across the spatial-temporal dimension. It should be noted that the gridded PM$_{2.5}$ distribution should not be applied to obtain the gridded health risk assessment results. Following the basic developing process and assumptions of the GEMM model, the exposure level scale should be consistent with the baseline mortality rate, here at city-level.

Data availability

Code availability

References

1. Anderson, J. O., Thundiyil, J. G. & Stolbach, A. Clearing the air: a review of the effects of particulate matter air pollution on human health. *J. Med. Toxicol.* **8**, 166–175 (2011).
2. Brook, R. D., Newby, D. E. & Rajagopalan, S. The global threat of outdoor ambient air pollution to cardiovascular health. *JAMA Cardiol.* **2**, 353 (2017).
3. Yin, P. et al. The effect of air pollution on deaths, disease burden, and life expectancy across China and its provinces, 1990–2017: an analysis for the Global Burden of Disease Study 2017. *Lancet Planet. Health* **4**, e386–e398 (2020).
4. WHO. Ambient (outdoor) air pollution (avaliable at: https://www.who.int/data/gho/data/themes/air-pollution). (2022).
5. Yu, P. et al. Cohort studies of long-term exposure to outdoor particulate matter and risks of cancer: A systematic review and meta-analysis. *Innovation* **2**, 3 (2021).
6. Pranata, R., Vania, R., Tondas, A. E., Setianto, B. & Santoso, A. A time-to-event analysis on air pollutants with the risk of cardiovascular disease and mortality: A systematic review and meta-analysis of 84 cohort studies. *J. Evid. Based Med.* **13**, 102–115 (2020).
7. Chen, J. & Hoek, G. Long-term exposure to PM and all-cause and cause-specific mortality: a systematic review and meta-analysis. *Env. Int.* **143**, 105974 (2020).
8. Burnett, R. T. et al. An integrated risk function for estimating the global burden of disease attributable to ambient fine particulate matter exposure. *Environ. Health Persp.* **122**, 397–403 (2014).
9. Burnett, R. & Spadaro, J. Global Mortality and Long-Term Ambient Exposure to Fine Particulate Matter: A New Relative Risk Estimator. *ISEE Conference Abstracts.* (Vol. 2018, No. 1).
10. Lall, R., Kendall, M., Ito, K. & Thurston, G. D. Estimation of historical annual $PM_{2.5}$ exposures for health effects assessment. *Atmos. Environ.* **38**, 5217–5226 (2004).
11. Laden, F., Schwartz, J., Speizer, F. E. & Dockery, D. W. Reduction in fine particulate air pollution and mortality: extended follow-up of the Harvard Six Cities study. *Am. J. Resp. Crit. Care* **173**, 667–672 (2006).
12. Kong, S. et al. Receptor modeling of $PM_{2.5}$, PM_{10} and TSP in different seasons and long-range transport analysis at a coastal site of Tianjin, China. *Sci. Total Env.* **408**, 4681–4694 (2010).
13. Feng, J. et al. The chemical composition and sources of $PM_{2.5}$ during the 2009 Chinese New Year's holiday in Shanghai. *Atmos. Res.* **118**, 435–444 (2012).
14. Cao, J.-J. et al. Winter and Summer $PM_{2.5}$ chemical compositions in fourteen Chinese cities. *J. Air Waste Manag. Assoc.* **62**, 1214–1226 (2012).
15. Pope, C. A. III et al. Lung cancer and cardiovascular disease mortality associated with ambient air pollution and cigarette smoke: shape of the exposure–response relationships. *Environ. Health Persp.* **119**, 1616–1621 (2011).
16. Turner, M. C. et al. Long-term ambient fine particulate matter air pollution and lung cancer in a large cohort of never-smokers. *Am. J. Resp. Crit. Care* **184**, 1374–1381 (2011).
17. Chen, R. et al. Coarse particles and mortality in three Chinese cities: the China Air Pollution and Health Effects Study (CAPES). *Sci. Total Environ.* **409**, 4934–4938 (2011).
18. Cao, J., Xu, H., Xu, Q., Chen, B. & Kan, H. Fine particulate matter constituents and cardiopulmonary mortality in a heavily polluted Chinese city. *Environ. Health Persp.* **120**, 373–378 (2012).
19. Lim, C.-H., Ryu, J., Choi, Y., Jeon, S. W. & Lee, W.-K. Understanding global $PM_{2.5}$ concentrations and their drivers in recent decades (1998–2016). *Environ. Int.* **144**, 106011 (2020).
20. Zhang, Q. et al. Drivers of improved $PM_{2.5}$ air quality in China from 2013 to 2017. *Proc. Natl Acad. Sci.* **116**, 24463–24469 (2019).
21. Kumar, N. et al. Contributions of international sources to $PM_{2.5}$ in South Korea. *Atmos. Environ.* **261**, 118542 (2021).
22. Sarkodie, S. A., Strezov, V., Jiang, Y. & Evans, T. Proximate determinants of particulate matter ($PM_{2.5}$) emission, mortality and life expectancy in Europe, Central Asia, Australia, Canada and the US. *Sci. Total Environ.* **683**, 489–497 (2019).
23. Zhang, Y. et al. Source apportionment of $PM_{2.5}$ pollution in the central six districts of Beijing, China. *J. Cleaner Prod.* **174**, 661–669 (2018).
24. Wang, Y., Shi, M., Lv, Z., Liu, H. & He, K. Local and regional contributions to $PM_{2.5}$ in the Beijing 2022 Winter Olympics infrastructure areas during haze episodes. *Front. Environ. Sci. Eng.* **15**, 1–11 (2021).
25. Wu, J. et al. Insights into particulate matter pollution in the North China Plain during wintertime: local contribution or regional transport? *Atmos. Chem. Phys.* **21**, 2229–2249 (2021).
26. Yang, X. et al. Effectively controlling hazardous airborne elements: insights from continuous hourly observations during the seasons with the most unfavorable meteorological conditions after the implementation of the APPCAP. *J. Hazard Mater.* **387**, 121710 (2020).
27. Eeftens, M. et al. Spatial variation of $PM_{2.5}$, PM_{10}, $PM_{2.5}$ absorbance and PM_{coarse} concentrations between and within 20 European study areas and the relationship with NO_2–Results of the ESCAPE project. *Atmos. Environ.* **62**, 303–317 (2012).
28. Edwards, L. et al. Personal exposure monitoring of $PM_{2.5}$ among US diplomats in Kathmandu during the COVID-19 lockdown, March to June 2020. *Sci. Total Environ.* **772**, 144836 (2021).
29. Zheng, T., Bergin, M. H., Hu, S., Miller, J. & Carlson, D. E. Estimating ground-level $PM_{2.5}$ using micro-satellite images by a convolutional neural network and random forest approach. *Atmos. Environ.* **230**, 117451 (2020).
30. Li, Z., Yang, X., Zhao, C. & Fan, T. Ratio of $PM_{2.5}$ to PM_{10} mass concentrations in Beijing and relationships with pollution from the North China Plain. *Asia-Pacific J. Atmos. Sci.* **57**, 421–434 (2021).
31. Al-sareji, O. J., Grmasha, R. A., Hashim, K. S., Salman, J. M. & Al-Juboori, R. A. Personal exposure and inhalation doses to PM_1 and $PM_{2.5}$ pollution in Iraq: An examination of four transport modes. *Building Environ.* **212**, 108847 (2022).
32. Madureira, J. et al. Assessment of indoor air exposure at residential homes: inhalation dose and lung deposition of PM_{10}, $PM_{2.5}$ and ultrafine particles among newborn children and their mothers. *Sci. Total Environ.* **717**, 137293 (2020).
33. Song, J., Zhou, S., Xu, J. & Su, L. From $PM_{2.5}$ exposure to $PM_{2.5}$ risks of inhaled dose in daily activities: Empirical evidence during workdays from Guangzhou, China. *Atmos. Environ.* **249**, 118224 (2021).
34. Ott, W. R. Total human exposure. *Environ. Sci. Technol.* **19**, 880–886 (1985).
35. Jerrett, M. et al. A review and evaluation of intraurban air pollution exposure models. *J. Exposure Sci. Environ. Epidemiol.* **15**, 185–204 (2005).
36. Pond, Z. A. et al. Cardiopulmonary mortality and fine particulate air pollution by species and source in a national US cohort. *Environ. Sci. Technol.* **56**, 7214–7223 (2021).
37. Xue, T. et al. Long-term $PM_{2.5}$ exposure and depressive symptoms in China: a quasi-experimental study. *Lancet Regional Health-Western Pacific* **6**, 100079 (2021).
38. García-Serna, A. M. et al. Cytokine profiles in cord blood in relation to prenatal traffic-related air pollution: The NELA cohort. *Pediatr. Allergy Immunol.* **33**, e13732 (2022).
39. Diao, M. et al. Methods, availability, and applications of $PM_{2.5}$ exposure estimates derived from ground measurements, satellite, and atmospheric models. *J. Air Waste Manag. Assoc.* **69**, 1391–1414 (2019).
40. Johnson, S., Haney, J., Cairone, L., Huskey, C. & Kheirbek, I. Assessing air quality and public health benefits of New York City's climate action plans. *Environ. Sci. Technol.* **54**, 9804–9813 (2020).
41. Lv, Z. et al. Source–receptor relationship revealed by the halted traffic and aggravated haze in Beijing during the COVID-19 lockdown. *Environ. Sci. Technol.* **54**, 15660–15670 (2020).
42. Luo, Z. et al. Impacts of vehicle emission on air quality and human health in China. *Sci. Total Environ.* **813**, 152655 (2022).
43. Cheewinsiriwat, P., Duangyiwa, C., Sukitpaneenit, M. & Stettler, M. E. Influence of land use and meteorological factors on $PM_{2.5}$ and PM_{10} concentrations in Bangkok, Thailand. *Sustainability* **14**, 5367 (2022).
44. Dai, H., Huang, G., Wang, J., Zeng, H. & Zhou, F. Spatio-temporal characteristics of $PM_{2.5}$ concentrations in China based on multiple sources of data and LUR-GBM during 2016–2021. *Int. J. Environ. Res. Publ. Health* **19**, 6292 (2022).
45. Geng, G. et al. Tracking air pollution in China: Near real-time $PM_{2.5}$ retrievals from multisource data fusion. *Environ. Sci. Technol.* **55**, 12106–12115 (2021).
46. Li, R., Guo, J., Geng, G., Xiao, Q. & Zhang, Q. Satellite-derived long-term estimates of full-coverage PM_1 concentrations across China based on a stacking decision tree model. *Atmos. Environ.* **255**, 118448 (2021).
47. Luo, Z., Xu, H., Zhang, Z., Zheng, S. & Liu, H. Year-round changes in tropospheric nitrogen dioxide caused by COVID-19 in China using satellite observation. *J. Environ. Sci-China* **132**, 162–168 (2023).
48. Roy, D. et al. Conterminous United States demonstration and characterization of MODIS-based Landsat ETM+ atmospheric correction. *Remote Sensing Environ.* **140**, 433–449 (2014).
49. Li, Z., Roy, D. P., Zhang, H. K., Vermote, E. F. & Huang, H. Evaluation of Landsat-8 and Sentinel-2A aerosol optical depth retrievals across Chinese cities and implications for medium spatial resolution urban aerosol monitoring. *Remote Sensing* **11**, 122 (2019).

50. Eeftens, M. et al. Development of land use regression models for PM$_{2.5}$, PM$_{2.5}$ absorbance, PM$_{10}$ and PM$_{coarse}$ in 20 European study areas; results of the ESCAPE project. *Environ. Sci. Technol.* **46**, 11195–11205 (2012).

51. Zhai, L. et al. Land use regression modeling of PM$_{2.5}$ concentrations at optimized spatial scales. *Atmosphere* **8**, 1 (2017).

52. Henderson, S. B., Beckerman, B., Jerrett, M. & Brauer, M. Application of land use regression to estimate long-term concentrations of traffic-related nitrogen oxides and fine particulate matter. *Environ. Sci. Technol.* **41**, 2422–2428 (2007).

53. Wu, C. D. et al. Land-use regression with long-term satellite-based greenness index and culture-specific sources to model PM$_{2.5}$ spatial-temporal variability. *Environ. Pollut.* **224**, 148–157 (2017).

54. Brokamp, C., Jandarov, R., Rao, M. B., LeMasters, G. & Ryan, P. Exposure assessment models for elemental components of particulate matter in an urban environment: A comparison of regression and random forest approaches. *Atmos. Environ.* **151**, 1–11 (2017).

55. Karimian, H. et al. Evaluation of different machine learning approaches to forecasting PM$_{2.5}$ mass concentrations. *Aerosol Air Quality Res.* **19**, 1400–1410 (2019).

56. Sun, H. et al. Spatial resolved surface ozone with Urban and rural differentiation during 1990–2019: a space–time bayesian neural network downscaler. *Environ. Sci. Technol.* **56**, 7337–7349 (2021).

57. Sun, H. Z. et al. Antagonism between ambient ozone increase and urbanization-oriented population migration on Chinese cardiopulmonary mortality. *Innovation* **4**, 100517 (2023).

58. Boylan, J. W. & Russell, A. G. PM and light extinction model performance metrics, goals, and criteria for three-dimensional air quality models. *Atmos. Environ.* **40**, 4946–4959 (2006).

59. Meng, C. et al. Contribution of meteorological factors to particulate pollution during winters in Beijing. *Sci. Total Environ.* **656**, 977–985 (2019).

60. Wang, J. D. et al. Impact of aerosol-meteorology interactions on fine particle pollution during China's severe haze episode in January 2013. *Environ. Res. Lett.* **9**, 094002 (2014).

61. Zhai, B. X., Chen, J. G., Yin, W. W. & Huang, Z. L. Relevance analysis on the variety characteristics of PM$_{2.5}$ concentrations in Beijing, China. *Sustainability* **10**, 3228 (2018).

62. Bai, H., Zheng, Z., Zhang, Y., Huang, H. & Wang, L. Comparison of satellite-based PM$_{2.5}$ estimation from aerosol optical depth and top-of-atmosphere reflectance. *Aerosol Air Quality Res.* **21**, 200257 (2021).

63. Wang, B. et al. Estimate hourly PM$_{2.5}$ concentrations from Himawari-8 TOA reflectance directly using geo-intelligent long short-term memory network. *Environ. Pollut.* **271**, 116327 (2021).

64. Shen, H., Li, T., Yuan, Q. & Zhang, L. Estimating regional ground-level PM$_{2.5}$ directly from satellite top-of-atmosphere reflectance using deep belief networks. *J. Geophys. Res.: Atmos.* **123**, 13875–13886 (2018).

65. Yang, L., Xu, H. & Yu, S. Estimating PM$_{2.5}$ concentrations in Yangtze River Delta region of China using random forest model and the Top-of-Atmosphere reflectance. *J. Environ. Manag.* **272**, 111061 (2020).

66. Mao, F. et al. Estimating hourly full-coverage PM$_{2.5}$ over China based on TOA reflectance data from the Fengyun-4A satellite. *Environ. Pollut.* **270**, 116119 (2021).

67. Liu, J., Weng, F. & Li, Z. Satellite-based PM$_{2.5}$ estimation directly from reflectance at the top of the atmosphere using a machine learning algorithm. *Atmos. Environ.* **208**, 113–122 (2019).

68. Yang, Y. B. et al. Source Apportionment and Influencing Factor Analysis of Residential Indoor PM$_{2.5}$ in Beijing. *Int. J. Environ. Res. Publ. Health* **15**, 686 (2018).

69. Zuo, J. X. et al. Using big data from air quality monitors to evaluate indoor PM$_{2.5}$ exposure in buildings: case study in Beijing. *Environ. Pollut.* **240**, 839–847 (2018).

70. Gonzalo, F. D. A., Griffin, M., Laskosky, J., Yost, P. & Li, M. S. Assessment of indoor air quality in residential buildings of New England through actual data. *Sustainability* **14**, 739 (2022).

71. U.S. EPA. *Exposure Factors Handbook 2011 Edition (Final Report)*. U.S. Environmental Protection Agency, Washington, DC, EPA/600/R-09/052F, 2011.

72. Bai, H., Wu, H., Gao, W., Wang, S. & Cao, Y. Influence of spatial resolution of PM$_{2.5}$ concentrations and population on health impact assessment from 2010 to 2020 in China. *Environ. Pollut.* **326**, 121505 (2023).

73. Chen, X. et al. Long-term exposure to urban air pollution and lung cancer mortality: a 12-year cohort study in Northern China. *Sci. Total Environ.* **571**, 855–861 (2016).

74. Jerrett, M. et al. A Cohort study of traffic-related air pollution and mortality in Toronto, Ontario, Canada. *Environ. Health Persp* **117**, 772–777 (2009).

75. Crouse et al. Ambient PM$_{2.5}$, O$_3$, and NO$_2$ exposures and associations with mortality over 16 years of follow-up in the Canadian census health and environment Cohort (CanCHEC). *Environ. Health Persp.* **123**, 1180–1186 (2015).

76. Hanigan, I. C. et al. All-cause mortality and long-term exposure to low level air pollution in the '45 and up study' cohort, Sydney, Australia, 2006-2015. *Environ. Int.* **126**, 762–770 (2019).

77. Cesaroni, G. et al. Long-term exposure to urban air pollution and mortality in a Cohort of more than a million adults in Rome. *Environ. Health Persp.* **121**, 324–331 (2013).

78. Dimakopoulou, K. et al. Air pollution and nonmalignant respiratory mortality in 16 Cohorts within the ESCAPE Project. *Am. J. Resp. Crit. Care* **189**, 684–696 (2014).

79. Hvidtfeldt, U. A. et al. Long-term residential exposure to PM$_{2.5}$, PM$_{10}$, black carbon, NO$_2$, and ozone and mortality in a Danish cohort. *Environ. Int.* **123**, 265–272 (2019).

80. Deng, G. F. et al. Indoor/outdoor relationship of PM$_{2.5}$ concentration in typical buildings with and without air cleaning in Beijing. *Indoor Built Environ.* **26**, 60–68 (2017).

81. Zhao, L. et al. Influence of atmospheric fine particulate matter (PM$_{2.5}$) pollution on indoor environment during winter in Beijing. *Building Environ.* **87**, 283–291 (2015).

82. Lv, Z. et al. Development and application of a multi-scale modeling framework for urban high-resolution NO$_2$ pollution mapping. *Atmos. Chem. Phys.* **22**, 15685–15702 (2022).

83. Yuan, Y., Luo, Z. W., Liu, J., Wang, Y. W. & Lin, Y. Y. Health and economic benefits of building ventilation interventions for reducing indoor PM exposure from both indoor and outdoor origins in urban Beijing, China. *Sci. Total Environ.* **626**, 546–554 (2018).

84. Zhao, J. et al. Comprehensive assessment for the Impacts of S/IVOC emissions from mobile sources on SOA formation in China. *Environ. Sci. Technol.* **56**, 16695–16706 (2022).

85. Kusaka, H., Kondo, H., Kikegawa, Y. & Kimura, F. A simple single-layer urban canopy model for atmospheric models: comparison with multi-layer and slab models. *Boundary-Layer Meteorol.* **101**, 329–358 (2001).

86. Chang, X. et al. Contributions of inter-city and regional transport to PM$_{2.5}$ concentrations in the Beijing-Tianjin-Hebei region and its implications on regional joint air pollution control. *Sci. Total Environ.* **660**, 1191–1200 (2019).

87. Zheng, M. et al. Seasonal trends in PM$_{2.5}$ source contributions in Beijing, China. *Atmos. Environ.* **39**, 3967–3976 (2005).

88. Wang, Y. et al. The ion chemistry and the source of PM$_{2.5}$ aerosol in Beijing. *Atmos. Environ.* **39**, 3771–3784 (2005).

89. Di, Q. et al. Assessing PM$_{2.5}$ exposures with high spatiotemporal resolution across the continental United States. *Environ. Sci. Technol.* **50**, 4712–4721 (2016).

90. Li, L. et al. Spatiotemporal imputation of MAIAC AOD using deep learning with downscaling. *Remote Sensing Environ.* **237**, 111584 (2020).

91. Pu, Q. & Yoo, E.-H. Ground PM$_{2.5}$ prediction using imputed MAIAC AOD with uncertainty quantification. *Environ. Pollut.* **274**, 116574 (2021).

92. Picornell, M. et al. Population dynamics based on mobile phone data to improve air pollution exposure assessments. *J. Exp. Sci. Environ. Epidemiol.* **29**, 278–291 (2019).

93. Shi, S., Chen, C. & Zhao, B. Air infiltration rate distributions of residences in Beijing. *Building Environ.* **92**, 528–537 (2015).

94. Xiao, Q. et al. The long-term trend of PM$_{2.5}$-related mortality in China: the effects of source data selection. *Chemosphere* **263**, 127894 (2021).

95. Burnett, R. et al. Global estimates of mortality associated with long-term exposure to outdoor fine particulate matter. *Proc. Natl Acad. Sci.* **115**, 9592–9597 (2018).

96. Baklanov, A. et al. Integrated systems for forecasting urban meteorology, air pollution and population exposure. *Atmos. Chem. Phys.* **7**, 855–874 (2007).

97. Valari, M., Markakis, K., Powaga, E., Collignan, B. & Perrussel, O. EXPLUME v1.0: a model for personal exposure to ambient O$_3$ and PM$_{2.5}$. *Geosci. Model Dev.* **13**, 1075–1094 (2020).

98. Bureau, C. *Statistical yearbook of China (2019)*. (China Statistics Press, 2019).

Author contributions

Y.W. and Q.L. contributed equally. Y.W. performed the simulation covering the whole Beijing, provided multiple analytical perspectives and conducted manuscript preparation. Q.L. constructed the simulation method and data preparation. Z.Luo, J.Z., and Z.Lv helped with model running and data analysis. Q.D., J.L., M.E., and J.B.

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/s43247-023-01119-3.

Correspondence and requests for materials should be addressed to Huan Liu or Kebin He.

Peer review information *Communications Earth & Environment* thanks Haitong Sun and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editors: Keiichiro Hara and Clare Davis. A peer review file is available.

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

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