Article

Substantially underestimated global health risks of current ozone pollution

Yuan Wang [1,11], Yuanjian Yang [1,11], Qiangqiang Yuan [2] ✉, Tongwen Li[3], Yi Zhou [1], Lian Zong[1], Mengya Wang[1], Zunyi Xie[4], Hung Chak Ho [5], Meng Gao [6], Shilu Tong [1,7,8], Simone Lolli[9] & Liangpei Zhang [10] ✉

Existing assessments might have underappreciated ozone-related health impacts worldwide. Here our study assesses current global ozone pollution using the high-resolution (0.05°) estimation from a geo-ensemble learning model, with key focuses on population exposure and all-cause mortality burden. Our model demonstrates strong performance, achieving a mean bias of less than -1.5 parts per billion against in-situ measurements. We estimate that 66.2% of the global population is exposed to excess ozone for short term (> 30 days per year), and 94.2% suffers from long-term exposure. Furthermore, severe ozone exposure levels are observed in Cropland areas, particularly over Asia. Importantly, the all-cause ozone-attributable deaths significantly surpass previous recognition from specific diseases worldwide. Notably, mid-latitude Asia (30°N) and the western United States show high mortality burden, contributing substantially to global ozone-attributable deaths. Our study highlights current significant global ozone-related health risks and may benefit the ozone-exposed population in the future.

Ambient ozone (O_3) posed a significant threat to human health and the ecological environment worldwide, particularly during warmer days[1–5]. Several countries, including China, the United States (US), Italy, and Japan, have established standards for ambient O_3 concentrations to improve air quality[6]. The World Health Organization's (WHO) latest air quality guideline (AQG), published in 2021, recommended that the maximum daily 8-hour averaged (MDA8) ambient O_3 concentrations should not exceed 100 μg m^{-3} or 50.11 parts per billion (ppb) for short-term exposure. Additionally, the AQG of warm-season MDA8 ambient O_3 concentrations (the maximum consecutive 6-month average) was suggested as 60 μg m^{-3} or 30.07 ppb for long-term exposure. In the last few years, global ambient O_3 concentrations have shown an upward trend[7–10]. It was estimated that O_3-attributable mortality due to chronic respiratory disease (CRD) increased by 46% from 2000 to 2019 worldwide[11]. Furthermore, global exposure to ambient O_3 resulted in 0.470 million (95% confidence interval [CI]: 0.100, 0.818) and 0.423 million (95% CI: 0.223, 0.659) deaths from chronic obstructive pulmonary disease (COPD)[12] and CRD[13] in 2019, respectively. Therefore, it is crucial to accurately and continuously monitor global ambient O_3 distribution and assess O_3-related health risks.

In-situ measurements can provide accurate and reliable ambient O_3 concentrations for estimating exposure levels and associated health impacts[14]. However, in situ stations were typically located in cities[15,16], which limited their representativeness for peri-urban and rural areas. In contrast, the chemical transport model offered an alternative method for simulating ambient O_3 concentrations over

[1]School of Atmospheric Physics, Nanjing University of Information Science and Technology, Nanjing, China. [2]School of Geodesy and Geomatics, Wuhan University, Wuhan, China. [3]School of Geospatial Engineering and Science, Sun Yat-sen University, Zhuhai, China. [4]College of Geography and Environmental Science, Henan University, Kaifeng, China. [5]Department of Public and International Affairs, The City University of Hong Kong, Hong Kong, China. [6]Department of Geography, Hong Kong Baptist University, Hong Kong, China. [7]National Institute of Environmental Health, Chinese Centre for Disease Control and Prevention, Beijing, China. [8]School of Public Health and Social Work, Queensland University of Technology, Brisbane, Australia. [9]CNR-IMAA, Contrada S, Loja, Italy. [10]The State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China. [11]These authors contributed equally: Yuan Wang, Yuanjian Yang. ✉e-mail: qqyuan@sgg.whu.edu.cn; zlp62@whu.edu.cn

large areas[17,18]. Unfortunately, it required extensive computational resources and relied on temporally lagging emission inventories with unexpected biases over local regions. These factors can significantly increase the time consumed for simulations and result in high uncertainties into the model output.

To date, data fusion algorithms based on statistical or machine learning (ML) techniques[7–13,15,19–25] have been emerging as preferred methods to acquire ambient O_3 distribution. For example, in national/regional studies, a geographically weighted regression model was proposed by Zhang et al. [19] to acquire monthly 0.25-degree ambient O_3 concentrations over Eastern China. Li et al. [20] developed an enhanced geographically and temporally weighted neural network for obtaining high-resolution (0.05°) MDA8 ambient O_3 concentrations across the Greater Bay Area in China. A daily ambient O_3 dataset at spatial resolution of 0.1° over China was established by Wei et al. [21] based on a space-time extremely randomized trees model. Wang et al. [22] employed the random forest model to generate 0.1-degree MDA8 ambient O_3 concentrations over California in the US. The light gradient boosting machine was adopted by Chen et al. [23] for estimating MDA8 ambient O_3 concentrations (0.1°) in Europe, which also analysed O_3 exposure with other air pollutants.

Regarding globe-scale works[7–13,24,25], a M^3Fusion model and its improved version with Bayesian maximum entropy (M3Fusion-BME) were devised by Chang et al. [7] and DeLang et al. [10], respectively, to establish warm-season 0.1-degree ambient O_3 datasets over the globe. Liu et al. [8] mapped globally distributed monthly ambient O_3 exposure at spatial resolution of 0.5° by developing a cluster-enhanced ML model. The random forest model was employed in Xu et al. [24] for estimating daily 0.25-degree ambient O_3 concentrations from landscape fire, which then systematically analysed relevant population exposure worldwide. Sun et al. [9] proposed a spatiotemporal Bayesian neural network to generate global monthly ambient O_3 dataset (0.1°) and assessed O_3-attributable mortality burden for respiratory diseases in 2010. Based on existing modelling algorithms, the global burden of disease (GBD) 2021[12] discussed the health risks of ambient O_3 for COPD across 204 countries and regions through a M^3Fusion-BME model[10]. The M^3Fusion-BME model[10] was also introduced by Malashock et al. (2022)[11,13] to assess O_3-attributable mortality burden for CRD in cities and peri-urban/rural areas globally. Xue et al. [25] investigated the exposure-response function between O_3 exposure and children mortality in 55 low-income and mid-income countries using a cluster-enhanced ML model[8].

Data fusion algorithms based on statistical or ML techniques[7–13,15,19–25] have significantly advanced the estimation of population exposure to ambient O_3 and its associated health risks. For national/regional studies, localized methods with individual O_3 modelling strategies[15,19–21] generally outperformed the holistic models that utilized all in situ stations simultaneously. The superior performance was attributed to the pronounced local heterogeneity of ambient O_3 distribution, driven by spatiotemporally varying surface emissions of O_3 precursors[5,26] and meteorological conditions[27,28]. Furthermore, the established ambient O_3 datasets mostly achieved high spatial (e.g., 0.05°) and temporal (e.g., daily) resolutions in national/regional studies. Conversely, the globe-scale works typically utilized the holistic models to acquire ambient O_3 concentrations, often with coarse spatial (e.g., 0.25°) or temporal (e.g., monthly) resolutions. As for health impacts, they mainly aimed at assessing long-term O_3 exposure referencing part of WHO standards. Meanwhile, the globe-scale works usually concentrated on investigating O_3-related health risks for specific diseases, selected countries, or single source.

Previous studies[7–13,15,19–25] have yielded robust results and significantly contributed to society. However, the severity of current global O_3-related health risks still remained underappreciated in the globe-scale works[7–13,24,25], which exhibited considerable shortcomings. To be specific, the geospatially local apriority of ambient O_3

distribution has not been adequately incorporated into global modelling, unlike in national/regional studies, potentially introducing greater uncertainties in the modelled results. In the meantime, previous globe-scale works generally adopted statistical or ML models with coarser spatial (e.g., 0.25°) or temporal (e.g., monthly) resolutions compared to national/regional studies. This likely led to cumulative errors or defects in the estimation of O_3 exposure levels and mortality burden[29,30]. For instance, triple coarser spatial resolution (e.g., 36 versus 12 km) might cause biased (> 10 %) O_3-attributable mortality burden in national regions[30]. The coarse temporal resolution (e.g., monthly) also cannot support the assessment of short-term population exposure to ambient O_3. Moreover, these globe-scale works have not sufficiently explored the connection between short- and long-term O_3-related health risks with multiple standards from WHO. Both short- and long-term health impacts should be considered to evaluate the acute and chronic risks due to O_3 exposure over the globe. Importantly, previous globe-scale works simply focused on limited conditions (e.g., diseases, countries, or O_3 sources), which might substantially underestimate current worldwide O_3-attributable mortality burden.

In our study, the objectives are threefold: to propose a geospatially dynamic ensemble ML model (global-local coupled ensemble forest, GL-CEF) (1) for the global modelling of daily seamless high-resolution (0.05°) ambient O_3 concentrations; to provide an in-depth assessment of global O_3 pollution, with prominent focuses on short-/long-term population exposure (2) and all-cause mortality burden (3). It is crucial to incorporate geospatially local apriority into the model due to the pronounced local heterogeneity in ambient O_3 distribution. Nevertheless, the spatial locations of in situ stations were highly non-uniform across the globe, with sparse or even nonexistent stations in many regions. To fix this issue, we devise three modules (global, local, and global-local) in the GL-CEF model for the estimation over station-sparse (or no-station) and station-dense regions, exhibiting superior performance to popular holistic ML models. The GL-CEF model can generate consistent global ambient O_3 distribution with previous globe-scale works[7–13,24,25], but performed at higher resolutions (daily and 0.05°). The high resolutions result in data volumes 25 to 1,460 times larger per year (yr⁻¹) compared to them, offering richer spatiotemporal details. By combining all three objectives, our study investigates current O_3-related health risks worldwide, considering the latest WHO standards and land use disparities. The ambient O_3 standards from WHO include the AQG, interim target 2 (IT2), and interim target 1 (IT1), with the values of 50.11/30.07, 60.13/35.08, and 80.18/50.11 ppb for short-/long-term O_3 exposure, respectively. Our ambient O_3 dataset supports for fine-scale assessment of health risks from both short- and long-term O_3 exposure, potentially reducing cumulative errors in global analyses. Our findings reveal that O_3-related health impacts might have been underappreciated and identify key polluted regions over the globe. Therefore, our study may have implications for investigating global O_3 pollution and benefiting the O_3-exposed population in the future.

Results
Multiscale model performance
Figure 1a, b depict the performance of GL-CEF model based on two cross validation schemes, which include the space-informed cross validation (SICV) and temporally extrapolated SICV (TESICV)[15,16,20,21,29]. The collocated samples are sufficient, 3.67 million for SICV and 1.15 million for TESICV, providing reliable validation metrics. All metrics are computed under the significance levels of $p < 0.01$. At daily scale, the GL-CEF model demonstrates satisfactory performance, with the coefficient of determination (R^2) of 0.87 and 0.73 for SICV and TESICV, respectively. At monthly (warm-season) scale (Supplementary Fig. 1a–d), the model errors decrease further, with the root mean square error (RMSE) improving by 2.200 (2.624) ppb for SICV and
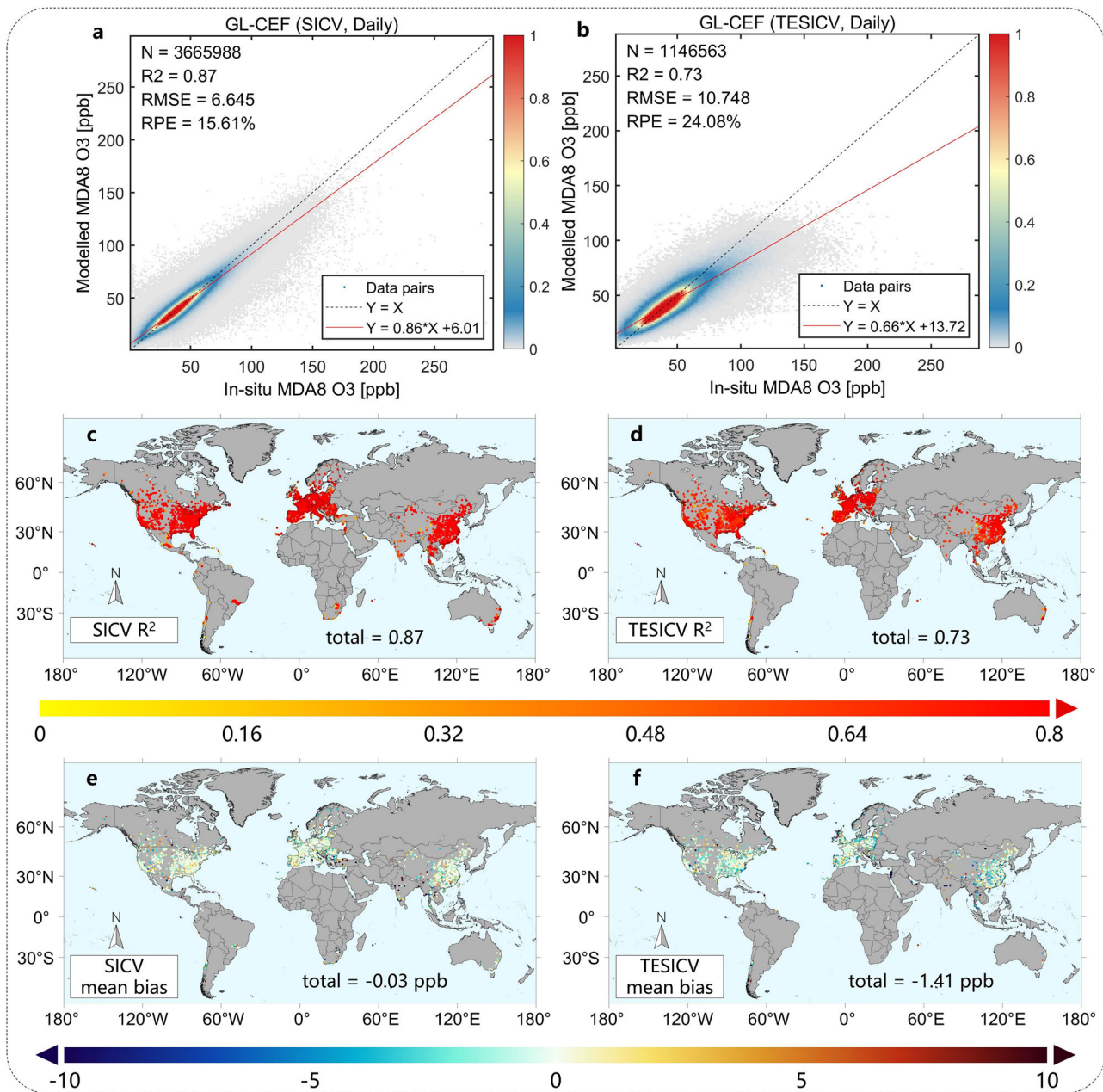
Fig. 1 | Overall model performance of the GL-CEF model. a, b Density scatter plots of cross-validation results and **c–f** spatial distribution of metrics on collocated grids worldwide at daily scale. Black dashed and red solid lines stand for 1:1 and fitted lines in **a, b**, respectively. Colour bars denote the normalized densities of data pairs in **a, b** and the values of metrics in **c–f**. Unit for RMSE and mean bias: ppb. Definitions of acronyms: GL-CEF (global-local coupled ensemble forest), N (number of collocated samples), SICV (space-informed cross-validation), TESICV (temporally extrapolated space-informed cross-validation), and RPE (relative percentage error).

4.537 (5.192) ppb for TESICV compared to daily scale. Additionally, the GL-CEF model generally yields favourable accuracy on collocated grids that involve in situ stations (Fig. 1c–f), with 82.59%/50.54% and 87.97%/81.26% of them showing the R^2 of > 0.7 and mean bias of < ± 5 ppb for SICV/TESICV, respectively. Meanwhile, the collocated grids with high R^2 (> 0.8) and near-zero mean bias are mostly located in densely populated regions, such as China, Europe, and the US. This indicates that the global MDA8 ambient O_3 concentrations established by the GL-CEF model are of both significant quality and practical utility.

Global population exposure to ambient O_3
The spatial patterns of annual MDA8 ambient O_3 are generally similar to those observed during the warm season (Fig. 2a, b). However, significant regional differences still exist. For example, there is much

higher MDA8 ambient O_3 during the warm season than the annual levels over Africa and southern Asia, likely due to massive surface emissions of O_3 precursors from biomass burning[31–33]. Meanwhile, the high values of MDA8 ambient O_3 are also observed in southern Europe and the western US during the warm season, exceeding annual levels. At country scale (Fig. 2c, d), the large population-weighted MDA8 ambient O_3 (> 48 ppb) is predominantly found in Asia and Africa, particularly in northern regions of middle latitudes. By contrast, the population-weighted MDA8 ambient O_3 is smaller (<32 ppb) across the countries of South America and Oceania.

Globally, the population-weighted MDA8 ambient O_3 are 42.39 ± 11.29 and 49.76 ± 13.75 ppb for annual and warm-season levels, respectively (Fig. 2e–j). In some Asian countries, annual and warm-season population-weighted MDA8 ambient O_3 are distinctly higher
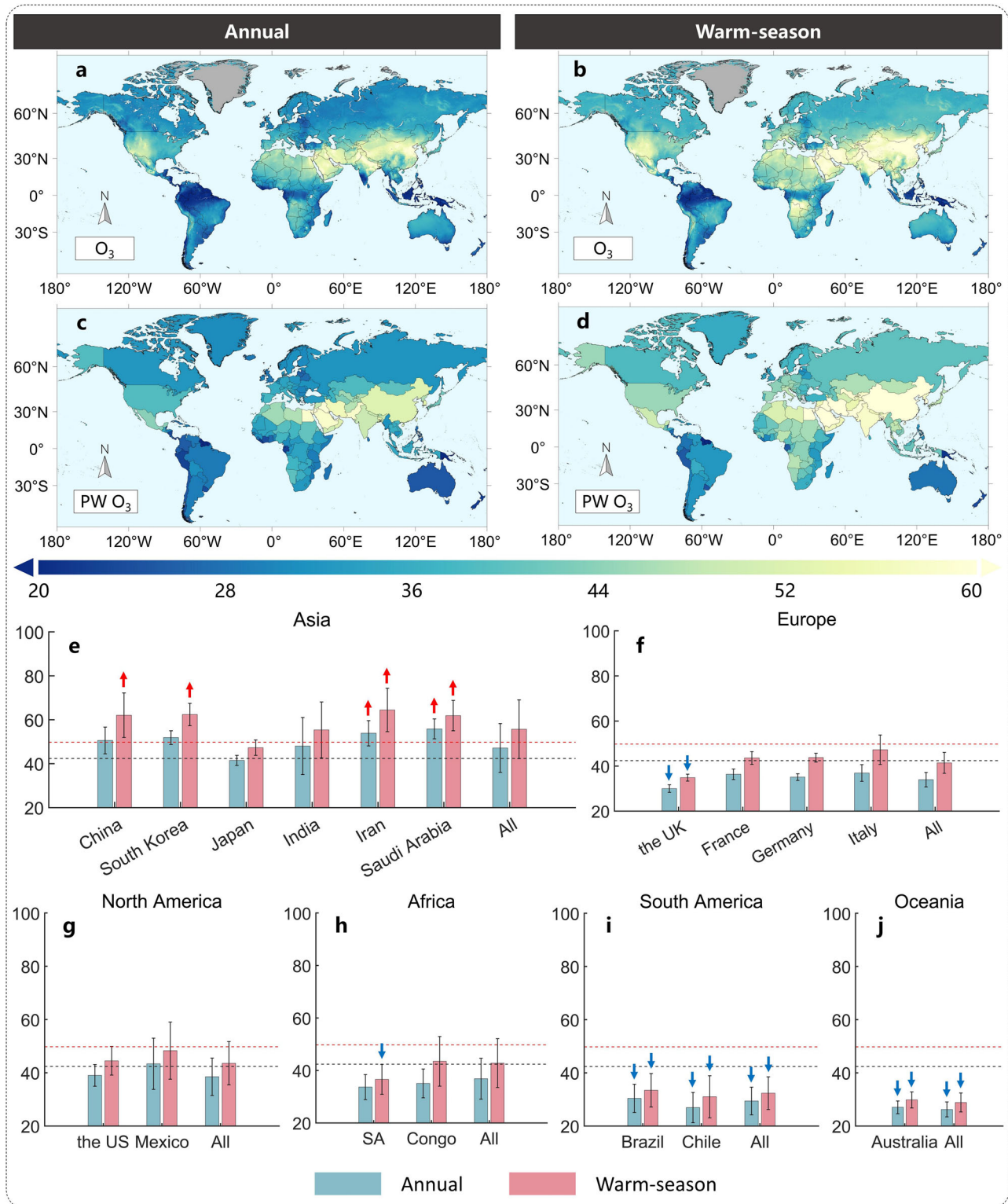
Fig. 2 | Maps and statistics of current global ambient O₃ distribution.
a, **b** Annual and warm-season globally gridded, **c**, **d** country-scale population-weighted, and **e**–**j** population-weighted statistics for typical countries of MDA8 ambient O₃ concentrations. Colour bar stands for the values of ambient O₃ concentrations (unit: ppb) in **a**–**d**. Light grey denotes the unavailable regions with snow, ice, and lake for the GL-CEF model in **a**–**d**. The black and red dashed lines represent population-weighted MDA8 ambient O₃ concentrations worldwide for annual (42.39 ppb) and warm season (49.76 ppb) in **e**–**j**, respectively. The vertical lines indicate the ranges of one standard deviation in **e**–**j**. The countries that present significantly larger (difference > 10 ppb) and smaller (difference < −10 ppb) population-weighted MDA8 ambient O₃ concentrations than the global average are marked with red up arrow and blue down arrow in **e**–**j**, respectively. Definitions of acronyms: PW (population-weighted) and SA (South Africa).
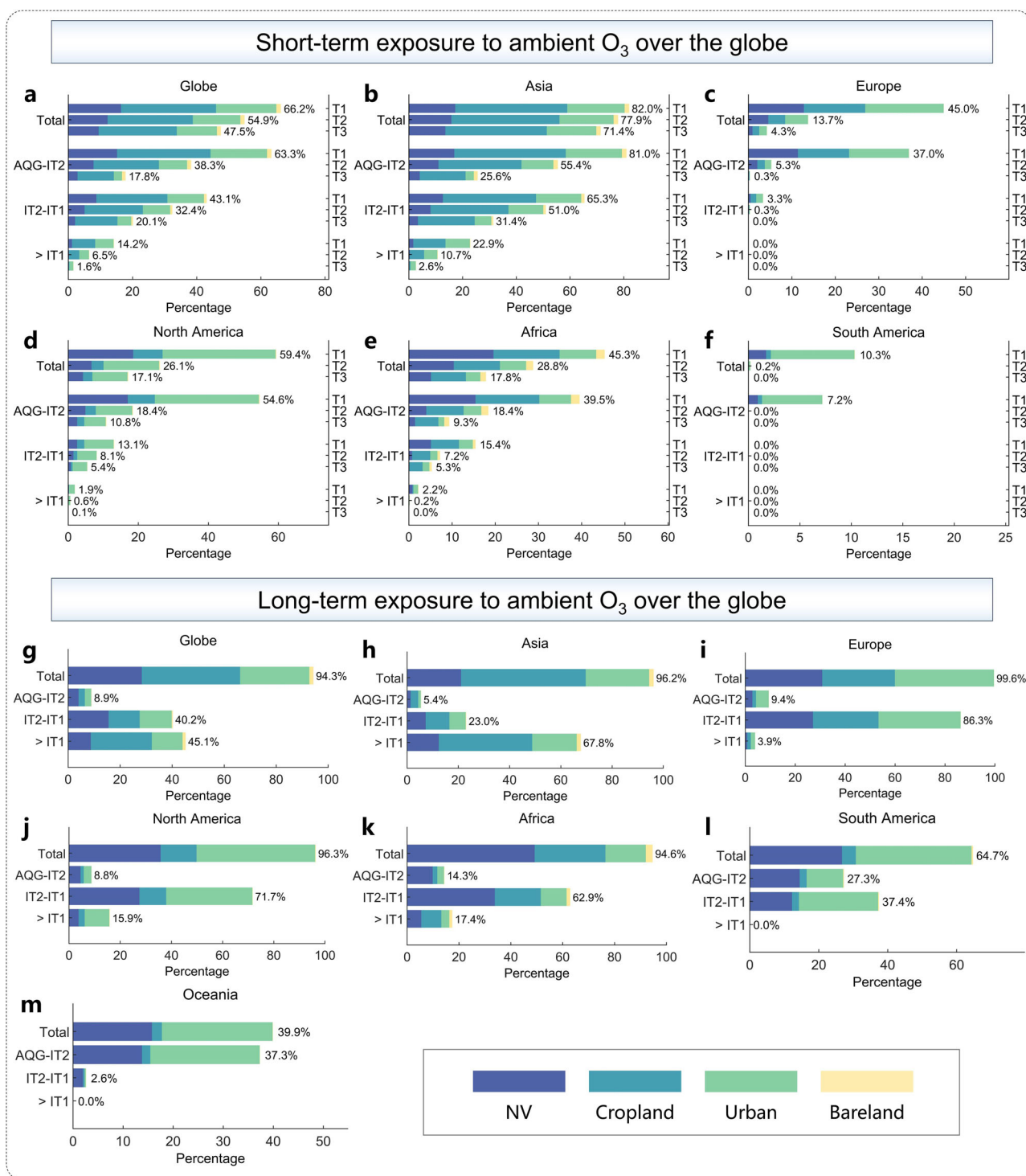
Fig. 3 | Current population exposure to ambient O₃ worldwide. a–f Population fractions divided by land use classes falling on total (> AQG) and three short-term O₃ exposure intervals (AQG-IT2, IT2-IT1, and > IT1). **g–m** Population fractions divided by land use classes falling on total (> AQG) and three long-term O₃ exposure intervals (AQG-IT2, IT2-IT1, and > IT1). T1, T2, and T3 represent the thresholds of more than 30, 60, and 90 days yr⁻¹ at the right Y-axis in (**a-f**), respectively. Definitions of acronyms: AQG (air quality guideline, 50.11 ppb for short term and 30.07 ppb for long term), IT2 (interim target 2, 60.13 ppb for short term and 35.08 ppb for long term), IT1 (interim target 1, 80.18 ppb for short term and 50.11 ppb for long term), and NV (Natural Vegetation).

than the global average (difference > 10 ppb), likely due to sufficient solar radiation intensity combined with favourable meteorological conditions[27,28] and surface emissions of O₃ precursors[5,26]. By contrast, a few countries in South America and Oceania exhibit much lower population-weighted MDA8 ambient O₃ compared to the global average (difference < −10 ppb), indicating milder O₃ pollution. Notably, severe O₃ pollution can occur at country scale for brief periods

(Supplementary Fig. 7). For instance, the population-weighted MDA8 ambient O₃ is observed extremely high over China and South Korea from March to May (MAM), which surpasses the IT1 standard (i.e., > 80.18 ppb).

Four land use classes (Natural Vegetation, Cropland, Urban, and Bareland) are adopted to assess the population exposure to ambient O₃ (Fig. 3). Natural Vegetation includes forests, shrublands, savannas,

and grasslands (see Supplementary Table 4 for details). The total (three) O_3 interval (intervals) represents (represent) > AQG (AQG-IT2, IT2-IT1, and > IT1). Regarding short-term O_3 exposure (Fig. 3a–f), 66.2%/54.9%/47.5% of the global population is exposed to excess ambient O_3 for more than 30/60/90 days yr^{-1}. The global O_3 exposure patterns are mainly influenced by those of Asia (large population), with similar land use fractions on various O_3 exposure intervals (e.g., large contribution from Cropland). In North America (Africa), the population exposed to O_3 pollution primarily lives in Urban (Natural Vegetation and Cropland) areas, especially for the MDA8 ambient O_3 exceeding IT1. Europe and South America both exhibit relatively low O_3 exposure levels, predominantly within the AQG-IT2 interval. The O_3 pollution is very mild in Oceania and consequently not analysed in detail. As for long-term O_3 exposure (Fig. 3g–m), 94.3% of the global population experiences excess ambient O_3, with nearly half (45.1%) exposed to severe O_3 pollution (> IT1). Asia again dominates the global O_3 exposure patterns, generally presenting consistent land use fractions across different O_3 exposure intervals. Specifically, the fraction of Cropland increases with higher O_3 pollution levels (Fig. 3g, h), which likely suggests that more people suffer from high MDA8 ambient O_3 during the warm season[11,13,34]. Europe, North America, and Africa exhibit similar O_3 exposure patterns, though varying by land use disparities. Severe O_3 pollution (> IT1) is not observed over South America and Oceania, indicating low O_3 exposure levels there.

Overall, the population amount exposed to short-term ambient O_3 gradually declines with the O_3 pollution increasing (Fig. 3a–f). However, the O_3 exposure interval of IT2-IT1 involves the largest global population (20.1%) for the duration of > 90 days yr^{-1} (Fig. 3a). This implies that more people worldwide are exposed to the MDA8 ambient O_3 within the IT2-IT1 interval for a long period, likely leading to higher short-term mortality risks. On the other hand, the areas of polluted regions for long-term O_3 exposure on IT2-IT1 (near 70%) distinctly exceed that on > IT1 (less than 20%) globally (Supplementary Fig. 5g, h). Nevertheless, more people (45.1%) are exposed to the warm-season MDA8 ambient O_3 of > IT1 (Fig. 3g), which also potentially results in higher long-term mortality risks.

Global O_3-attributable mortality burden
Our study adopts a log-linear exposure-response function to obtain the global O_3-attributable mortality burden, based on the pooled relative risks of short- and long-term O_3 exposure for all-cause deaths from two worldwide epidemiological researches[35,36]. We estimate that the short- and long-term population attributable fraction (PAF) are 3.42×10^{-3} (95% CI: 2.71×10^{-3}, 4.12×10^{-3}) and 0.0272 (95% CI: 0.176, 0.365) globally during 2019–2021 (Supplementary Fig. 5i, j), respectively. This indicates that 0.34% (95% CI: 0.27%, 0.41%) and 2.72% (95% CI: 1.76%, 3.65%) of the global total deaths are attributed to O_3 pollution. The long-term PAF substantially exceeds that of the short-term (eight times), suggesting its much higher mortality risks. This is potentially attributed to that short-term O_3 exposure may stimulate the production of antioxidant enzymes, which is deemed as a compensatory mechanism[37]. The upregulation of antioxidant enzymes can decrease oxidative stress and is likely related to the suppression of proinflammatory responses. Conversely, long-term O_3 exposure probably engenders the upregulation of internal redox homoeostasis, which will result in an increment of systematic inflammation[37].

The total (three) exposure interval (intervals) stands (stand) for > AQG (AQG-IT2, IT2-IT1, and > IT1) in Fig. 4. For short-term O_3-attributable mortality (Fig. 4a–d), the high values of PAF ($> 3 \times 10^{-3}$) are predominantly observed in Asian countries of middle latitudes (30°N), while other countries generally present relatively smaller PAF ($< 2 \times 10^{-3}$). Within different O_3 exposure intervals, the high values of country-scale PAF ($> 3 \times 10^{-3}$) typically emerge on IT2-IT1. Notably, China also demonstrates large PAF (3.46×10^{-3} [95% CI: 2.75×10^{-3}, 4.17×10^{-3}]) for the > IT1 interval,

indicating more frequent occurrences of extremely high MDA8 ambient O_3 (i.e., > 80.18 ppb) than other countries. Furthermore, the four land use classes (Natural Vegetation, Cropland, Urban, and Bareland) are applied to assess the O_3-attributable mortality burden (Supplementary Fig. 5i). Cropland shows the largest total PAF (4.68×10^{-3} [95% CI: 3.71×10^{-3}, 5.64×10^{-3}]), reflecting significant mortality risks due to heavy O_3 pollution. This is likely because surface emissions of biogenic volatile organic compounds (O_3 precursor) from vegetation and carbon monoxide (which contributes to radical formation) from incomplete biomass burning are likely pronounced over Cropland areas[34]. Additionally, nitrogen oxides (another O_3 precursor) can also transport from Urban to Cropland areas[34]. Favourable solar radiation intensity further accelerates photochemical reactions, leading to the accumulation of ambient O_3 concentrations[5,26–28]. Among three O_3 exposure intervals, the distribution of PAF based on land use classes is diverse. Generally, the PAF on IT2-IT1 is larger compared to other O_3 exposure intervals, which results from its higher short-term O_3 exposure levels over the globe (Fig. 3a).

As for long-term O_3-attributable mortality (Fig. 4e–h), the high values of country-scale PAF (> 0.03) also primarily concentrate in Asia of middle latitudes (30°N), which signifies the substantial mortality risks from both short- and long-term O_3 exposure over these regions. Other countries generally show moderate PAF of > 0.01. Among three O_3 exposure intervals, the high values of country-scale PAF (> 0.03) are predominantly observed for > IT1. By contrast, most countries exhibit significantly small PAF on AQG-IT2 ($< 1 \times 10^{-3}$), suggesting that the ambient O_3 standard of long-term IT2 (35.08 ppb) from WHO may need to be revised upward, which has shown a similar effect with AQG (30.07 ppb) across various countries. Moreover, Cropland and Bareland both present large total PAF (0.0332 [95% CI: 0.0216, 0.0446] and 0.0355 [95% CI: 0.0230, 0.0477]) (Supplementary Fig. 5j). However, excess ambient O_3 is more hazardous for Cropland worldwide due to its larger population amount (2.885 billion) compared to Bareland (0.114 billion). Concerning various O_3 exposure intervals, the PAF gradually increases with O_3 pollution rising, since more people are exposed to higher warm-season MDA8 ambient O_3 over the globe (Fig. 3g).

We estimate that a total of 0.177 million yr^{-1} (95% CI: 0.139, 0.214) and 1.407 million yr^{-1} (95% CI: 0.909, 1.896) deaths globally stem from short- and long-term O_3 exposure, respectively (Supplementary Table 5). The mortality burden attributed to long-term O_3 exposure distinctly exceeds that from short term. Meanwhile, most O_3-attributable deaths due to short- and long-term exposure occur in the IT2-IT1 (53.01%) and > IT1 (73.04%) intervals, respectively. This aligns with the above analyses of O_3 exposure. Notably, the polluted regions on IT2-IT1 for short-term exposure present a similar distribution with that on > IT1 for long-term exposure (Supplementary Fig. 5c, h). Considering the top-10 countries simultaneously (Supplementary Table 2, 3), we find that Asia of middle latitudes (30°N) and the western US experience high mortality burden attributed to short- and long-term O_3 exposure.

Discussion
This study provides an assessment of current global O_3 pollution. We propose a geo-ensemble ML model (GL-CEF) to acquire MDA8 ambient O_3 concentrations worldwide. The establishment of global ambient O_3 dataset integrates remote sensing observations, chemical apriori data, meteorological fields, geographic elements, and in-situ measurements from > 7000 stations across more than 100 countries. The GLC-CEF model is rigorously validated and stably displays global daily seamless high-resolution (0.05°) patterns of O_3 pollution during 2019–2021, achieving the R^2 of 0.87 and 0.73 for SICV and TESICV, respectively. Based on the modelled dataset, we characterize short- and long-term
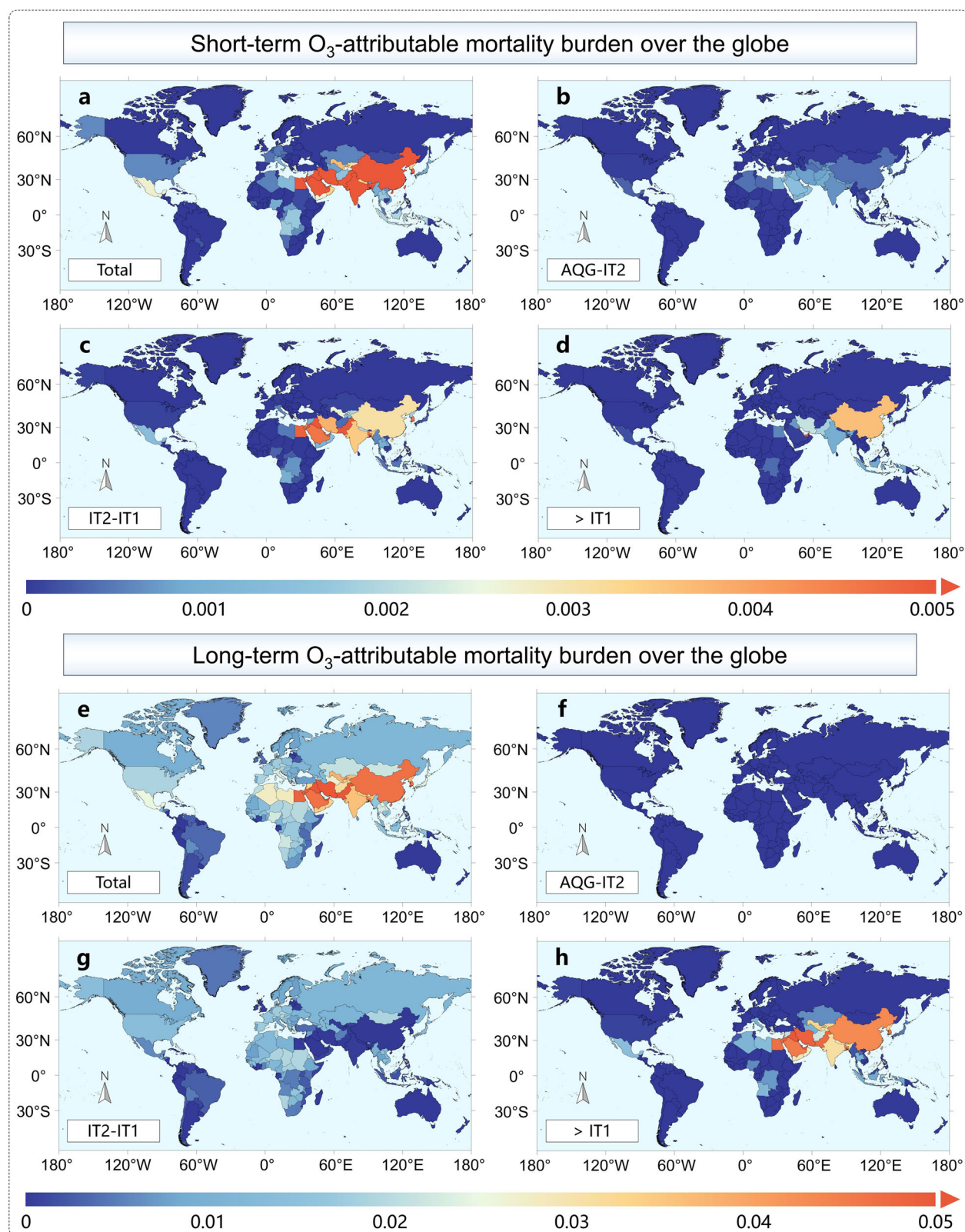
Fig. 4 | Current O₃-attributable mortality burden worldwide. a–d Country-scale PAF for total (> AQG) and three short-term O₃ exposure intervals (AQG-IT2, IT2-IT1, and > IT1). **e–h** Country-scale PAF for total (> AQG) and three long-term O₃ exposure intervals (AQG-IT2, IT2-IT1, and > IT1). The short- and long-term gridded PAF with more details at a high spatial resolution can be referred to in Supplementary Fig. 8. Colour bars stand for the values of PAF. Definitions of acronyms: AQG (air quality guideline, 50.11 ppb for short term and 30.07 ppb for long term), IT2 (interim target 2, 60.13 ppb for short term and 35.08 ppb for long term), and IT1 (interim target 1, 80.18 ppb for short term and 50.11 ppb for long term).

O_3 population exposure and their associated all-cause mortality burden globally, with consideration of the latest ambient O_3 standards from WHO and land use disparities. The assessment of population exposure and mortality burden is conducted at various scales (the globe, continent, country, and grid), with results that corroborate across different levels.

Our study improves the modelling algorithm to derive global ambient O_3 concentrations. The spatial distribution of ambient O_3 generally presented strong local heterogeneity, making it essential to incorporate geospatially local apriority into the model. Nevertheless, there were sparse or even no in situ stations established in plenty of regions worldwide, which rendered previous localized methods[15,19–21] challenging for global modelling tasks. To address this issue, the developed GL-CEF model comprises three components: global module (employing global features), local module (incorporating geospatially local apriority), and global-local module (connecting global and local modules). Notably, the local module consisted of two strategies: sliding block strategy (to accelerate model convergence) and variable local sub-model (to improve model robustness). The GL-CEF model can adopt global and local knowledge over station-sparse (or no-station) and station-dense regions, respectively. Additionally, we introduce a spatiotemporal-based imputation method for recovering missing information in the model inputs based on their self-correlation. Validation results demonstrate that the GL-CEF model achieves favourable accuracy, with the mean bias of −0.03 and −1.41 ppb for SICV and TESICV, respectively. The performance of the GL-CEF model also surpasses those of widely-used holistic ML models (Supplementary Fig. 2), such as random forest and light gradient boosting machine.

Using the GL-CEF model, we generate a daily spatiotemporally continuous 0.05-degree dataset for assessing short- and long-term O_3-related health risks. This dataset exhibits improved performance (an increment of 0.39 in R^2) and spatial resolution five times higher than the GEOS-CF product from the National Aeronautics and Space Administration (NASA) (Supplementary Fig. 3). The modelled dataset captures the seasonality of current ambient O_3 worldwide, showing high values during the warm season over Asia, Africa, and North America, which is consistent with previous reports[7–10,16]. In addition, the ambient O_3 concentrations from previous globe-scale works[7–13,24,25] had coarse spatial (e.g., 0.25°) or temporal (e.g., monthly) resolutions (Supplementary Table 1), potentially introducing spatially or temporally cumulative errors in final analyses[29,30]. Conversely, our study considers O_3 precursors[3,5,26] and employed high-resolution (~ 5 km) remote sensing observations as key inputs[8,19–22] of the model, which include nitrogen dioxide (NO_2) and formaldehyde (HCHO) from the tropospheric monitoring instrument (TROPOMI). By integrating various multisource datasets, the modelled dataset achieves spatial and temporal resolutions of 0.05° and daily, respectively. The high resolutions lead to massive data volumes, ranging from 25 to 1460 times yr⁻¹ compared to previous globe-scale works[7–13,24,25], which provide richer spatiotemporal details and information.

Our study refines the land use disparities in global O_3-related health risks assessment. Land use disparities potentially signified different sources of ambient O_3[38,39], which affected O_3 exposure levels and mortality burden worldwide. Previous reports[9,11,13,34] implied that O_3-related health risks were generally higher in peri-urban or rural areas compared to those in cities. We further divide global land use disparities into four classes, including Natural Vegetation, Cropland, Urban, and Bareland. Although more people experience short- and long-term O_3 exposure in Cropland areas over Asia, the population exposed to O_3 pollution primarily reside in Urban areas over North and South America. In addition, a large amount of population suffers from short- and long-term O_3 exposure in Natural Vegetation areas over Africa. We reveal the geographically diverse impacts of land use disparities on O_3-related health risks globally.

Our study increases the understanding of ambient O_3 standards in the context of global O_3-related health risks assessment. The ambient O_3 standards from WHO were applied in some previous globe-scale works[8,11,24,25]. Liu et al. [8] reported that 37.13% of global population lives in the regions beyond the long-term IT1 suggested by WHO during the warm season. The short-term WHO AQG was adopted in Xu et al. [24] to define the events of substantial fire-sourced air pollution. Malashock et al. [11] implied that the number of cities with warm-season MDA8 ambient O_3 of > long-term AQG from WHO increased by 865 from 2000 to 2019. A threshold value near the long-term WHO IT1 was designed by Xue et al. [25] for obtaining the non-linear exposure-response function of O_3 and under-5 mortality. However, these globe-scale works only considered partial WHO standards and did not account for the different O_3 exposure levels and mortality burden defined by AQG, IT2, and IT1. Additionally, they focused solely on either short- or long-term WHO standards, without exploring the connection between short- and long-term O_3-related health risks.

By contrast, our study provides an assessment of both short- and long-term health risks associated with current O_3 pollution globally, with consideration of all ambient O_3 standards from WHO. We estimate that a large amount of global population (20.1%) is exposed to the MDA8 ambient O_3 of between IT2 and IT1 standards for more than 90 days yr⁻¹, likely leading to higher mortality risks from short-term exposure. Meanwhile, nearly half of the people worldwide (45.1%) are exposed to the warm-season MDA8 ambient O_3 exceeding the IT1 standard, which also potentially results in elevated mortality risks due to long-term exposure. As a consequence, the majority of O_3-attributable deaths from short- and long-term exposure are associated with the IT2-IT1 (53.01%) and > IT1 (73.04%) intervals, respectively. Importantly, this analysis reveals that mid-latitude Asia (30°N) and the western US experience high mortality burden due to short- and long-term O_3 exposure, likely driven by intense solar radiation, favourable meteorological conditions[27,28], surface emissions of O_3 precursors[5,26]. Our study identifies the similar key polluted regions at grid level for global short- and long-term O_3-related health risks.

A highlight of our study is the assessment of both short- and long-term exposure levels and all-cause mortality burden due to current all-source O_3 pollution across > 200 countries or regions, rather than focusing on specific diseases, limited countries, or single-source O_3. Especially for specific diseases, previously reported global O_3-attributable mortality burden primarily involved respiratory diseases, such as CRD and COPD[9–13,17,18,40]. The O_3 exposure could stimulate the respiratory system to generate a large amount of inflammatory cell hormones and then accumulate toxic lipid oxidation products, which finally engendered respiratory diseases. Nevertheless, a recent review[41] has provided the epidemiological evidence and biological mechanisms linking O_3 exposure to cardiovascular diseases. Specifically, the O_3 exposure could trigger chain reactions including respiratory and systemic inflammation, oxidative stress, disruption of autonomic nervous and neuroendocrine systems, impairment of coagulation function, glucose, and lipid metabolism. These reactions can ultimately lead to vascular dysfunction and the progression of cardiovascular diseases. Furthermore, a number of existing meta-analyses have indicated that the mortality of all-cause diseases was positively associated with short-[35,42,43] and long-term[36,44,45] O_3 exposure globally. Therefore, focusing solely on respiratory diseases might significantly underestimate the broader mortality burden associated with ambient O_3 worldwide, which ignored potential influences from other diseases.

Specifically, we estimate that 1.407 million yr⁻¹ (95% CI: 0.909, 1.896) all-cause deaths are attributed to long-term O_3 exposure globally, which substantially exceeds estimates from the GBD 2021[12] (COPD, 0.470 million [95% CI: 0.100, 0.818]) and Malashock et al. [11,13] (CRD, 0.423 million [95% CI: 0.223, 0.659]). Nevertheless, Malley et al. [17] and Chowdhury et al. [18] claimed that 1.04 million (95% CI: 0.72, 1.37)

and 1.30 million (95% CI: 0.93, 1.68) respiratory deaths were attributed to long-term O_3 exposure worldwide in 2010 and 2015, respectively. The O_3-attributable mortality burden from respiratory diseases might be overestimated in these two globe-scale reports[17,18]. This was mainly because they adopted chemical transport models (e.g., GEOS-Chem) with large errors against in-situ measurements worldwide (e.g., a mean bias of 10.8 ppb in Malley et al.[17]), which also had coarse spatial resolutions (e.g., 2.5°). Conversely, the global ambient O_3 dataset established via our model is performed at high spatial resolution of 0.05°, achieving the mean bias of less than -1.5 ppb, suggesting more reasonable results of O_3-attributable mortality burden. In addition, we find that global short-term O_3 exposure results in 0.177 million yr^{-1} (95% CI: 0.139, 0.214) all-cause deaths. By adopting a comprehensive perspective, our assessment reveals that the severity of current worldwide O_3-related health risks is fairly greater than previously recognized from specific diseases.

However, there are some limitations in this study. The global module in the GL-CEF model is trained with all collocated samples over station-sparse regions, where only a few in situ stations can be applied in the validation. This probably leads to insufficient representativeness of the validation results compared to other regions. Meanwhile, our study is restricted by the availability of satellite observations from TROPOMI (operating since 2018), preventing the comparison with the historical sequence (e.g., 2000-2018). Additionally, our study assumes even personal O_3 exposure on each grid, likely resulting in misclassification bias due to the population migration and diverse time patterns of anthropogenic activities (e.g., time spent outdoors versus indoors). Moreover, the calculation of premature deaths uses a log-linear exposure-response function between ambient O_3 concentrations and premature deaths from epidemiological researches[35,36]. Limited by existing meta-analyses, it is challenging to account for all conditions (e.g., ages and socioeconomic differences) of worldwide population in the computation of pooled relative risks due to O_3 exposure. The calculation of O_3-attributable premature deaths also ignores other related influencing factors on mortality burden, such as heat waves. Heat waves are favourable meteorological conditions for photochemical reactions, which can exacerbate O_3 pollution[46,47]. The synergistic effects between heat waves and ambient O_3 may lead to greater health risks. Finally, the impacts of O_3 exposure on the mortality burden attributed to the coronavirus disease 2019 (COVID-19) are not evaluated in our study. There was literature claiming that higher mortality risks of COVID-19 were associated with short-/long-term O_3 exposure[48,49]. Nevertheless, the exact mechanisms of chronic and acute exposure to ambient O_3 on COVID-19 mortality are unclear and required to be elucidated, which need further exploration in the future. These limitations potentially introduce some uncertainties in the estimation of O_3-attributable health risks.

In conclusion, we conduct a globe-scale study on current O_3 pollution, utilizing a daily seamless high-resolution (0.05°) perspective. This study features a geospatially dynamic modelling algorithm and assesses short- and long-term O_3 exposure levels as well as their associated all-cause mortality burden. Our ambient O_3 dataset offers insights into a fine-scale assessment of both short- and long-term O_3-related risks worldwide. Notably, our study reveals significant global O_3-related health impacts that might have been underappreciated in recent years. Effective mitigation of global O_3 pollution, particularly over the regions identified in this study, may lead to a reduction of significant O_3-related health risks and benefit the population exposed to O_3 pollution in the future.

Methods

In this study, all variates are first imputed for missing values and re-sampled before being considered as inputs. Next, the processed data and ground truths (output) should be spatiotemporally collocated and fed into the GL-CEF model for training. A total of two validation

schemes are then exploited to verify the performance of modelled results. Eventually, the global exposure levels and all-cause mortality burden due to ambient O_3 are carefully assessed and discussed.

Datasets

Previous works widely considered O_3 precursors as key inputs of the model[8,19–22], including NO_2 and HCHO. In our study, the high-resolution (~ 5 km) NO_2 and HCHO tropospheric vertical column density (TroVCD) from TROPOMI[50,51] are adopted as the primary variates. The O_3 profile from CAMS[52] is also introduced as a primary variate to provide the apriori information of ambient O_3. Meanwhile, multiple frequently used factors are selected as the auxiliary variates to improve the performance of the model, which consist of solar radiation intensity (necessary conditions for photochemical reactions[53–55]), meteorological fields, and geographic elements. The in-situ measurements of MDA8 O_3 from the Open Air Quality (OpenAQ), China National Environmental Monitoring Center, US Environmental Protection Agency, and European Environment Agency are used as the ground truths (output). At last, the latest replay ambient O_3 product from GEOS-CF[56] is applied for comparison to the modelled results. More specific details are given as follows.

OpenAQ can provide globally distributed ambient concentrations of major air pollutants, which came from various sources over more than 100 countries. The air quality records from OpenAQ have been broadly utilized in worldwide studies of recent years[24,56–59]. In the present study, the global in-situ O_3 measurements during 2019–2021 are collected from OpenAQ and calculated to MDA8 ambient O_3 concentrations (regarded as the ground truths). It's worth noting that the duration of in-situ measurements per day should exceed 20 h (from the first to the last). Furthermore, only the source of "governments" is employed to guarantee the data quality. The units of all values are transformed to ppb based on corresponding references. Supplementary Fig. 9a illustrates the spatial locations of in situ stations in the globe, using the symbols of red circles. A total of > 7,000 in situ stations (by 2021) are considered in this study, which densely cover China, Europe, the US, etc. Since the data from OpenAQ could be irregularly missing, the in-situ O_3 measurements from the China National Environmental Monitoring Center, US Environmental Protection Agency, and European Environment Agency are selected for supplement.

TROPOMI devised a self-appropriate atmospheric NO_2 retrieval algorithm based on that from the Ozone Monitoring Instrument, which can generate NO_2 TroVCD worldwide[51]. The differential optical absorption spectroscopy method, a chemical transport model (or data reanalysis system), and an air-mass factor lookup table were all introduced in the TROPOMI NO_2 retrieval algorithm. Related steps are as below: (1) Retrieve NO_2 total slant column density via the differential optical absorption spectroscopy method. (2) Separate the stratospheric and tropospheric parts using the chemical transport model (or data reanalysis system). (3) Transform the tropospheric slant column density to TroVCD according to the air-mass factor lookup table. In our study, the record of "nitrogendioxide_tropospheric_column" is applied as a primary variate for modelling global ambient O_3 concentrations from 2019 to 2021. Supplementary Table 6 lists more information about the TROPOMI NO_2 TroVCD.

TROPOMI applied a method based on differential optical absorption spectroscopy and combined ultraviolet spectral bands to produce HCHO TroVCD globally[50]. The detailed procedures of TROPOMI HCHO retrieval algorithm were similar to those of NO_2. In the present study, the global record of "formaldehyde_tropospheric_vertical_column" during 2019–2021 is adopted as a primary variate. Supplementary Table 6 provides more details about the TROPOMI HCHO TroVCD.

CAMS was the fourth generation of globally gridded atmospheric reanalysis product from the European Centre for Medium-Range

Weather Forecasts[52]. Based on the mechanisms of physics and chemistry, CAMS can provide multiple chemical components by integrating simulations of chemical transport model with worldwide measured data. The general spatial and temporal resolutions for CAMS were 0.75° and 3-hour, respectively. CAMS reanalysis product has been extensively exploited for previous atmospheric works over the globe[52], which suggested its reliable data quality. In this study, the global record of "ozone_mass_mixing_ratio" (O_3 profile) from 2019 to 2021 is selected as a primary variate to introduce the apriori information of ambient O_3 into the model. Considering that ground truths were MDA8 ambient O_3 concentrations, the daily maximum 9-hour averaged O_3 profile is employed. Supplementary Table 6 shows the specific information of the CAMS O_3 profile.

Similar to CAMS, ERA5 reanalysis product was also devised by the European Centre for Medium-Range Weather Forecasts[60], which involved simulations of chemical transport model and actual measured data. In general, ERA5 can generate atmospheric/surficial parameters with spatial and temporal resolutions of 0.25° and hourly, respectively. In our study, the records of solar radiation intensity (i.e., "surface_solar_radiation_downwards") and several meteorological fields are regarded as auxiliary variates for modelling global ambient O_3 concentrations during 2019–2021. Attributed to that ground truths are MDA8 ambient O_3 concentrations, the MDA8 solar radiation intensity, air temperature, and dew point temperature are utilized. As for other meteorological fields, the hourly values for each day are averaged in this study. Detailed information can be referred to in Supplementary Table 6.

Previous works broadly adopted the geographic elements as auxiliary inputs of the model[8,9,20,21,61] due to their significant association with the spatial distribution of ambient O_3. In this study, the normalized differential vegetation index (NDVI)[62] and land use classes[63] from the moderate resolution imaging spectroradiometer (MODIS) with LandScan population density[64] worldwide are deemed as auxiliary variates to increase the robustness of the model. Supplementary Table 6 lists the specific details of geographic elements.

GEOS-CF was developed by NASA in 2021, including two versions: forecast and replay (improved through reanalysed meteorological fields)[56]. It can produce various global atmospheric chemical components based on simulations of chemical transport model, with spatial and temporal resolutions of 0.25° and hourly, respectively. In the present study, the replay record of "surface_ozone" is exploited in the validation for comparison with the modelled MDA8 ambient O_3 concentrations globally. More information about GEOS-CF can be found in Keller et al. [56].

Data preprocessing
The surface emissions of O_3 precursors had continuous and complicated impact on ambient O_3 distribution[65–67]. Therefore, the NO_2 and HCHO TroVCD are averaged to acquire monthly data, which reflect the conditions for surface emissions of O_3 precursors in our study. Meanwhile, the average by month can reduce data noise (especially for HCHO TroVCD[68]) and improve the coverage of available values.

Afterward, the data interpolating empirical orthogonal functions (DINEOF) method[69] is employed to recover the missing information in monthly NO_2 TroVCD, HCHO TroVCD, and NDVI, relying on their spatiotemporal self-correlation. The brief procedures of the DINEOF method are as follows: 1) Initialize missing values and unfold the 3-dimensional origin data to a 2-dimensional matrix (**M**) along the spatial dimension; 2) Decompose **M** using the singular value decomposition; 3) Reconstruct a matrix **M$_r$** with top-k singular values and replace the missing values of **M** with those of **M$_r$**; and 4) Repeat the procedures of 2) and 3) until the errors reach a pre-determined threshold and reshape **M** to final results according to the origin dimensions. More details of the DINEOF method can be referred to in Alvera-Azcárate et al. [69]. Supplementary Table 7 shows the simulated

experiment results of the DINEOF method. The missing masks for validation in the simulation are acquired from the real scenes of origin products. As listed, the imputed results in monthly NO_2 TroVCD, HCHO TroVCD, and NDVI present expected accuracy via the DINEOF method, with the correlation coefficient (CC) of 0.85, 0.8, and 0.91, respectively.

Furthermore, the spatial resolutions of various variates need to be consistent in our study. Considering the spatial resolution of TROPOMI products (~ 5 km), a global grid of 3600 × 7200 (0.05°) is adopted. To be specific, the monthly NO_2 and HCHO TroVCD are re-sampled to 0.05° through the nearest neighbouring interpolation[70]. The re-sampling methods for other variates are inverse distance weighted interpolation[71] and area-weighted aggregation[72] (Supplementary Table 6).

Remote-sensing and reanalysis products were gridded data with various spatial and temporal resolutions, while the applicable ranges of in-situ measurements only focused on small regions. Therefore, it is required to unify the spatial and temporal dimensions between gridded data and ground truths. Initially, the variates with multiple temporal resolutions, such as monthly NO_2 TroVCD and daily meteorological fields, are mutually aligned. Next, all the ground truths falling on the same grid are averaged to collocate with the aligned gridded data.

Model description
The GL-CEF model includes global, local, and global-local modules, which can adopt global and local knowledge over station-sparse (or no-station) and station-dense regions, respectively. Specific information about the three modules is shown in the following parts.

Global module: as depicted in Supplementary Fig. 9b, a certain number of collocated grids are firstly given (blue circles). Next, a transition zone is set in the regions that are $[l_{min}, l_{max}]$ from the collocated grids. If the distance between the target grid i (yellow square) and its h-th nearest collocated grid is greater than l_{max}, i is deemed as station-sparse, and the global module should be exploited for the estimation (G-value). In our study, the global module utilizes all collocated samples in the modelling, which adopts the deep forest[73] (see the **Supplementary Methods** for details) as the global sub-model. The general expression of the global sub-model is defined in Eq. (1).

$$VG_{O3} = F_G(V_{THCHO}, V_{TNO2}, V_{CAMS}, V_{ESRI}, V_{EMF}, V_{GE}, V_{TC}) \qquad (1)$$

where VG_{O3} represents the modelled MDA8 ambient O_3 concentrations through the global sub-model (G-value). F_G stands for the global sub-model based on the deep forest. V_{THCHO}, V_{TNO2}, V_{CAMS}, V_{ESRI}, and V_{EMF} indicate the TROPOMI HCHO TroVCD, TROPOMI NO_2 TroVCD, CAMS O_3 profile, ERA5 solar radiation intensity, and EAR5 meteorological fields. V_{GE} denotes geographic elements, including NDVI, land use classes, and population density. V_{TC} signifies the temporal encoding[74]. Supplementary Table 8 lists the parameters of the global sub-model designed in this study.

Local module: as illustrated in Supplementary Fig. 9c, if the distance between the target grid i and its h-th nearest collocated grid is less than l_{min}, i is regarded as station-dense, and the local module ought to be adopted for the estimation. The local module involves two highlights: sliding block strategy and variable local sub-model.

Sliding block strategy. Previous localized methods[15,19,20] normally built independent sub-model for each target grid, which yielded commendable performance but required large time consumptions. These methods were likely unfit for the global modelling task. As a result, the sliding block strategy is proposed for the fast training of local sub-models in our study. Related procedures are as follows. A sliding block with the radius of r is first selected and then traverses all the target grids with the step size of s. To ensure that collocated samples can smoothly change in adjacent sliding blocks, a buffer

distance of db is introduced, which is larger than r. If the collocated samples in some sliding blocks are too few ($<cm$), they should be discarded. Next, the collocated samples in each sliding block are utilized for the independent modelling, which can generate several well-trained local sub-models. The local sub-models trained by the nearest N sliding blocks from the target grids are adopted for their intermediate estimation (Ls-values). Finally, the Ls-values are aggregated to acquire the modelled result (L-value). In addition, if the collocated samples included in a few adjacent sliding blocks remain unchanged, they need to be discarded. This can avoid multiple identical local sub-models that likely lead to discontinuous modelled results. Supplementary Table 8 shows the parameters of the sliding block strategy devised in this study.

Variable local sub-models. Generally, the complexity of the model is supposed to be positively correlated with the number of collocated samples. In our study, the variable local sub-models are developed deriving from the light gradient boosting machine[75] (see the Supplementary Methods for details). The parameters of variable local sub-models can automatically vary depending on the number of collocated samples, whose key parameters (see bold fonts in Supplementary Table 8) are determined as provided in Eq. (2).

$$p = \begin{cases} p_{min}, & cou < q_{min} \\ Rou\left(\frac{cou - q_{min}}{q_{max} - q_{min}} \times (p_{max} - p_{min}) + p_{min}\right), & q_{min} < cou < q_{max} \\ p_{max}, & cou \geq q_{max} \end{cases} \quad (2)$$

where p_{min} and p_{max} stand for the minimum and maximum of key parameters, respectively. q_{min} and q_{max} represent the minimum and maximum of number thresholds, respectively. cou is the number of collocated samples. $Rou(\bullet)$ indicates the rounding function. The general expression of the variable local sub-models is defined in Eq. (3).

$$VL_{O3} = \frac{1}{N} \sum_{n=1}^{N} F_{VLn}(V_{THCHO}, V_{TNO2}, V_{CAMS}, V_{ESRI}, V_{EMF}, V_{TC}) \quad (3)$$

where N denotes the number of variable local sub-models. VL_{O3} signifies the modelled MDA8 ambient O_3 concentrations (L-value) aggregated from the intermediate estimation through the variable local sub-models (Ls-values). F_{VLn} indicates the n-th VLSM deriving from the light gradient boosting machine. It is worth noting that the inputs of variable local sub-models discard geographic elements, which ensures their collocated samples can present sufficient temporal differences.

Global-local module: as displayed in Supplementary Fig. 9d, a transition zone is set to smooth the modelled results in the connections of global and local modules. Regarding the target grid i in the transition zone, the modelled results of global (G-value) and local (L-value) modules are merged using a geospatial weighting method. Related weights are defined in Eqs. (4)–(6).

$$VW_{O3} = VG_{O3} \times w_1 + VL_{O3} \times w_2 \quad (4)$$

$$w_1 = \frac{l_1^2}{l_1^2 + l_2^2} \quad (5)$$

$$w_2 = \frac{l_2^2}{l_1^2 + l_2^2} \quad (6)$$

where VW_{O3} indicates the MDA8 ambient O_3 concentrations after geospatial weighting (W-value). l_1 and l_2 stand for the distances between i and two boundaries (l_{min} and l_{max}) of the transition zone, respectively.

Validation scheme

The spatial performance of the GL-CEF model need to be emphatically validated. Hence, the SICV and TESICV schemes of 5 folds are utilized in the present study, which focus on spatial accuracy and spatiotemporally predictive ability, respectively. As depicted in Supplementary Fig. 6a, all the collocated samples are first divided into 5 folds based on spatial locations in the SICV scheme. Next, the GL-CEF model will be trained and validated using 80% (4 folds) and 20% (1 fold) of the collocated samples, respectively. Finally, the above step should be repeated 4 times until each fold has been adopted. As for the TESICV scheme, the only difference is that the training and validation sets came from 2020–2021 and 2019, respectively. In our study, the global imputed and modelled results are verified with the help of 5 metrics: R^2, as defined in Eq. (7); CC, as defined in Eq. (8); RMSE, as defined in Eq. (9); relative percentage error (RPE), as defined in Eq. (10); and mean bias, as defined in Eq. (11).

$$R^2 = 1 - \frac{\sum(\hat{v} - v)^2}{\sum(\bar{v} - v)^2} \quad (7)$$

$$CC = \frac{k\sum v\hat{v} - \sum v \sum \hat{v}}{\sqrt{k\sum v^2 - (\sum v)^2}\sqrt{k\sum \hat{v}^2 - (\sum \hat{v})^2}} \quad (8)$$

$$RMSE = \sqrt{\frac{1}{k}\sum(\hat{v} - v)^2} \quad (9)$$

$$RPE = \frac{RMSE}{\bar{v}} \quad (10)$$

$$\text{mean bias} = \frac{1}{k}\sum(\hat{v} - v) \quad (11)$$

where k represents the number of collocated samples. $\hat{v}$, v, and $\bar{v}$ stand for the modelled, benchmark, and mean benchmark values, respectively.

Population exposure estimation

In this study, the population-weighted MDA8 ambient O_3 concentrations are acquired using Eq. (12)[13,76].

$$v_{pw} = \frac{\sum pop_i \times \hat{v}_i}{\sum pop_i} \quad (12)$$

where $\hat{v}_i$ indicates the modelled MDA8 ambient O_3 concentrations on grid i. pop_i stands for the population density on grid i. v_{pw} represents the population-weighted MDA8 ambient O_3 concentrations over target region. In the meantime, the standard deviation of v_{pw} (i.e., std_{pw}) is computed with Eq. (13)[57,77].

$$std_{pw} = \sqrt{\frac{\sum pop_i \times (\hat{v}_i - v_{pw})}{\sum pop_i \times (e-1)/e}} \quad (13)$$

where e signifies the grid count of the population density exceeding 0. Moreover, the exposure levels during a period are calculated by Eq. (14)[77,78], which temporally cumulates the MDA8 ambient O_3 concentrations on each grid.

$$EL_{U,D}(U_x, D_y) = \frac{\sum pop_i \times I_{U,D}(U_i > U_x, D_i > D_y)}{\sum pop_i} \times 100\% \quad (14)$$

where $EL_{U,D}(U_x, D_y)$ reflects the fraction of the population exposed to MDA8 ambient O_3 concentrations $> U_x$ ppb for more than D_y days.

$I_{U,D}(U_i > U_x, D_i > D_y)$ denotes positive as the day count of MDA8 ambient O_3 concentrations $> U_x$ ppb surpasses D_y days. Otherwise, it is configured to negative.

Mortality burden estimation

In the present study, the long-term all-cause mortality burden due to O_3 exposure is computed according to a log-linear exposure-response function between ambient O_3 and premature deaths from epidemiological researches. The all-cause diseases are defined by A00 to R99 in the International Classification of Diseases 10. The full procedures can be expressed as Eqs. (15)–(18)[2,11,34,79].

$$RR_i = e^{\gamma(v_{mi} - v_s)} \tag{15}$$

where RR_i represents the pooled relative risk on grid i. γ indicates the pooled effect value of long-term ambient O_3 exposure for all-cause deaths from a worldwide meta-analysis involving 226 million participants[36], with a value of 1.39×10^{-3} (95% CI: 8.9597×10^{-4}, 1.8822×10^{-3}) per ppb. This suggests that an increment of 10 ppb in warm-season MDA8 ambient O_3 concentrations is associated with a pooled relative risk of 1.014 (95% CI: 1.009, 1.019) for long-term exposure. v_{mi} is the warm-season modelled MDA8 ambient O_3 concentrations on grid i. v_s stands for the long-term threshold of counterfactual concentrations advised by WHO in 2021[80], which denotes its AQG (30.07 ppb) as the starting point.

$$paf_i = 1 - \frac{1}{RR_i} \tag{16}$$

where paf_i reflects the PAF on grid i, indicating the proportion of mortality burden that will be eliminated when ambient O_3 is decreased to the threshold of counterfactual concentrations (30.07 ppb for long term).

$$tpaf = \frac{\sum pop_i \times paf_i}{\sum pop_i} \tag{17}$$

$$dea = \sum pop_i \times tpaf \times mor \tag{18}$$

where $tpaf$ signifies the PAF over target region. dea denotes the annual long-term O_3-attributable deaths over target region. mor indicates the annual baseline mortality from the GBD 2021[12] over target region. For different land use classes, $tpaf$ is acquired using the population of a single class.

As for short-term O_3 exposure, the associated all-cause deaths are also calculated through Eqs. (15)–(18). By comparison, γ can be obtained from another worldwide meta-analysis for short-term all-cause deaths, with a value of 8.5627×10^{-4} (95% CI: 6.7736×10^{-4}, 1.035×10^{-3})[35] per ppb. This demonstrates that an increment of 10 ppb in MDA8 ambient O_3 concentrations is associated with a pooled relative risk of 1.0086 (95% CI: 1.0068, 1.0104) for short-term exposure. v_{mi} stands for the modelled MDA8 ambient O_3 concentrations on grid i. v_s indicates the short-term threshold of counterfactual concentrations provided by WHO in 2021[80], which deems its AQG (50.11 ppb) as the starting point. The daily PAF is first accumulated and then averaged to acquire the annual value for the short-term O_3 exposure.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Data of in-situ O_3 measurements is available from the Open Air Quality (https://openaq.org), China National Environmental Monitoring Center (http://www.cnemc.cn), US Environmental Protection Agency (https://www.epa.gov), and European Environment Agency (https://www.eea.europa.eu/en/); Data of NO_2 and HCHO TroVCD from the TROPOMI can be accessed at the Copernicus Open Access Hub (https://scihub.copernicus.eu); Data of CAMS and ERA5 atmospheric reanalysis products are available at the Copernicus Atmosphere Data Store (https://ads.atmosphere.copernicus.eu) and Copernicus Climate Data Store (https://cds.climate.copernicus.eu), respectively; Data of LandScan population density and MODIS NDVI/land use classes are open-access at the Oak Ridge National Laboratory (https://landscan.ornl.gov) and NASA Level-1 and Atmosphere Archive & Distribution System Distributed Active Archive Center (https://ladsweb.modaps.eosdis.nasa.gov), respectively; Data of ambient O_3 product from GEOS-CF is available at the NASA Goddard Space Flight Center (https://gmao.gsfc.nasa.gov); Results of the GBD 2021 can be accessed at the Institute for Health Metrics and Evaluation (https://vizhub.healthdata.org/gbd-results/). Modelled daily spatiotemporally continuous 0.05-degree ambient O_3 dataset worldwide generated in this study has been deposited on the Zenodo database with public availability (https://doi.org/10.5281/zenodo.13386462)[81]. Source data are provided with this paper.

Code availability

Relevant code can be accessed from the corresponding authors on request or from the GitHub with public availability (https://github.com/nuist-wy/global_ozone_pollution)[82].

References

1. Springmann, M. et al. The global and regional air quality impacts of dietary change. *Nat. Commun.* **14**, 6227 (2023).
2. Dedoussi, I. C., Eastham, S. D., Monier, E. & Barrett, S. R. H. Premature mortality related to United States cross-state air pollution. *Nature* **578**, 261–265 (2020).
3. Wang, W. et al. Ozone pollution mitigation strategy informed by long-term trends of atmospheric oxidation capacity. *Nat. Geosci.* 1–6 https://doi.org/10.1038/s41561-023-01334-9 (2023).
4. Wang, Y. et al. Sustained emission reductions have restrained the ozone pollution over China. *Nat. Geosci.* **16**, 967–974 (2023).
5. Ou, J. et al. Role of export industries on ozone pollution and its precursors in China. *Nat. Commun.* **11**, 5492 (2020).
6. Akimoto, H. Global air quality and pollution. *Science* **302**, 1716–1719 (2003).
7. Chang, K.-L. et al. A new method (M³Fusion v1) for combining observations and multiple model output for an improved estimate of the global surface ozone distribution. *Geosci. Model Dev.* **12**, 955–978 (2019).
8. Liu, X., Zhu, Y., Xue, L., Desai, A. R. & Wang, H. Cluster-enhanced ensemble learning for mapping global monthly surface ozone from 2003 to 2019. *Geophys. Res. Lett.* **49**, e2022GL097947 (2022).
9. 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 (2022).
10. DeLang, M. N. et al. Mapping yearly fine resolution global surface ozone through the Bayesian maximum entropy data fusion of observations and model output for 1990–2017. *Environ. Sci. Technol.* **55**, 4389–4398 (2021).
11. Malashock, D. A. et al. Global trends in ozone concentration and attributable mortality for urban, peri-urban, and rural areas between 2000 and 2019: a modelling study. *Lancet Planet. Health* **6**, e958–e967 (2022).
12. Brauer, M. et al. Global burden and strength of evidence for 88 risk factors in 204 countries and 811 subnational locations, 1990–2021: a systematic analysis for the Global Burden of Disease Study 2021. *Lancet* **403**, 2162–2203 (2024).

13. Malashock, D. A. et al. Estimates of ozone concentrations and attributable mortality in urban, peri-urban and rural areas worldwide in 2019. *Environ. Res. Lett.* **17**, 054023 (2022).

14. Jerrett, M. et al. Long-term ozone exposure and mortality. *N. Engl. J. Med.* **360**, 1085–1095 (2009).

15. Wang, Y., Yuan, Q., Zhu, L. & Zhang, L. Spatiotemporal estimation of hourly 2-km ground-level ozone over China based on Himawari-8 using a self-adaptive geospatially local model. *Geosci. Front.* **13**, 101286 (2022).

16. Liu, R. et al. Spatiotemporal distributions of surface ozone levels in China from 2005 to 2017: a machine learning approach. *Environ. Int.* **142**, 105823 (2020).

17. Malley, C. S. et al. Updated global estimates of respiratory mortality in adults ≥ 30years of age attributable to long-term ozone exposure. *Environ. Health Perspect.* https://doi.org/10.1289/EHP1390 (2017).

18. Chowdhury, S., Pozzer, A., Dey, S., Klingmueller, K. & Lelieveld, J. Changing risk factors that contribute to premature mortality from ambient air pollution between 2000 and 2015. *Environ. Res. Lett.* **15**, 074010 (2020).

19. Zhang, X. Y., Zhao, L. M., Cheng, M. M. & Chen, D. M. Estimating ground-level ozone concentrations in eastern China using satellite-based precursors. *IEEE Trans. Geosci. Remote Sens.* **58**, 4754–4763 (2020).

20. Li, T., Wu, J., Chen, J. & Shen, H. An enhanced geographically and temporally weighted neural network for remote sensing estimation of surface ozone. *IEEE Trans. Geosci. Remote Sens.* **60**, 1–13 (2022).

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

22. Wang, W., Liu, X., Bi, J. & Liu, Y. A machine learning model to estimate ground-level ozone concentrations in California using TROPOMI data and high-resolution meteorology. *Environ. Int.* **158**, 106917 (2022).

23. Chen, Z.-Y. et al. Population exposure to multiple air pollutants and its compound episodes in Europe. *Nat. Commun.* **15**, 2094 (2024).

24. Xu, R. et al. Global population exposure to landscape fire air pollution from 2000 to 2019. *Nature* **621**, 521–529 (2023).

25. Xue, T. et al. Estimating the exposure–response function between long-term ozone exposure and under-5 mortality in 55 low-income and middle-income countries: a retrospective, multicentre, epidemiological study. *Lancet Planet. Health* **7**, e736–e746 (2023).

26. Lu, X. et al. The underappreciated role of agricultural soil nitrogen oxide emissions in ozone pollution regulation in North China. *Nat. Commun.* **12**, 5021 (2021).

27. Kou, W. et al. High downward surface solar radiation conducive to ozone pollution more frequent under global warming. *Sci. Bull.* **68**, 388–392 (2023).

28. Mousavinezhad, S., Choi, Y., Pouyaei, A., Ghahremanloo, M. & Nelson, D. L. A comprehensive investigation of surface ozone pollution in China, 2015–2019: separating the contributions from meteorology and precursor emissions. *Atmos. Res.* **257**, 105599 (2021).

29. Wei, J. et al. First close insight into global daily gapless 1 km $PM_{2.5}$ pollution, variability, and health impact. *Nat. Commun.* **14**, 8349 (2023).

30. Punger, E. M. & West, J. J. The effect of grid resolution on estimates of the burden of ozone and fine particulate matter on premature mortality in the USA. *Air Qual. Atmosphere Health* **6**, 563–573 (2013).

31. Rowlinson, M. J. et al. Tropospheric ozone radiative forcing uncertainty due to pre-industrial fire and biogenic emissions. *Atmos. Chem. Phys.* **20**, 10937–10951 (2020).

32. Kumar, A., Bali, K., Singh, S., Naja, M. & Mishra, A. K. Estimates of reactive trace gases (NMVOCs, CO and NO_x) and their ozone forming potentials during forest fire over Southern Himalayan region. *Atmos. Res.* **227**, 41–51 (2019).

33. Anderson, D. C. et al. A pervasive role for biomass burning in tropical high ozone/low water structures. *Nat. Commun.* **7**, 10267 (2016).

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

35. Orellano, P., Reynoso, J., Quaranta, N., Bardach, A. & Ciapponi, A. Short-term exposure to particulate matter (PM_{10} and $PM_{2.5}$), nitrogen dioxide (NO_2), and ozone (O_3) and all-cause and cause-specific mortality: systematic review and meta-analysis. *Environ. Int.* **142**, 105876 (2020).

36. Sun, H. Z. et al. Cohort-based long-term ozone exposure-associated mortality risks with adjusted metrics: a systematic review and meta-analysis. *Innovation* **3**, 100246 (2022).

37. Hu, X. et al. Inflammatory and oxidative stress responses of healthy adults to changes in personal air pollutant exposure. *Environ. Pollut.* **263**, 114503 (2020).

38. Wang, L. et al. Impacts of future land use and land cover change on mid-21st-century surface ozone air quality: distinguishing between the biogeophysical and biogeochemical effects. *Atmos. Chem. Phys.* **20**, 11349–11369 (2020).

39. Erickson, L. E., Newmark, G. L., Higgins, M. J. & Wang, Z. Nitrogen oxides and ozone in urban air: a review of 50 plus years of progress. *Environ. Prog. Sustain. Energy* **39**, e13484 (2020).

40. Pozzer, A. et al. Mortality attributable to ambient air pollution: a review of global estimates. *GeoHealth* **7**, e2022GH000711 (2023).

41. Hua, Q. et al. Ozone exposure and cardiovascular disease: a narrative review of epidemiology evidence and underlying mechanisms. *Fundam. Res.* (2024).

42. Vicedo-Cabrera, A. M. et al. Short term association between ozone and mortality: global two stage time series study in 406 locations in 20 countries. *BMJ* **368**, m108 (2020).

43. Liu, C. et al. Interactive effects of ambient fine particulate matter and ozone on daily mortality in 372 cities: two stage time series analysis. *BMJ* **383**, e075203 (2023).

44. Huangfu, P. & Atkinson, R. Long-term exposure to NO_2 and O_3 and all-cause and respiratory mortality: a systematic review and meta-analysis. *Environ. Int.* **144**, 105998 (2020).

45. Peng, M. et al. Long-term ozone exposure and all-cause mortality: cohort evidence in China and global heterogeneity by region. *Ecotoxicol. Environ. Saf.* **270**, 115843 (2024).

46. Ban, J., Lu, K., Wang, Q. & Li, T. Climate change will amplify the inequitable exposure to compound heatwave and ozone pollution. *One Earth* **5**, 677–686 (2022).

47. Gao, M. et al. Large-scale climate patterns offer preseasonal hints on the co-occurrence of heat wave and O_3 pollution in China. *Proc. Natl Acad. Sci.* **120**, e2218274120 (2023).

48. Ranzani, O. et al. Long-term exposure to air pollution and severe COVID–19 in Catalonia: a population-based cohort study. *Nat. Commun.* **14**, 2916 (2023).

49. Musonye, H. A. et al. Exploring the association between ambient air pollution and COVID-19 risk: a comprehensive meta-analysis with meta-regression modelling. *Heliyon* (2024).

50. De Smedt, I. et al. Algorithm theoretical baseline for formaldehyde retrievals from S5P TROPOMI and from the QA4ECV project. *Atmos. Meas. Tech.* **11**, 2395–2426 (2018).

51. Van Geffen, J., Eskes, H. J., Boersma, K. F., Maasakkers, J. D. & Veefkind, J. P. TROPOMI ATBD of the total and tropospheric NO_2 data products. *DLR Doc.* (2019).

52. Inness, A. et al. The CAMS reanalysis of atmospheric composition. *Atmos. Chem. Phys.* **19**, 3515–3556 (2019).

53. Yeung, L. Y. et al. Isotopic constraint on the twentieth-century increase in tropospheric ozone. *Nature* **570**, 224–227 (2019).

54. Ivatt, P. D., Evans, M. J. & Lewis, A. C. Suppression of surface ozone by an aerosol-inhibited photochemical ozone regime. *Nat. Geosci.* **15**, 536–540 (2022).

55. Qin, M. et al. Criteria pollutant impacts of volatile chemical products informed by near-field modelling. *Nat. Sustain.* **4**, 129–137 (2021).

56. Keller, C. A. et al. Description of the NASA GEOS composition forecast modeling system GEOS-CF v1.0. *J. Adv. Model. Earth Syst.* **13**, e2020MS002413 (2021).

57. Wang, Y., Yuan, Q., Li, T. & Zhu, L. Global spatiotemporal estimation of daily high-resolution surface carbon monoxide concentrations using Deep Forest. *J. Clean. Prod.* **350**, 131500 (2022).

58. Yang, Q. et al. Global air quality change during COVID-19: a synthetic analysis of satellite, reanalysis and ground station data. *Environ. Res. Lett.* **16**, 074052 (2021).

59. Sokhi, R. S. et al. A global observational analysis to understand changes in air quality during exceptionally low anthropogenic emission conditions. *Environ. Int.* **157**, 106818 (2021).

60. Hersbach, H. et al. The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc.* **146**, 1999–2049 (2020).

61. Li, T. & Cheng, X. Estimating daily full-coverage surface ozone concentration using satellite observations and a spatiotemporally embedded deep learning approach. *Int. J. Appl. Earth Obs. Geoinf.* **101**, 102356 (2021).

62. Fensholt, R. & Proud, S. R. Evaluation of earth observation based global long term vegetation trends — comparing GIMMS and MODIS global NDVI time series. *Remote Sens. Environ.* **119**, 131–147 (2012).

63. Friedl, M. A. et al. Global land cover mapping from MODIS: algorithms and early results. *Remote Sens. Environ.* **83**, 287–302 (2002).

64. Meng, X., Jiang, Z., Wang, X. & Long, Y. Shrinking cities on the globe: evidence from LandScan 2000–2019. *Environ. Plan. Econ. Space* **53**, 1244–1248 (2021).

65. Pavel, M. R. S., Zaman, S. U. & Jeba, F. & Salam, A. Long-term (2011–2019) trends of O_3, NO_2, and HCHO and sensitivity analysis of O_3 chemistry over the GBM (Ganges–Brahmaputra–Meghna) Delta: spatial and temporal variabilities. *ACS Earth Space Chem.* **5**, 1468–1485 (2021).

66. Ren, J., Guo, F. & Xie, S. Diagnosing ozone–NO_x–VOC sensitivity and revealing causes of ozone increases in China based on 2013–2021 satellite retrievals. *Atmos. Chem. Phys.* **22**, 15035–15047 (2022).

67. Wang, W., van der A, R., Ding, J., van Weele, M. & Cheng, T. Spatial and temporal changes of the ozone sensitivity in China based on satellite and ground-based observations. *Atmos. Chem. Phys.* **21**, 7253–7269 (2021).

68. Vigouroux, C. et al. TROPOMI–Sentinel-5 Precursor formaldehyde validation using an extensive network of ground-based Fourier-transform infrared stations. *Atmos. Meas. Tech.* **13**, 3751–3767 (2020).

69. Alvera-Azcárate, A., Barth, A., Sirjacobs, D. & Beckers, J.-M. Enhancing temporal correlations in EOF expansions for the reconstruction of missing data using DINEOF. *Ocean Sci.* **5**, 475–485 (2009).

70. Rukundo, O. & Cao, H. Nearest neighbor value interpolation. *Int. J. Adv. Comput. Sci. Appl.* **3**, (2012).

71. Setianto, A. & Triandini, T. Comparison of kriging and inverse distance weighted (IDW) interpolation methods in lineament extraction and analysis. *J. Appl. Geol.* **5**, (2013).

72. Liu, D. et al. Estimating ground-level CO concentrations across China based on the national monitoring network and MOPITT: potentially overlooked CO hotspots in the Tibetan Plateau. *Atmos. Chem. Phys.* **19**, 12413–12430 (2019).

73. Zhou, Z.-H. & Feng, J. Deep forest. *Natl Sci. Rev.* **6**, 74–86 (2019).

74. Yang, N., Shi, H., Tang, H. & Yang, X. Geographical and temporal encoding for improving the estimation of $PM_{2.5}$ concentrations in China using end-to-end gradient boosting. *Remote Sens. Environ.* **269**, 112828 (2022).

75. Ke, G. et al. LightGBM: a highly efficient gradient boosting decision tree. In *Advances in Neural Information Processing Systems* vol. 30 (Curran Associates, Inc., 2017).

76. Wang, Y. et al. Health impacts of long-term ozone exposure in China over 2013–2017. *Environ. Int.* **144**, 106030 (2020).

77. Zhan, Y. et al. Spatiotemporal prediction of daily ambient ozone levels across China using random forest for human exposure assessment. *Environ. Pollut.* **233**, 464–473 (2018).

78. Chen, L. et al. Increases in ozone-related mortality in China over 2013–2030 attributed to historical ozone deterioration and future population aging. *Sci. Total Environ.* **858**, 159972 (2023).

79. Xiao, Q. et al. Tracking $PM_{2.5}$ and O_3 pollution and the related health burden in China 2013–2020. *Environ. Sci. Technol.* **56**, 6922–6932 (2022).

80. WHO. WHO global air quality guidelines: particulate matter ($PM_{2.5}$ and PM_{10}), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide. *World Health Organization*. (2021).

81. Wang, Y. et al. Substantially underestimated global health risks of current ozone pollution. *Zenodo*. https://doi.org/10.5281/zenodo.13386462 (2024).

82. Wang, Y. et al. Substantially underestimated global health risks of current ozone pollution. *Zenodo*. https://doi.org/10.5281/zenodo.14141984 (2024).

Acknowledgements

Author contributions

Y.W., Q.Y., and Lp.Z. conceive and conceptualize the study. Y.W. and Y.Y. devise the methodology. Y.W. collects and processes the data, analyses the results, and writes the first manuscript. T.L. performs the validation of the GL-CEF model. Y.Z. collects and processes the results of GBD 2021. L.Z. and M.W. collect and process the ambient O_3 product of GEOS-CF. Z.X., M.G., and S.T. provide constructive comments and perform critical editing of the manuscript. H.C.H. and S.L. revise and polish the manuscript.

Competing interests

The authors declare no competing interests.

Additional information