**Urban-rural differentiated ambient O3 tracking**

The core basis ambient O3 concentration tracking database with urban-rural distinguishment was developed by a two-stage space-time Bayesian neural network framework, consisting of first-stage multi-model ensembler (BayNNE)1 and second-stage downscaler (BayNND)2. BayNNE integrated eight fully-coupled free-running simulations from CMIP6 (Coupled Model Intercomparison Project Phase 6) endorsed Earth system models with interactive chemistry and chemistry-climate feedbacks, assisted with over 40 auxiliary predictors including sociodemographic, ecological, and emission features2, improved from the previously published version (see details in Supplementary Method 1). The target spatial resolution was set at 1°×1°, capturing the cell-average ambient O3 concentrations with intra-cell variabilities smoothed. Predictions of cell-average concentrations () followed Equation 1, which were the basis for further downscaling. In equation, *M*(*i*) refer to simulations by different models, and subscripts *loc* and *t* represent spatial locations (by coordinates) and temporal nodes (by month), respectively.

|  |  |
| --- | --- |
|  | Equation 1 |

BayNND predicted ambient O3 concentrations from BayNNE-generated cell-level averages concentrations in 1/8°×1/8° spatial resolution with stacked urban-rural differentiation. The “stacked” downscaling algorithm encapsulated urban- and rural-averaged ambient O3 concentrations into each spatial cell, assigning all urban (or rural) population in each cell uniformly with a cell-specific urban (or rural) prediction. The schematic diagram of two-stage Bayesian neural network algorithms was conceptualised in Supplementary Fig. X, and mathematical forms of BayNND are demonstrated in Equation 2 and Equation 3, whererepresents Bayesian neural network regressor, *e* for Bayesian estimation ensemble member, *res* for urban/rural classification, *si* for 3 spatial indicators, *ti* for 3 temporal indicators, and *a* for auxiliary predictors. The parameter family ***θ*** including *αi*, *β*, *σ*, *k*, and *δ* were predicted from ensemble averages by Markov Chain Monte Carlo method for Bayesian neural network.

|  |  |
| --- | --- |
|  | Equation 2 |
|  | Equation 3 |

BayNND could reach accuracy R2 =0.91, RMSE =4.5 ppb for urban, and R2 =0.89, RMSE =5.2 ppb for rural sites under 30-year overall global-scale evaluation, and R2 =0.90, RMSE =4.6 ppb for Chinese urban, and R2 =0.94, RMSE =5.5 ppb for rural sites during 2014–2019.

**Data fusion**

Besides the BayNND, we fused two additional peer-reviewed high-quality data products3,4 to realise an enhanced 30-year historical monthly averaged ambient O3 concentration database spanning 1990–2019. One ambient O3 products was constructed using cluster-enhanced ensemble machine learning (CEML), training region-exclusive algorithms to retain the geographical variability.3 CEML mixed the results from chemistry reanalysis and remote sensing, with over 80 supplemental geographical and meteorological features, to realise 0.5°×0.5° monthly resolved ambient O3 concentrations across 2003–2019, with overall accuracy R2 =0.92, RMSE =4.1 ppb.

The other base dataset supported by the team of Tracking Air Pollution in China (TAP), was produced by random forest regressor with stochastic spatial auto-correlation signal compensation4. TAP utilised CTM simulations and satellite remote-sensing measurements to realise near real-time 0.1°×0.1° daily prediction since 2013, achieving accuracy as R2 =0.70, RMSE =13.3 ppb. All 3 data products measured the ambient O3 in metric of daily maximum 8-hour average. Detailed procedures were precisely delineated in the original literatures2-4.

Fusing multiple databases supervised by *in situ* observations can restrict biases from any single approach. As all three ambient O3 tracking products had achieved high consistency with the observations, we used an elastic net regressor to fuse BayNND, CEML, and TAP, assisted with 3 spatial and 3 temporal indicators2, to avoid overfitting. The 10-fold methodological cross-validation tests on Chinese sites during 2014–2019 revealed R2 ≥0.82, RMSE ≤7.0 ppb, and 30-year global overall accuracy of the final dataset was R2 = 0.92, RMSE = 4.4 ppb. Detailed phased procedures for data fusion were illustrated in Supplementary Method 2, and performance evaluations were stated in Supplementary Table 6*.*

Finally, by highlighting the peak exposure (April to September), 6-month ozone-season daily maximum 8-hour average (OSDMA8) was calculated for mortality estimation. The Bayesian neural networks were constructed on Python-package *TensorFlow* (version 2.3.1), and elastic net regression was performed by *scikit-learn* (version 0.23.2).

**Ground-level observations for supervised training and validation**

We used stationary observations as labels for all-stage supervised model training and accuracy evaluation. The urban-rural distinguished *in situ* observations were obtained from the Tropospheric Ozone Assessment Report archives (TOAR)5 and China National Environmental Monitoring Centre (CNEMC)6. In the first-stage multi-model fusion, 1°×1° gridded cell-average concentrations including all available sites excluding CNEMC stations (cell-average levels could be urban-biased due to disproportional deployment in urban and rural environments) were used as supervision labels for model training. The global-scale overall fitting accuracy was R2 =0.96, RMSE =2.1 ppb by metric of OSDMA8, and the evaluation of 10-fold cross-validation test showed R2 =0.92. In the second-stage 1/8°×1/8° gridded downscaling with urban-rural differentiation and third-stage data fusion, we used urban- and rural-labelled observations for model training. Throughout the studied 30 years globally, accuracy of urban predictions was R2 =0.92, RMSE =2.8 ppb (cross-validation R2 =0.89) and R2 =0.94, RMSE =3.3 ppb (cross-validation R2 =0.91) for rural predictions in the second-stage Bayesian neural network downscaler. For the latest six years (2014–2019), prediction accuracies were evaluated with observations in China, as R2 =0.91, RMSE =4.2 ppb (cross-validation R2 =0.82) for urban, and R2 =0.89, RMSE =5.2 ppb (cross-validation R2 =0.86) for rural predictions by the third-stage data fusion algorithm.

TOAR clearly recognised 1,704 urban and 1,274 rural sites based on population density by remote-sensing; CNEMC identified 1,777 urban and 251 suburban sites by administrative district division, whereas 251 suburban-labelled sites were reclassified as rural sites throughout this study, as i) the observed “suburban”-labelled ambient O3 concentrations were closer to the predicted rural concentrations (R2 =0.84, NMB =3.2%) than urban predictions (R2 =0.56, normalised mean bias, NMB =15.7%, see evaluation details in Supplementary Fig. X), and ii) the projected population density of 2019 of the “suburban”-labelled sites were way lower than 1,500 people per km2, the urbanisation standard (see full list in Supplementary Table X).

**Cell-level urbanisation and population migration**

The urban-rural binary classification for each cell resided with habitants was based on the population density of each 30″×30″ fine cell: >1,500 people per km2 as urban, and <1,500 people per km2 as rural. When upscaling to 1/8°×1/8° coarser cell, the urban and rural residents were summed up separately and stacked in each coarse cell. The reason of gridded population upscaling is the spatial resolution limitation of ambient O3 tracking (approximately 10×10 km2). A schematic illustration was attached in Supplementary Fig. X. The definition of urbanisation throughout the study is cell-level proportion of urban residents among all population. Due to data unavailability, we did not track the individual-level migration behaviour, whereby rural-to-urban population migration was reflected in a cross-sectional level by change of the urban-rural population structure, as illustrated in Supplementary Fig. X.

**Population-weighted exposure assignment**

REFERENCES

1. Sun Z, Archibald AT. Multi-stage ensemble-learning-based model fusion for surface ozone simulations: A focus on CMIP6 models. *Environmental Science and Ecotechnology* 2021; **8**: 100124.

2. Sun H, Shin YM, Xia M, et al. Spatial Resolved Surface Ozone with Urban and Rural Differentiation during 1990-2019: A Space-Time Bayesian Neural Network Downscaler. *Environ Sci Technol* 2022; **56**(11): 7337-49.

3. Liu X, Zhu YJ, Xue L, Desai AR, Wang HK. Cluster-Enhanced Ensemble Learning for Mapping Global Monthly Surface Ozone From 2003 to 2019. *Geophys Res Lett* 2022; **49**(11): e2022GL097947.

4. Xue T, Zheng Y, Geng G, et al. Estimating Spatiotemporal Variation in Ambient Ozone Exposure during 2013-2017 Using a Data-Fusion Model. *Environ Sci Technol* 2020; **54**(23): 14877-88.

5. Schultz MG, Schroder S, Lyapina O, et al. Tropospheric Ozone Assessment Report: Database and metrics data of global surface ozone observations. *Elementa-Sci Anthrop* 2017; **5**(58): 1-26.

6. Lu X, Hong JY, Zhang L, et al. Severe Surface Ozone Pollution in China: A Global Perspective. *Environ Sci Tech Let* 2018; **5**(8): 487-94.