



HAL
open science

Using Regionalized Air Quality Model Performance and Bayesian Maximum Entropy data fusion to map global surface ozone concentration

Jacob Becker, Marissa Delang, Kai-Lan Chang, Marc Serre, Owen Cooper, Hantao Wang, Martin Schultz, Sabine Schröder, Xiao Lu, Lin Zhang, et al.

► To cite this version:

Jacob Becker, Marissa Delang, Kai-Lan Chang, Marc Serre, Owen Cooper, et al.. Using Regionalized Air Quality Model Performance and Bayesian Maximum Entropy data fusion to map global surface ozone concentration. *Elementa: Science of the Anthropocene*, 2023, 11 (1), 10.1525/elementa.2022.00025 . hal-04251795

HAL Id: hal-04251795

<https://hal.science/hal-04251795v1>

Submitted on 20 Oct 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.


L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

RESEARCH ARTICLE

Using Regionalized Air Quality Model Performance and Bayesian Maximum Entropy data fusion to map global surface ozone concentration

Jacob S. Becker¹, Marissa N. DeLang¹, Kai-Lan Chang^{2,3}, Marc L. Serre¹, Owen R. Cooper^{2,3}, Hantao Wang¹, Martin G. Schultz⁴, Sabine Schröder⁴, Xiao Lu⁵, Lin Zhang⁶, Makoto Deushi⁷, Beatrice Josse⁸, Christoph A. Keller^{9,10}, Jean-François Lamarque¹¹, Meiyun Lin^{12,13}, Junhua Liu^{9,10}, Virginie Marécal⁸, Sarah A. Strode^{9,10}, Kengo Sudo^{14,15}, Simone Tilmes¹¹, Li Zhang^{12,13,16}, Michael Brauer^{17,18}, and J. Jason West^{1,*} 

Estimates of ground-level ozone concentrations have been improved through data fusion of observations and atmospheric chemistry models. Our previous global ozone estimates for the Global Burden of Disease study corrected for bias uniformly across continents and then corrected near monitoring stations using the Bayesian Maximum Entropy (BME) framework for data fusion. Here, we use the Regionalized Air Quality Model Performance (RAMP) framework to correct model bias over a much larger spatial range than BME can, accounting for the spatial inhomogeneity of bias and nonlinearity as a function of modeled ozone. RAMP bias correction is applied to a composite of 9 global chemistry-climate models, based on the nearest set of monitors. These estimates are then fused with observations using BME, which matches observations at measurement stations, with the influence of observations declining with distance in space and time. We create global ozone maps for each year from 1990 to 2017 at fine spatial resolution. RAMP is shown to create unrealistic discontinuities due to the spatial clustering of ozone monitors, which we overcome by applying a weighting for RAMP based on the number of monitors nearby. Incorporating RAMP before BME has little effect on model performance near stations, but strongly increases R^2 by 0.15 at locations farther from stations, shown through a checkerboard cross-validation. Corrections to estimates differ based on location in space and time, confirming heterogeneity. We quantify the likelihood of exceeding selected ozone levels, finding that parts of the Middle East, India, and China are most likely to exceed 55 parts per billion (ppb) in 2017. About 96% of the global population was exposed to ozone levels above the World Health Organization guideline of $60 \mu\text{g m}^{-3}$ (30 ppb) in 2017. Our annual fine-resolution ozone estimates may be useful for several applications including epidemiology and assessments of impacts on health, agriculture, and ecosystems.

Keywords: Ozone, Data fusion, Ground-level concentrations, Exposure, Air pollution

¹ Department of Environmental Sciences and Engineering, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

² Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO, USA

³ NOAA Chemical Sciences Laboratory, Boulder, CO, USA

⁴ Jülich Supercomputing Centre (JSC), Forschungszentrum Jülich, Jülich, Germany

⁵ School of Atmospheric Sciences, Sun Yat-Sen University, Zhuhai, Guangdong, China

⁶ Laboratory for Climate and Ocean-Atmosphere Studies, Department of Atmospheric and Oceanic Sciences, School of Physics, Peking University, Beijing, China

⁷ Meteorological Research Institute (MRI), Tsukuba, Japan

⁸ Centre National de Recherches Météorologiques, Université de Toulouse, Météo-France, CNRS, Toulouse, France

⁹ NASA Goddard Space Flight Center, Greenbelt, MD, USA

¹⁰ Universities Space Research Association, Columbia, MD, USA

¹¹ National Center for Atmospheric Research, Boulder, CO, USA

¹² NOAA Geophysical Fluid Dynamics Laboratory, Princeton, NJ, USA

¹³ Program in Atmospheric and Oceanic Sciences, Princeton University, Princeton, NJ, USA

¹⁴ Graduate School of Environmental Studies, Nagoya University, Nagoya, Japan

¹⁵ Japan Agency for Marine-Earth Science and Technology (JAMSTEC), Yokosuka, Japan

¹⁶ Department of Meteorology and Atmospheric Science, Pennsylvania State University, University Park, PA, USA

¹⁷ Institute for Health Metrics and Evaluation, University of Washington, Seattle, WA, USA

¹⁸ School of Population and Public Health, University of British Columbia, Vancouver, British Columbia, Canada

* Corresponding author:
Email: jasonwest@unc.edu

1. Introduction

Ground-level ozone is a pervasive air pollutant that detrimentally affects human health and vegetation. Ozone can cause a wide range of health problems in humans and has been associated with premature mortality from daily exposures (Bell et al., 2004; Di et al., 2017; U.S. Environmental Protection Agency [U.S. EPA], 2020) and from long-term chronic exposure (Jerrett et al., 2009; Turner et al., 2016; U.S. EPA, 2020). Ozone concentrations above roughly 35 parts per billion (ppb) are associated with higher respiratory and cardiovascular mortality, with every 10 ppb increase in ozone concentration increasing all-cause mortality by 2% (Turner et al., 2016). Ambient ozone is estimated to have caused about 365,000 deaths globally in 2019 or 0.65% of all global deaths (GBD 2019 Risk Factors Collaborators, 2020). Recently, the World Health Organization (WHO) established new guidelines for ambient ozone of $100 \mu\text{g m}^{-3}$ (50 ppb) for 8-h average ozone and of $60 \mu\text{g m}^{-3}$ (30 ppb) for the annual metric used in this study (WHO, 2021). These guidelines suggest that ozone affects health even at concentrations frequently observed in relatively unpolluted regions. Unlike other air pollutants, ozone is purely secondary, which is created through photochemical reactions involving nitrogen oxides (NO_x), volatile organic compounds (VOCs), carbon monoxide, and methane in the presence of sunlight. Ozone concentrations are typically higher in the daytime and during summer months.

Understanding of ozone impacts on human health and plants has been limited in part by our understanding of ground-level ozone distributions in space and time. Estimates of surface ozone distributions rely on monitoring station observations and chemical transport models, but both have limitations. While the United States, Europe, and Japan have dense monitoring networks that began prior to 1990 and China recently created a large network, station observations of ozone elsewhere are extremely limited (Schultz et al., 2017; Fleming et al., 2018; Lu et al., 2020). Models can help fill in these gaps in space and time but have biases (Cooper et al., 2014; Young et al., 2018; Archibald et al., 2020).

In our previous work, we conducted a data fusion of ozone observations and multiple global atmospheric models, in 2 phases, to estimate global ground-level ozone concentrations at fine spatial resolution (Chang et al., 2019; DeLang et al., 2021). These ozone maps were used to estimate global premature deaths from exposure to ambient ozone in the Global Burden of Diseases, Injuries, and Risk Factors (GBD) 2017 and 2019 Studies (GBD 2017 Risk Factors Collaborators, 2018; GBD 2019 Risk Factors Collaborators, 2020). GBD conducts a comparative risk assessment to estimate the global health burden caused by various risk factors from 1990 to the present.

Prior to GBD 2017, ozone in previous GBD studies was estimated solely by a single model with no observational bias correction (Brauer et al., 2016). The first study to combine ozone observations and output from global atmospheric chemistry models applied the new M³Fusion method to correct model bias, improving global ozone estimates from purely observation- or model-based

approaches (Chang et al., 2019), in support of GBD 2017. M³Fusion bias corrects and combines multiple chemical transport models by finding an optimal linear combination of models for each world region, using weighting based on performance when compared to available observations. This multimodel composite was then corrected within 2° of a monitoring station using a spatial interpolation of observations, creating fine resolution output.

To support GBD 2019, we then improved this method using a novel combination of Bayesian Maximum Entropy (BME) along with M³Fusion (DeLang et al., 2021). BME is a framework for nonlinear geostatistics that performs the fusion of data from multiple sources (Serre and Christakos, 1999; Christakos et al., 2001; Christakos et al., 2004). BME used observations to correct the M³Fusion multimodel composite smoothly in both space and time, so that ozone estimates match the observations at station locations. The influence each station exerts diminishes over time and space based on a calculated spatiotemporal covariance function. BME had been used before on smaller scales to fuse ozone observations and models (Christakos et al., 2004; de Nazelle et al., 2010; Xu et al., 2016), but DeLang et al. (2021) were the first to apply it globally. The ability of observations in BME to influence estimates across time was also shown to be useful in informing earlier years before new stations were added (DeLang et al., 2021).

DeLang et al. (2021) showed that ozone estimates improved markedly over purely model- or observation-based estimates and produced a global fine resolution (0.1°) dataset for each year from 1990 to 2017. However, like Chang et al. (2019) and DeLang et al. (2021) used the M³Fusion method, which uses linear model bias corrections that are homogeneous across continents. BME then corrected estimates based on nearby observations, allowing for spatial nonhomogeneous bias corrections. The spatial scale over which BME has an influence is not bounded, but it is determined by the spatial covariance function based on observations from which we find that BME ozone bias corrections are small beyond about 0.5° from a monitoring location (DeLang et al., 2021). Here, we implement a second method that allows spatial nonhomogeneous and nonlinear bias corrections that have influence beyond 0.5° from a monitoring location.

Previous research has shown that air pollution model performance and biases are nonhomogeneous (vary by location) and nonlinear (vary as a nonlinear function of the model estimate) (Reyes et al., 2017). For example, chemical transport models generally overpredict fine particulate matter ($\text{PM}_{2.5}$) when predicting high ($>25 \mu\text{g m}^{-3}$) values and underpredict elsewhere (Reyes et al., 2017). Model errors for ozone stem from uncertainties in inputs, especially emissions of ozone precursors (NO_x and VOCs) from anthropogenic and natural sources, and in model processes including chemistry, model resolution, transport, and deposition (Young et al., 2018). Previous model evaluations have found that models have errors that vary by season and latitude (von Kuhlman et al., 2003), reflecting uncertainties in emissions inputs and in physical and

chemical processes within the models. These errors could lead to overestimates in some locations and underestimates in others, so that model performance is heterogeneous (Liang and Jacobson, 2003). For example, emission estimates used in an atmospheric model may be too high or too low for a nation, and if monitors are present, it may be possible to identify and correct biases over that nation. Ozone production is also known to change nonlinearly with emissions (Cohan et al., 2005). Consequently, model-simulated ozone concentrations exhibit a nonlinear bias with respect to observations. The Community Multiscale Air Quality model, for example, has been shown to overestimate maximum 8-h average ozone levels where observations are below 35 ppb and underestimate where observations are above 85 ppb (Appel et al., 2007).

The goal of this study is to improve the work of DeLang et al. (2021) to map global surface ozone concentrations each year from 1990 to 2017 at fine spatial resolution by adding a nonlinear and heterogeneous bias correction before BME data fusion and then evaluate the improvement in performance. We use the Constant Air Quality Model Performance (CAMP) method, which allows for nonlinear bias correction as a function of the modeled ozone concentration, and the Regionalized Air Quality Model Performance (RAMP) method (Reyes et al., 2017), which applies CAMP regionally at individual points based on the nearest observations. CAMP and RAMP corrections are applied to the M³Fusion multimodel composite prior to BME data fusion. While BME is not restricted spatially, ozone's steep covariance curve means that observations have little influence beyond about 0.5° from a monitoring station (DeLang et al., 2021). In contrast, CAMP and RAMP corrections can influence estimates over a larger spatial range. Specifically, we aim to use regional differences in model underestimation/overestimation to correct the M³Fusion results regionally and increase the fidelity of our estimation in areas with sparse or no ozone observation stations, and we apply CAMP and RAMP here to improve estimates in particular beyond about 0.5° from a monitor. These geostatistical methods differ from other recent efforts to map ozone concentrations based on machine learning (Seltzer et al., 2020; Kleinert et al., 2021; Sun et al., 2021; Betancourt et al., 2022; Liu et al., 2022), and so there is value in further developing geostatistical methods for comparison with these recent advancements.

Both CAMP and RAMP bias correct models by comparing observed and modeled values at collocated points in space and time. CAMP applies a nonlinear correction as a function of the modeled ozone concentration, assuming that nonlinear function is constant across the study region. RAMP improves on this by giving each model grid cell its own nonlinear bias correction based on nearby observations; RAMP is both nonlinear and nonhomogeneous spatially. Here, the RAMP method (Reyes et al., 2017) is applied globally for the first time, with each model grid cell correction based on a unique area that includes the nearest points in space/time. These areas are much smaller than the continental regions used in

M³Fusion, allowing us to better correct biases in the M³Fusion multimodel composite at points far from observations, while BME then applies corrections near observations. In applying RAMP at a global scale, we also make a novel modification of the RAMP method because station observations are sparse in some regions and clustered in others. This modification prevents sharp spatial changes in corrections when transitioning between 2 different regions with dense observation networks. The CAMP- and RAMP-corrected estimates are then each used as global background ozone levels (the global offset) for BME data fusion with observations, as was done by DeLang et al. (2021) with the uncorrected M³Fusion multimodel composite.

We aim for the improved global ozone database to be useful for researchers in a variety of fields, including atmospheric science, epidemiology and health impact assessment, and ecosystem and crop impacts. As this is the first global application of CAMP and RAMP, we document the methods and results to inform future global data fusion efforts for ozone and other air pollutants. Finally, by further developing geostatistical methods for global ozone estimates in this study, we also aim to provide ozone estimates that can be compared with recent machine learning applications in future studies.

2. Methods

2.1. Ozone observations and model estimates

We use the ozone season daily maximum 8-h mixing ratio (OSDMA8) as the annual ozone metric, as it is used for calculating health outcomes from ozone pollution by GBD (GBD 2019 Risk Factors Collaborators, 2020) using concentration–mortality relationships from Turner et al. (2016) and other studies. OSDMA8 is the maximum 6-month running mean of monthly averages of the daily 8-h maximum mixing ratios. Since ozone is usually highest in the summer, each defined year includes up to March of the following year to capture the Southern Hemisphere summer. All reported ozone values, including observations, modeled values, and estimates, are OSDMA8 values.

Ground-level ozone measurements are taken from databases compiled by the Tropospheric Ozone Assessment Report (TOAR) and the Chinese National Environmental Monitoring Center (CNEMC) (Schultz et al., 2017; Lu et al., 2020). The TOAR database is the largest collection of global hourly surface ozone concentrations and spans 1970–2015. To support this project, some national datasets were extended for 2015–2017 (DeLang et al., 2021). While observations are dense in North America, Europe, Japan, and South Korea, they are sparse to nonexistent elsewhere (Figure S1). CNEMC provides surface ozone observations in China beginning in 2013, and data through 2017 are used in this study (Lu et al., 2020). Both datasets were quality controlled with the same algorithm developed for the TOAR database. The number of observation locations in the combined dataset is least in 1990 (with 1,190) while 2015 has the most (4,999).

We used surface concentration output from 9 atmospheric chemistry models to create our M³Fusion multimodel composite (DeLang et al., 2021). Models include 4

from the chemistry-climate model initiative (CCMI; Morgenstern et al., 2017) that simulate 1990–2010, 2 additional CCMI models that extend the simulation beyond 2010, 2 CMIP6 models that cover years after 2010 (Collins et al., 2017), and MERRA2-GMI that covers 1990–2017 (Strode et al., 2019; Anderson et al., 2021). The compilations of observations and models used here are the same as used by DeLang et al. (2021).

2.2. Data fusion methods

M³Fusion was used to evaluate model performance and create a bias-corrected multimodel composite for each year 1990–2017 (Chang et al., 2019). This method finds a linear combination of the 9 models in each year and continental domain that minimizes the mean square error (MSE) compared to interpolated observations, creating a single bias-corrected global composite for each year. This is the same composite used by DeLang et al. (2021), and it reflects model expectations about variations in ozone due to emissions, geography, meteorology, and chemistry. However, M³Fusion does not capture the nonlinearity of model performance with respect to model value, nor how model performance varies within a continental domain (heterogeneity), both of which we address using RAMP.

The CAMP method (de Nazelle et al., 2010) is a precursor to RAMP that bias corrects for nonlinear model performance but does not account for nonhomogeneity as a single correction applies globally. It matches each observation point with the model estimate at that location. These matched pairs are then binned by the model estimate, and an average of model estimates and observations is set for each bin. The M³Fusion composite is then corrected by interpolating between these values. Since CAMP is closely related to RAMP, we describe the method in depth in the following paragraphs. While CAMP works well for local applications in a single year, RAMP allows us to account for the heterogeneity in model performance at a global scale by performing the model correction based on the nearest observations only.

RAMP is a method to visualize and evaluate model performance that can be used to bias correct models (Reyes, 2016; Reyes et al., 2017). The correction accounts for nonlinear and nonhomogeneous model performance (de Nazelle et al., 2010), in which the RAMP correction is not limited to a linear function with respect to model value, and it may correct differently in different geographic regions. Here, we apply RAMP to the M³Fusion composites so that we address residual nonlinear and nonhomogeneous biases. While previous studies have used RAMP to bias correct model estimates of air pollutants (de Nazelle et al., 2010; Xu et al., 2016; Reyes et al., 2017), none has done so at a global scale.

We describe RAMP by letting $\check{y}(\mathbf{p})$ be the M³Fusion multimodel composite prediction of ozone at space/time coordinate $\mathbf{p} = (s, t)$, where s is the spatial location in longitude/latitude degrees, and t is the time in years. Let $\hat{y}_i = \hat{y}(\mathbf{p}_i)$ be the ozone observation (TOAR or CNEMC) at space/time monitoring points \mathbf{p}_i . M³Fusion predictions are available throughout our entire global study domain, whereas observations are only available at certain

locations. We match each observation \hat{y}_i with the underlying model prediction $\check{y}_i = \check{y}(\mathbf{p}_i)$, so that (\hat{y}_i, \check{y}_i) are the paired observation-model values. We let $\mathcal{R}(\mathbf{p})$ be the space/time region around \mathbf{p} containing the $N = 250$ spatially closest stations in years $t, t - 1$, and $t + 1$ (1990 does not use $t - 1$ and 2017 does not use $t + 1$). After testing other numbers of stations, we found that $N = 250$ is ideal, with enough stations to maintain consistent patterns and prevent outliers from having significant effects, while giving a narrow enough spatial range to correlate with local trends. As we use 3 years, $\mathcal{R}(\mathbf{p})$ contains up to 750 collocated (\hat{y}_i, \check{y}_i) pairs. We sort these pairs by increasing model value and stratify them in 10 bins corresponding to increasing model decile values $\check{y}_k, k = 1, \dots, 10$. Then, we calculate the average observed value λ_1 for model decile value \check{y}_k in region $\mathcal{R}(\mathbf{p})$ as

$$\lambda_1(\check{y}_k, \mathcal{R}(\mathbf{p})) = \frac{1}{n(\check{y}_k, \mathcal{R}(\mathbf{p}))} \sum_{j=1}^{n(\check{y}_k, \mathcal{R}(\mathbf{p}))} \hat{y}_j,$$

where $n(\check{y}_k, \mathcal{R}(\mathbf{p}))$ is the number of paired observed/model values (\hat{y}_i, \check{y}_i) for which \check{y}_i is in the k th decile of modeled values, and \hat{y}_j is the j th observation in these pairs.

The abovementioned steps follow those outlined by Reyes et al. (2017); in this article, we further improve RAMP by ensuring that the slope between λ_1 s does not become negative, or in other words, that the λ_1 RAMP curve for any $\mathcal{R}(\mathbf{p})$ is monotonically increasing. To do this, we define the mean value of all observed values \hat{y}_i in $\mathcal{R}(\mathbf{p})$ as λ_{mean} . We compare λ_{mean} with $\lambda_1(\check{y}_5, \mathcal{R}(\mathbf{p}))$, the λ_1 in the fifth decile bin. If $\lambda_{\text{mean}} < \lambda_1(\check{y}_5, \mathcal{R}(\mathbf{p}))$, we set $\lambda_1(\check{y}_5, \mathcal{R}(\mathbf{p})) = \lambda_{\text{mean}}$. We then compare the fifth and fourth bin in the same way, and so on, ensuring that $\lambda_1(\check{y}_k, \mathcal{R}(\mathbf{p})) \geq \lambda_1(\check{y}_{k-1}, \mathcal{R}(\mathbf{p}))$, by setting them as equal when necessary. We do the same for bins $k = 6$ through 10, first comparing bin 6 to λ_{mean} and setting the value of $\lambda_1(\check{y}_6, \mathcal{R}(\mathbf{p}))$ equal to λ_{mean} if $\lambda_{\text{mean}} > \lambda_1(\check{y}_6, \mathcal{R}(\mathbf{p}))$. This is a novel improvement to Reyes et al. (2017) as it maintains the ordinality of estimates from the original model with the same $\mathcal{R}(\mathbf{p})$.

By plotting $\lambda_1(\check{y}_k, \mathcal{R}(\mathbf{p}))$ with respect to \check{y}_k , we obtain the RAMP curve at location \mathbf{p} showing how the average observation changes with respect to model values. **Figure 1** visualizes the nonlinear performance of the M³Fusion composite, and by changing location \mathbf{p} , we can see how that performance varies spatially and where it is nonlinear. This visualization can, for example, be used to detect regions where the M³Fusion composite overpredicts high ozone values and underpredicts low ozone values. These plots also allow us to correct the model value by interpolating along $\lambda_1(\check{y}_k, \mathcal{R}(\mathbf{p}))$ and selecting new model values based on the value of λ_1 evaluated at the original $\check{y}_k = \text{M}^3\text{Fusion value}$. Therefore, the RAMP-corrected model value is $\lambda_1(\check{y}(\mathbf{p}_i), \mathcal{R}(\mathbf{p}_i))$.

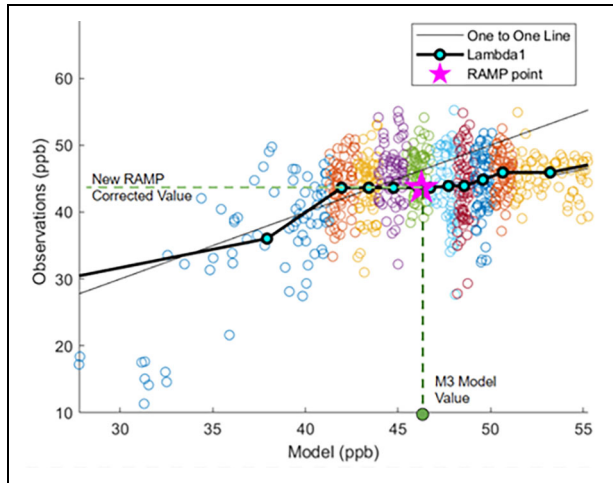


Figure 1. A visualization of the Regionalized Air Quality Model Performance (RAMP) correction at a single point in North America for 2015. Three years of ozone season daily maximum 8-h mixing ratio (OSDMA8) data (2014, 2015, and 2016) from the 250 nearest observation locations are used. Paired M³Fusion/observation points are divided into deciles by the model value, indicated by color, and the M³Fusion estimate at this gridpoint (x -axis) is corrected with RAMP to a new value (y -axis) using the λ_1 line.

A novel challenge posed by the implementation of the RAMP method at the global scale is that station locations are clustered in some countries or continents (e.g., the United States, China, Japan, and Europe) and are sparse in large areas in between. Previous applications of RAMP had more uniform distributions of observations over a smaller scale (Reyes et al., 2017). As a result, globally the region $\mathcal{R}(\mathbf{p})$ containing the $N = 250$ stations closest to \mathbf{p} can change dramatically over a short distance, for example, when shifting from a domain dominated by European observations to one dominated by China. This abrupt change in $\mathcal{R}(\mathbf{p})$ can result in a discontinuity in the RAMP-corrected value $\lambda_1(\mathbf{p})$. To reduce this discontinuity, we introduce the RAMP–M³Fusion weighted-average $\lambda_1^*(\mathbf{p})$ calculated as

$$\lambda_1^*(\mathbf{p}) = w(\mathbf{p}) * \lambda_1(\mathbf{p}) + (1 - w(\mathbf{p})) * \check{y}(\mathbf{p}),$$

where $\lambda_1(\mathbf{p})$ and $\check{y}(\mathbf{p})$ are the RAMP and M³Fusion values, respectively, and $w(\mathbf{p})$ is the weight for RAMP at location \mathbf{p} . We want a weight that is high when a large fraction of the N stations used to construct the RAMP curve are close to \mathbf{p} and low when this fraction is low. We therefore set the weight using the following equation:

$$w(\mathbf{p}) = \frac{N_q(\mathbf{p})}{N(\mathbf{p})},$$

where $N(\mathbf{p})$ is the number of stations used to calculate the RAMP curve at location \mathbf{p} (i.e., 250), $N_q(\mathbf{p})$ is the number of these stations that are within a radius q of \mathbf{p} , and $w(\mathbf{p})$ is the fraction of RAMP stations (between 0 and 1)

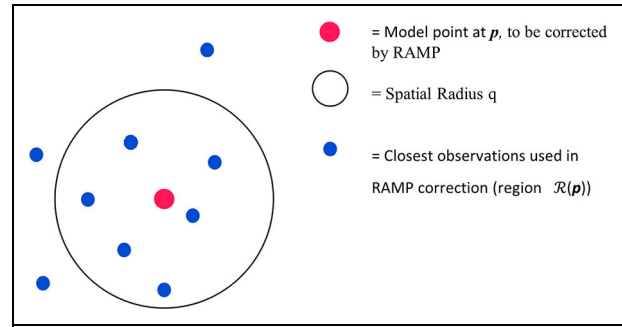


Figure 2. An example of Regionalized Air Quality Model Performance (RAMP) weight at a model point. In this case, $N_q(\mathbf{p}) = 6$, as 6 observations are within radius q . The weighted RAMP (wRAMP) estimate at this location would be 6/250 times the RAMP-corrected composite value plus 244/250 times the M³Fusion composite value without RAMP correction.

that are within q degrees of \mathbf{p} (Figure 2). We chose a radius $q = 25^\circ$, so that the RAMP weight $w(\mathbf{p})$ allows RAMP to exert influence beyond the range of BME without extending into areas lacking representative observations. Areas that are more than 25° away from these station clusters, like the area at the midpoint between China and Europe, will have a RAMP weight close to 0 and $\lambda_1^*(\mathbf{p}) \approx \check{y}(\mathbf{p})$, thereby mitigating any RAMP discontinuity. We call the global output of λ_1^* values as weighted RAMP (wRAMP).

BME data fusion is then applied after RAMP correction to fuse model prediction and observations, using the approach described by DeLang et al. (2021). Each BME estimate uses a different background assumption for global ozone levels at every grid cell, which we call the global offset, based on either the M³Fusion composite, CAMP-corrected M³Fusion, or wRAMP-corrected M³Fusion. This global offset is corrected using BME so that the final BME estimate matches observed values at each station location. Observations are treated as “hard data,” meaning they are perfectly accurate, neglecting measurement uncertainty. Each station exerts an influence based on the difference between the station estimate and the global offset, which decreases as the space/time distance from observations increases, eventually matching the offset prediction away from observations. The rate at which this influence falls is based on a derived covariance function. BME has been used previously for the fusion of ozone observations and models (Christakos, 2000; Christakos et al., 2004; de Nazelle et al., 2010), though only once before at the global scale (DeLang et al., 2021). While these papers provide the details of BME, we give here the main BME steps.

The fundamental step in BME data fusion is the definition of an offset function $o(\mathbf{p})$ at all points \mathbf{p} across the study space/time domain. Here, we set $o(\mathbf{p})$ equal to either $\check{y}(\mathbf{p})$ (M³Fusion), $\lambda_1(\mathbf{p})$ (RAMP), or $\lambda_1^*(\mathbf{p})$ (wRAMP). We calculate the offset-removed observations \hat{x}_i as

$$\hat{x}_i = \hat{y}_i - o(\mathbf{p}_i), i = 1, \dots, n,$$

where $\hat{y}_i = \hat{y}(\mathbf{p}_i)$ are the observations at point $\mathbf{p}_i, i = 1, \dots, n$. We define $X(\mathbf{p})$ as a homogeneous/stationary space/time random field (STRF) with realizations $\hat{x}_i, i = 1, \dots, n$. $X(\mathbf{p})$ is a STRF representing the residual uncertainty and variability that is left in the offset-removed observations, and therefore its covariance function changes with the offset considered (either M³Fusion, RAMP or wRAMP). Finally, we define the STRF $Y(\mathbf{p})$ representing the ozone concentration as the sum of the residual field and the offset, that is,

$$Y(\mathbf{p}) = X(\mathbf{p}) + o(\mathbf{p}).$$

We implement BME on the residual STRF $X(\mathbf{p})$ to obtain the BME estimate and associated estimation error variance of $X(\mathbf{p}_k)$ at estimation points \mathbf{p}_k across a global estimation grid. The general knowledge base characterizing $X(\mathbf{p})$ consists of a mean assumed to equal 0 within the estimation neighborhood, and a covariance function obtained from a variogram analysis (see Supplementary information for details on the covariance model and its parameters). The site-specific knowledge consists of the offset-removed observations treated as hard data (data with no assumed uncertainty). We numerically implement BME using the BMelib library written in the MATLAB programming language (Serre and Christakos, 1999; Christakos et al., 2001), and as shown by DeLang et al. (2021), in this case, the BME posterior probability density function of $X(\mathbf{p}_k)$ is Gaussian with a mean \tilde{x}_k and associated error variance $\text{var}(\tilde{x}_k)$ equal to that of simple kriging. Finally, the BME estimate \tilde{y}_k of $Y(\mathbf{p}_k)$, representing ozone at the estimation point, and associated variance $\text{var}(\tilde{x}_k)$ are obtained as

$$\tilde{y}_k = \tilde{x}_k + o(\mathbf{p}_k),$$

and $\text{var}(\tilde{y}_k) = \text{var}(\tilde{x}_k)$, where $o(\mathbf{p}_k)$ is the (M³Fusion, RAMP or wRAMP) offset at the estimation point. Estimation points are set on a 0.5° grid, giving a final BME estimation at 0.5° resolution. All data fusion methods are the same globally, as we do not attempt to alter the methods for particular regions, although the effects of the methods differ in space and time reflecting observations and model output.

The variance estimated here is a statistical measure of uncertainty in the BME framework, but the variance does not quantify the uncertainty fully as it omits other sources of uncertainty. In this study, measurements are assumed to be accurate, ignoring measurement uncertainty, and the lack of measurements over large world regions hinders our understanding of ozone in those regions. Models involve additional uncertainties in model physical and chemical processes and in emission inventories and other model inputs. Finally, there is uncertainty in the BME framework as other data fusion methods may also be used.

2.3. Cross validation

Leave-one-out cross-validation (LOOCV) was done by removing each observation one at a time and using various estimation methods to evaluate our ability to predict this observation based on the remaining data. LOOCV was performed by predicting ozone at each 0.5° grid cell

containing an observation point and comparing it with the observations ($\hat{y}(\mathbf{p}_i)$) in the grid cell, following the same protocols as DeLang et al. (2021). This was done for M³Fusion, CAMP, and wRAMP both before and after data fusion with BME. For LOOCV of BME, BME was used to estimate each removed point in turn, and the aggregated errors were used to calculate MSE and the Pearson correlation coefficient squared (R^2) that is bounded between 0 and 1. For LOOCV on the offsets, the difference between the offset and observation point at each station location was used.

Whereas LOOCV tests the ability to predict based on nearby clustered observations, we use checkerboard cross-validation (CBCV) to better test each estimation method, especially farther from nearby observations. This method derives from the radius-based validation methods of Xu et al. (2016) and Cleland et al. (2020). In CBCV, we create a “checkerboard” of boxes over the world with each box having a side length s latitude and $2*s$ longitude. For each box, we remove all observed values ($\hat{y}(\mathbf{p}_i)$) within the box and use BME to reestimate the ozone values at the location $\mathbf{y}^*(\mathbf{p}_i)$ of the removed observations within the box, using only observations outside of it. The validation error is defined as $e_b = \hat{y}(\mathbf{p}_i) - \mathbf{y}^*(\mathbf{p}_i)$, which is then used to calculate R^2 values to quantify error for each observation in every box. We test CBCV with s ranging from 0.5° to 50°. BME relies on observations to make corrections within the covariance range (less than about 1° for ozone), so as s increases, observed values will have a smaller influence on correction. CBCV simulates the effect of sparse observations, while still having observations to validate the estimate. As most of the world lacks dense observation networks, the ability to correct away from observations is valuable to global estimations of air pollution.

3. Results

The M³Fusion multimodel composite (Chang et al., 2019; DeLang et al., 2021) (Figure S1) is used here as the basis for CAMP and RAMP bias corrections based on the TOAR and CNEMC observations (Figure 1). Using CAMP, we see that M³Fusion performance is nonlinear, tending to overpredict ozone where it estimates high values and slightly underpredict low values (Figure S2). This has been demonstrated for individual models in previous studies of surface ozone (de Nazelle et al., 2010), but we are not aware that it has been demonstrated previously at the global scale. Using RAMP, we confirm that M³Fusion bias varies at a finer scale than the continental regions used in M³Fusion, supporting the value of RAMP’s localized (non-homogenous) bias correction.

When applying RAMP bias correction to each model grid point, corrections vary each year and at each location, but the largest changes generally occur where the M³Fusion estimate is above about 55 ppb or below 35 ppb. While the M³Fusion composite generally overestimates when it predicts high ozone and underestimates where it predicts low, this is not true for all regions. Figure 3a shows an area where the model consistently underpredicts ozone, and the RAMP correction has a steeper slope at high values. Figure 3b shows a nearby region in the same

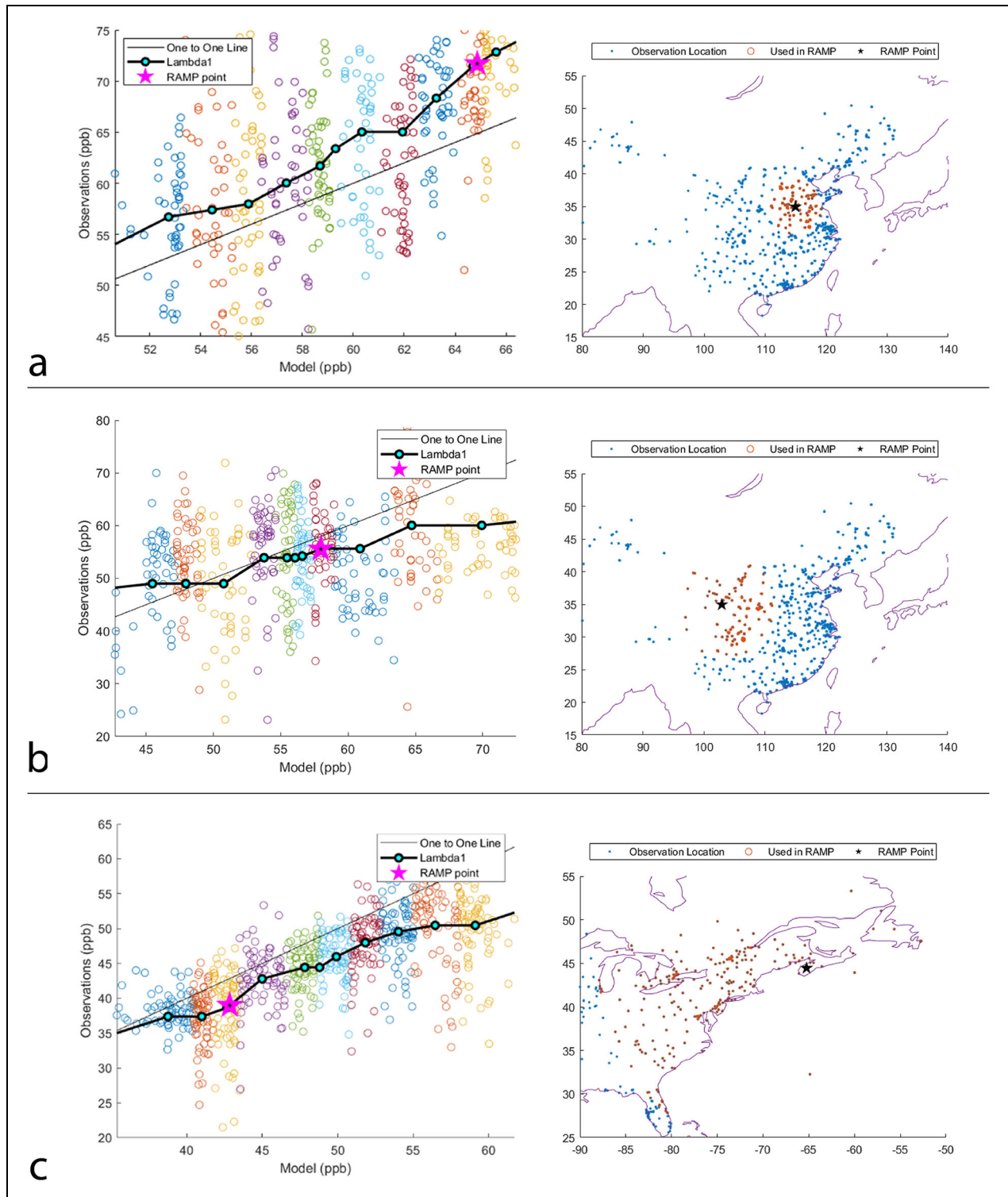


Figure 3. Examples of Regionalized Air Quality Model Performance (RAMP) correction at 3 specific locations and years. The RAMP curve shows paired M³Fusion composite ozone season daily maximum 8-h mixing ratio (OSDMA8) values $\check{y}(p)$ and observations \check{y}_i , and the locations of the nearest 250 observation locations are also shown. The RAMP-corrected value $\lambda_1(p)$, the star, is estimated by replacing the M³Fusion prediction $\check{y}(p)$ with its RAMP-corrected value, that is, $\lambda_1(p) = \lambda_1(\check{y}(p), \mathcal{R}(p))$. Each colored circle is a paired model/observation value (\hat{y}_i), with the colors denoting which bin each falls into. If $\lambda_1(p)$ is below the 1:1 line, it indicates that M³Fusion overpredicts ozone. (a) An increase in estimation due to RAMP. (b) A nonlinear correction. (c) A decrease in estimation.

year where M³Fusion overpredicts ozone at all but the lowest levels, and the ozone estimate at the specific point is lowered by the RAMP correction. The locations of

Figure 3a and **b** are relatively close to one another, showing that the M³Fusion bias varies at finer spatial scales than continental. **Figure 4** shows the heterogeneity in

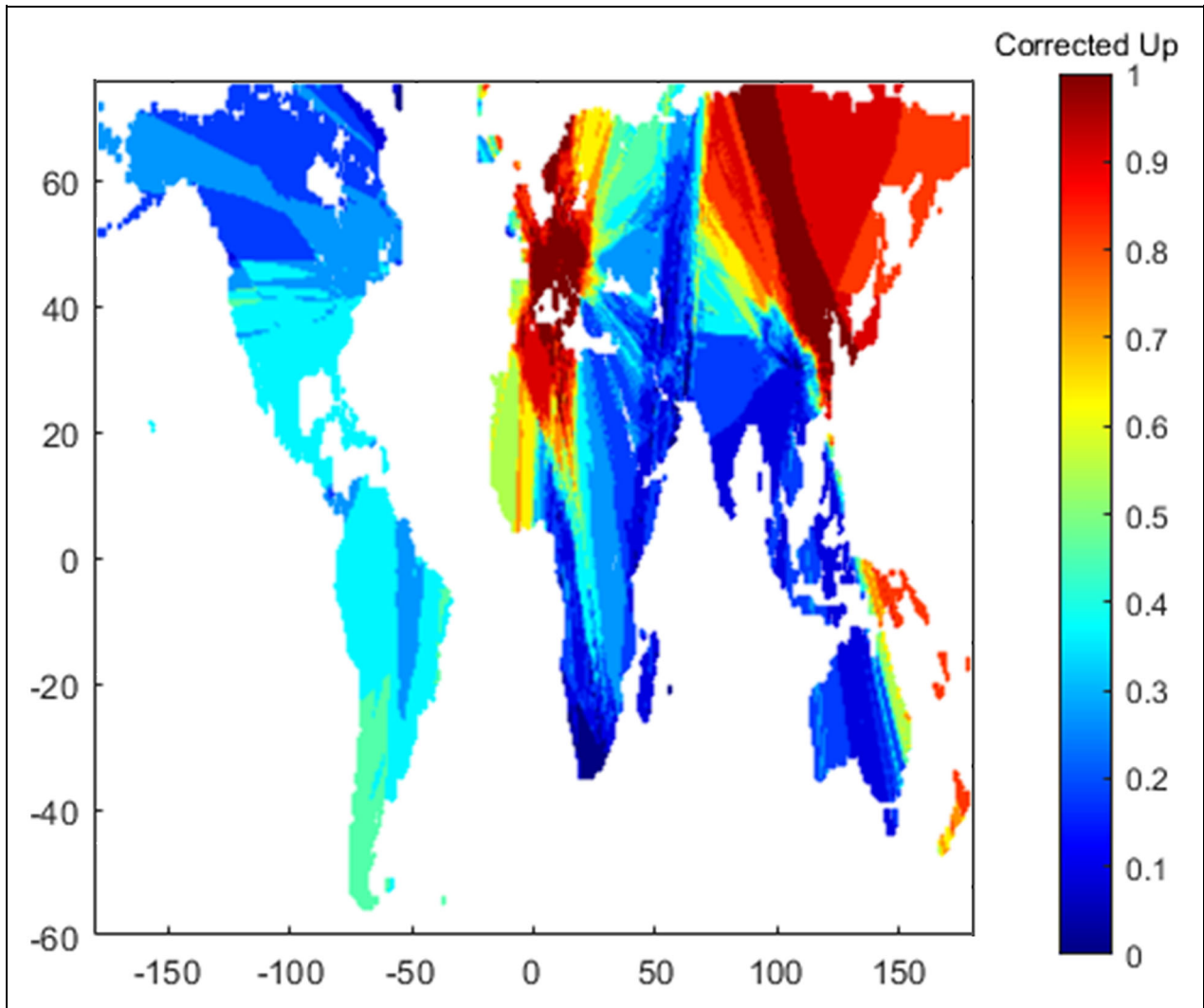


Figure 4. Fraction of λ_1 points above the one-to-one line in 2017. A higher number indicates a higher likelihood that a model point in this location would be increased with a Regionalized Air Quality Model Performance (RAMP) correction, while a number <0.5 indicates a greater chance that RAMP lowers the M³Fusion estimate. Results vary geographically, showing that the performance of the M³Fusion composite is heterogeneous spatially, in some places varying strongly over a short distance.

model performance and bias correction globally. While some areas, like the Americas, primarily have a RAMP correction in a single direction, others like northern Africa and eastern Europe have regions which are corrected upward and bordering regions corrected downward.

Figure 3b also shows a region where model performance is nonlinear with respect to estimations, overpredicting high values and underpredicting low values. In these areas, the M³Fusion bias does not vary linearly with respect to the M³Fusion estimate, and therefore our correction is not a linear function. This shows the value of RAMP over a linear bias correction, as a linear correction could not replicate these nonlinear curves. **Figure 3c** shows an example region where M³Fusion consistently overestimates ozone. These RAMP curves show the trends in model performance in the region, as a function of modeled concentration, as well as correcting the individual points (the pink star) based on this evaluation.

At a global scale, RAMP creates “streaks” where the observations used to correct the model change from being dominated by one region (eastern Europe) to another far away region (Japan and South Korea) over a short spatial distance (**Figure 5**). This situation arises because there are no/few local observations for the RAMP correction in this area. To avoid these discontinuities and to avoid RAMP bias corrections based only on measurements far away, we weight RAMP (**Figure 2**) to allow a smooth transition between regions, using weights for RAMP and M³Fusion that vary spatially and temporally (**Figure 6**). wRAMP heavily favors RAMP over M³Fusion in areas with high density of observations, and RAMP maintains some influence up to 25° from any station. This distance is long enough to give RAMP an influence in areas not reached by the BME correction, but short enough that it creates smooth transitions between regions and lessens the discontinuities seen in pure RAMP.

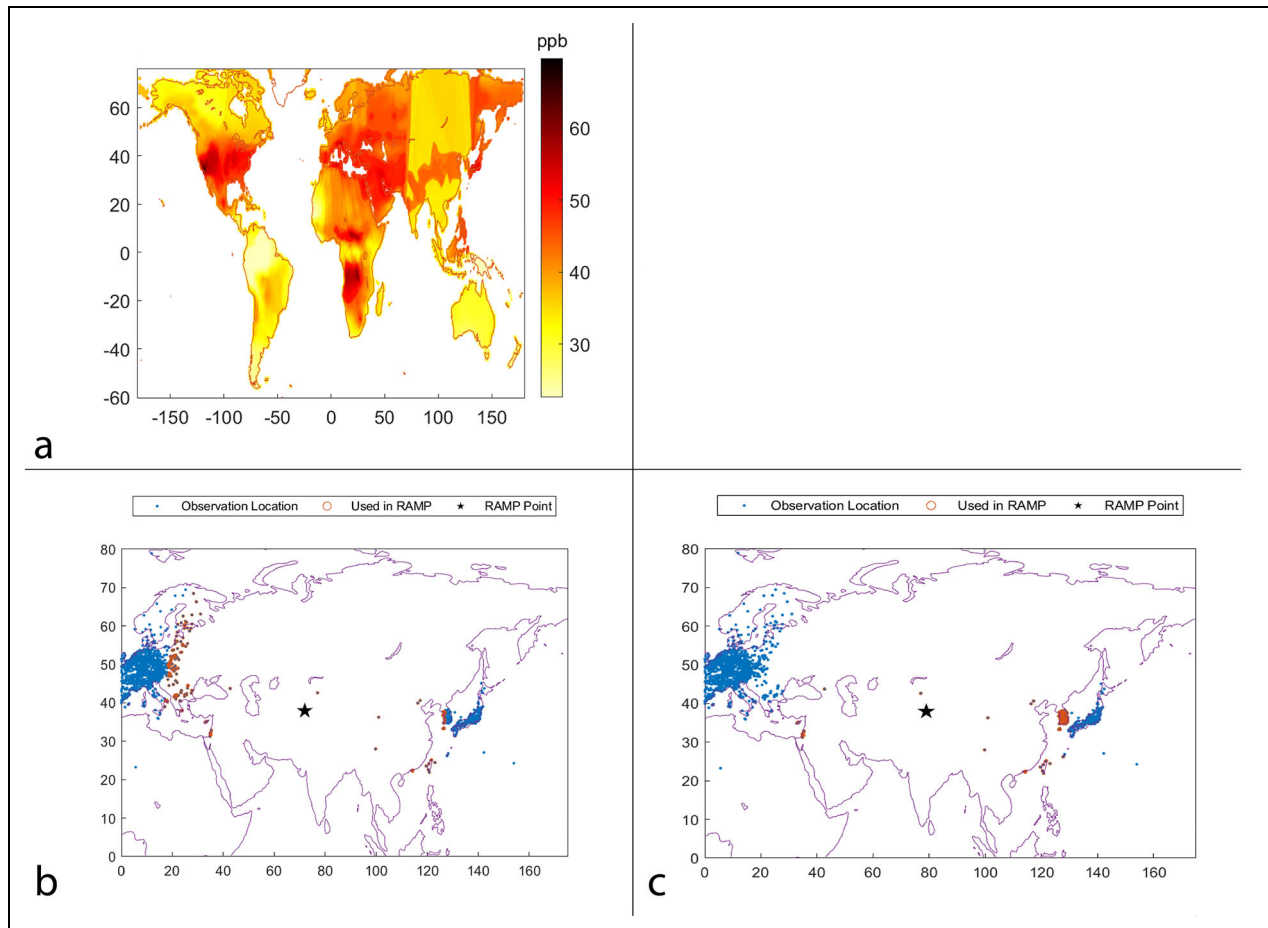


Figure 5. Regionalized Air Quality Model Performance (RAMP)-corrected estimate of ozone season daily maximum 8-h mixing ratio (OSDMA8) ozone for 2005 with a streak in central Asia. Applying a RAMP correction to the M^3 Fusion multimodel composite produces discontinuities that appear as streaks in central Asia (a). Examination of which observations are used for RAMP correction (\hat{y} , in $\mathcal{R}(p)$) at 2 nearby model points ($\check{y}(p)$) (b and c) shows a large shift in points comprising $\mathcal{R}(p)$, which causes these large changes over short spatial ranges. Weighting RAMP prevents this from occurring, as areas far from the stations used for RAMP corrections will default to the M^3 Fusion composite.

Using wRAMP as our global offset and station observations as hard data for BME, we obtain yearly estimates of global ozone and variance at 0.5° resolution (Figure 6). The ozone estimates match observations at any space/time location with an observation, as the observations are assumed to be accurate. The influence of observations decreases as a function of space/time distance as the estimate moves further from an observation, based on the derived covariance (see Supplementary Materials equations and Figure S3). The influence of an observation over multiple years in BME is valuable when correcting areas with inconsistent observations. DeLang et al. (2021) explore the significance of the temporal factor in more detail. Variance is strongly influenced by proximity to observations, which are the only source of hard data in the BME estimate. Variance drops to 0 at measurement locations and quickly rises as distance from stations increases, since measurements are assumed to be accurate. Therefore, variance is low in Europe, North America, Japan, South Korea, and in some years parts of China, and high elsewhere. As variance approaches 60 ppb^2 , the BME estimation approaches wRAMP.

While in our previous work, the BME data fusion of DeLang et al. (2021) improved markedly over the M^3 Fusion composite, Figure 7 shows the difference between our BME estimate using the wRAMP-corrected model as our global offset and the BME estimate of DeLang et al. (2021) which uses the M^3 Fusion composite as the global offset. The 2 methods differ most at distances more than 0.5° from stations, because BME matches observations exactly at measurement locations, with that influence approaching 0 beyond 0.5° . But the influence of wRAMP generally drops off with distance from stations because of the weighting used, and wRAMP has no influence beyond 25° from a station. Thus, wRAMP often has the greatest influence in regions between monitors, even in regions that are densely monitored. Whether RAMP increases or decreases estimates varies in time and space, and even nearby areas can have different signs of the correction. Changes in specific regions also vary year to year. General trends include decreases in the Korean peninsula, large changes in China once local data become available in 2014, overall increases in eastern China prior to 2014, increases in the northeastern United States in

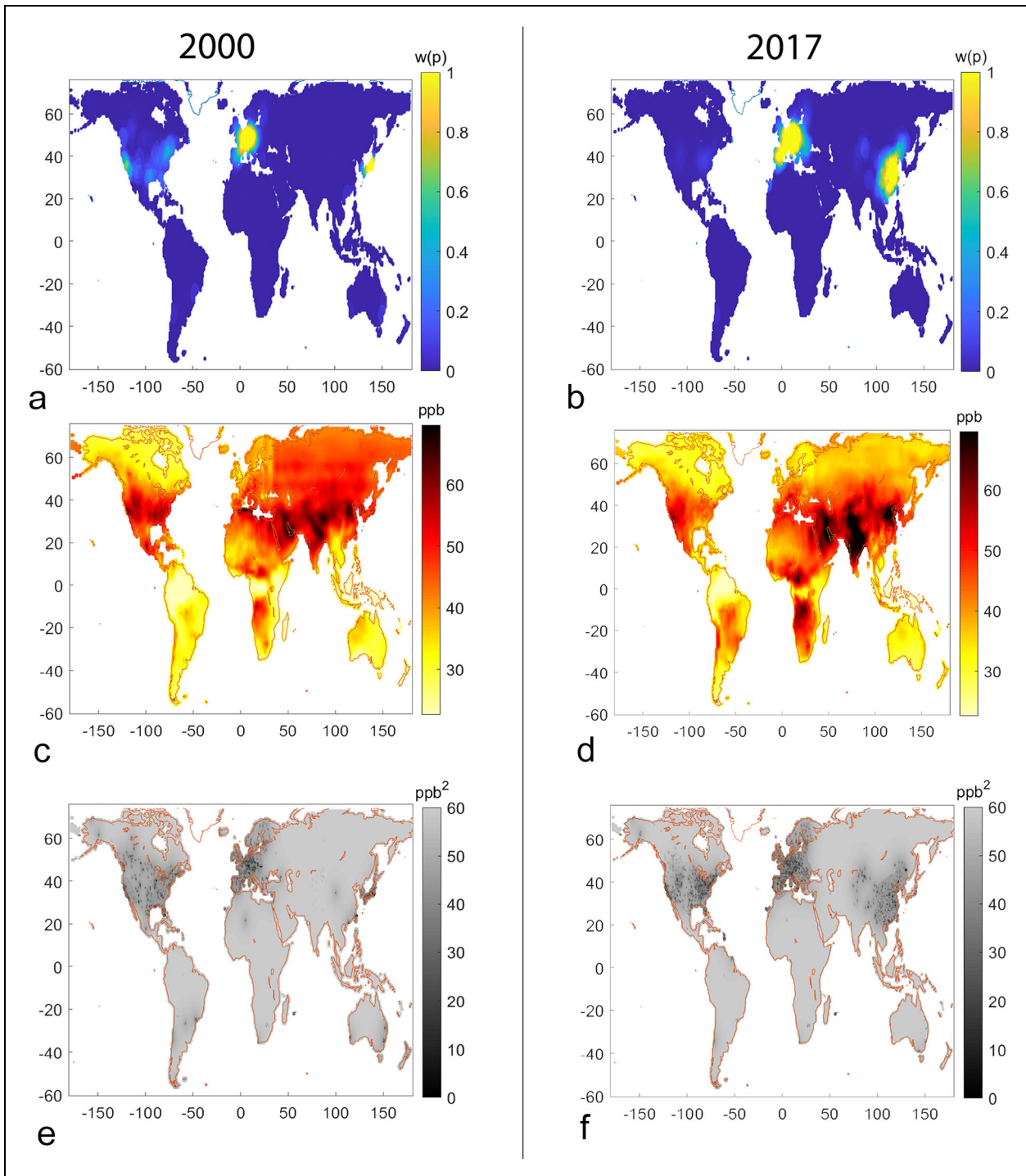


Figure 6. Weights applied in weighted RAMP (wRAMP) and final wRAMP estimates of ozone and variance. The weights applied in wRAMP, $w(p)$, for 2000 (a) and 2017 (b) correspond to the percent of the estimate at that location that is based on RAMP, with $1 - w(p)$ being the weight applied to the M³Fusion composite. wRAMP ozone season daily maximum 8-h mixing ratio (OSDMA8) estimates, for 2000 (c) and 2017 (d), are obtained using the multimodel composite bias corrected with wRAMP as the global offset in Bayesian Maximum Entropy (BME) data fusion. Ozone values match observations (Figure S1) at each station location, with an observational influence that decreases as space/temporal distance from the observations increases. BME with wRAMP eliminates the streaks in **Figure 5**, as those areas have little or no RAMP weight. The variance of BME estimates for 2000 (e) and 2017 (f) is obtained as a function of spatial/temporal distance from observation locations. Variance is 0 at any observation location. Ozone estimates and variance are shown for other years in Figures S1 and S4.

most years, slight increases in southeastern Europe, and overall better model performance in inland United States and European Union (EU) than on the coasts indicated by

smaller corrections in those regions. In regions with few observation stations, especially those further than 25° from the nearest observation, changes caused by including

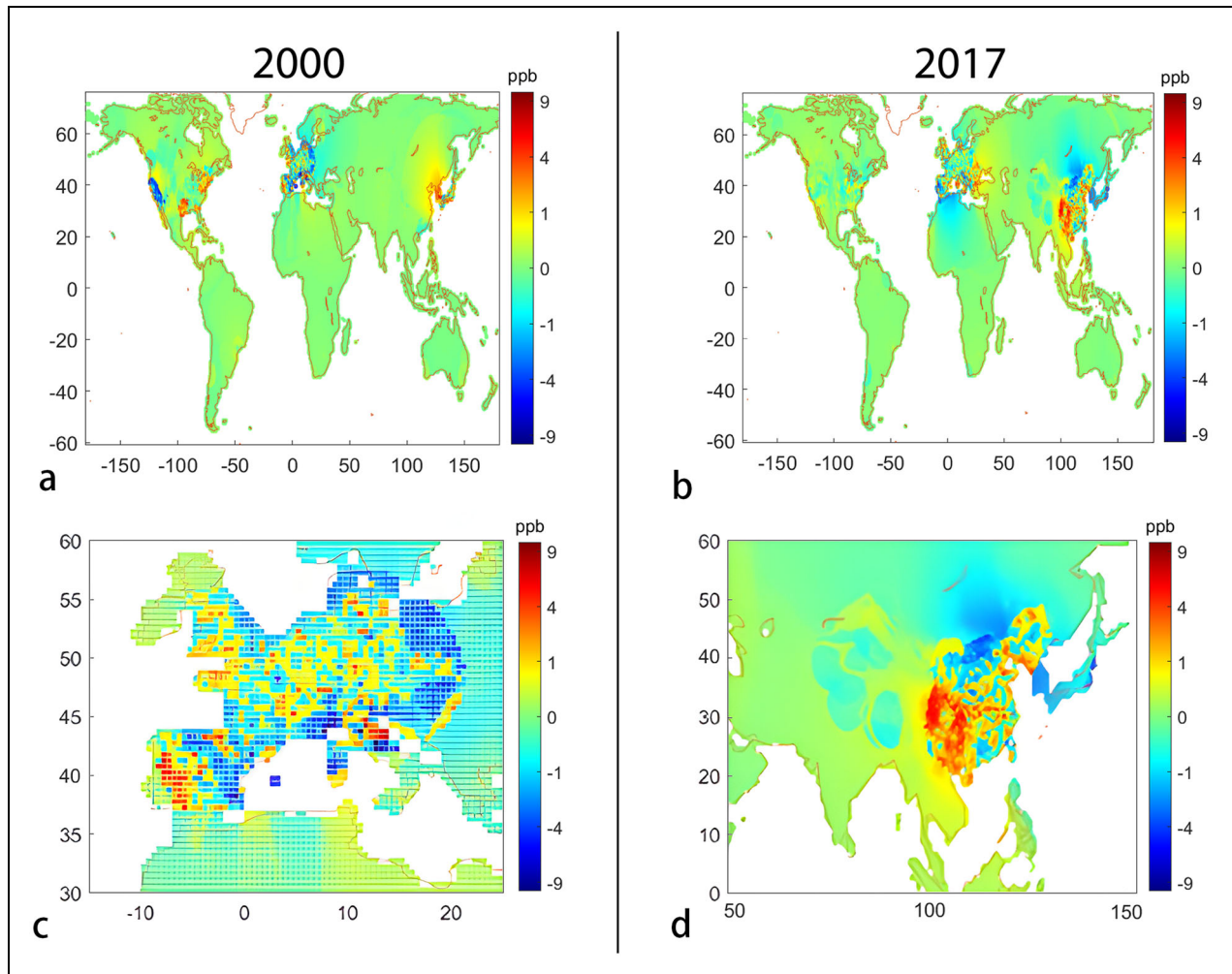


Figure 7. Differences in final Bayesian Maximum Entropy (BME) estimates caused by applying weighted RAMP (wRAMP). Results show the difference in ozone season daily maximum 8-h mixing ratio (OSDMA8) when using BME with wRAMP as the global offset minus BME with the M³Fusion composite as the offset (the results of DeLang et al., 2021), for 2000 (a) and 2017 (b). Also shown are differences zoomed in on regions of interest for 2000 (c) and 2017 (d). Red indicates that RAMP corrected the M³Fusion composite value up, while blue indicates that RAMP lowered the M³Fusion value.

wRAMP are often less than 0.2 ppb in individual grid cells (Table S1). Larger changes are observed in regions with more monitors. In Europe in 2000, for example, the population-weighted average ozone concentration decreases by 0.205 ppb (about 0.5% of the regional population-weighted mean ozone) with some grid cells changing by more than 3 ppb (Table S1; **Figure 7**). Since there are compensating positive and negative changes, we also report a population-weighted mean of the absolute value of changes, which is 0.863 ppb (about 2% of the mean). Similarly, in East Asia in 2017, the population-weighted mean ozone increases by 0.217 ppb (about 0.4%), some grid cells change by more than 3 ppb, and the average absolute change is 0.897 ppb (about 1.7%).

We evaluate our results relative to DeLang et al. (2021) and other methods by using LOOCV and CBCV, for 7 cases:

- Simple Model Mean: An average of all models used in M³Fusion where each is given equal weight.

- M³Fusion: Multimodel composite of 9 models using the M³Fusion method.
- CAMP: CAMP-corrected M³Fusion composite.
- wRAMP: RAMP-corrected M³Fusion composite, weighted based on proximity to observations.
- BME using M³Fusion as Offset: BME data fusion using the M³Fusion multimodel composite as the global offset, matching the results of DeLang et al. (2021).
- BME using CAMP as Offset: BME data fusion using CAMP as the global offset.
- BME using wRAMP as Offset: BME data fusion using wRAMP as the global offset.

In LOOCV, M³Fusion, CAMP, and wRAMP all improve performance over the simple model mean (Table S2), and CAMP and wRAMP provide clear improvements with reduced MSE and improved R^2 over M³Fusion. Using BME, LOOCV performance does not differ whether M³Fusion, CAMP, and wRAMP are used as the global offset. This

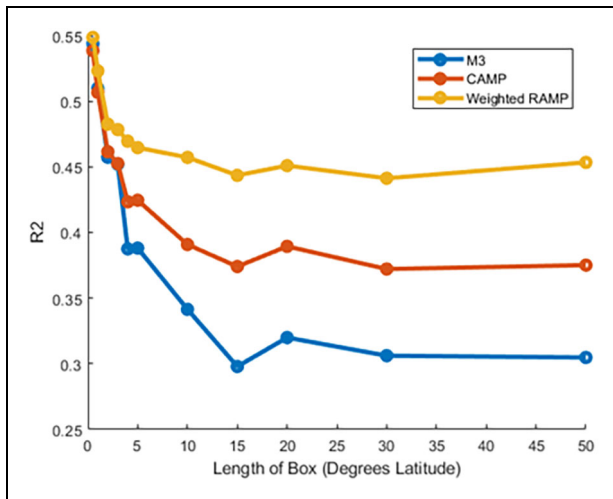


Figure 8. Checkerboard cross-validation (CBCV) results. Correlation coefficient (R^2) values are shown for the CBCV, in which stations are removed from individual boxes of size s in latitude on the x -axis and $2*s$ in longitude, while performing Bayesian Maximum Entropy (BME). Results are shown for BME data fusion using the M³Fusion composite, and that composite corrected with Constant Air Quality Model Performance (CAMP) and weighted Regionalized Air Quality Model Performance (wRAMP) as the global offsets, pooling results over all years. While the differences between M³Fusion and wRAMP are small at small box sizes (similar to Leave-one-out cross-validation [LOOCV]), RAMP greatly outperforms M³Fusion and CAMP at large box sizes, where BME has less influence as there are fewer nearby observations to aid estimation.

result is expected because most observations are clustered, and BME predicts accurately when observations are close together, similar to kriging on the observations. In an LOOCV, removing a single monitor in many cases leaves a sufficient number of nearby stations to accurately predict ozone.

Methods other than LOOCV are therefore important to test the improvement in ozone estimates when wRAMP bias correction is used before BME data fusion, and, for this, we use CBCV. In particular, CBCV more meaningfully tests each method's ability to estimate ozone where there is not a dense network of observations (Figure 8). At small box sizes, CBCV approximates LOOCV and all BME methods have similar R^2 (though a smaller R^2 than in LOOCV since CBCV removes observations in all years). As the box size increases, R^2 for wRAMP decreases less than the other cases, maintaining a minimum of 0.45 at large box sizes. It also does not experience the dramatic performance drop-off that CAMP and M³Fusion have at about 4°. CAMP also has consistently higher R^2 than M³Fusion at box sizes greater than 4°. BME with wRAMP shows great improvement in estimating where observations have been removed compared to the base M³Fusion BME estimates. That is, even when all nearby measurements are removed (at large box sizes) wRAMP is beneficial, by using

information from monitors farther away to inform the bias correction.

Using BME estimates and variance, we evaluate the likelihood that ozone values exceed specific thresholds. Specifically, using our best estimates using BME with wRAMP as the global offset, we analyze the likelihood that a given location surpasses selected ozone levels (Figure 9), where 30 ppb corresponds to the WHO guideline (WHO, 2021). Note that we do not estimate the likelihood of exceeding daily 8 h standards, such as the U.S. National Ambient Air Quality Standards for ozone, which cannot be estimated directly from OSDMA8. Areas with ozone estimates near the threshold and areas with high variance (few observations) are most likely to fall in the uncertain range. Certainty in exceeding or not exceeding a given value comes from extreme estimates and/or dense observations. For example, very few parts of the world are definitively below 45 ppb in 2017, but only areas with high estimates (central Africa, India, the Middle East, and parts of China) and areas with dense sensor networks (EU and western United States) are clearly above it. Similarly, comparing the likelihood of exceedance with our ozone estimates, we see some areas which have the same level of estimated ozone but have different likelihood of exceeding thresholds due to the difference in nearby observations (and therefore variance). For example, the hotspots in southern Africa are estimated to be above 65 ppb, but we are less than 90% certain that this area exceeds 55 ppb. Meanwhile the hotspot centered around Beijing, which has nearby observations, is above 55 ppb with near certainty, and even above 65 ppb with 90% certainty in some areas.

We estimate that 96% of the world's population was exposed to ambient air that exceeds the new WHO guideline for ozone in 2017 (Figure 10), and that this percentage has increased from about 91% in 1990. Here, we use 2019 population data from GBD 2019 for all years, so all changes are due purely to ozone changes not population changes. Among individual world regions, North America, Europe, Russia, and South Central Asia have very nearly 100% of their population exceeding the guideline. Regions with relatively less ozone pollution, Oceania and South America, have over 50% of the population exceeding the guideline in 2017.

Finally, following DeLang et al. (2021), we use global population data from GBD 2019 to analyze annual population-weighted ozone in different regions (Figure S6). Trends in regions and years with sparse observations are less certain. Although there are small differences in individual years and regions, trends overall follow the same pattern as for DeLang et al. (2021). This reflects both compensating positive and negative bias corrections within regions by wRAMP (Figure 7; Table S1), and the fact that the population-weighted averages plotted here are weighted toward large urban areas, some of which have dense monitoring networks in which BME estimates ozone accurately whether or not wRAMP is used. Asia has a large upward trend, which along with a large increase in Africa drives an overall upward global trend in population weighted ozone. North America and Europe trend

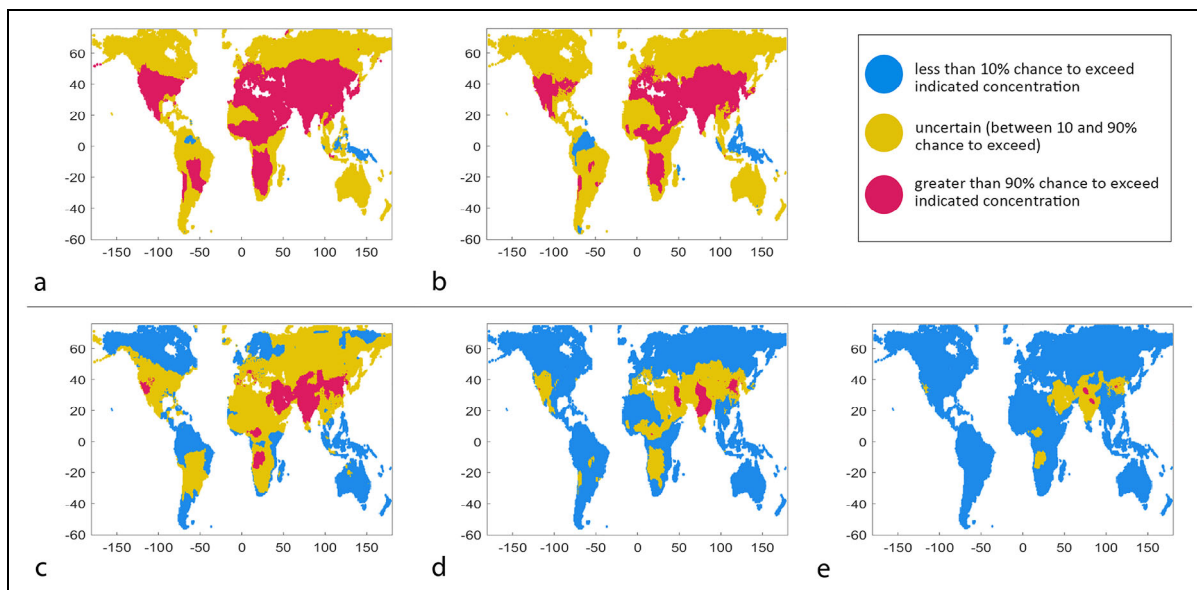


Figure 9. Likelihood of exceedance of 5 ozone levels in 2017. Results are shown for Bayesian Maximum Entropy (BME) with weighted RAMP (wRAMP) as the global offset, relative to 5 ozone season daily maximum 8-h mixing ratio (OSDMA8) ozone levels: the World Health Organization (WHO) guideline of 30 ppb (a), 35 (b), 45 (c), 55 (d), and 65 (e). Corresponding ozone estimates and uncertainty can be seen in **Figure 6**. Areas with low variance (near station observations) have more certainty, as do areas where BME estimates are very high or low compared to the levels selected. Likelihood of exceedance is shown for other years in Figure S5.

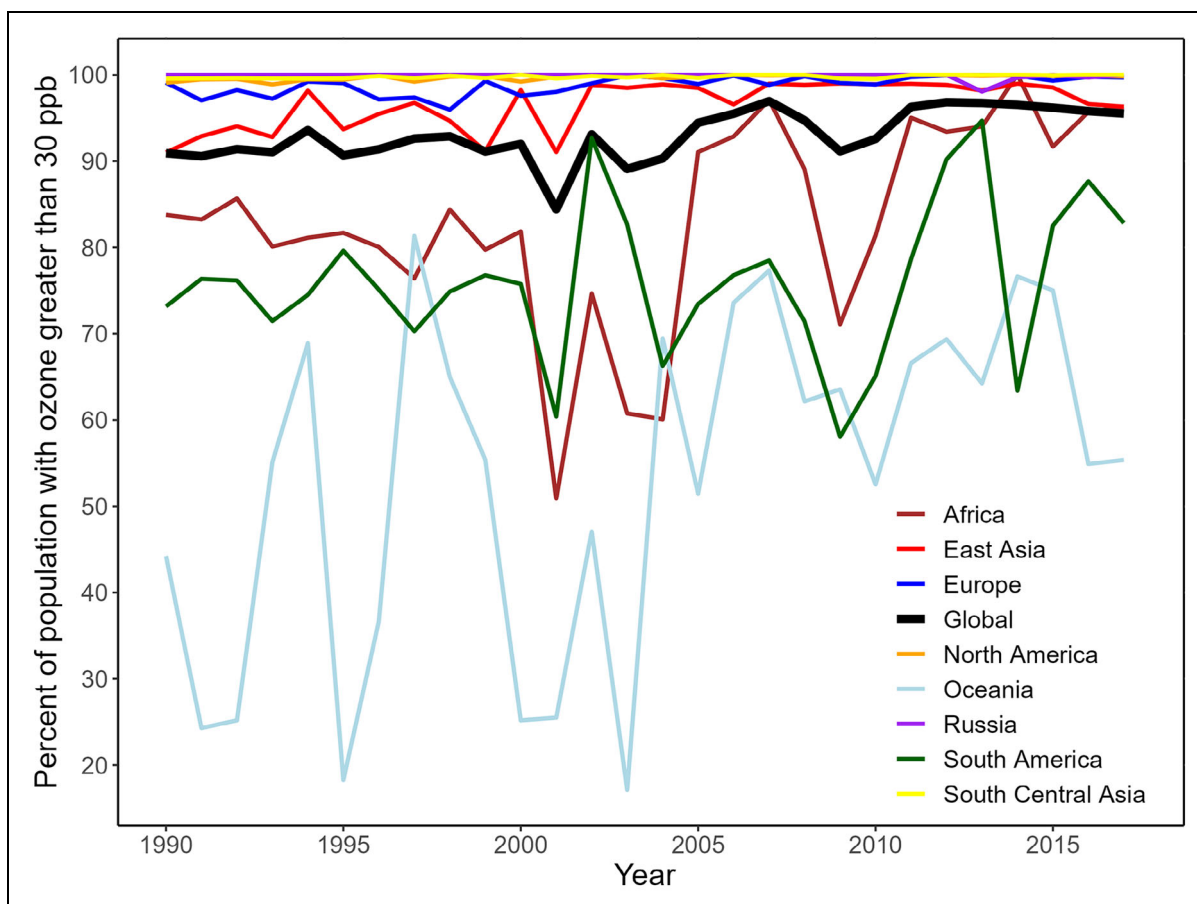


Figure 10. The percentage of population exposed to ozone greater than the World Health Organization (WHO) guideline (30 ppb). Results are shown for ozone season daily maximum 8-h mixing ratio (OSDMA8) estimated with Bayesian Maximum Entropy (BME) with weighted RAMP (wRAMP) as the global offset, for each year 1990 to 2017, for the global population, and for 8 world regions defined by Chang et al. (2019).

downward, though the European trend is much weaker. Russia begins to trend down in 2010, while South America and Oceania fluctuate but have no clear trend. TOAR observational studies support the downward trend in North America from 2000 to 2014 (Chang et al., 2017), and a study of CNEMC observations supports the increase in East Asia based on observational trends in China (Lu et al., 2020).

4. Conclusions

Here, we improve upon the global ozone estimates of Chang et al. (2019) and DeLang et al. (2021) by providing an additional bias correction step to the M³Fusion model composite before BME data fusion. This RAMP correction provides a local, nonlinear, nonhomogenous model bias correction, in which each point receives a different bias correction based on the M³Fusion composite performance at the nearest stations. Using this corrected model as the basis for BME data fusion leads to improvements in ozone estimates in regions beyond the distance over which observations have a substantial influence in BME (about 0.5° from monitoring stations), but less than the 25° limit in the weighting scheme used for RAMP. Including RAMP therefore can help at intermediate distances to fill in values between monitors that are not very closely spaced. We found that performance of M³Fusion varies by space/time location and is often nonlinear, making RAMP the ideal tool to further improve this composite. This method also takes full advantage of TOAR and CNEMC observations, as it allows them to both directly correct estimates locally through time with BME data fusion and inform model corrections at a larger regional scale through M³Fusion and RAMP. Our final estimates provide yearly fine resolution global ozone estimates for 1990–2017, involving a data fusion of surface observations from global monitoring stations and 9 chemistry-climate models. We also provide estimates of variance generated by BME, which may be useful for some further applications, but we caution that the variance estimated here does not represent the full uncertainty in ozone concentration.

The RAMP method demonstrates that model performance and biases have local variations, even after a uniform continental bias correction is applied in the M³Fusion multimodel composite. RAMP therefore improves estimates over M³Fusion or the global CAMP in accounting for heterogeneous model performance. RAMP also shows that model performance is nonlinear with respect to observations in many areas, which often manifests as an overprediction at one extreme and an underprediction at the other. Overall, the multimodel composite is better at estimating ozone values near the average (often 40–55 ppb) and poorer at the extremes. RAMP's ability to account for nonlinear model performance allows greater corrections where M³Fusion predictions are worse.

As this is the first application of RAMP at a global scale, we find that RAMP alone creates “streaks” where the observations being used to inform the correction change over a short distance, showing the difficulty of using a single method over areas both rich and sparse in data. RAMP

could potentially encounter this issue for any dataset where there are 2 or more heavily clustered regions of observations separated by areas with sparse observations. Therefore, we choose to weight RAMP to create smooth transitions between regions, giving a much greater weight to the multimodel composite when corrections are far from observations, while very close to observations corrections are incorporated by BME data fusion. In areas with a more uniform distribution of observations, such as those from previous studies using RAMP (Xu et al., 2016; Reyes et al., 2017), weighting RAMP would not be necessary. Weighting RAMP by distance from observations avoids such streaks, and limits RAMP's influence in regions very far from monitors. It also allows a smooth transition between the regional bias corrections of RAMP, and M³Fusion, which bias corrects within each continent. In future work on different scales, different weighting schemes for RAMP may be chosen, including no down-weighting at all in regional applications where there are many monitors.

Overall, RAMP is more accurate than CAMP and M³Fusion at estimating global ozone. When used in conjunction with BME, RAMP does not appreciably improve estimates in LOOCV and within 1° of another station. BME alone can correct the model within the range where observations co-vary with each other, especially if it can draw on observations at the same location in other years. The advantage of RAMP is seen in the CBCV, which is greatest beyond about 0.5° from the nearest monitoring station but less than 25°. The improvements of CAMP over M³Fusion shows the value of a nonlinear model correction alone, while RAMP's improvements over CAMP show the value of accounting for regional heterogeneity in model performance. Changes in ozone estimates due to including RAMP before BME data fusion are as large as 3 ppb in individual grid cells and absolute values of changes are up to 1 ppb when averaged over regions where ozone monitors are dense. These changes are smaller (<0.2 ppb in individual grid cells) in regions far from monitors, where the weighting put on RAMP goes to near 0.

Because the BME method provides both ozone estimates and the associated variance, we can evaluate the confidence that ozone is above or below selected values. We find that most of the world's population lives in areas very likely above 35 ppb in OSDMA8 and even above 45 ppb. Some regions estimated to have the highest ozone, including much of India, are very likely above 55 ppb. In the case of India, model estimates suggest high ozone that may be above 65 ppb, but the lack of ground observations decreases confidence in this region. Regions with high modeled ozone but low confidence in results because of the distance from observations can be among those prioritized for increased monitoring. While RAMP improves estimation far from monitors, additional monitoring capacity in regions currently lacking monitors would be valuable for improving ozone estimates. Currently much of the world's population lives in low- and middle-income nations, which are located far from ozone monitors. In these regions, large populations are likely exposed to high and growing ozone levels. Based on our analysis,

enhanced surface ozone monitoring in these regions would have the greatest impact for improving our understanding of global scale exposure to ozone pollution.

Future work should aim to update these methods using newer atmospheric models, including results from chemical reanalyses, and to the forthcoming TOAR-II measurement database (<https://toar-data.org/surface-data>), which will include new monitoring sites in populated regions such as India. This analysis focuses on a single annual metric relevant for health, but this analysis could be reproduced monthly and could be reproduced for other ozone metrics such as those relevant for plants. These methods are also applicable to other air pollutants. Finally, these methods offer an interesting counterpoint to recent machine learning estimates (Seltzer et al., 2020; Kleinert et al., 2021; Sun et al., 2021; Betancourt et al., 2022; Liu et al., 2022). Machine learning has advantages of incorporating a wide range of datasets as predictors of ozone and in finding nonlinear relationships that can be difficult to capture in the geostatistical methods used here. However, machine learning has possibilities of spurious results—especially where relationships found over densely monitored regions are assumed to hold globally—and machine learning results can be difficult to understand and interpret. The geostatistical methods used here can be adapted to incorporate a wider range of input datasets, including machine learning estimates, and using machine learning and geostatistical methods in combination may further improve global ozone estimates.

Data accessibility statement

Input data from TOAR are publicly available through the TOAR web interface (<https://join.fz-juelich.de/accounts/login/>) and through PANGAEA (<https://doi.pangaea.de/10.1594/PANGAEA.876108>). Modeling output from CCMI used as input to this data fusion project can be obtained from CCMI (<http://blogs.reading.ac.uk/ccmi/>). MERRA2-GMI output is publicly available (<https://acd-ext.gsfc.nasa.gov/Projects/GEOSCCM/MERRA2GMI/>). Codes for the M³Fusion method to create a multi-model composite were made available through the journal publication by Chang et al. (2019) (<https://gmd.copernicus.org/articles/12/955/2019/>). BME codes are publicly available through the BME library (https://mserre.sph.unc.edu/BMELib_web/index.htm). Files containing full gridded results for our annual ozone maps produced by data fusion in this project, and their variance, are available to the public through Zenodo (<https://zenodo.org/record/7573881>).

Supplemental files

The supplemental files for this article can be found as follows:

SuppInfo_Final2.Pdf

Funding

The authors acknowledge funding from the NASA Health and Air Quality Applied Sciences Team (NNX16AQ30G), NASA (80NSSC23K0930), and the National Institute for Occupational Safety and Health (T42-OH008673). ORC and KLC were supported by NOAA cooperative

agreements NA17OAR4320101 and NA22OAR4320151. MD was supported by the Japan Society for the Promotion of Science (JP20K04070). The MERRA-2 GMI Replay simulation was supported by the NASA Modeling, Analysis and Prediction (MAP) program, with computer resources provided by the NASA Center for Climate Information.

Competing interests

The authors have declared that no competing interests exist.

Author contributions

Contributed to conception and design: JSB, MND, KLC, MLS, ORC, MB, JJW.

Contributed to acquisition of data: JSB, MND, KLC, ORC, MGS, SS, XL, LZ, MD, BJ, CAK, JFL, ML, JL, VM, SAS, KS, ST, LZ.

Contributed to analysis and interpretation of data: JSB, MND, KLC, MLS, ORC, HW, JJW.

Drafted and/or revised the article: JSB, MND, KLC, MLS, ORC, HW, JJW.

Approved the submitted version for publication: All authors.

References

- Anderson, DC, Duncan, BN, Fiore, AM, Baublitz, CB, Follette-Cook, MB, Nicely, JM, Wolfe, GM.** 2021. Spatial and temporal variability in the hydroxyl (OH) radical: Understanding the role of large-scale climate features and their influence on OH through its dynamical and photochemical drivers. *Atmospheric Chemistry and Physics* **21**: 6481–6508. DOI: <http://dx.doi.org/10.5194/acp-21-6481-2021>.
- Appel, KW, Gilliland, AB, Sarwar, G, Gilliam, RC.** 2007. Evaluation of the Community Multiscale Air Quality (CMAQ) model version 4.5: Sensitivities impacting model performance: Part I—Ozone. *Atmospheric Environment* **41**: 9603–9615.
- Archibald, AT, Neu, JL, Elshorbany, Y, Cooper, OR, Young, PJ, Akiyoshi, H, Cox, RA, Coyle, M, Derwent, R, Deushi, M, Finco, A, Frost, GJ, Galbally, IE, Gerosa, G, Granier, C, Griffiths, PT, Hossaini, R, Hu, L, Jöckel, P, Josse, B, Lin, MY, Mertens, M, Morgenstern, O, Naja, M, Naik, V, Oltmans, S, Plummer, DA, Revell, LE, Saiz-Lopez, A, Saxena, P, Shin, YM, Shaahid, I, Shallcross, D, Tilmes, S, Trickl, T, Wallington, TJ, Wang, T, Worden, HM, Zeng, G.** 2020. Tropospheric ozone assessment report: A critical review of changes in the tropospheric ozone burden and budget from 1850 to 2100. *Elementa: Science of the Anthropocene* **8**(1): 034. DOI: <https://dx.doi.org/10.1525/elementa.2020.034>.
- Bell, ML, McDermott, A, Zeger, SL, Samet, JM, Dominici, F.** 2004. Ozone and short-term mortality in 95 US urban communities, 1987–2000. *Journal of the American Medical Association* **292**: 2372–2378.
- Betancourt, C, Stomberg, TT, Edrich, A-K, Patnala, A, Schultz, MG, Roscher, R, Kowalski, J, Stadtler, S.** 2022. Global, high-resolution mapping of tropospheric ozone—Explainable machine learning and

- impact of uncertainties. *Geoscientific Model Development* **15**: 4331–4354. DOI: <http://dx.doi.org/10.5194/gmd-15-4331-2022>.
- Brauer, M, Freedman, G, Frostad, J, van Donkelaar, A, Martin, RV, Dentener, F, van Dingenen, R, Estep, K, Amini, H, Apte, JS, Balakrishnan, K, Barre-gard, L, Broday, D, Feigin, V, Ghosh, S, Hopke, PK, Knibbs, LD, Kokubo, Y, Liu, Y, Ma, S, Morawska, L, Sangrador, JLT, Shaddick, G, Anderson, HR, Vos, T, Forouzanfar, MH, Burnett, RT, Cohen, A.** 2016. Ambient air pollution exposure estimation for the Global Burden of Disease 2013. *Environmental Science & Technology* **50**(1): 79–88. DOI: <http://dx.doi.org/10.1021/acs.est.5b03709>.
- Chang, K-L, Cooper, OR, West, JJ, Serre, ML, Schultz, MG, Lin, M, Marecal, V, Josse, B, Deushi, M, Sudo, K, Liu, J, Keller, CA.** 2019. A new method (M³Fusion v1) for combining observations and multiple model output for an improved estimate of the global surface ozone distribution. *Geoscientific Model Development* **12**: 955–978. DOI: <http://dx.doi.org/10.5194/gmd-12-955-2019>.
- Chang, K-L, Petropavlovskikh, I, Cooper, OR, Schultz, MG, Wang, T.** 2017. Regional trend analysis of surface ozone observations from monitoring networks in eastern North America, Europe and East Asia. *Elementa: Science of the Anthropocene* **5**: 50.
- Christakos, G.** 2000. *Modern spatiotemporal geostatistics*. Oxford, UK: Oxford University Press.
- Christakos, G, Bogaert, P, Serre, M.** 2001. *Temporal GIS: Advanced functions for field-based applications*. Berlin, Germany: Springer-Verlag. DOI: <http://dx.doi.org/10.1007/978-3-642-56540-3>.
- Christakos, G, Kolovos, A, Serre, ML, Vukovich, F.** 2004. Total ozone mapping by integrating databases from remote sensing instruments and empirical models. *IEEE Transactions on Geoscience and Remote Sensing* **42**: 991–1008.
- Cleland, SE, West, JJ, Jia, Y, Reid, S, Raffuse, S, O'Neill, S, Serre, ML.** 2020. Estimating wildfire smoke concentrations during the October 2017 California fires through BME space/time data fusion of observed, modeled, and satellite-derived PM_{2.5}. *Environmental Science & Technology* **54**: 13439–13447. DOI: <http://dx.doi.org/10.1021/acs.est.0c03761>.
- Cohan, DS, Hakami, A, Hu, Y, Russell, AG.** 2005. Non-linear response of ozone to emissions: Source apportionment and sensitivity analysis. *Environmental Science & Technology* **39**: 6739–6748.
- Collins, WJ, Lamarque, J-F, Schulz, M, Boucher, O, Eyring, V, Hegglin, MI, Maycock, A, Myhre, G, Prather, M, Shindell, D, Smith, SJ.** 2017. AerChemMIP: Quantifying the effects of chemistry and aerosols in CMIP6. *Geoscientific Model Development* **10**: 585–607. DOI: <http://dx.doi.org/10.5194/gmd-10-585-2017>.
- Cooper, OR, Parrish, DD, Ziemke, J, Balashov, NV, Cupeiro, M, Galbally, IE, Gilge, S, Horowitz, L, Jensen, NR, Lamarque, J-F, Naik, V, Oltmans, SJ, Schwab, J, Shindell, DT, Thompson, AM, Thouret, V, Wang, Y, Zbinden, RM.** 2014. Global distribution and trends of tropospheric ozone: An observation-based review. *Elementa: Science of the Anthropocene* **2**: 000029. DOI: <http://dx.doi.org/10.12952/journal.elementa.000029>.
- de Nazelle, A, Arunachalam, S, Serre, ML.** 2010. Bayesian maximum entropy integration of ozone observations and model predictions: An application for attainment demonstration in North Carolina. *Environmental Science & Technology* **44**: 5707–5713.
- DeLang, MN, Becker, JS, Chang, K-L, Serre, ML, Cooper, OR, Schultz, MG, Schroder, S, Lu, X, Zhang, L, Deushi, M, Josse, B, Keller, CA, Lamarque, J-F, Lin, M, Liu, J, Marecal, V, Strode, SA, Sudo, K, Tilmes, S, Zhang, L, Cleland, S, Collins, E, Brauer, M, West, JJ.** 2021. Mapping yearly fine resolution global surface ozone through the Bayesian Maximum Entropy data fusion of observations and model output for 1990–2017. *Environmental Science & Technology* **55**: 4389–4398. DOI: <http://dx.doi.org/10.1021/acs.est.0c07742>.
- Di, Q, Dai, L, Wang, Y, Zanobetti, A, Choirat, C, Schwartz, JD, Dominici, F.** 2017. Association of short-term exposure to air pollution with mortality in older adults. *Journal of the American Medical Association* **318**: 2446–2456. DOI: <http://dx.doi.org/10.1001/jama.2017.17923>.
- Fleming, ZL, Doherty, RM, von Schneidmesser, E, Malley, CS, Cooper, OR, Pinto, JP, Colette, A, Xu, X, Simpson, D, Schultz, MG, Lefohn, AS, Hamad, S, Moolla, R, Solberg, S, Feng, Z.** 2018. Tropospheric ozone assessment report: Present-day ozone distribution and trends relevant to human health. *Elementa: Science of the Anthropocene* **6**: 12.
- GBD 2017 Risk Factors Collaborators.** 2018. Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks for 195 countries and territories, 1990–2017: A systematic analysis for the Global Burden of Disease Study 2017. *The Lancet* **392**: 1923–1994.
- GBD 2019 Risk Factors Collaborators.** 2020. Global burden of 87 risk factors in 204 countries and territories, 1990–2019: A systematic analysis for the Global Burden of Disease Study 2019. *The Lancet* **396**: 1223–1249. DOI: [http://dx.doi.org/10.1016/S0140-6736\(20\)30752-2](http://dx.doi.org/10.1016/S0140-6736(20)30752-2).
- Jerrett, M, Burnett, RT, Pope, CA, Ito, K, Thurston, G, Krewski, D, Shi, Y, Calle, E, Thun, M.** 2009. Long-term ozone exposure and mortality. *New England Journal of Medicine* **360**: 1085–1095. DOI: <http://dx.doi.org/10.1056/NEJMoa0803894>.
- Kleinert, F, Leufen, LH, Schultz, MG.** 2021. IntelliO3-Ts v1.0: A neural network approach to predict near-surface ozone concentrations in Germany. *Geoscientific Model Development* **14**: 1–25.
- Liang, J, Jacobson, MZ.** 2000. Effects of subgrid segregation on ozone production efficiency in a chemical model. *Atmospheric Environment* **34**: 2975–2982.
- Liu, X, Zhu, Y, Xue, L, Desai, AR, Wang, H.** 2022. Cluster-enhanced ensemble learning for mapping global

- monthly surface ozone from 2003 to 2019. *Geophysical Research Letters* **49**: e2022GL097947. DOI: <http://dx.doi.org/10.1029/2022GL097947>.
- Lu, X, Zhang, L, Wang, X, Gao, M, Li, K, Zhang, Y, Yue, X, Zhang, Y.** 2020. Rapid increases in warm-season surface ozone and resulting health impact in China since 2013. *Environmental Science & Technology Letters* **7**: 240–247.
- Morgenstern, O, Hegglin, M, Rozanov, E, O'Connor, F, Abraham, NL, Akiyoshi, H, Archibald, AT, Bekki, S, Butchart, N, Chipperfield, MP, Deushi, M, Dhomse, SS, Garcia, RR, Hardiman, SC, Horowitz, LW, Jöckel, P, Josse, B, Kinnison, D, Lin, M, Mancini, E, Manyin, ME, Marchand, M, Maréchal, V, Michou, M, Oman, LD, Pitari, G, Plummer, DA, Revell, LE, Saint-Martin, D, Schofield, R, Stenke, A, Stone, K, Sudo, K, Tanaka, TY, Tilmes, S, Yamashita, Y, Yoshida, K, Zeng, G.** 2017. Review of the global models used within phase 1 of the Chemistry–Climate Model Initiative (CCMI). *Geoscientific Model Development* **10**: 639–671. DOI: <http://dx.doi.org/10.5194/gmd-10-639-2017>.
- Reyes, JM.** 2016. Geostatistical data fusion estimation methods of ambient PM_{2.5} and polycyclic aromatic hydrocarbons [Dissertation]. Chapel Hill, NC: University of North Carolina at Chapel Hill.
- Reyes, JM, Xu, Y, Vizuete, W, Serre, ML.** 2017. Regionalized PM_{2.5} Community Multiscale Air Quality model performance evaluation across a continuous spatio-temporal domain. *Atmospheric Environment* **148**: 258–265.
- Schultz, MG, Schröder, S, Lyapina, O, Cooper, O, Galbally, I, Petropavlovskikh, I, Von Schneidmeyer, E, Tanimoto, H, Elshorbany, Y, Naja, M, Seguel, RJ, Dauert, U, Eckhardt, P, Feigenspan, S, Fiebig, M, Hjellbrekke, A-G, Hong, Y-D, Kjeld, PC, Koide, H, Lear, G, Tarasick, D, Ueno, M, Wallasch, M, Baumgardner, D, Chuang, M-T, Gillett, R, Lee, M, Molloy, S, Moolla, R, Wang, T, Sharps, K, Adame, JA, Ancellet, G, Apadula, F, Artaxo, P, Barlasina, ME, Bogucka, M, Bonasoni, P, Chang, L, Colomb, A, Cuevas-Agulló, E, Cupeiro, M, Degorska, A, Ding, A, Fröhlich, M, Frolova, M, Gadhavi, H, Gheusi, F, Gilge, S, Gonzalez, MY, Gros, V, Hamad, SH, Helmig, D, Henriques, D, Hermansen, O, Holla, R, Hueber, J, Im, U, Jaffe, DA, Komala, N, Kubistin, D, Lam, K-S, Laurila, T, Lee, H, Levy, I, Mazzoleni, C, Mazzoleni, LR, McClure-Begley, A, Mohammad, M, Murovec, M, Navarro-Comas, M, Nicodim, F, Parrish, D, Read, KA, Reid, N, Ries, L, Saxena, P, Schwab, JJ, Scorgie, Y, Senik, I, Simmonds, P, Sinha, V, Skorokhod, AI, Spain, G, Spangl, W, Spoor, R, Springston, SR, Steer, K, Steinbacher, M, Suharguniyawan, E, Torre, P, Trickl, T, Weili, L, Weller, R, Xiaobin, X, Xue, L, Zhiqiang, M.** 2017. Tropospheric Ozone Assessment Report: Database and metrics data of global surface ozone observations. *Elementa: Science of the Anthropocene* **5**: 58. DOI: <http://dx.doi.org/10.1525/elementa.244>.
- Seltzer, KM, Shindell, DT, Kasibhatla, P, Malley, CS.** 2020. Magnitude, trends, and impacts of ambient long-term ozone exposure in the United States from 2000 to 2015. *Atmospheric Chemistry and Physics* **20**: 1757–1775. DOI: <http://dx.doi.org/10.5194/acp-20-1757-2020>.
- Serre, ML, Christakos, G.** 1999. Modern geostatistics: Computational BME analysis in the light of uncertain physical knowledge—The Equus Beds study. *Stochastic Environmental Research and Risk Assessment* **13**: 1–26.
- Strode, SA, Ziemke, JR, Oman, LD, Lamsal, LN, Olsen, MA, Liu, J.** 2019. Global changes in the diurnal cycle of surface ozone. *Atmospheric Environment* **199**: 323–333.
- Sun, H, Shin, YM, Xia, M, Ke, S, Wan, M, Yuan, L, Guo, Y, Archibald, AT.** 2021. Spatial resolved surface ozone with urban and rural differentiation during 1990–2019: A space-time Bayesian neural network downscaler. *Environmental Science & Technology* **56**(11): 7337–7349.
- Turner, MC, Jerrett, M, Pope, CA, Krewski, D, Gapstur, SM, Diver, WR, Beckerman, BS, Marshall, JD, Su, J, Crouse, DL, Burnett, RT.** 2016. Long-term ozone exposure and mortality in a large prospective study. *American Journal of Respiratory and Critical Care Medicine* **193**: 1134–1142.
- U.S. Environmental Protection Agency.** 2020. Integrated science assessment for ozone and related photochemical oxidants. Washington, DC: U.S. Environmental Protection Agency. Report No.: EPA/600/R-20/012.
- von Kuhlmann, R, Lawrence, MG, Crutzen, PJ, Rasch, PJ.** 2003. A model for studies of tropospheric ozone and nonmethane hydrocarbons: Model evaluation of ozone-related species. *Journal of Geophysical Research: Atmospheres* **108**: 4729. DOI: <https://doi.org/10.1029/2002JD003348>.
- World Health Organization.** 2021. *WHO global air quality guidelines, particulate matter (PM_{2.5} and PM₁₀), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide*. Geneva, Switzerland: World Health Organization.
- Xu, Y, Serre, ML, Reyes, J, Vizuete, W.** 2016. Bayesian maximum entropy integration of ozone observations and model predictions: A national application. *Environmental Science & Technology* **50**: 4393–4400.
- Young, PJ, Naik, V, Fiore, AM, Gaudel, A, Guo, J, Lin, MY, Neu, JL, Parrish, DD, Rieder, HE, Schnell, JL, Tilmes, S, Wild, O, Zhang, L, Ziemke, J, Brandt, J, Delcloo, A, Doherty, RM, Geels, C, Hegglin, MI, Hu, L, Im, U, Kumar, R, Luhar, A, Murray, L, Plummer, D, Rodriguez, J, Saiz-Lopez, A, Schultz, MG, Woodhouse, MT, Zeng, G.** 2018. Tropospheric Ozone Assessment Report: Assessment of global-scale model performance for global and regional ozone distributions, variability, and trends. *Elementa: Science of the Anthropocene* **6**: 10.

How to cite this article: Becker, JS, DeLang, MN, Chang, K-L, Serre, ML, Cooper, OR, Wang, H, Schultz, MG, Schröder, S, Lu, X, Zhang, L, Deushi, M, Josse, B, Keller, CA, Lamarque, J-F, Lin, M, Liu, J, Marécal, V, Strode, SA, Sudo, K, Tilmes, S, Zhang, L, Brauer, M, West, JJ. 2023. Using Regionalized Air Quality Model Performance and Bayesian Maximum Entropy data fusion to map global surface ozone concentration. *Elementa: Science of the Anthropocene* 11(1). DOI: <https://doi.org/10.1525/elementa.2022.00025>

Domain Editor-in-Chief: Detlev Helmig, Boulder AIR LLC, Boulder, CO, USA

Associate Editor: Ian Faloona, University of California Davis, Davis, CA, USA

Knowledge Domain: Atmospheric Science

Part of an Elementa Special Feature: Tropospheric Ozone Assessment Report (TOAR)

Published: October 13, 2023 **Accepted:** August 02, 2023 **Submitted:** February 11, 2022

Copyright: © 2023 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.



Elem Sci Anth is a peer-reviewed open access journal published by University of California Press.

OPEN ACCESS 