Mapping Global Surface Ozone Concentrations through the Statistical Fusion of Observations and Models

Marissa DeLang[1], Jacob Becker[1], Kai-Lan Chang[2], Owen Cooper[2,3], Stephanie Cleland[1], Martin Schultz[4], Sabine Schroder[4], Jason West[1], Marc Serre[1], CCMI & NASA modelers

[1]Department of Environmental Sciences and Engineering, University of North Carolina Chapel Hill, NC

[2]NOAA Earth System Research Laboratory, Boulder, CO

[3]Cooperative Institute for Research in Environmental Studies, University of Colorado, Boulder, CO

[4]Forschungszentrum Jülich, Jülich, Germany

UNC GILLINGS SCHOOL OF GLOBAL PUBLIC HEALTH

HAQAST

INTRODUCTION

We estimated global surface ozone concentrations by fusing ozone measurements and global models. Our motivation includes the following:

- Tropospheric ozone is an EPA criteria pollutant known to be detrimental to health.
- The Global Burden of Disease (GBD) Assessment estimated 472,000 deaths from ambient ozone pollution in 2017 [1].
- GBD requests fine resolution global surface ozone estimations for 1990-present.
- Ozone measurements are common in North America, Europe, and Japan, but other regions have very few measurements.
- Models provide information on ozone levels in data sparse regions, but model predictions are inaccurate.

To map ozone for the forthcoming GBD-2019, we improve upon the previous methods used for GBD-2017 [2] by:

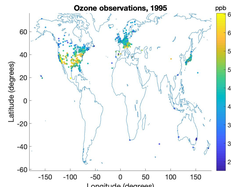
- Estimating global surface ozone for each year from 1990-2017
- Adding new observations in China
- Using the Bayesian Maximum Entropy method for space-time data fusion
- Adding fine spatial structure

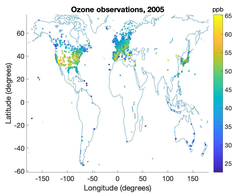
METHODS

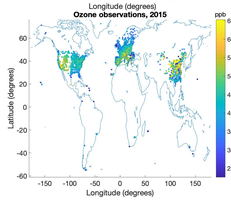
We produce global ozone maps annually for the maximum six month average of the daily maximum eight hour average, or mda8 (in ppb):

Observations

Ozone measurements were obtained from the Tropospheric Ozone Assessment Report (TOAR) for 1990-2017 and the China National Environmental Monitoring Center (CNEMC) Network for 2013-2017 [3]. In total, there were 8,834 monitoring sites over the time period, with a minimum of 1,199 stations in 1990 and a maximum of 4,999 stations in 2015.







M^{3}Fusion Model

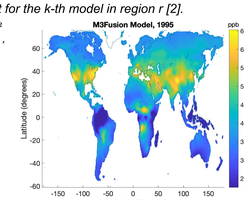
Modeled ozone concentrations were used from the following models, many from the Chemistry-Climate Model Initiative (CCMI)[4] : CHASER (1990-2010), MOCAGE (1990-2016), MRI-ESM (1990-2017), MERRA2-GMI (1990-2017), NCAR CESM-Chem (1990-2010), NCAR WACCM (1990-2010), GFDL AM3 (1990-2014), and GFDL AM4 (2010-2016). Using the M^{3}Fusion method, models were weighted in each region to minimize the difference between the bias-corrected multi-model average and observations:

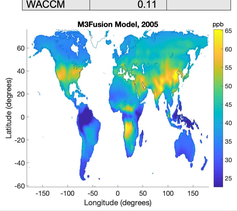
Let s_g be the grid cell at resolution 0.5° × 0.5°, $\hat{y}(s_g)$ be the interpolated observations, $\{\eta_k(s_g); k = 1, \ldots, n\}$ be the model output registered onto the same grid from the n models available in a given year. αr is a constant that allows adjustment to the overall (regional) underestimation or overestimation and β_{rk} is an optimal weight for the k-th model in region r [2].

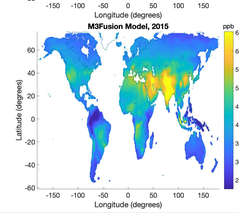
$$\underset{\{\alpha_r, \beta_{rk}; k=1,..,n\}}{minimize} \sum_{s_g \in Region\ r} \left(\hat{y}(s_g) - \alpha_r - \sum_{k=1}^{n} \beta_{rk} \eta_k(s_g) \right)^2,$$

$$subject\ to \sum_{k=1}^{n} \beta_{rk} = 1 \ and \ \beta_{rk} \geq 0$$

	North America		Europe	
Year	*1995*	*2005*	*1995*	*2005*
CESM-Chem	0.38	0.29		
CHASER	0.04			
GFDL-AM3			0.18	0.62
MERRA		0.32	0.82	0.38
MOCAGE	0.36	0.01		
MRI-ESM	0.22	0.27		
WACCM		0.11		



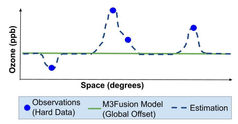




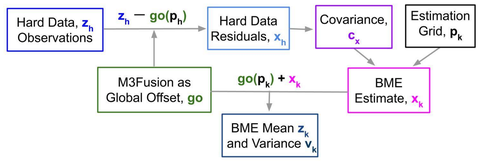
METHODS

Bayesian Maximum Entropy Framework

BME is a geostatistical modeling tool that we use to estimate of ozone pollution by combining site specific observations with modeled concentrations using the correlations between measurement locations. BME also estimates a variance, which assesses estimation confidence spatially. The steps are:



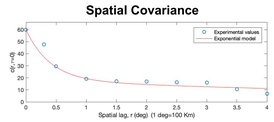
1. Let **Z(p)** be a field of ozone concentration estimations in space and time
2. Assign the M^{3}Fusion output as the "global offset": **go(p)**
3. Remove global offset from measurements (hard data), $\mathbf{z_h}$, to obtain residuals x_h=$\mathbf{z_h}$-**go**($\mathbf{p_h}$).
4. Model the covariance (the correlation between locations in space and time) c_x based on the residuals x_h.
5. Combine hard data residuals (x_h), covariance (c_x), and estimation parameters to get the BME estimation (x_k)
6. Obtain final values (z_k) by adding back the global offset **go**($\mathbf{p_k}$) to BME estimation (x_k).

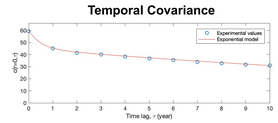


For the theory, derivation, and details of BME, see Christakos 1990 [5].
Implemented with BMElib [6].

Covariance

The covariance in space and time is calculated from ozone observations:



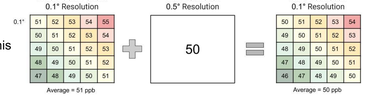


$$C_x(r,\tau) = 59.9938\ ppb^2 (0.7 \exp\left(-\frac{3r}{1.2\ degrees}\right)\exp\left(-\frac{3\tau}{80\ years}\right) + 0.05\exp\left(-\frac{3r}{25\ degrees}\right)\exp\left(-\frac{3\tau}{80\ years}\right) + 0.25\exp\left(-\frac{3r}{25\ degrees}\right)\exp\left(-\frac{3\tau}{1.5\ years}\right)$$

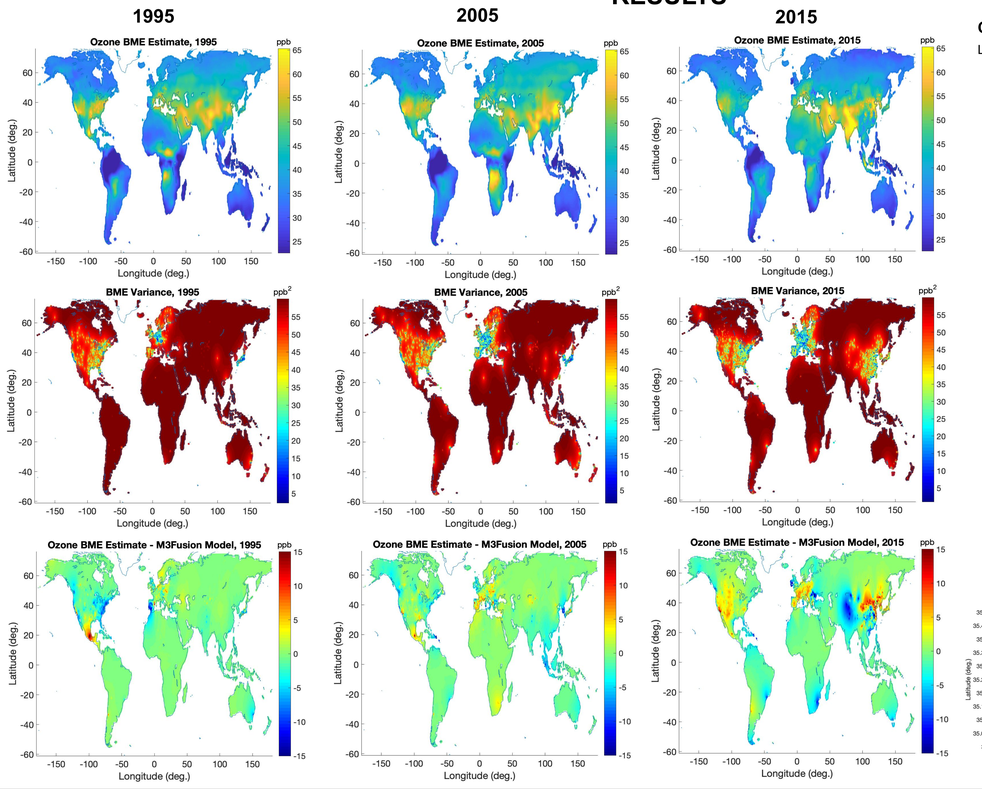
where C_x is the covariance (ppb^2), r is the spatial lag (degrees), and τ is the temporal lag (years). The covariance is used to give the range of influence of a measurement to predict other concentrations. The spatial covariance, and therefore the spatial influence of an observation, drops off steeply at one degree. The temporal covariance, however, remains high, meaning that observations can influence ozone estimates over several years.

Fine Resolution

Most models used are approximately 2° resolution, but finer resolution is required for GBD. We add fine spatial resolution using estimates from the NASA G5NR-Chem model [7] at 0.125° resolution for July 2013 to June 2014. We regridded this to 0.1° resolution and used it to provide fine spatial patterns, keeping the average of each 0.5° grid cell from the BME output, for each year 1990-2017.



RESULTS



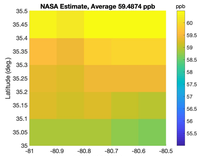
Cross Validation

Leave one out cross validation was performed for the following:

- Multimodel Mean: Equally weighted mean of all available models in each year
- M^{3}Fusion Model: Bias corrected, observation weighted mean of all available models in each year
- Space Only Correction: BME mean estimate allowing observations to correct M^{3}Fusion Model across space
- Space Time Correction: BME mean estimate allowing observations to correct M^{3}Fusion Model across space and time
- Fine Resolution: Space time corrected M^{3}Fusion Model with the fine resolution spatial pattern of the NASA G5NR-Chem model

Scenario	MSE (ppb^2)	R^2 (ppb^2)
Multimodel Mean	189.2331	0.2814
M3Fusion Model	61.1408	0.2996
Space Only Correction	31.1659	0.6301
Space Time Correction	15.7267	0.8163
Fine Resolution	30.0853	0.6399

Fine resolution results are shown for a single 0.5° grid cell over Charlotte, NC in 2005.



FUTURE WORK

Constant Air Quality Model Performance

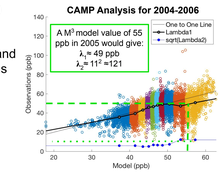
The M^{3}Fusion output can be directly evaluated using observations, in which each modeled value is given a variance based on how well the model matches observations (a lower variance is weighed higher in the final estimation). CAMP also removes model bias by basing new mean offset values on both models and observations. [8]

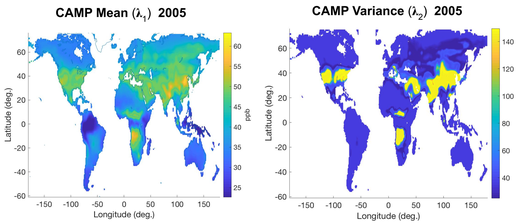
1) Match station measurement values (observations) with model estimations at each measurement location in space and time
2) Sort each paired value into 10 equally sized bins and calculate λ_1=mean and λ_2=variance of observations in each bin:

$$\lambda_1(\bar{s}_i) = \frac{1}{n(\bar{s}_i)} \sum_{j=1}^{n(\bar{s}_i)} \hat{s}_j \qquad \lambda_2(\bar{s}_i) = \frac{1}{n(\bar{s}_i) - 1} \sum_{j=1}^{n(\bar{s}_i)} \left(\hat{s}_j - \lambda_1(\bar{s}_i) \right)^2$$

Where $n(\bar{s}_i)$ is the number of collocated modeled ($\hat{s}_i$) and observed ($\bar{s}_i$) values in each bin.

3) Find λ_1 and λ_2 for each model point, using three years of data to create the $\boldsymbol{\lambda}$s for a given year





Regionalized Air Quality Model Performance

Using the Regionalized Air Quality Model Performance (RAMP) method [9] would account for non-homogeneity in model performance across space, which could provide better model correction in data sparse areas.

CONCLUSION

We successfully created global surface ozone estimations for 1990-2017 at fine resolution with the following features:

- **Observations were combined with models that are bias corrected and weighted based on their performance.**
- **Observations take precedence in estimation near observation locations, but BME weight falls off based on covariance and space/time distance**
- **Observations not only correct the model in the year they were observed, but also influence other years according to the temporal covariance. This is important for regions that were not monitored every year over the entire 1990-2017 time period, such as China.**
- **The spatial and temporal corrections greatly reduce MSE and improve R^2.**
- **The BME method provides a variance for each ozone estimation based on space/time distance from observations.**
- **The fine spatial structure of the final product follows the spatial distribution of the 0.125° NASA G5NR-Chem model.**

Annual results at 0.1° resolution were submitted for the 2019 Global Burden of Disease Assessment.

ACKNOWLEDGEMENTS

We thank NASA HAQAST for funding, the Tropospheric Ozone Assessment Report, the Global Burden of Disease Assessment, the Chinese National Environmental Monitoring Network, and modeling teams who provided data: CHASER, MOCAGE, MRI-ESM, NASA G5NR-Chem, NASA MERRA2-GMI, NCAR CESM-Chem, NCAR WACCM, GFDL AM3,and GFDL AM4.

REFERENCES

[1] Global Burden of Disease. (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. Lancet, 392(10159), 1923.

[2] Chang, Kai-Lan, Owen R. Cooper, J. Jason West, Marc L. Serre, Martin G. Schultz, Meiyun Lin, Virginie Marticorl, et al. 2019. "A New Method (M3 Fusion v1) for Combining Observations and Multiple Model Output for an Improved Estimate of the Global Surface Ozone Distribution." *Geoscientific Model Development* 12 (3): 955–78. https://doi.org/10.5194/gmd-12-955-2019.

[3] Schultz, Martin G., Sabine Schröder, Olga Lyapina, Owen Cooper, Ian Galbally, Irina Petropavlovskikh, Erika Von Schneidemesser, et al. 2017. "Tropospheric Ozone Assessment Report Database and Metrics Data of Global Surface Ozone Observations." Elem Sci Anth 5 (0): 58. https://doi.org/10.1525/elementa.244.

[4] Morgenstern, O., Hegglin, M., Rozanov, E., O'Connor, F., Luke Abraham, N., Akiyoshi, H., . . . Zeng, G. (2017). Review of the global models used within phase 1 of the chemistry-climate model initiative (CCMI). *Geoscientific Model Development*, 10(2), 639-671. doi:10.5194/gmd-10-639-2017

[5] Christakos, George. 1990. "A Bayesian/Maximum-Entropy View to the Spatial Estimation Problem." *Mathematical Geology* 22 (7): 763–77. https://doi.org/10.1007/BF00890661.

[6] Serre, M. L., and G. Christakos. 1999. "Modern Geostatistics: Computational BME Analysis in the Light of Uncertain Physical Knowledge - the Equus Beds Study." *Stochastic Environmental Research and Risk Assessment (SERRA)* 13 (1–2): 1–26. https://doi.org/10.1007/s004770050029.

[7] Hu, L., Keller, C. A., Long, M. S., Sherwen, T., Auer, B., Da Silva, A., . . . Jacob, D. J. (2018). Global simulation of tropospheric chemistry at 12.5 km resolution: Performance and evaluation of the GEOS-chem chemical module (v10-1) within the NASA GEOS earth system model (GEOS-5 ESM). Geoscientific Model Development, 11(11), 4603-4620. doi:10.5194/gmd-11-4603-2018

[8] Nazelle, Audrey de, Saravanan Arunachalam, and Marc L. Serre. 2010. "Bayesian Maximum Entropy Integration of Ozone Observations and Model Predictions: An Application for Attainment Demonstration in North Carolina." *Environmental Science & Technology* 44 (15): 5707–13. https://doi.org/10.1021/es1002294e.

[9] Reyes, Jeanette M., Yadong Xu, William Vizuete, and Marc L. Serre. 2017. "Regionalized PM2.5 Community Multiscale Air Quality Model Performance Evaluation across a Continuous Spatiotemporal Domain." *Atmospheric Environment* 148 (January): 258–65. https://doi.org/10.1016/J.ATMOSENV.2016.10.048.