Supplemental information:
Comparisons of simple and complex methods for quantifying exposure to point source air pollution emissions

Lucas RF Henneman1, Irene C Dedousi2,3, Joan A Casey4,5, Christine Choirat6, Steven RH Barrett3, and Corwin M Zigler7

1Department of Biostatistics, Harvard T. H. Chan School of Public Health, Boston, U.S.A.
2Faculty of Aerospace Engineering, Delft University of Technology, Delft, The Netherlands
3Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, Boston, U.S.A.
4School of Public Health, University of California, Berkeley, U.S.A.
5Columbia University Mailman School of Public Health, New York, U.S.A.
6Swiss Data Science Center, ETH Zürich and EPFL, Lausanne, Switzerland
7Department of Statistics and Data Sciences and Department of Women’s Health, University of Texas, Austin, U.S.A.
SI-1 Reduced complexity approaches as PM$_{2.5}$

For HyADS and IDWE exposure fields to all emissions sources ($exposure_{ij}^m$), we projected raw exposure fields to match the CMAQ-DDM Hybrid grid and trained multiple models over the continental United States. Along with model defined in equation (4) in the main paper, we trained two additional linear models:

\[ PM_{2.5}^{CMAQ-DDM} = \beta_0^m + \beta_{\text{exp}}^m \sum_{j=1}^J exposure_j^m + \epsilon^m \]  

(SI-1)

\[ PM_{2.5}^{CMAQ-DDM} = \beta_0^m + \beta_{\text{exp}}^m \sum_{j=1}^J exposure_j^m + \beta_X^m \tilde{X} + \beta_{\text{exp,exp}}^m \tilde{X} \sum_{j=1}^J exposure_j^m + \beta_s^m s(x,y) + \epsilon^m \]  

(SI-2)

where $PM_{2.5}^{CMAQ-DDM}$ is PM$_{2.5}$ coal impacts from CMAQ-DDM Hybrid, $\tilde{X}$ is the vector of meteorological variables from the North American Reanalysis1, and $s(x,y)$ is a bivariate spline of latitude and longitude (in meters) with 100 knots. $\epsilon$ is assumed iid normal with no spatial structure. We employed average temperature, accumulated precipitation, relative humidity, and $x$ and $y$ wind vectors for meteorological inputs.

As a fourth model, we employed a Z-score adjustment of $exposure^m$ to match that of $PM_{2.5}^{CMAQ-DDM}$. For conversions of $exposure_j^m$ to $PM_{2.5}^m$, we employed this equation:

\[ PM_{2.5,j}^m = sd(PM_{2.5}^{CMAQ-DDM}) \times \left( \frac{exposure_j^m - \text{mean}(exposure^m)}{sd(exposure^m)} \right) + \text{mean}(PM_{2.5}^{CMAQ-DDM}) \]  

(SI-3)

where $sd(\bullet)$ represents the standard deviation and $\text{mean}(\bullet)$ represents the mean.

SI-1.1 Annual evaluation

We trained the models using total PM$_{2.5}$ coal source impacts in 2005 and evaluated them by predicting 2006 total PM$_{2.5}$ coal source impacts (Figure SI-1). The linear model formulation in the main document was found to have the best performance and the least complex formulation; therefore, we present results from this model throughout the main results and the remainder of this document.
Figure SI-1: Evaluation statistics for total annual coal PM$_{2.5}$ source impacts PM$_{2.5}^{\text{PM}}$ evaluated against PM$_{2.5}^{\text{CMAQ–DDM}}$. 
SI-1.2 Monthly evaluation

Figure SI-2: Evaluation statistics for total monthly coal PM$_{2.5}$ source impacts PM$_{2.5}^m$ evaluated against PM$_{2.5}^{CMAQ-DDM}$. Models were trained in each month in 2005 and evaluated in 2006.

SI-1.3 Total source impact fields as PM$_{2.5}$

Raw HyADS and IDWE exposure from all coal power plants ($\sum_{j=1}^J exposure_{i,j}^{HyADS}$ and $\sum_{j=1}^J exposure_{i,j}^{IDWE}$) were highly correlated with CMAQ-DDM in 2006 (Pearson R of 0.94 for both). $PM_{2.5}^{IDWE}$ year 2006 model predictions trained on 2005 exposure$_{i}^{IDWE}$ and $PM_{2.5}^{CMAQ-DDM}$ yielded lower bias and error than comparable results for $PM_{2.5}^{HyADS}$. 

PM$_{2.5}$
Figure SI-3: Total annual $\text{PM}_{2.5}^{\text{CMAQ-DDM}}$, $\text{PM}_{2.5}^{\text{HyADS}}$, and $\text{PM}_{2.5}^{\text{IDWE}}$ in 2006. * denotes converted metrics from exposure $\text{HyADS}$ and exposure $\text{IDWE}$.

Figure SI-4: Spatial bias of total annual $\text{PM}_{2.5}^{\text{HyADS}}$ and $\text{PM}_{2.5}^{\text{IDWE}}$ relative to $\text{PM}_{2.5}^{\text{CMAQ-DDM}}$ in 2006. * denotes converted metrics from exposure $\text{HyADS}$ and exposure $\text{IDWE}$. 
Figure SI-5: Population-emissions weighted distance ($D_{pew}$) calculated for each grid cell in the contiguous United States.
SI-2 Source impact evaluation metrics

This section presents expanded annual evaluations of $PWSI_{LP}^{HyADS}$ and $PWSI_{LP=US}^{IDWE}$ against $PWSI_{LP=US}^{Adjoint}$. These figures supplement the evaluation metrics presented in Figure 3.

SI-2.1 Annual evaluations

Figure SI-6: Scatterplot of $PWSI_{LP}^{HyADS}$ and $PWSI_{LP}^{IDWE}$ against $PWSI_{LP}^{Adjoint}$ for each coal-fired power plant.
Figure SI-7: Spearman R (rank-ordered correlation), Normalized Mean Error (0% < NME < +∞) and Mean Bias (MB) of PWSI\textsubscript{HADS}\textsuperscript{P} and PWSI\textsubscript{IDWE}\textsuperscript{P} compared to GEOS-Chem adjoint sensitivities. IDWE* for CA are omitted from this plot because they are many times higher than the NME in other states. The removed values range from 3,600% to 6,200%.
SI-2.1 Monthly evaluations

Figure SI-8: Monthly linear (Pearson R) and rank-ordered (Spearman R) correlations between PWSI_{HyADS}^{P,j} and PWSI_{IDWE}^{P,j} source impacts evaluated against PWSI_{Adjoint}^{P,j} on individual states and entire United States (US). States are ordered east to west descending.
Figure SI-9: Normalized Mean Error ($0\% < \text{NME} < +\infty$) of $\text{PWSI}_{P,j}^{\text{IDWE}}$ evaluated against $\text{PWSI}_{P,j}^{\text{HYADS}}$. The values in Colorado (CO) range up to 18,000% and in California range from 800% to greater than 2,000,000%.
Figure SI-10: Mean bias (MB) of $\text{PWSI}_{P,j}^{\text{IDWE}}$ evaluated against $\text{PWSI}_{P,j}^{\text{HyADS}}$. 
Figure SI-11: Mean error (ME) and root mean square error (RMSE) of $PWSI_{P,j}^{IDWE}$ evaluated against $PWSI_{P,j}^{HyADS}$.
Figure SI-12: Linear (Pearson R) and rank-order (Spearman R) correlations of raw HyADS and IDWE individual source exposure metrics ($\text{exposure}_{ij}^{\text{HyADS}}$ and $\text{exposure}_{ij}^{\text{IDWE}}$) compared to $\text{PWSI}_{p,j}^{\text{Adjacent}}$. 