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

Lucas RF Henneman[1], Irene C Dedoussi[2,3], Joan A Casey[4,5], Christine Choirat[6], Steven RH Barrett[3], and Corwin M Zigler[7]

[1]Department of Biostatistics, Harvard T. H. Chan School of Public Health, Boston, U.S.A.

[2]Faculty of Aerospace Engineering, Delft University of Technology, Delft, The Netherlands

[3]Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, Boston, U.S.A.

[4]School of Public Health, University of California, Berkeley, U.S.A.

[5]Columbia University Mailman School of Public Health, New York, U.S.A.

[6]Swiss Data Science Center, ETH Zürich and EPFL, Lausanne, Switzerland

[7]Department 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 ($exposure_{i,j}^{m}$), 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:

$$PM_{2.5}^{CMAQ-DDM} = \beta_0^m + \beta_{exp}^m \sum_{j=1}^{J} exposure_j^m + \epsilon^m \quad \text{(SI-1)}$$

$$PM_{2.5}^{CMAQ-DDM} = \beta_0^m + \beta_{exp}^m \sum_{j=1}^{J} exposure_j^m + \beta_{\vec{X}}^m \vec{X} + \beta_{exp,\vec{X}}^m \vec{X} * \sum_{j=1}^{J} exposure_j^m + \beta_s^m s(x,y) + \epsilon^m \quad \text{(SI-2)}$$

where $PM_{2.5}^{CMAQ-DDM}$ is $PM_{2.5}$ coal impacts from CMAQ-DDM Hybrid, $\vec{X}$ is the vector of meteorological variables from the North American Reanalysis[1], and $s(x,y)$ is a bivariate spline of latitude and longitude (in meters) with 100 knots. ϵ 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:

$$PM_{2.5,j}^m = sd\left(PM_{2.5}^{CMAQ-DDM}\right) * \left(\frac{exposure_j^m - mean(exposure^m)}{sd(exposure^m)} + mean\left(PM_{2.5}^{CMAQ-DDM}\right)\right) \quad \text{(SI-3)}$$

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

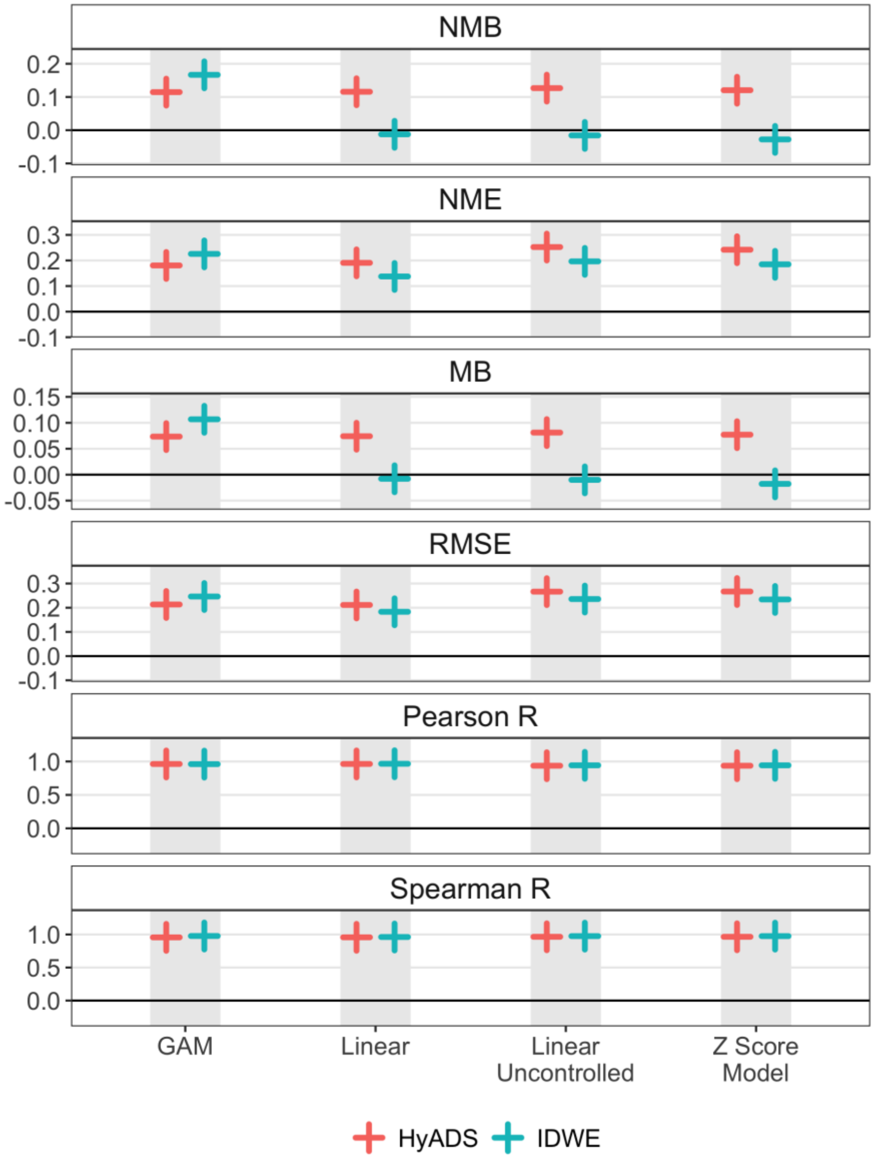

Figure SI-1: Evaluation statistics for total annual coal PM2.5 source impacts $PM_{2.5}^{m}$ evaluated against $PM_{2.5}^{CMAQ-DDM}$.

SI-1.2 Monthly evaluation

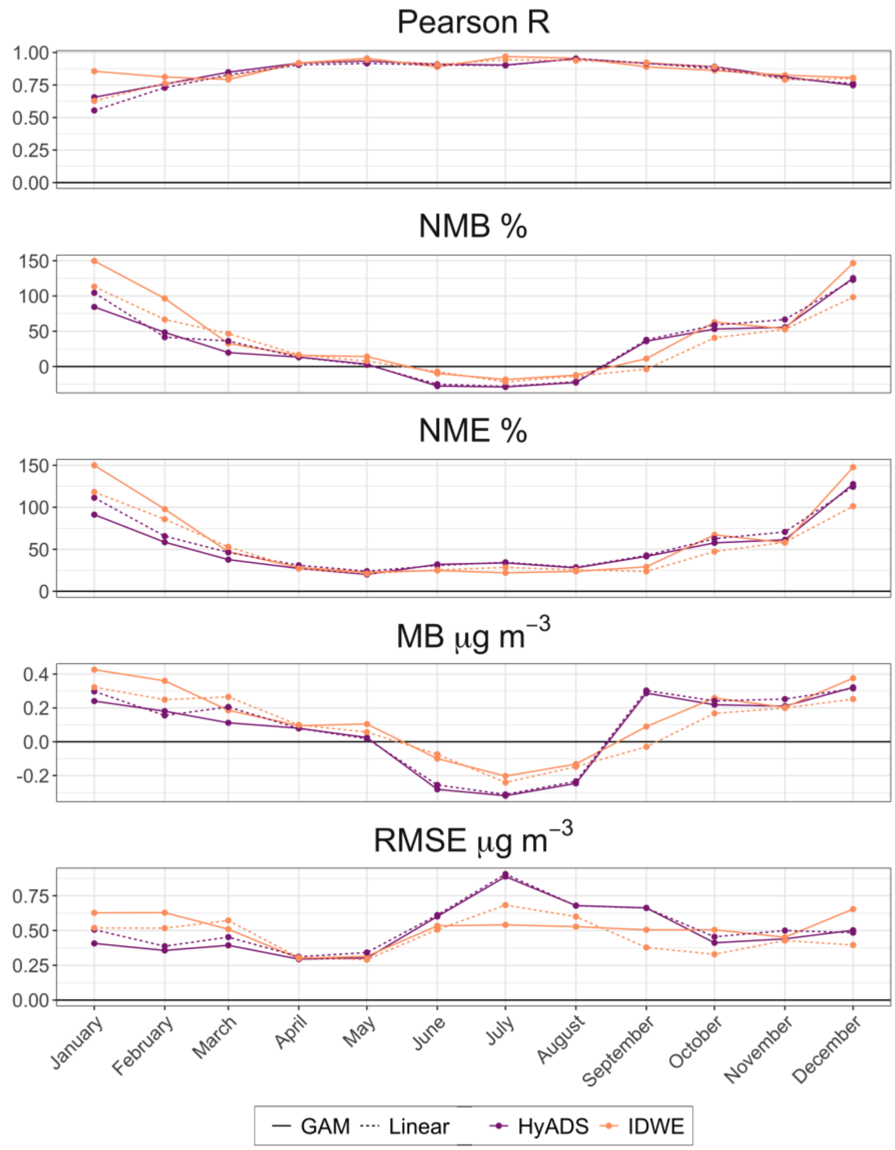

Figure SI-2: Evaluation statistics for total monthly coal PM2.5 source impacts $PM_{2.5}^{m}$ evaluated against $PM_{2.5}^{CMAQ-DDM}$. Models were trained in each month in 2005 and evaluated in 2006.

SI-1.3 Total source impact fields as PM2.5

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 $exposure^{IDWE}$ and $PM_{2.5}^{CMAQ-DDM}$ yielded lower bias and error than comparable results for $PM_{2.5}^{HyADS}$.

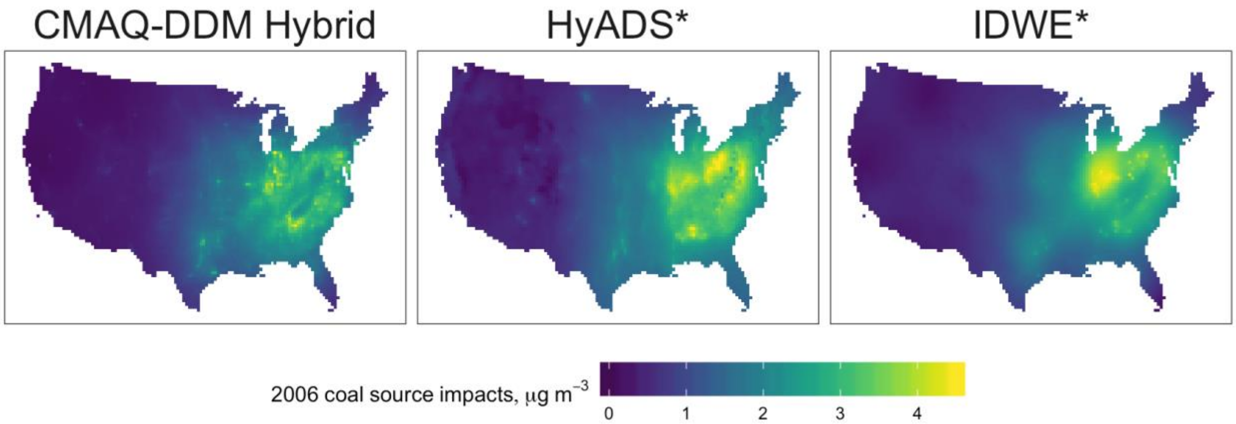

Figure SI-3: Total annual $PM_{2.5}^{CMAQ-DDM}$, $PM_{2.5}^{HyADS}$, and $PM_{2.5}^{IDWE}$ in 2006. * denotes converted metrics from $exposure^{HyADS}$ and $exposure^{IDWE}$.

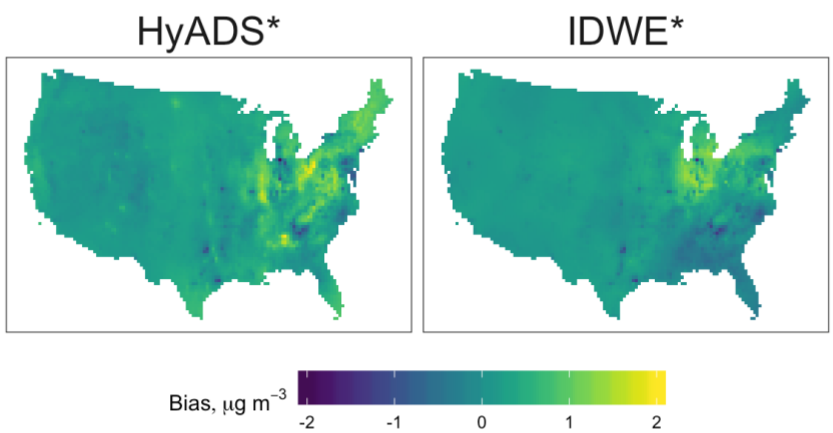

Figure SI-4: Spatial bias of total annual $PM_{2.5}^{HyADS}$ and $PM_{2.5}^{IDWE}$ relative to $PM_{2.5}^{CMAQ-DDM}$ in 2006. * denotes converted metrics from $exposure^{HyADS}$ and $exposure^{IDWE}$.

SI-2 Additional supplemental figures

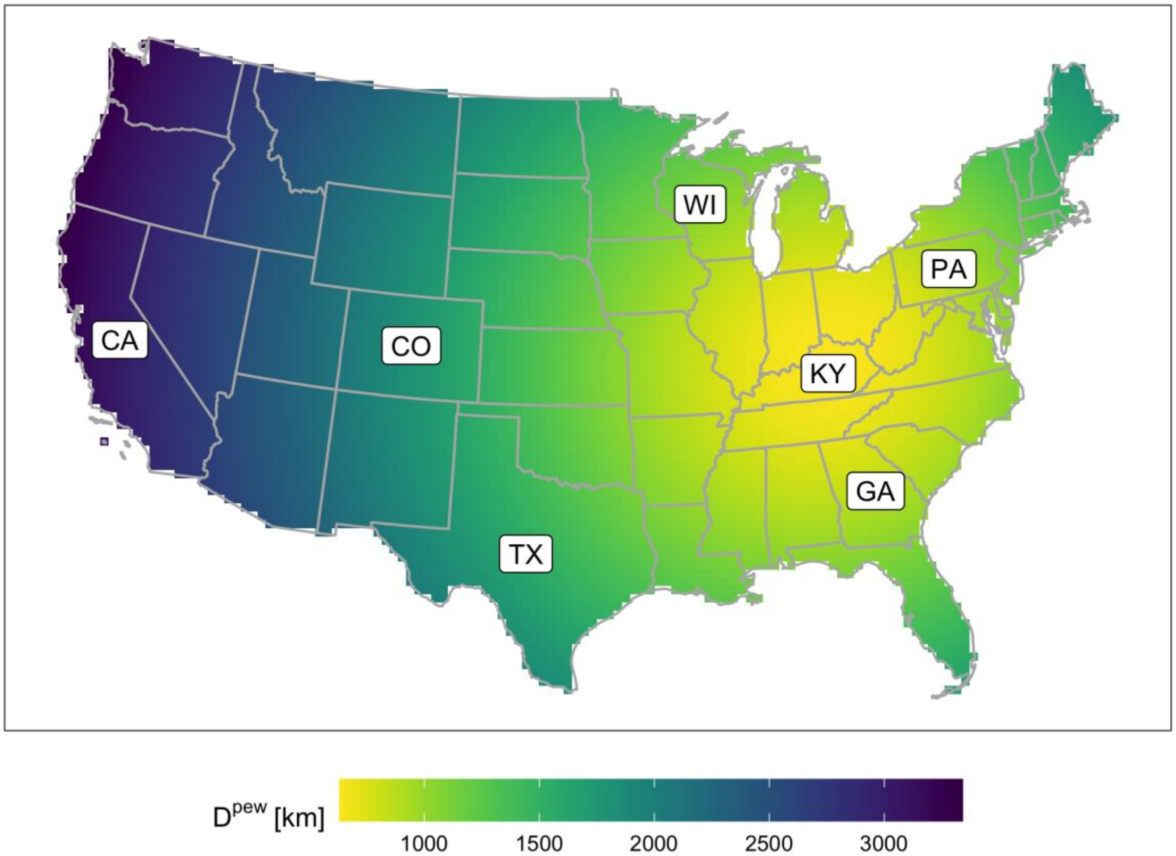

Figure SI-5: Population-emissions weighted distance (D_{pew}) calculated for each grid cell in the contiguous United States.

SI-2 Source impact evaluation metrics

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

SI-2.1 Annual evaluations

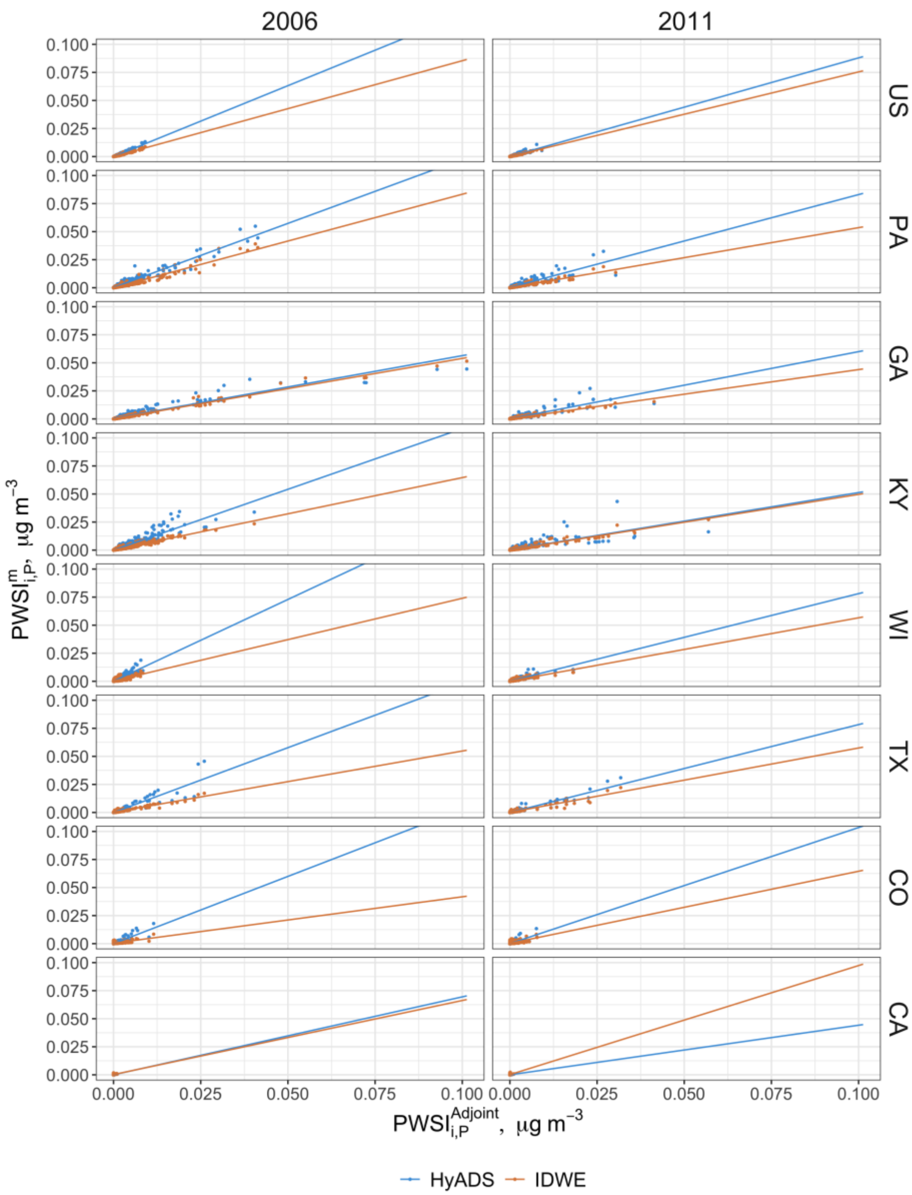

Figure SI-6: Scatterplot of $\mathrm{PWSI}_{i,P}^{HyADS}$ and $\mathrm{PWSI}_{i,P}^{IDWE}$ against $\mathrm{PWSI}_{i,P}^{Adjoint}$ for each coal-fired power plant.

Figure SI-7: Spearman R (rank-ordered correlation), Normalized Mean Error (0% < NME < +∞) and Mean Bias (MB) of $PWSI_{i,P}^{HyADS}$ and $PWSI_{i,P}^{IDWE}$ 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

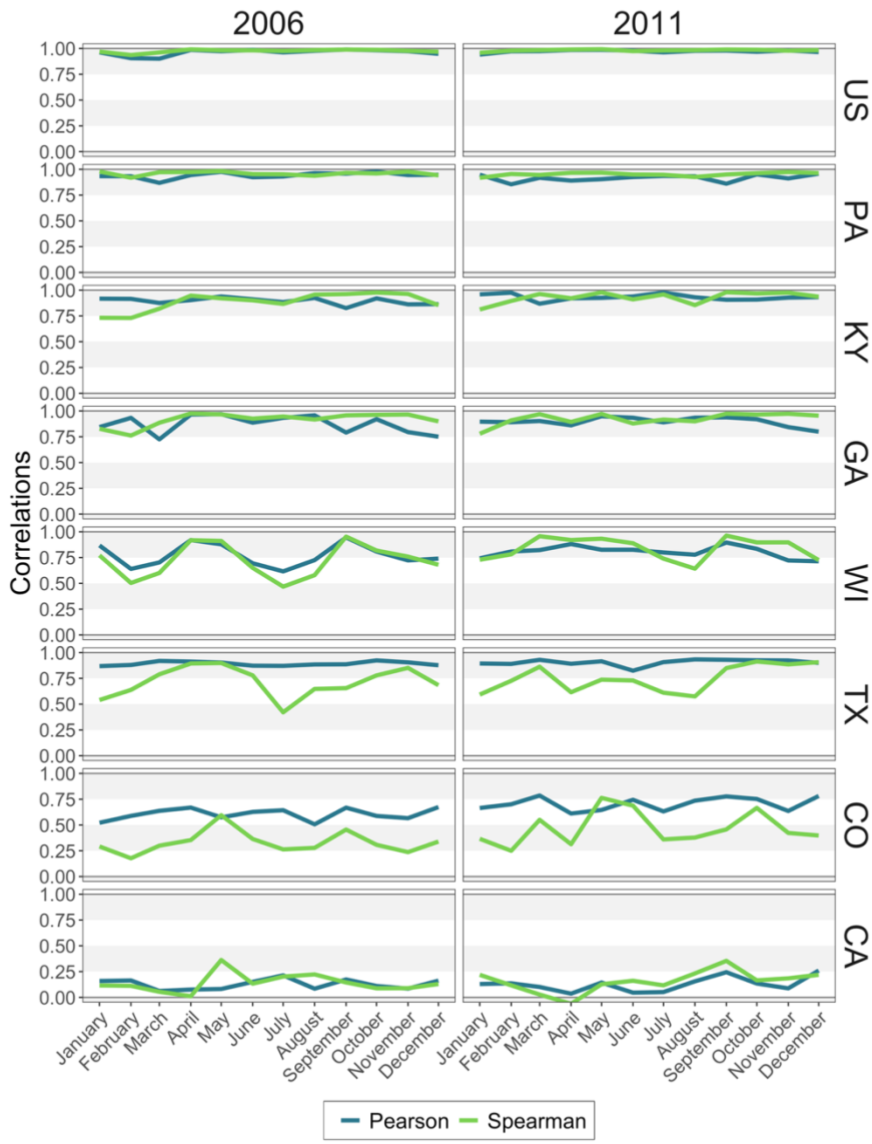

Figure SI-8: Monthly linear (Pearson R) and rank-ordered (Spearman R) correlations between $\mathrm{PWSI}_{\mathrm{P,j}}^{\mathrm{HyADS}}$ and $\mathrm{PWSI}_{\mathrm{P,j}}^{\mathrm{IDWE}}$ source impacts evaluated against $\mathrm{PWSI}_{\mathrm{P,j}}^{\mathrm{Adjoint}}$ on individual states and entire United States (US). States are ordered east to west descending.

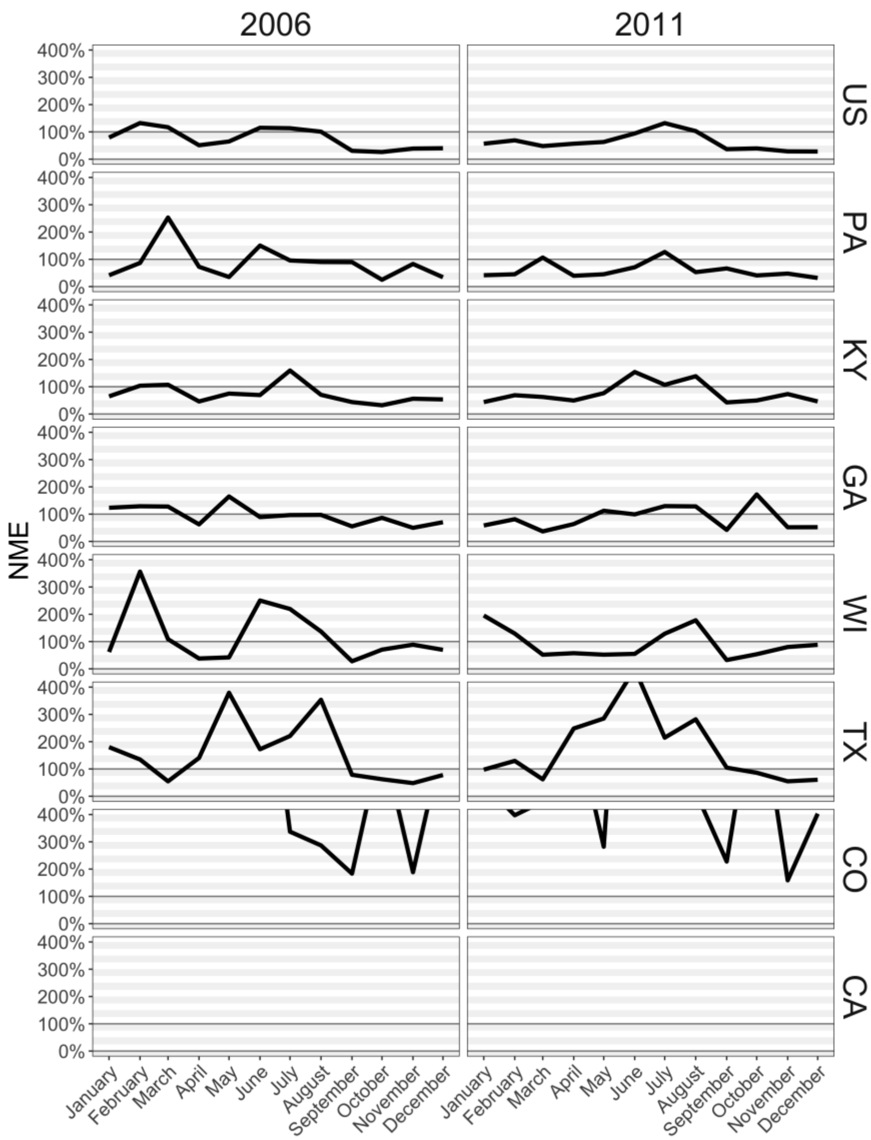

Figure SI-9: Normalized Mean Error (0% < NME < +∞) of $PWSI_{P,j}^{IDWE}$ evaluated against $PWSI_{P,j}^{HyADS}$. The values in Colorado (CO) range up to 18,000% and in California range from 800% to greater than 2,000,000%.

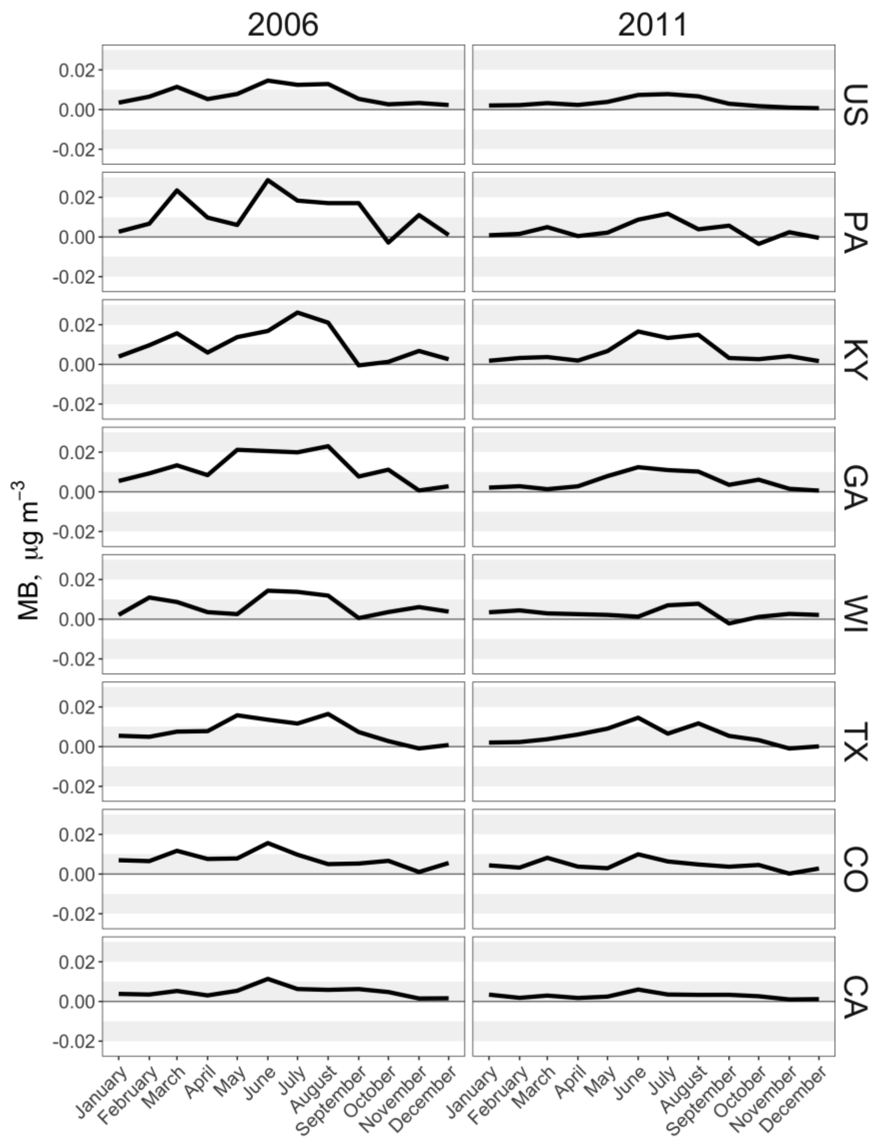

Figure SI-10: Mean bias (MB) of $PWSI_{P,j}^{IDWE}$ evaluated against $PWSI_{P,j}^{HyADS}$.

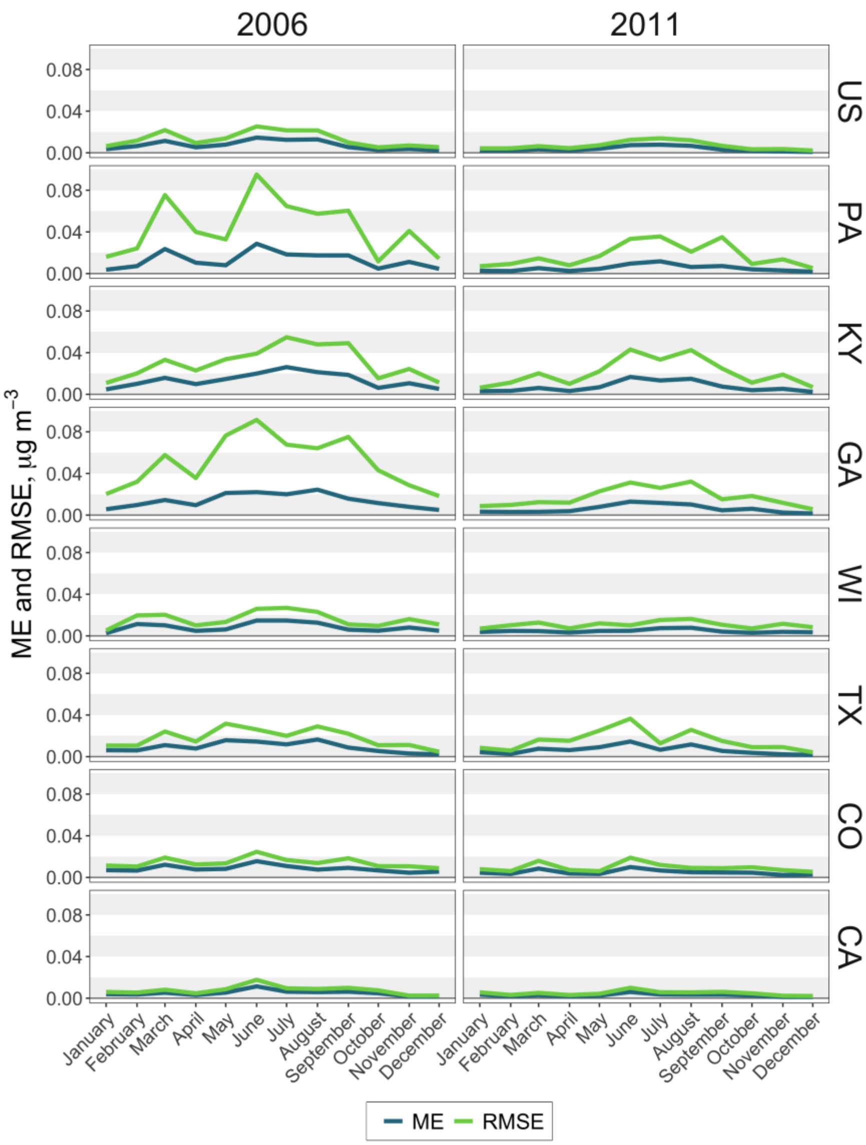

Figure SI-11: Mean error (ME) and root mean square error (RMSE) of $PWSI_{P,j}^{IDWE}$ evaluated against $PWSI_{P,j}^{HyADS}$.

Figure SI-12: Linear (Pearson R) and rank-order (Spearman R) correlations of raw HyADS and IDWE individual source exposure metrics ($exposure_{i,j}^{HyADS}$ and $exposure_{i,j}^{IDWE}$) compared to $\mathrm{PWSI}_{\mathrm{P,j}}^{Adjoint}$.