Supplement for "Comparison of multiple $PM_{2.5}$ exposure products for estimating health benefits of emission controls over New York State, USA"

Jin et al.

1 Methods for data comparison

Multi-product comparison

We collected seven gridded $PM_{2.5}$ products, which give 21 pairs of products for comparison. For each data pair, we first calculate the spatial correlation coefficient (R_S) and root mean squared difference ($RMSD_S$) on the 11-year average and annual average $PM_{2.5}$ from 2002 to 2012. We re-gridded all products to a common grid of 0.1˚ × 0.1˚ resolution. We apply linear interpolation for those products with coarse resolution (i.e. CMAQ, FAQSD, CDC WONDER). For the Dalhousie and Emory products, whose resolutions are an order of magnitude finer than the targeted resolution, we also calculate the average of all grid cells falling in the given coarse grid cell. Compared to the linear interpolation approach, we find that this averaging approach shows a smoother distribution of $PM_{2.5}$, and the resulting gridded product (especially for the Emory product) shows a higher spatial correlation with the coarse products. Next, we calculate the temporal correlation coefficient (R_T) and $RMSD_T$ at monthly scales for both the state average and each grid cell at 0.1˚ resolution.

Comparison with ground-based observations (AQS, SRMT, NYCCAS)

We sample the products that are available daily by matching the spatial coordinates and the date of each daily ground-based observation. The daily average ground-based observation at each SRMT site is calculated from hourly data. For comparison with NYCCAS data that are available as two-week averages, we sample the daily $PM_{2.5}$ products for each NYCCAS period, and calculate the two-week average. We then construct monthly, annual and 11-year averages from the sampled $PM_{2.5}$ data for comparison with the ground-based observations to avoid the discrepancies introduced by limited versus continuous sampling. For comparison with $PM_{2.5\_Dal\_NA}$ (or $PM_{2.5\_Dal\_GL}$), which are only available at monthly (or annual) resolution, we calculate monthly (or annual) averages of the ground-based observations, and then sample $PM_{2.5\_Dal\_NA}$ consistently.

Spatial RMSD ($RMSD_S$):

$$RMSD_S = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(Conc_{xi} - Conc_{yi})^2} \quad \text{(S1)}$$

where $Conc_{xi}$ and $Conc_{yi}$ are annual average (or multi-year average) $PM_{2.5}$ (μg/m^3) estimated from product *x* and product *y* for grid cell (or site) *i*; *N* is the total number of grid cells (or sites).

Population weighted spatial RMSD (PW $RMSD_S$):

$$RMSD_S = \sqrt{\frac{\sum_{i=1}^{N}(Conc_{xi} - Conc_{yi})^2 \times pop_i}{\sum_{i=1}^{N} pop_i}} \quad \text{(S2)}$$

where $Conc_{xi}$ and $Conc_{yi}$ are annual average (or multi-year average) $PM_{2.5}$ (μg/m^3) estimated from product *x* and product *y* for grid cell *i*; pop_i is population density at grid cell *i*; *N* is the total number of grid cells (or sites).

Temporal RMSD ($RMSD_T$):

$$RMSD_T = \sqrt{\frac{1}{M}\sum_{t=1}^{M}(Conc_{xt} - Conc_{yt})^2} \quad \text{(S3)}$$

where $Conc_{xt}$ and $Conc_{yt}$ are monthly average $PM_{2.5}$ (μg/m^3) estimated from product *x* and product *y* for time *t*; *M* is the total number of months for the comparison period (132 months at most).

Spatial Pearson Correlation Coefficient (R_S):

$$R_S = \frac{\sum_{i=1}^{N}(Conc_{x_i} - \overline{Conc_x})(Conc_{yi} - \overline{Conc_y})}{\sqrt{\sum_{i=1}^{N}(Conc_{x_i} - \overline{Conc_x})^2 \sum_{i=1}^{N}(Conc_{yi} - \overline{Conc_y})^2}} \quad \text{(S4)}$$

where $Conc_{xi}$ and $Conc_{yi}$ are annual average (or multi-year average) $PM_{2.5}$ ($\mu g/m^3$) estimated from product *x* and product *y* at grid cell (or site) *i* (the overbar indicates domain average); *N* is the total number of grid cells (or sites).

Temporal Pearson Correlation Coefficient (R_T):

$$R_T = \frac{\sum_{t=1}^{M}(Conc_{xt} - \overline{Conc_x})(Conc_{yt} - \overline{Conc_y})}{\sqrt{\sum_{t=1}^{M}(Conc_{xt} - \overline{Conc_x})^2 \sum_{t=1}^{M}(Conc_{yt} - \overline{Conc_y})^2}} \quad \text{(S5)}$$

where $Conc_{xt}$ and $Conc_{yt}$ are monthly average $PM_{2.5}$ ($\mu g/m^3$) estimated from product *x* and product *y* for time *t* (the overbar indicates temporal average) averaged for either the state or a single grid cell; *M* is the total number of months for the comparison period (132 months at most).

2 Characterizing uncertainty

Uncertainty in $PM_{2.5}$ estimate:

We define two metrics to characterize the variations in $PM_{2.5}$ across multiple products: the normalized range (NR) and the uncertainty (δ_{PM}). *NR* describes the spread of $PM_{2.5}$ across all products:

$$NR = \frac{1}{M}\sum_{t=1}^{M}\left(\frac{\max_{k\in K} C_k - \min_{k\in K} C_k}{\bar{C}}\right) \quad \text{(S6)}$$

where *C* is the quantity to be evaluated (e.g. NYS average $PM_{2.5}$, PWA $PM_{2.5}$, annual mortality burden); *k* is the product number; K is the total number of products; the ensemble maximum, minimum and mean ($\bar{C}$) are evaluated by comparing across different products at time *t*; M is the total number of time periods.

For a small sample size (K = 7), we assume the variations in $PM_{2.5}$ across multiple products follows the t statistical distribution with the mean being the ensemble average. The confidence interval (CI) for the ensemble mean at a given time *t* is calculated as:

$$CI_t = \bar{C} \pm t^* \frac{SD_t}{\sqrt{K}} \quad \text{(S7)}$$

where $\bar{C}$ is the ensemble average of the quantity to be evaluated at time t; t^* is the upper (1-CI)/2 critical value for the t distribution with K-1 degrees of freedom. For K = 7, t^* for the 95% double tailed confidence level is 2.45. SD_t is the sample standard deviation at time t:

$$SD_t = \sqrt{\frac{\sum_{k=1}^{K}(C_{k,t} - \bar{C_t})^2}{K-1}} \qquad \text{(S8)}$$

We define an overall estimate of uncertainty (δ_{PM}) as follows:

$$\delta_{PM} = \frac{2}{M}\sum_{t=1}^{M}\left(t^* \frac{SD_t}{\sqrt{K}\bar{C}_t}\right) \qquad \text{(S9)}$$

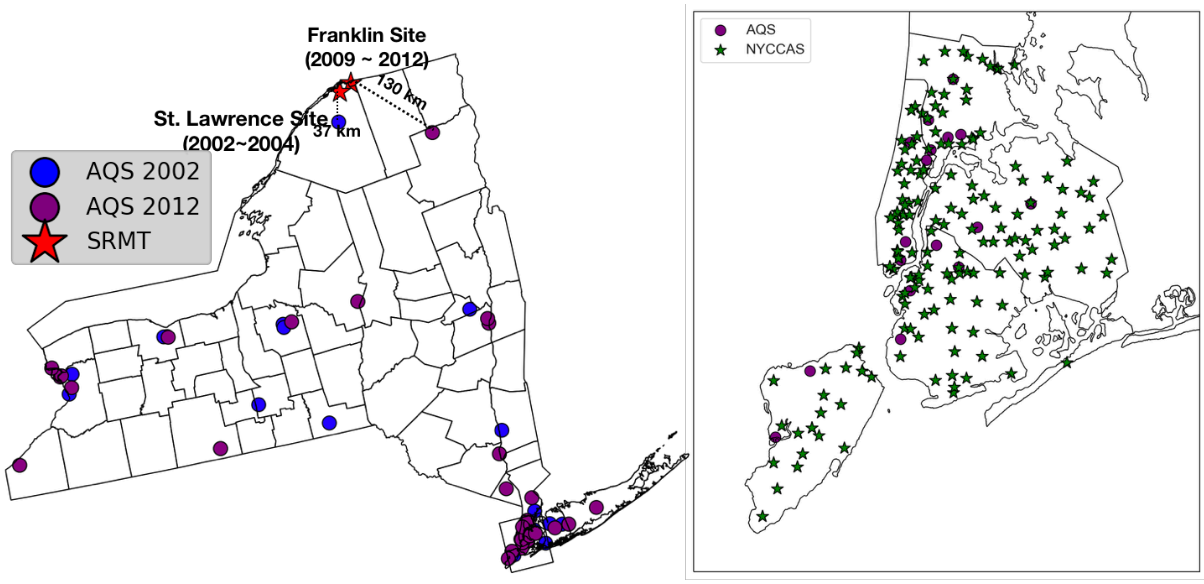
Uncertainty in exposure response function:

We use the 95% CI of the relative risk factors provided by the Global Burden of Disease Collaborative Network as a measure of the uncertainty in exposure-response function. The integrated exposure-response function relies on pooling relative risk factors from the available literature. The integrated exposure-response function is subject to uncertainties in the function shape, the counterfactual concentration (the level below which no additional risk is assumed), and the exposure estimate of $PM_{2.5}$ (Burnett *et al* 2014). The uncertainty bounds are estimated through 1000 realizations of the relative risk factors assuming a normal distribution (Burnett *et al* 2014). We define an overall uncertainty in the mortality burden attributed to uncertainty in the exposure response function (δ_{ER}) as follows:

$$\delta_{ER} = \frac{1}{M}\sum_{t=1}^{M}\left(\frac{\Delta Mort_{upper,t} - \Delta Mort_{lower,t}}{\Delta Mort_t}\right) \qquad \text{(S10)}$$

where $\Delta Mort_t$, $\Delta Mort_{upper,t}$ and $\Delta Mort_{lower,t}$ are the excess mortality burden at year t calculated using the relative risk factor and its upper and lower limits of the 95% CI.

Supplementary Figures



85

Figure S1 Locations of AQS (circles), SRMT (red stars), and NYCCAS monitors (green stars) over NYS (left) and NYC (right).

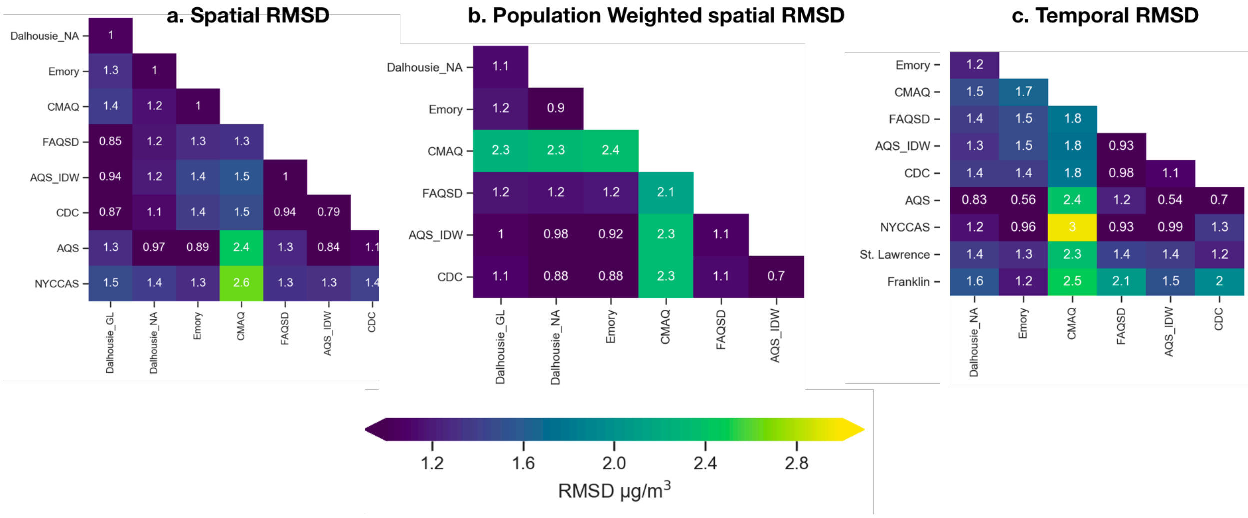


Figure S2 Spatial (a, equation (S1)), population weighed spatial (b, equation (S2)) and temporal RMSD (c, equation (S3)) for different pairs of $PM_{2.5}$ data.

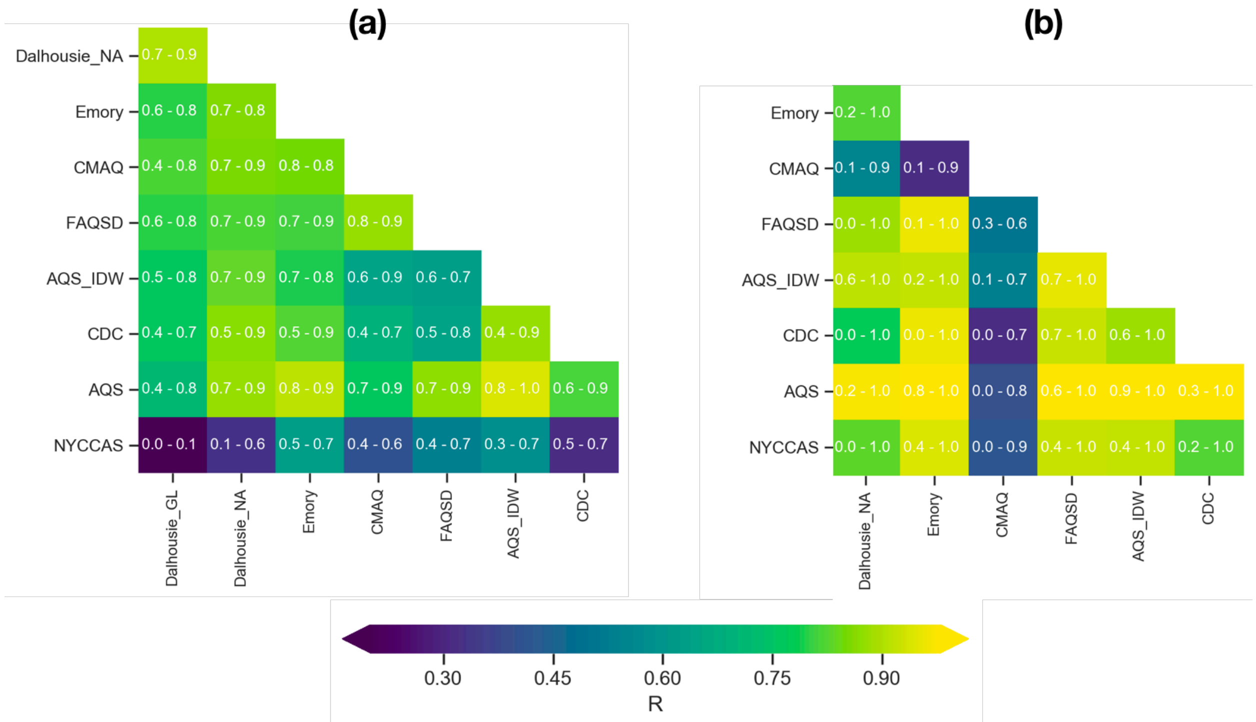


Figure S3 (a) Minimum (to the left of the dash in each square) and maximum (to the right of the dash in each square) spatial correlation coefficients (R_s) of annual average $PM_{2.5}$ from 2002 to 2012. (b) Minimum and maximum temporal correlation coefficients (R_T) for each grid cell at 0.1˚ resolution (for AQS and NYCCAS data, statistics correspond to each monitoring site).

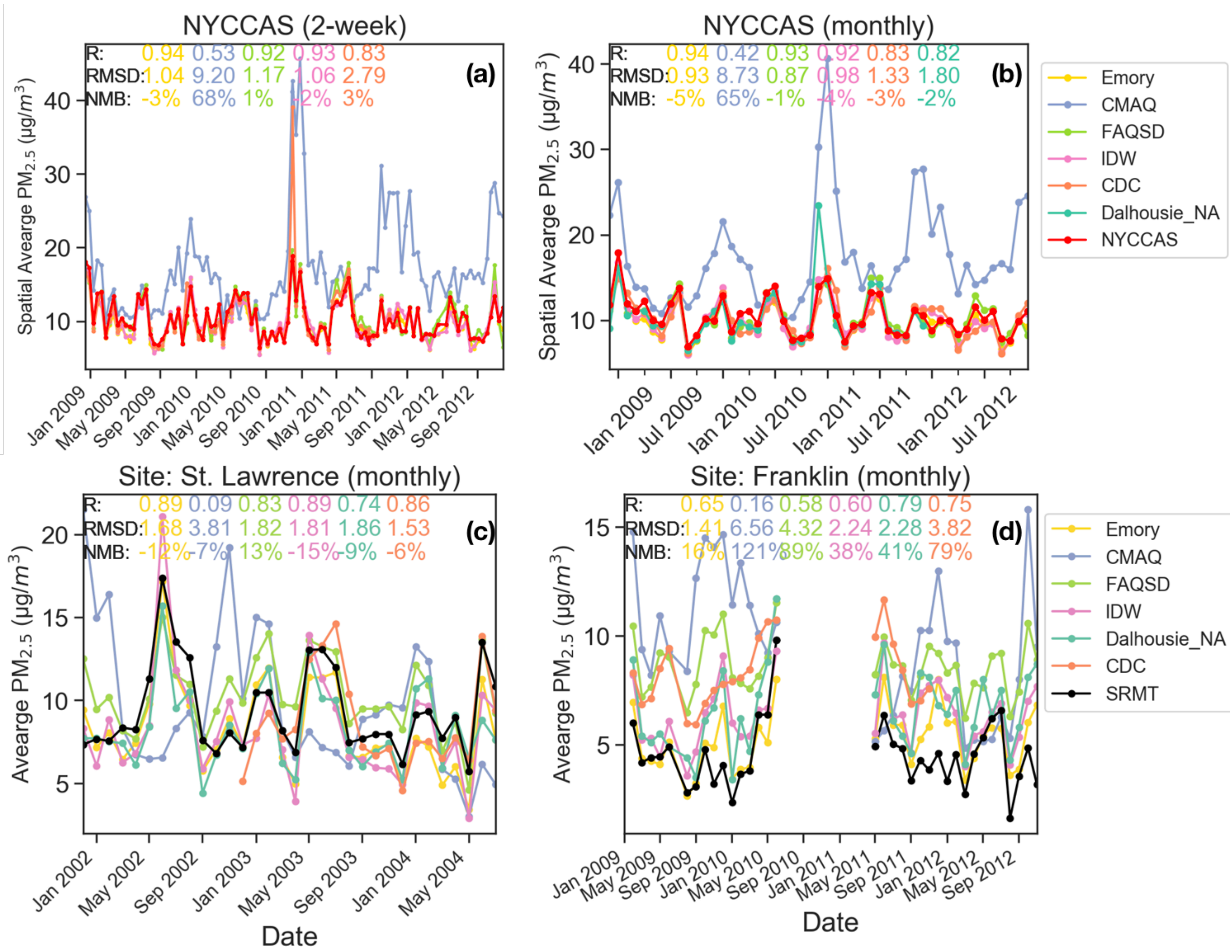


Figure S4 Comparison of (a) 2-week average, (b) monthly average $PM_{2.5}$ from multiple $PM_{2.5}$
100 products versus $PM_{2.5\_CAS}$ averaged across all sites; monthly average $PM_{2.5}$ from multiple $PM_{2.5}$ products versus $PM_{2.5\_SRMT}$ at the St. Lawrence and Franklin sites.

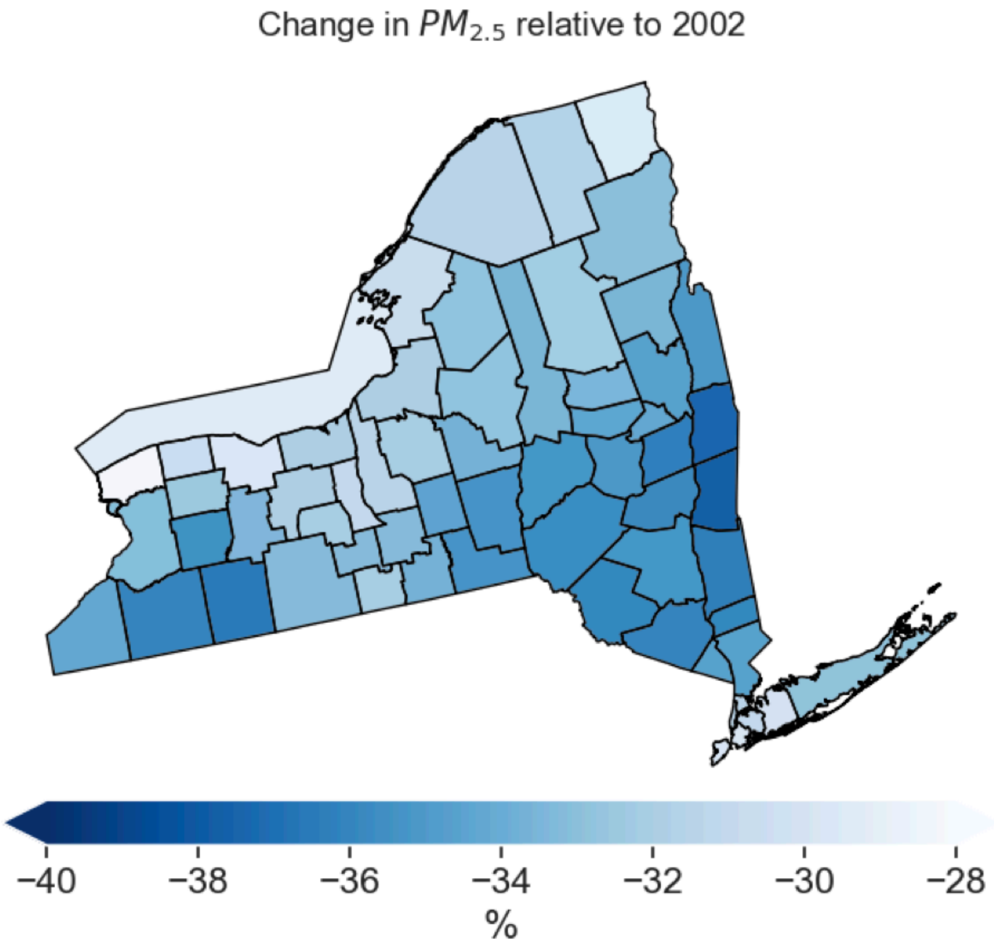


Figure S5 Change in ensemble mean $PM_{2.5}$ in 2012 relative to 2002 in each county over NYS.

105

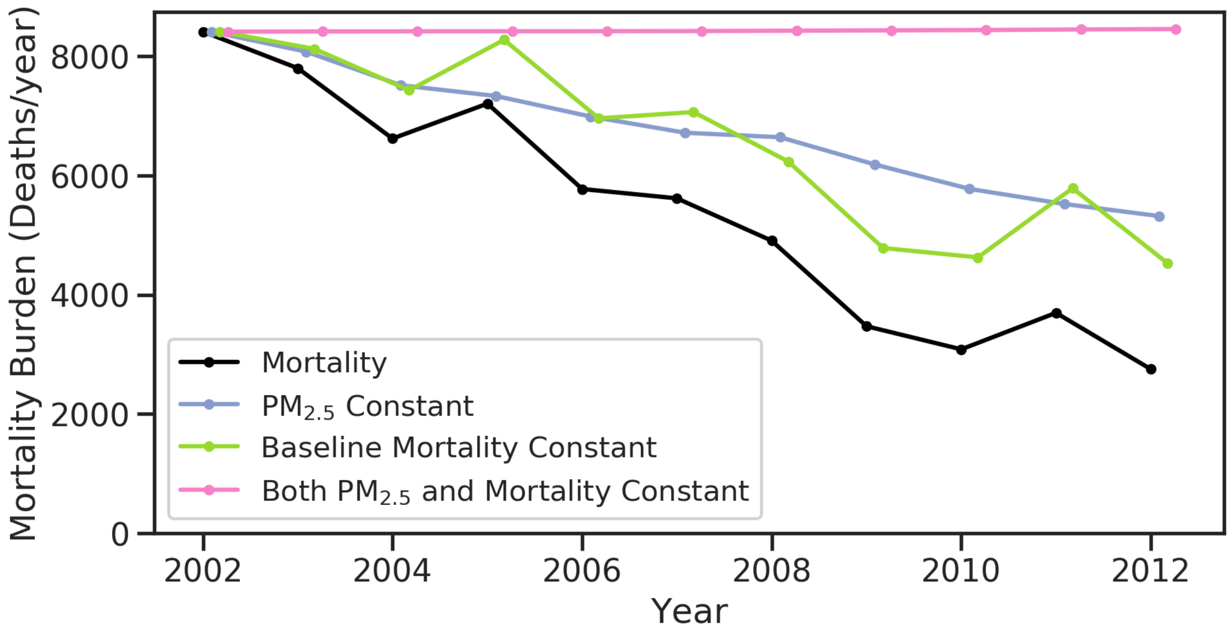


Figure S6 Trends in the ensemble mean annual NYS $PM_{2.5}$-related mortality burden (black), the mortality burden with $PM_{2.5}$ concentration kept constant at the 2002 level (blue), the mortality burden with baseline mortality kept constant at the 2002 level (green), and the mortality burden with both $PM_{2.5}$ concentration and baseline mortality kept constant (pink).

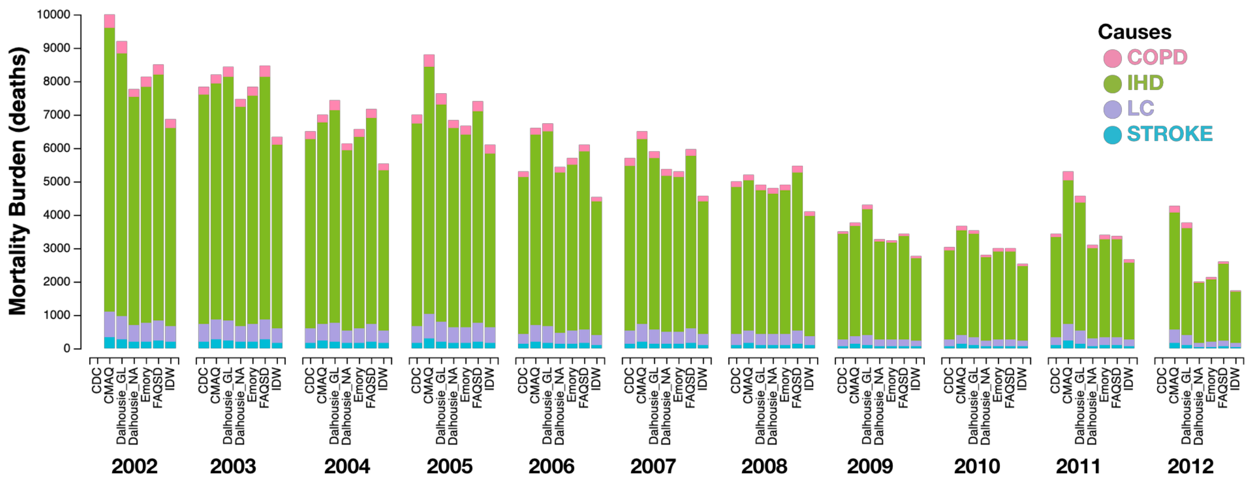
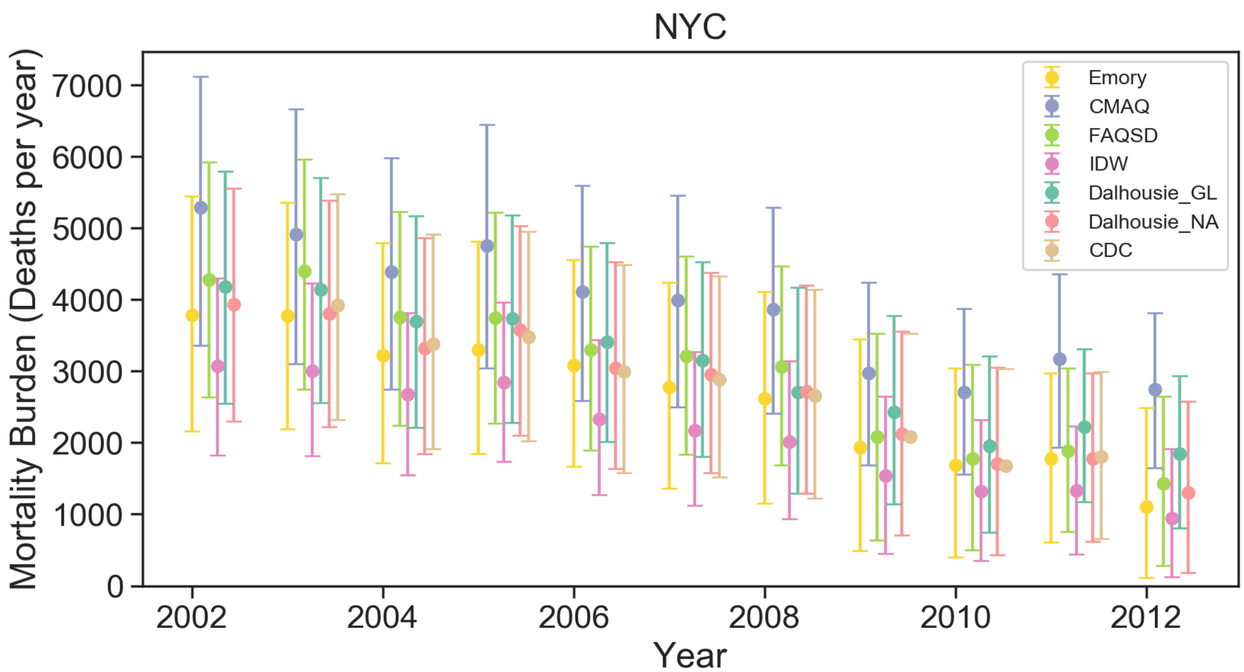


Figure S7 Annual $PM_{2.5}$-related mortality burden by causes (COPD, IHD, LC, STROKE) from 2002 to 2012 using multiple $PM_{2.5}$ products over NYS.



120 *Figure S8* Same as Figure 3(a) but for New York City (including New York, Bronx, Kings, Queens and Richmond counties).

Table S1 Summary of normalized uncertainties (i.e. NR and δ_{PM}) over NYS and NYC at different temporal scales. The numbers in parenthesis are estimated uncertainties that remove outlier product (in which one or two products lead to >10% increase in NR or δ_{PM}).

Quantity to be evaluated	NYS		NYC[e]	
	NR	δ_{PM}	NR	δ_{PM}
Uncertainty of Daily Average	55% (38% [a])	42%	61% (21% [a])	50% (18% [a])
Uncertainty of Monthly Average $PM_{2.5}$	43% (33% [a])	32%	50% (17% [a])	36% (12% [a])
Uncertainty of Annual Average $PM_{2.5}$	30%	22%	50% (14% [a])	32% (10%[a])
Uncertainty of Annual Population Weighed Average (PWA) $PM_{2.5}$	44% (10% [ab])	26% (8% [ab])	77% (34% [a])	44% (22%[a])
Uncertainty of Relative Change in Annual Average between 2002 and 2012	28% (12% [cd])	24%	46% (28% [a])	34% (20%[a])
Uncertainty of Relative Change in PWA $PM_{2.5}$ between 2002 and 2012	31% (18% [a])	20%	53% (33% [a])	38% (24%[a])
Uncertainty of Premature Mortality Burden due to choice of $PM_{2.5}$ products	43% (27% [b])	28%	66% (39% [ab])	38% (14% [ab])
Uncertainty of Changes in Premature Mortality Burden due to choice of $PM_{2.5}$ products	26%	20%	36% (22% [a])	26% (16% [a])

a. $PM_{2.5\_CMAQ}$ removed
b. $PM_{2.5\_IDW}$ removed
c. $PM_{2.5\_FAQSD}$ removed
d. $PM_{2.5\_Dal\_NA}$ removed
e. New York City includes New York, Bronx, Kings, Queens and Richmond counties.

Table S2 Qualitative summary of the strengths and limitations of each $PM_{2.5}$ product in terms of the accuracy (for both urban and remote environments), availability (i.e. spatial and temporal coverage) and resolution. The product is qualitatively assessed on a scale of 1 to 5 stars, with a 5-star being the best among all the products. Evaluation of accuracy is based on comparison with independent observations (Section 3.2). The evaluation of the availability and resolution is based on the original spatial and temporal coverage or resolution of the product (table 1). The products with the highest resolution/availability among these products are rated with 5 stars.

	Accuracy		Availability		Resolution	
	Urban	Remote	Spatial	Temporal	Spatial	Temporal
Global Geophysical Satellite-Based $PM_{2.5}$ (Dalhousie_GL)	★★•••	★★★★★	★★★★★	★★★★•	★★★★★	★••••
North America Geophysical Satellite-Based PM2.5 (Dalhousie_NA)	★★★••	★★★★★	★★★★•	★★★★•	★★★★★	★★•••
Statistical Satellite-Based $PM_{2.5}$ (Emory)	★★★★★	★★★★•	★••••	★★★••	★★★★★	★★★★•
AQS and remote sensing merged $PM_{2.5}$ (CDC WONDER)	★★★••	★★★••	★★★★•	★★•••	★★★••	★★★★•
Fused Air Quality Surface using Downscaling (FAQSD)	★★★★•	★★•••	★★★••	★★★••	★★★••	★★★★•
CMAQ	★••••	★••••	★★★★★	★★★★•	★★★••	★★★★★
Inverse Distance Weighed $PM_{2.5}$ (IDW)[a]	★★★★•	★★•••	★★•••	★★★★★	★••••	★★★••

a. The IDW data is given 1-star for spatial resolution because in effect the level of spatial detail is determined only by the density of AQS observations.