Appendix

Section S1:

Fire emission inventories

FINN aggregates 1-km MODIS active fire detections and produces a daily emission estimate given these detections and MODIS retrieved land cover (Wiedinmyer et al., 2011). GFED4.1s relies primarily on monthly MODIS MCD64A1 500-m burned area maps, derived from observed changes in surface reflectance, to generate emission estimates. The dataset then adds 1-km active fire information to incorporate the influence of small fires that may have not been accounted for by the burned area product (Randerson et al., 2012). GFED4.1s derives a daily emission estimate by applying the ratio of daily to total monthly active fire counts in each grid cell. GFAS also aggregates the 1-km MODIS FRP and uses emission factors similar to those in GFED4.1s. To account for FRP obscured by sub-grid clouds and other interferences, GFAS assumes that an obscured FRP pixel is equivalent in value to its adjacent non-obscured pixel, as long as they are not over a body of water (Kaiser et al., 2012). GFAS further uses a Kalman filtering method of data assimilation, in which the optimal estimate of FRP for a given day is a weighted average of the optimal FRP estimate from the previous day and the FRP estimate for the current day (Kaiser et al., 2012). Like GFAS, QFED uses information from adjacent pixels to estimate obscured thermal anomalies; QFED also relies on FRP estimates from the previous day. However, the QFED algorithm weights adjacent pixel information via the error covariance between pixels, and it allows the estimate of the previous day's FRP to decay according to a characteristic timescale derived for each land type (Darmenov and da Silva, 2015). GFAS, QFED, and FINN are available in near real time, whereas GFED4.1s requires several months of processing before public release. We include another inventory, called GFED+Agriculture, where increase the GFED4.1s emission factors associated with agricultural burning by a factor of three. The factor of three scaling is based on the laboratory findings of Oanh et al. (2010), who found that the particulate matter emissions of rice straw approximately tripled when the straw was piled instead of spread evenly on the ground. In reality, we expect a mixture of partial and whole field burning (Vadrevu et al., 2008), so the GFED + Agriculture emissions represent a high derived upper bound on agricultural burning in the region.

Section S2:

STILT model

STILT traces an ensemble of theoretical particles or air-mass trajectories from a receptor site backwards in time and computes the sensitivity of $PM_{2.5}$ concentration at the receptor to emissions in the surrounding region. The resulting flux footprint reveals those regions where emissions likely influenced $PM_{2.5}$ at the receptor. We simulate daily footprints of the sensitivities of Delhi pollution to fire emissions upwind by sending 500 simultaneous air-mass trajectories backwards in time for 5 days. We choose 500 ensembles in order to account for random turbulence air-masses experience, especially in the boundary layer (Lin et al., 2003). We choose five days as this timeframe should allow an air mass to traverse the approximately 800 km between Delhi and the farthest upwind burning regions even under the weak wind conditions prevalent at this time of year, which is often less than 5 m s^{-1} according to GDAS.

Liu et al. (2018) performed a back-trajectory analysis using the Hybrid Single Particle Lagrangian Integrated Trajectory Model (HYSPLIT; Stein et al., 2015) to create an airshed region upwind of Delhi. The boundaries of this airshed determined a region where agricultural fire emissions could potentially influence downwind air pollution in Delhi. Through the use of STILT, we make this relationship more explicit by quantifying explicitly how much those upwind emissions contribute to a particular downwind pollution observation. In other words, each STILT footprint can be coupled to an emissions inventory in order to simulate surface $PM_{2.5}$ concentrations. The footprint for the i^{th} receptor location and time can be expressed as a vector $\mathbf{k}_i = (\partial y_{stilt,i} / \partial \mathbf{x})^T$, where $\mathbf{x}$ is a vector of upwind emissions from the previous 5 days (units of μmol m^{-2} s^{-1}), and $y_{stilt,i}$ is the modeled $PM_{2.5}$ enhancement due to those emissions. If we couple this footprint to an emissions estimate (e.g., FINN, QFED, etc.) from the previous 5 days, we can simulate surface $PM_{2.5}$ enhancement from fires using the relation $y_{stilt,i} = \mathbf{k}_i \bullet \mathbf{x}$. These simulated surface concentrations can then be compared to the observed network average of $PM_{2.5}$ observations.

GEOS-Chem

GEOS-Chem is here driven by assimilated meteorological data from the Goddard Earth Observing System (GEOS-5) at the NASA Global Modeling Assimilation Office (GMAO). The aerosol simulation in GEOS-Chem includes sulfate, nitrate, ammonium, dust, and black and organic carbon (Kim et al., 2015), and many previous studies have examined $PM_{2.5}$ pollution in Asia using GEOS-Chem (e.g., Wang et al., 2013; Mu and Liao, 2014; Geng et al., 2015). Here we utilize the emission inventory for Model Inter-Comparison Study for Asia (MIX) for anthropogenic aerosol precursor emissions (Li et al., 2015). We follow Bond et al. (2007) for anthropogenic emissions of primary black and organic carbon. For this study, we perform nested grid simulations for the 2012 burning season at 0.50° x 0.67° resolution over most of eastern Asia, with lateral boundary conditions provided by a global simulation at 2.0° x 2.5° horizontal resolution.

Section S3:

Statistical modeling

Our statistical prediction of surface $PM_{2.5}$ from fires takes the form:

$$y_{stilt,i} = \mathbf{h}_i \bullet \mathbf{w}$$

where $\mathbf{h}_i$ is a *1xd* vector consisting of meteorological parameters and the STILT- driven $PM_{2.5}$ prediction ($y_{stilt,i}$), and $\mathbf{w}$ is a *dx1* vector of coefficient weights that represent the relative importance of each predictor in $\mathbf{h}_i$ to the prediction of $PM_{2.5}$. The optimal value of these weights is solved for empirically. For example, if we aggregate all daily observed network-averaged surface observations above the anthropogenic baseline ($\mathbf{y}_{obs}$), the traditional ordinary least square setup determines the optimal value of coefficient weights ($\mathbf{w}$) by the following relation:

$$\mathbf{w}^* = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{y}_{obs}$$

where **H** is an *nxd* matrix and *n* is the number of observations. Each column of **H** represents the time series of daily mean values of a particular predictor. To avoid overfitting in solving for **w***, we follow the method of the least absolute shrinkage and selection operator (LASSO; Tibshirani, 1996), which reduces the magnitude of the coefficients of correlated predictors and those predictors offering little information. Here the optimal coefficients are determined through the following algorithm:

$$\mathbf{w}^* = \min_{\mathbf{w} \in \mathbb{R}^D} \{\|\mathbf{H}\mathbf{w} - \boldsymbol{y}_{obs}\|_2^2 + \lambda\|\mathbf{w}\|_1\}$$

In the above equation, the first term on the right-hand side of the equation penalizes mismatch between model and observations using the square loss function, hence the "2" subscript. The second term of the equation regularizes the fit (i.e., reduces overfitting) by penalizing the magnitude of **w** via the absolute loss, also known as the L_1 norm, hence the "1" subscript. The algorithm is optimized over a grid of λ values (to control the degree of regularization), using three-fold cross validation. This method randomly separates the data into three sets and fits the statistical model on two of these sets, and then these fitted coefficients are applied to the remaining set (called the validation set), yielding the root mean squared error (RMSE) between the prediction ($\mathbf{y}_{stat}$) for that set and the observations ($\mathbf{y}_{obs}$). This process is repeated three times, and the value of λ that provides the best RMSE on the reserved validation sets is retained. In addition to $\mathbf{y}_{stilt}$, the array of local meteorological variables at Delhi used as predictors include wind speed and wind direction from the surface to the boundary layer and from the boundary layer to 500 hPa, as well as boundary layer height, precipitation, surface temperature, and surface pressure. All variables are taken from the Integrated Global Radiosonde Archive (Durre et al., 2006) and the Global Historical Climatology Network (Menne et al., 2012). IGRA estimates boundary layer heights over the Safdarjung airport (28.58°N, 77.2°E) using the parcel method, which locates the altitude where virtual potential temperature is equivalent to surface virtual potential temperature (Seibert et al., 2000).

Section S4:

Two step data cleaning procedure

First, we compare daily averaged CPCB $PM_{2.5}$ with corresponding MODIS AOD for each site during the burning season, using all available observations during 2012-2016. Then we select only those sites whose correlation with the AOD timeseries exceeds R=0.5 and is statistically significant ($p < 0.05$). The purpose of this step is to consider only those sites whose variability corresponds to regionally influenced pollution. Hence, if a surface site's daily averaged $PM_{2.5}$ correlates reasonably with the coarser 1° MODIS Deep Blue AOD retrieval, we assume that site to be sensitive to pollution from regional sources. Next, for each CPCB site that meets this correlation criterion, we calculate the mean absolute difference of daily-averaged $PM_{2.5}$ at that site compared to the network average of daily-averaged $PM_{2.5}$ at the other sites in the network. We also compute the standard deviation of that difference across the five years. This step produces a metric revealing how much daily $PM_{2.5}$ at each site tends to deviate from $PM_{2.5}$ at the other sites on average. We find that these deviations are distributed normally, so that for any given day, if the absolute difference in $PM_{2.5}$ at a particular site deviates more than ±2.5 standard deviations from the mean absolute difference associated with that site, we exclude that $PM_{2.5}$ observation from the

network average for that day. Thus for any given day, we remove from consideration those sites that either experience instrument malfunction and/or appear to be heavily influenced by strong local sources. For each station, we could alternatively consider removing observations that deviate too much from that station's mean $PM_{2.5}$ concentration during the post-monsoon burning season. However, we did not follow this approach because surface $PM_{2.5}$ varies widely during large fire episodes. For example, the Mandir Marg CPCB site recorded daily-averaged surface $PM_{2.5}$ concentrations ranging between 120 – 692 μg m^{-3} during Nov. 1-15, 2016. The higher $PM_{2.5}$ enhancements could erroneously be marked as outliers from local sources and/or instrument malfunction, when in fact all sites experienced large fluctuations in $PM_{2.5}$ during this time. Therefore, we wish to consider outliers as a function of the network average of monitors, as we expect all surface monitors to jointly respond to the regional signal of fire emissions and transport.

Section S5:

Baseline methods

Method 1: This method relies on the daily variability of the GFAS fire emissions. We choose GFAS due to its assimilation properties which account for some missing or obscured fire pixels. For each fire season, we analyze the time series of these emissions summed over all grid cells in the burning regions upwind of Delhi. We specify low-fire days as those days when total fire emissions fall below a specified threshold at the low end of the frequency distribution for that season – e.g., below the 10^{th} percentile. On days when fire emissions fall below that threshold, we assume that Punjab and Haryana are not burning significantly. If emissions remain below the threshold during the next N days, we tag the observation for that N^{th} day as representative of the baseline. The baseline is then the average of all tagged days during the fire season. We vary N between 1-5 and the emission percentile threshold between 10-30% to check the robustness of our baseline estimate. We assume that N represents the transport time for smoke from fires to ventilate out of the IGP.

Method 2: In this method, we take advantage of STILT sensitivity estimates. For each day of the fire season, STILT provides gridded sensitivities to upwind emissions for each observation in Delhi. If the map of sensitivity overlaps with cells containing fire emissions, the model predicts a pollution enhancement due to fire downwind in Delhi in the subsequent days. To compute the anthropogenic baseline, we count the number of fire emitting pixels for a particular day. We then count the number of those fire pixels that overlap with the STILT sensitivity map. If the ratio of overlapping pixels to total fire pixels is sufficiently low (e.g., less than a threshold of 0.1) on a given day, we assume that the urban pollution for that day has little influence from fires. We collect each of these non-fire days during each fire season and take their average as the baseline. We vary the ratio threshold between 0.1-0.7 to assess the sensitivity of this method to its underlying assumptions.

Method 3: For this method, we compute the weekly block average of $PM_{2.5}$ within the city for each week (Sunday through Saturday) of the burning season. We then average the M lowest weekly averages to determine the baseline. We vary M between 1-4 to check the sensitivity of this method to this parameter.

Section S6

Diwali

Though Diwali lasts a week, most firecrackers are lit on the first night of the festival (Singh et al., 2010). Without controlling for other factors, Singh et al. (2010) found PM_{10} concentrations to increase by a factor of 2-6 before and after Diwali in Delhi during 2002-2007, and found the effect to be strongest at night. In Figure 6, Nov. 3rd total $PM_{2.5}$ is observed to be especially high (338 μg m^{-3}); however, the STILT model simulations predict a small, near-zero $PM_{2.5}$ enhancement and the observed AOD is also relatively low. Thus an alternative explanation for the observed/modeled $PM_{2.5}$ mismatch on Nov. 3rd could instead be the effects of Diwali, which may not be captured in fire emission inventories and the coarser AOD product. Diwali generally occurs during the post-monsoon season, though not always during peak agricultural burning. In Figure S4, we show the post-monsoon time series of observed and modeled $PM_{2.5}$ for 2012 and 2014-2016. In 2012, Diwali occurred a week after peak burning and peak observed $PM_{2.5}$. In 2016, Diwali occurred a week before peak burning and peak observed $PM_{2.5}$. Though potentially a factor in 2013, the incongruous timing of post-monsoon burning and Diwali during the other seasons implies that observed $PM_{2.5}$ results from Figure 5 and Table 2 may sometimes be influenced, but are not driven principally by Diwali.

References

Bond, T. C., Bhardwaj, E., Dong, R., Jogani, R., Jung, S., Roden, C., Streets, D. G., and Trautmann, N. M.: Historical emissions of black and organic carbon aerosol from energy-related combustion, 1850–2000, Global Biogeochemical Cycles, 21, 2007.

Darmenov, A. and da Silva, A.: The quick fire emissions dataset (QFED)–documentation of versions 2.1, 2.2 and 2.4, NASA Technical Report Series on Global Modeling and Data Assimilation, NASA TM-2013-104606, 32, 183, 2013.

Durre, I., Vose, R. S., and Wuertz, D. B.: Overview of the integrated global radiosonde archive, Journal of Climate, 19, 53–68, 2006.

Geng, G., Zhang, Q., Martin, R. V., van Donkelaar, A., Huo, H., Che, H., Lin, J., and He, K.: Estimating long-term PM 2.5 concentrations in China using satellite-based aerosol optical depth and a chemical transport model, Remote Sensing of Environment, 166, 262–270, 2015.

Kaiser, J., Heil, A., Andreae, M., Benedetti, A., Chubarova, N., Jones, L., Morcrette, J.-J., Razinger, M., Schultz, M., Suttie, M., et al.: Biomass burning emissions estimated with a global fire assimilation system based on observed fire radiative power, Biogeosciences, 9, 527, 2012.

Kim, P. S., Jacob, D. J., Fisher, J. A., Travis, K., Yu, K., Zhu, L., Yantosca, R. M., Sulprizio, M., Jimenez, J. L., Campuzano-Jost, P., et al.: Sources, seasonality, and trends of southeast US aerosol: an integrated analysis of surface, aircraft, and satellite observations with the GEOS-Chem chemical transport model, Atmospheric Chemistry and Physics, 15, 10411–

10 433, 2015.

Li, M., Zhang, Q., Kurokawa, J.-i., Woo, J.-H., He, K., Lu, Z., Ohara, T., Song, Y., Streets, D. G., Carmichael, G. R., et al.: MIX: a mosaic Asian anthropogenic emission inventory under the international collaboration framework of the MICS-Asia and HTAP, Atmospheric Chemistry and Physics, 17, 935, 2017.

Lin, J., Gerbig, C., Wofsy, S., Andrews, A., Daube, B., Davis, K., and Grainger, C.: A near-field tool for simulating the upstream influence of atmospheric observations: The Stochastic Time-Inverted Lagrangian Transport (STILT) model, Journal of Geophysical Research: Atmospheres, 108, 2003.

Liu, T., Marlier, M. E., DeFries, R. S., Westervelt, D. M., Xia, K. R., Fiore, A. M., Mickley, L. J., Cusworth, D. H., and Milly, G.: Seasonal impact of regional outdoor biomass burning on air pollution in three Indian cities: Delhi, Bengaluru, and Pune, Atmospheric Environment, 172, 83– 92, 2018.

Menne, M., Durre, I., Korzeniewski, B., McNeal, S., Thomas, K., Yin, X., Anthony, S., Ray, R., Vose, R., Gleason, B., et al.: Global historical climatology network-daily (GHCN-Daily), Version 3, NOAA National Climatic Data Center, 2012.

Mu, Q. and Liao, H.: Simulation of the interannual variations of aerosols in China: role of variations in meteorological parameters, Atmospheric Chemistry and Physics, 14, 9597–9612, 2014.

Oanh, N. T. K., Ly, B. T., Tipayarom, D., Manandhar, B. R., Prapat, P., Simpson, C. D., and Liu, L.-J. S.: Characterization of particulate matter emission from open burning of rice straw, Atmospheric Environment, 45, 493–502, 2011.

Randerson, J., Chen, Y., Werf, G., Rogers, B., and Morton, D.: Global burned area and biomass burning emissions from small fires, Journal of Geophysical Research: Biogeosciences, 117, 2012.

Seibert, P., Beyrich, F., Gryning, S.-E., Joffre, S., Rasmussen, A., and Tercier, P.: Review and intercomparison of operational methods for the determination of the mixing height, Atmospheric environment, 34, 1001–1027, 2000.

Singh, D., Gadi, R., Mandal, T., Dixit, C., Singh, K., Saud, T., Singh, N., and Gupta, P. K.: Study of temporal variation in ambient air quality during Diwali festival in India, Environmental monitoring and assessment, 169, 1–13, 2010.

Stein, A., Draxler, R. R., Rolph, G. D., Stunder, B. J., Cohen, M., and Ngan, F.: NOAA?s HYSPLIT atmospheric transport and dispersion modeling system, Bulletin of the American Meteorological Society, 96, 2059–2077, 2015.

Tibshirani, R.: Regression shrinkage and selection via the lasso, Journal of the Royal Statistical Society. Series B (Methodological), pp. 267–288, 1996.

Vadrevu, K. P., Badarinath, K., and Anuradha, E.: Spatial patterns in vegetation fires in the Indian

region, Environmental monitoring and assessment, 147, 1–13, 2008.

Wang, Y., Zhang, Q., He, K., Zhang, Q., and Chai, L.: Sulfate-nitrate-ammonium aerosols over China: response to 2000–2015 emission changes of sulfur dioxide, nitrogen oxides, and ammonia, Atmospheric Chemistry and Physics, 13, 2635–2652, 2013.

Wiedinmyer, C., Akagi, S., Yokelson, R. J., Emmons, L., Al-Saadi, J., Orlando, J., and Soja, A.: The Fire INventory from NCAR (FINN): a high resolution global model to estimate the emissions from open burning, Geoscientific Model Development, 4, 625, 2011.

Supplementary Figures

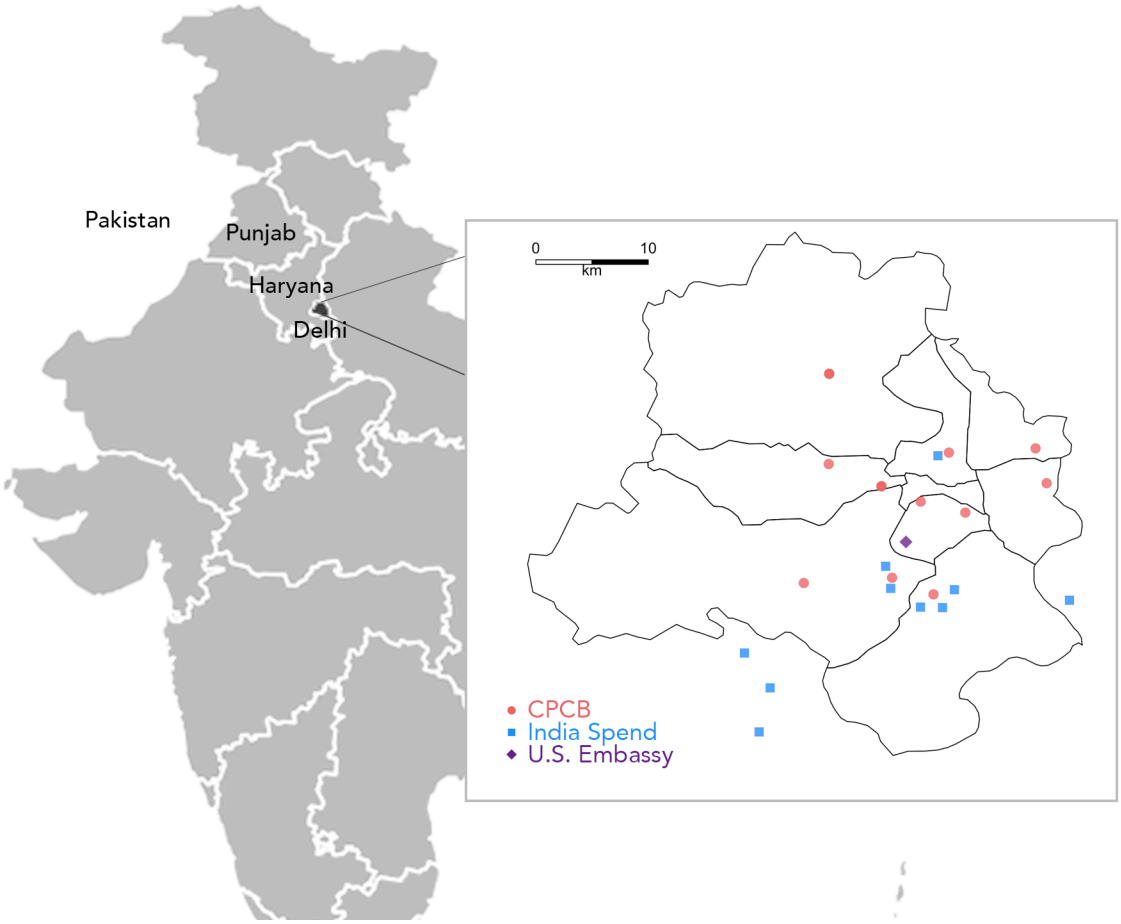

Figure S1: Distribution of Central Pollution Control Board (CPCB), India Spend, and U.S. Embassy $PM_{2.5}$ monitoring sites located in and around the Delhi. Solid lines represent districts within the National Capital Territory of Delhi.

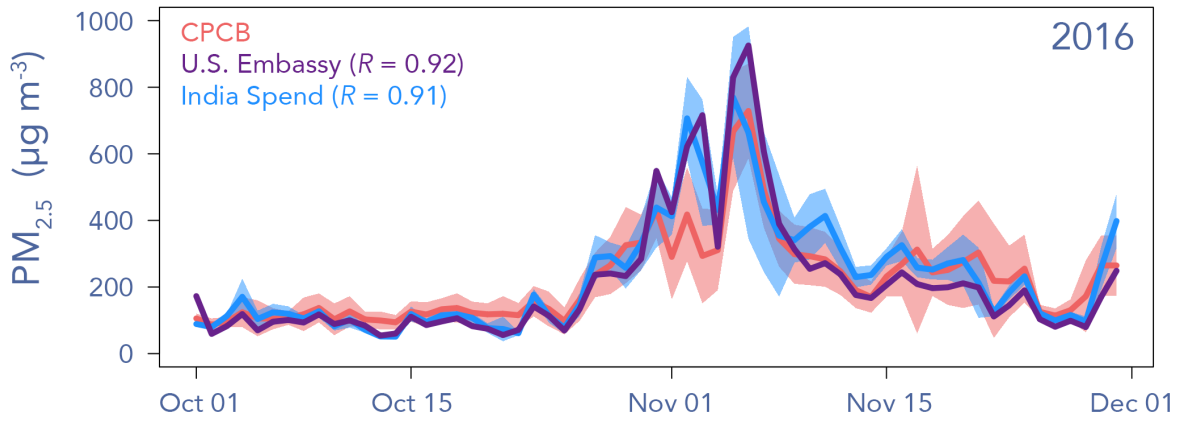

Figure S2. Daily-averaged $PM_{2.5}$ observations for CPCB (pink), U.S. Embassy (purple), and India Spend (blue) for 2016. Shading for the CPCB and India Spend curves represent one standard deviation from the network-averaged $PM_{2.5}$. The U.S. Embassy observations correlate well with the CPCB network average (R = 0.92), and the India Spend observations correlate well with the CPCB network (R = 0.91).

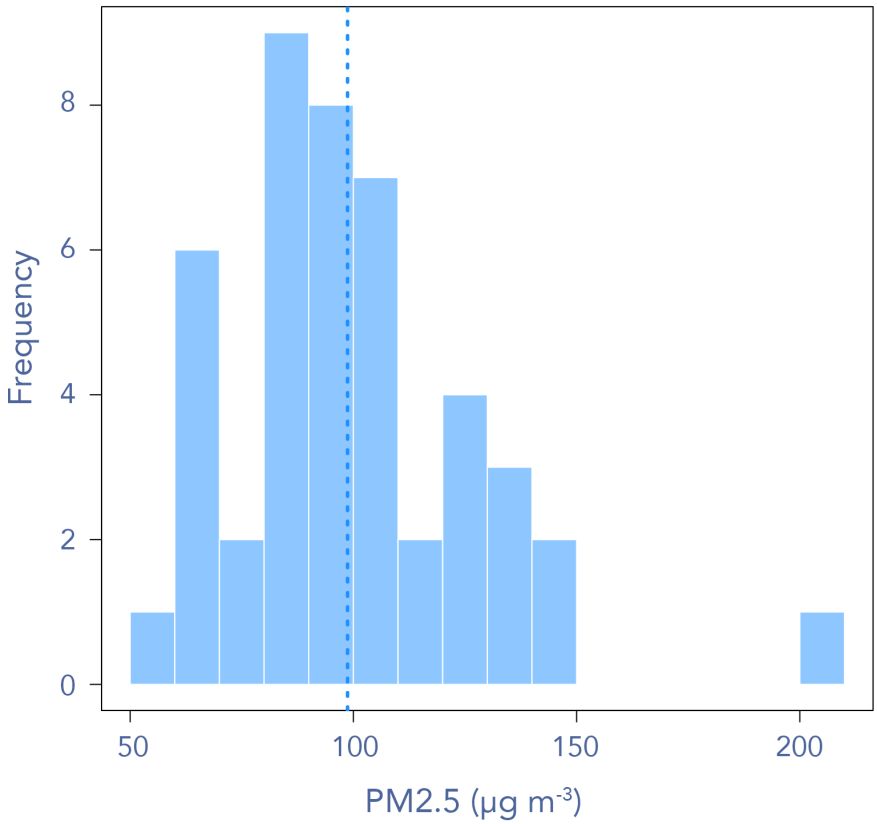

Figure S3: Frequency of $PM_{2.5}$ daily observations in Delhi derived from a GEOS-Chem simulation performed at 0.5° x 0.667° horizontal resolution during the burning season of 2012 (Oct 17. – Nov. 30). The dashed vertical line represents the mean of the distribution.

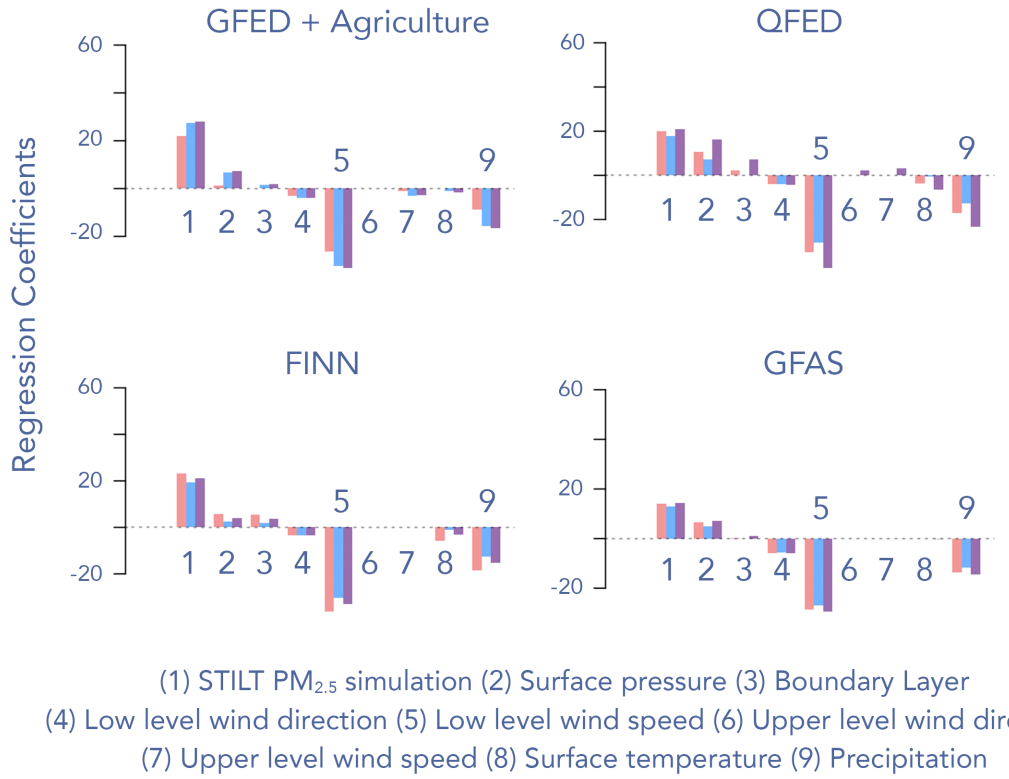

Figure S4. Standardized regression coefficients (μg m^{-3} standard $deviation^{-1}$) fit to daily $PM_{2.5}$ enhancements, derived from three different baseline methods. See text for description of these methods. The first term, labeled "1, STILT $PM_{2.5}$ simulation," represents simulated $PM_{2.5}$ using one of the four fire emission inventories – GFED + Agriculture, which assumes 100% agricultural landcover and emission factors increased by a factor 3; QFED; FINN; and GFAS.

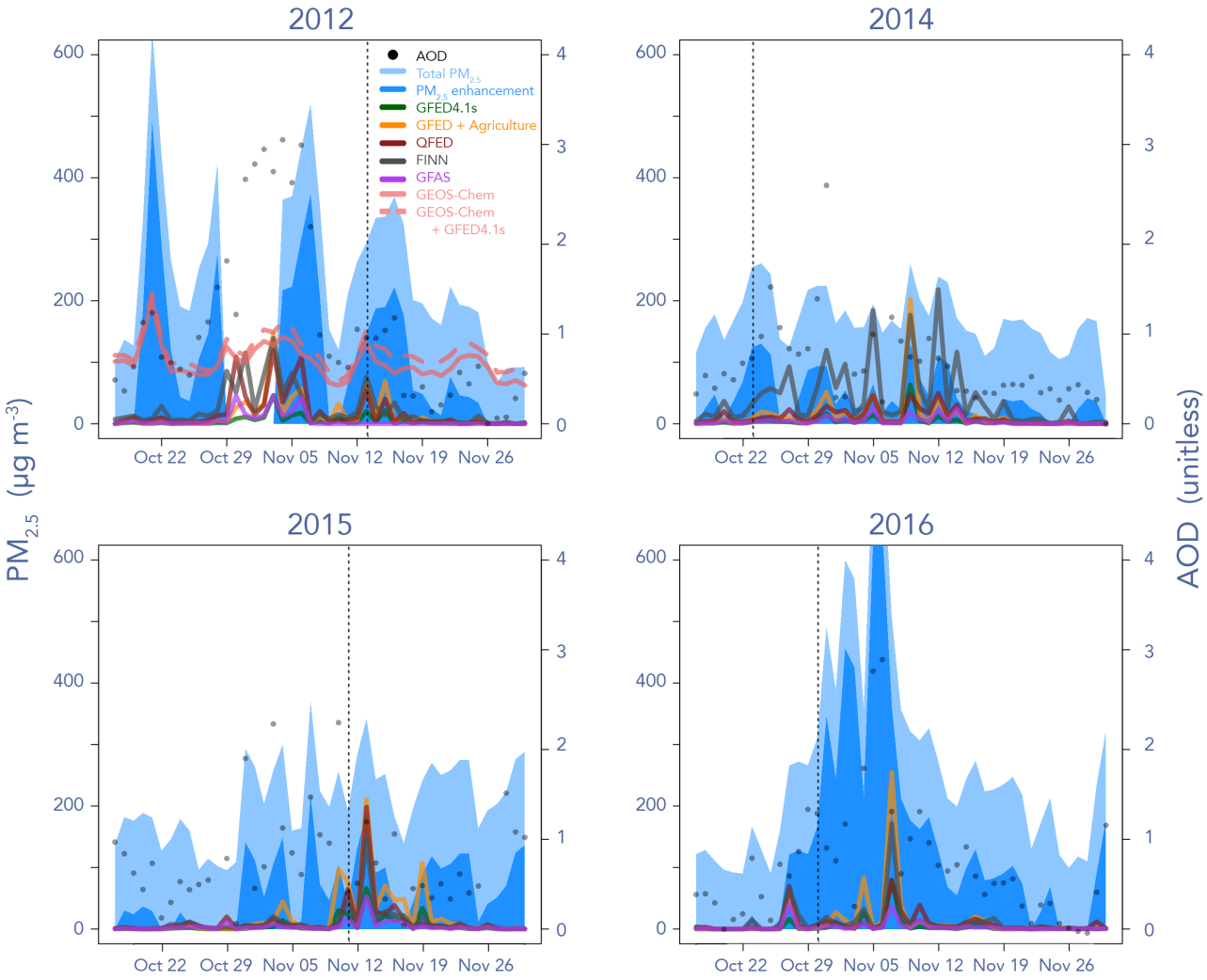

Figure S5: Time series of observed and modeled $PM_{2.5}$ during the 2012, 2014-2016 post-monsoon burning seasons. The blue envelopes represent the observed total $PM_{2.5}$ and the $PM_{2.5}$ enhancement derived by subtracting the daily $PM_{2.5}$ by the mean $PM_{2.5}$ of the lowest week during the season. Each colored line represents a model simulation with a different fire emission inventory. The black dots are the MODIS AOD retrievals during the burning season. The dashed vertical lines represent the first day of the Diwali festival. For 2012, we add two additional lines GEOS-Chem + GFED4.1s which represents GEOS-Chem $PM_{2.5}$ from both anthropogenic and fire sources and GEOS-Chem which represents $PM_{2.5}$ from anthropogenic sources alone.