Supplement information of

Improved estimates of smoke exposure during Australia fire seasons: Importance of quantifying plume injection heights

Xu Feng[1], Loretta J. Mickley[1], Michelle L. Bell[2], Tianjia Liu[3], Jenny A. Fisher[4], Maria Val Martin[5]

[1] John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, 02138, USA
[2] School of Environment, Yale University, New Haven, CT, 06511, USA
[3] Department of Earth System Science, University of California, Irvine, CA, 92697, USA
[4] Centre for Atmospheric Chemistry, School of Earth, Atmospheric and Life Sciences, University of Wollongong, Wollongong, 2522, Australia
[5] School of Biosciences, University of Sheffield, Sheffield, S10 2TN, UK

Section S1:

Algorithm of the STILT model

The STILT algorithm releases an ensemble of air parcels at a receptor and tracks the trajectory of each air parcel backwards in time for a specified number of days. As shown in Eq. (S1), the concentration change (ΔC, in units of ppm) at the receptor (x_r, y_r) at time t_r due to the surface emission flux (F, in units of μmol m^{-2} s^{-1}) from a source point (x_i, y_j) is determined by the total amount of time $(\Delta t_{p,i,j,k})$ that each air parcel stays in the volume element (i, j, k) during the time step t_m, the total number of air parcels (N_{tot}), and the diluting height (h). STILT assumes that surface emission fluxes are instantaneously diluted into an atmospheric column of h due to turbulent mixing. The average density of all air parcels below h is $\overline{\rho}$, and m_{air} is the air molecular weight. The diluting height h is generally set to be within the PBL, and STILT assumes that emission fluxes distributed above h do not affect surface concentrations downwind (Lin et al., 2003; Gerbig et al., 2003). The definition of sensitivity footprints is given in Eq. (S2) with units of ppm μmol^{-1} m^{2} s.

$$\Delta C\left(x_r, y_r, t_r \mid x_i, y_j, t_m\right) = \begin{cases} \dfrac{m_{air}}{h\overline{\rho}(x_i, y_j, t_m)} \dfrac{1}{N_{tot}} \displaystyle\sum_{p=1}^{N_{tot}} \Delta t_{p,i,j,k} \cdot F(x_i, y_i, t_m), & z \leq h \\ 0, & z > h \end{cases} \quad (S1)$$

$$f\left(x_r, y_r, t_r \mid x_i, y_j, t_m\right) = \frac{m_{air}}{h\overline{\rho}(x_i, y_j, t_m)} \frac{1}{N_{tot}} \sum_{p=1}^{N_{tot}} \Delta t_{p,i,j,k} \quad (S2)$$

Table S1. Statistics for seasonal-mean $PM_{2.5}$ concentrations simulated by three simulations, compared to the ground-based observations at 12 receptors. CTL is the control simulation, INJ-CLIM represents the climatological method of calculating plume height, and INJ-RF represents the method using random forest. The observation site name in each city or town is given in parentheses.

Cities (site)	**Observation periods (Locations)**	R^a			**NMB**			**RMSE (μg m^{-3})**		
		CTL	***INJ-CLIM***	***INJ-RF***	***CTL***	***INJ-CLIM***	***INJ-RF***	***CTL***	***INJ-CLIM***	***INJ-RF***
Darwin [b] (Palmerston)	2011-2020 (130.94°E, 12.50°S)	0.78	0.54	0.77	16.7%	-18.0%	-2.5%	2.8	2.7	1.7
Gladstone [c] (South Gladstone)	2009-2020 (151.27°E, 23.86°S)	0.86	0.88	0.88	-5.8%	-11.4%	-11.7%	1.0	1.1	1.1
Brisbane [c] (Springwood)	2009-2020 (153.13°E, 27.61°S)	0.60	0.54	0.61	6.3%	-5.6%	-10.0%	1.4	1.3	1.3
Newcastle [d] (Wallsend)	2009-2020 (151.66°E, 32.89°S)	0.91	0.92	0.91	4.5%	-5.1%	-11.2%	3.5	1.8	1.5
Sydney [d] (Liverpool)	2009-2020 (150.90°E, 33.93°S)	0.96	0.95	0.95	3.3%	-2.6%	-7.2%	1.9	1.1	1.1
Wollongong [d] (Wollongong)	2009-2020 (150.88°E, 34.41°S)	0.86	0.85	0.80	-4.4%	-8.6%	-11.2%	1.2	1.3	1.6
Melbourne [e] (Footscray)	2009-2020 (144.87°E, 37.80°S)	0.48	0.47	0.40	2.3%	1.7%	-1.8%	1.3	1.3	1.3
Melbourne [e] (Alphington)	2009-2020 (145.03°E, 37.77°S)	0.61	0.61	0.67	14.7%	14.1%	8.5%	1.7	1.6	1.3
Albury [d] (Albury)	2017-2020 (146.93°E, 36.05°S)	0.98	0.98	0.97	-25.5%	-26.6%	-32.2%	5.3	5.5	6.5
Canberra [f] (Florey)	2014-2020 (149.04°E, 35.22°S)	0.99	0.99	0.99	16.1%	-7.4%	-13.6%	7.0	1.4	1.8
Sydney [d] (Prospect)	2014-2020 (150.91°E, 33.79°S)	0.99	0.99	0.99	1.7%	-4.7%	-9.2%	1.6	0.69	0.98
Newcastle [d] (Newcastle)	2014-2020 (151.75°E, 32.93°S)	0.91	0.89	0.91	19.0%	4.8%	-3.3%	4.2	2.0	1.4

[a] Temporal correlation coefficient between the observed and simulated annual mean total $PM_{2.5}$ concentrations during the fire seasons (April to December for Darwin and Gladstone; August to December for other cities).

[b] Observation data source: Northern Territory Environment Protection Authority (http://ntepa.webhop.net/NTEPA/Default.ltr.aspx)

[c] Queensland Government Open Data Portal (https://apps.des.qld.gov.au/air-quality/download/)

[d] New South Wales Department of Planning and Environment (https://www.dpie.nsw.gov.au/air-quality/air-quality-data-services/data-download-facility)

[e] Victoria Environment Protection Authority (https://www.epa.vic.gov.au/for-community/airwatch)

[f] Australian Capital Territory Government Open Data Portal (https://www.data.act.gov.au/Environment/Air-Quality-Monitoring-Data/94a5-zqnn

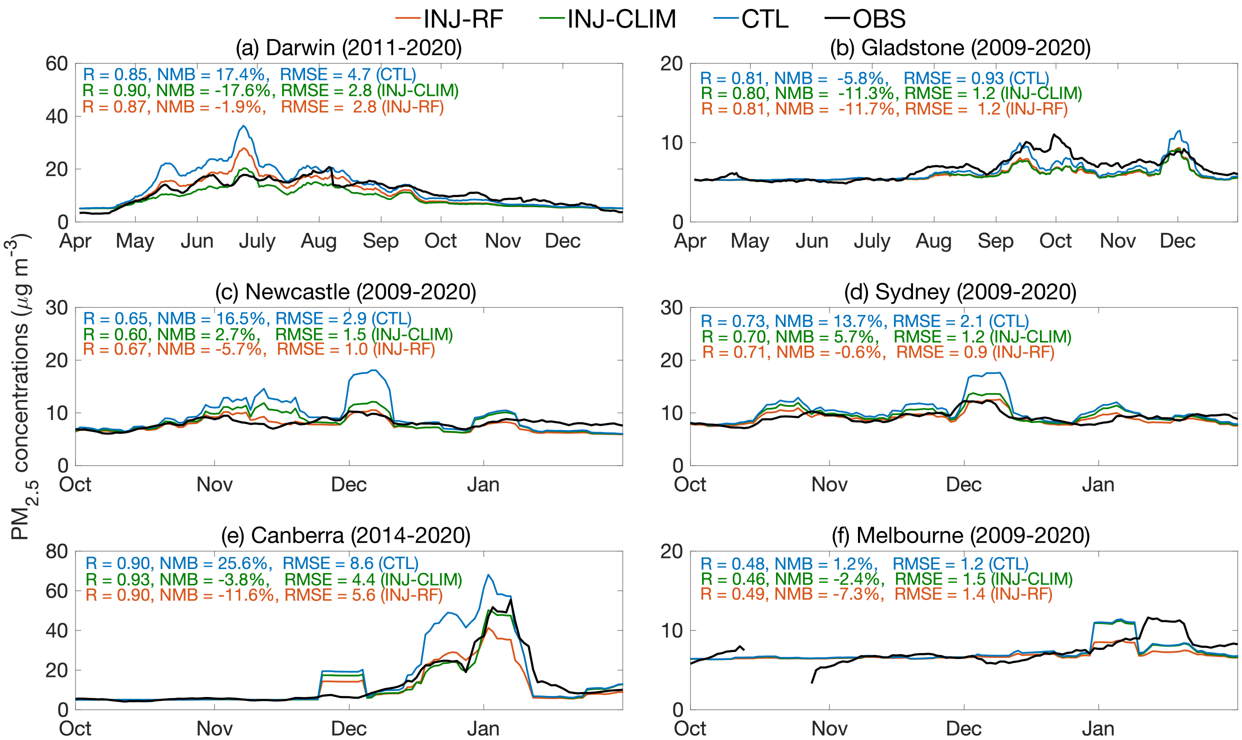


Figure S1. Time series of 10-day moving average of observed and simulated total $PM_{2.5}$ concentrations from the CTL (blue), INJ-CLIM (green), and INJ-RF (red) experiments during the fire seasons at representative receptors in six cities: (a) Darwin (Palmerston), (b) Gladstone (South Gladstone), (c) Newcastle (Wallsend), (d) Sydney (Liverpool), (e) Canberra (Florey), and (f) Melbourne (Footscray). (Given in parentheses are the names of the observation sites.) The 10-day moving averages are calculated over each receptor's observing period, as indicated above the panel. Shown inset are the temporal correlation coefficients *R*, NMBs, and RMSEs of daily total $PM_{2.5}$ concentrations compared to the surface measurements.

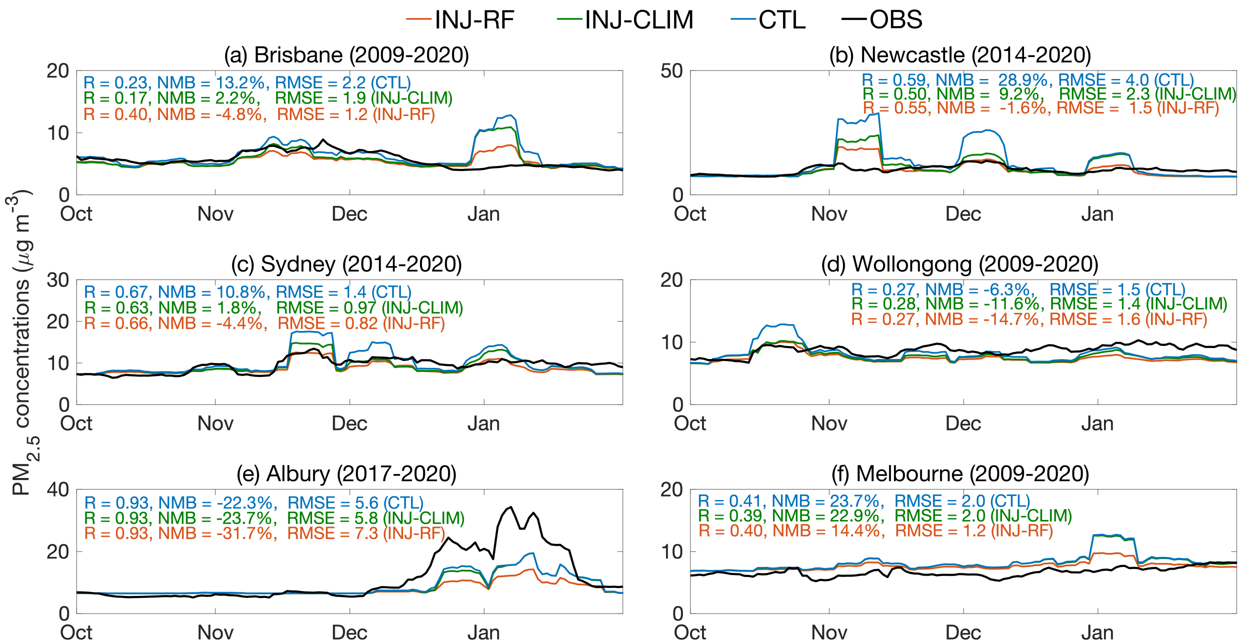


Figure S2. As in Figure S1 but for six other representative receptors: (a) Brisbane (Springwood), (b) Newcastle (Newcastle), (c) Sydney (Prospect), (d) Wollongong (Wollongong), (e) Albury (Albury), and (f) Melbourne (Alphington). (Given in parentheses are the names of the observation sites.) The 10-day moving averages are calculated over each receptor's observing period, as indicated above the panel. Shown inset are the temporal correlation coefficients *R*, NMBs, and RMSEs of daily total $PM_{2.5}$ concentrations compared to the surface measurements.

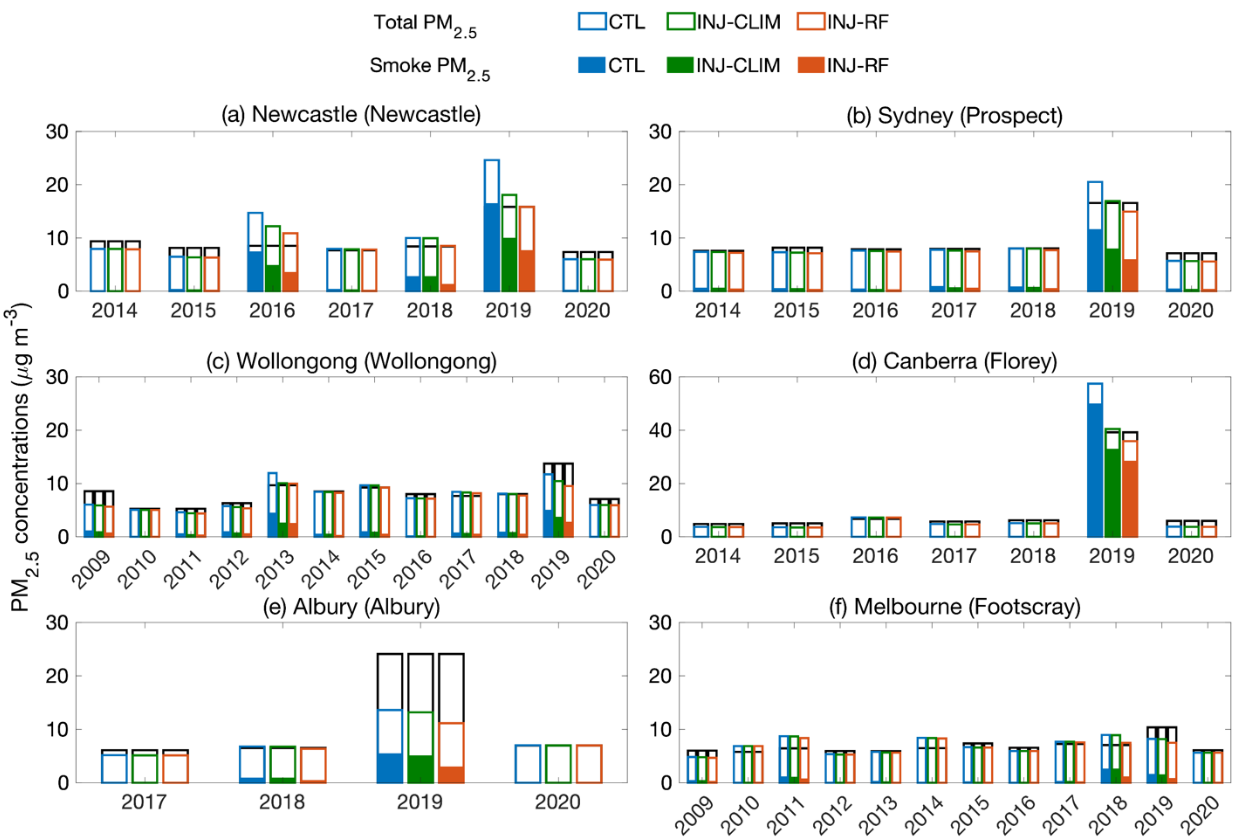


Figure S3. Mean simulated concentrations of smoke $PM_{2.5}$ and background $PM_{2.5}$ from the three sensitivity experiments (blue: CTL, green: INJ-CLIM, red: INJ-RF), as well as observed total $PM_{2.5}$ concentrations (black: OBS) in (a) Newcastle (Newcastle), (b) Sydney (Prospect), (c) Wollongong (Wollongong), (d)
65 Canberra (Florey), (e) Albury (Albury), and (f) Melbourne (Footscray). (The names of the observation sites are given in parentheses.) The different receptors have different observation periods. The modeled total $PM_{2.5}$ concentrations are designated by the height of the colored bars, consisting of smoke $PM_{2.5}$ (color-filled bars) and the background $PM_{2.5}$ (empty bars) in units of μg m^{-3}.

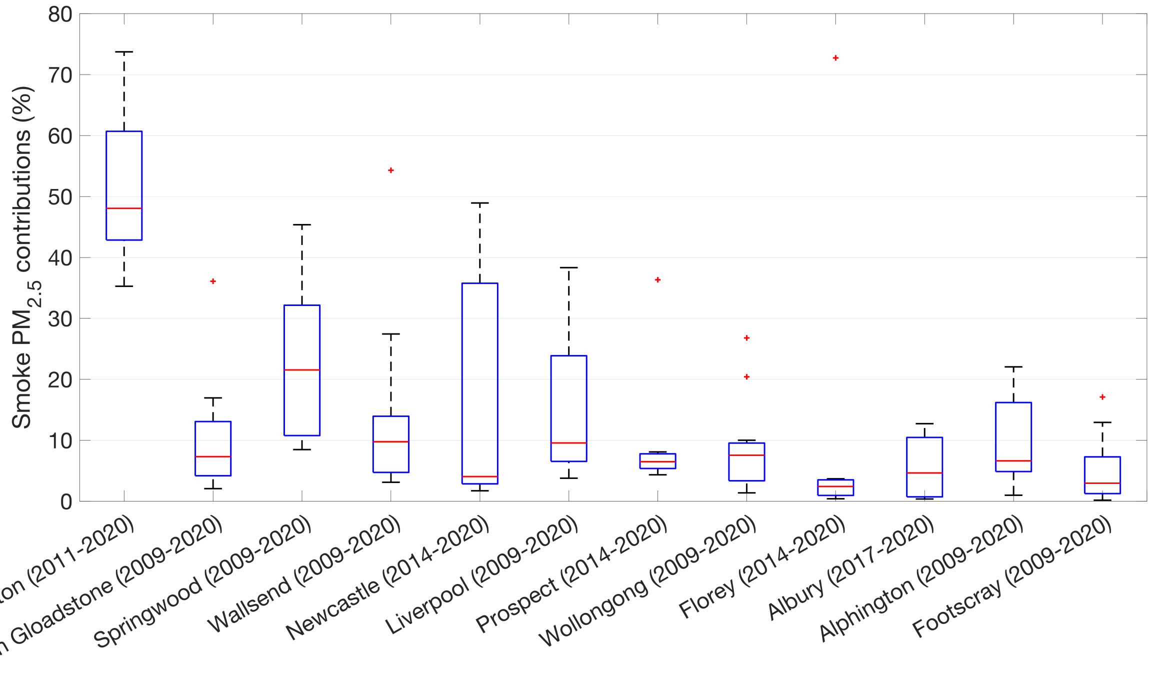
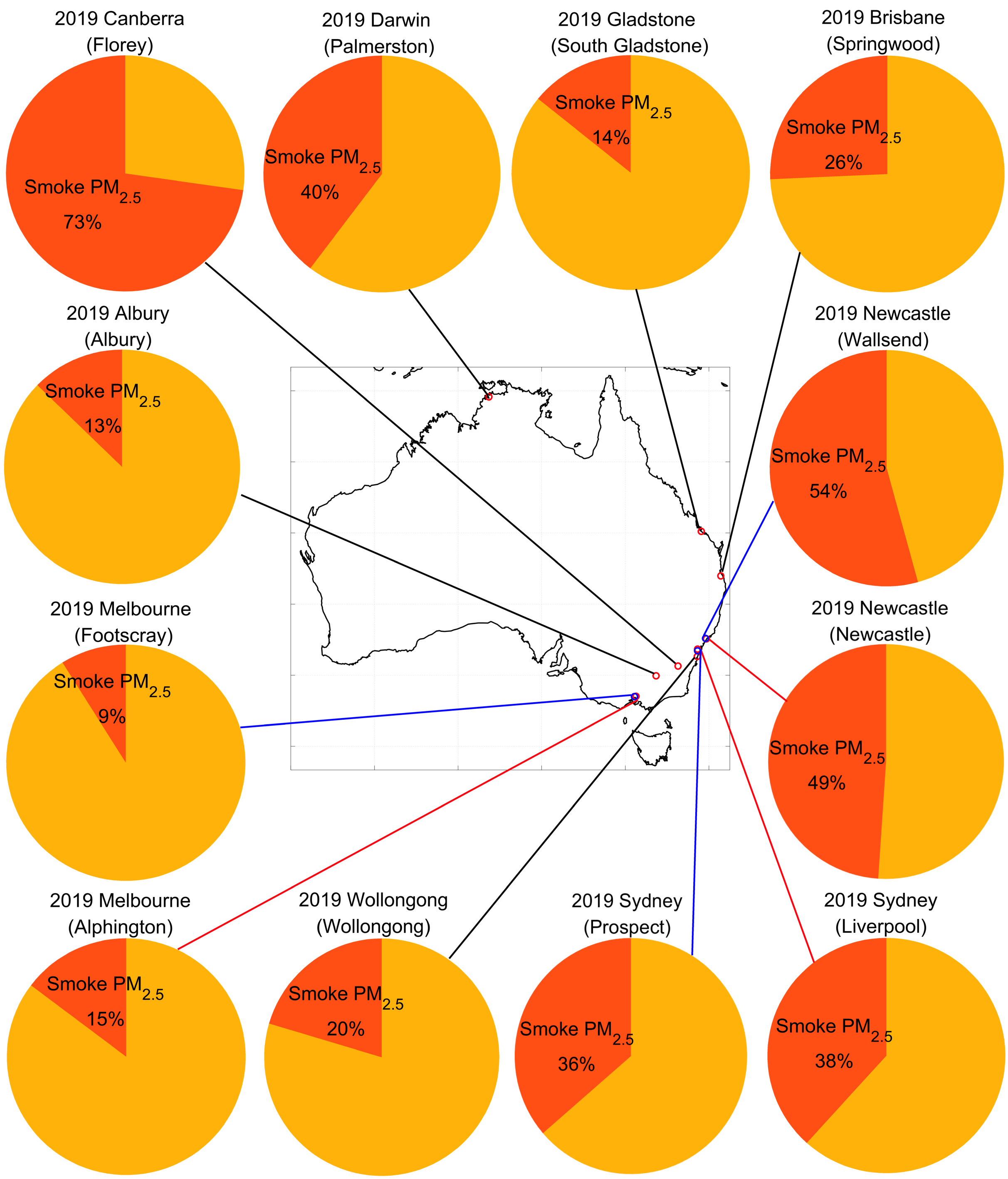


Figure S4. Boxplot of annual contributions of simulated smoke $PM_{2.5}$ concentrations from the INJ-RF experiment to observed total $PM_{2.5}$ concentrations at 12 receptors during the fire seasons of respective observations periods. The order of 12 receptors in this figure is based on the locations from north to south in Australia. The bottom, top, and red lines in the middle of each box are the 25th and 75th percentiles, as well as the median of all data. The distance between the 75th and 25th percentiles is the interquartile range. The lower and upper whisker limits represent the most extreme data values within 1.5 times the interquartile range. The data greater than 1.5 times outside the interquartile range are considered outliers and are shown as red crosses.



80 Figure S5. Contributions of simulated smoke $PM_{2.5}$ concentrations from the INJ-RF experiment to the observed total $PM_{2.5}$ concentrations at 12 receptors averaged over the fire seasons of 2019. Names of the observation sites are given in parentheses. Red sectors represent smoke contributions, while dark yellow sectors signify the differences between observed total $PM_{2.5}$ and simulated smoke $PM_{2.5}$ concentrations – i.e., the non-fire $PM_{2.5}$. Small circles on map represent the locations of these receptors. Different colors (red,
85 blue, and black) are used to distinguish adjacent receptors.

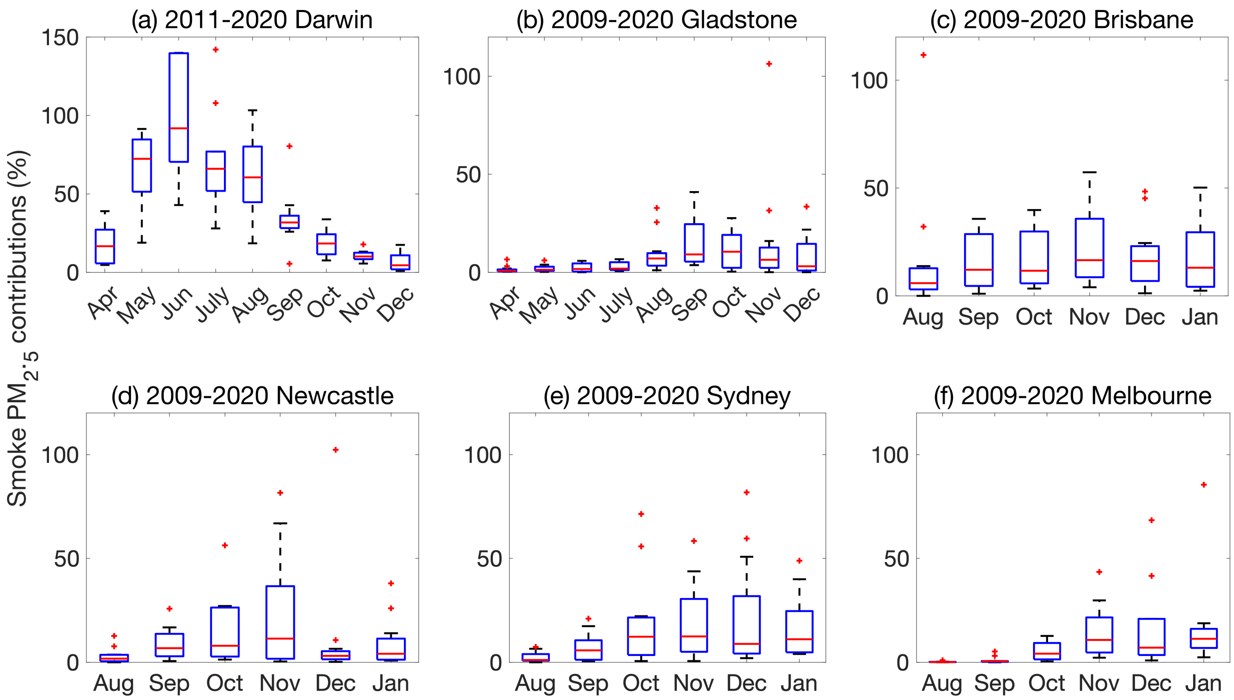


Figure S6. Boxplot of monthly contributions of simulated smoke $PM_{2.5}$ concentrations from the INJ-RF experiment to the observed total $PM_{2.5}$ concentrations during respective observation periods in (a) Darwin (Palmerston), (b) Gladstone (South Gladstone), (c) Brisbane (Springwood), (d) Newcastle (Wallsend), (e) Sydney (Liverpool), and (f) Melbourne (Alphington). The bottom, top, and red line in the middle of each box are the 25th and 75th percentiles, as well as the median of all data. The distance between the 75th and 25th percentiles is the interquartile range. The lower and upper whisker limits represent the most extreme data values within 1.5 times the interquartile range. The data greater than 1.5 times outside the interquartile range are considered outliers and are shown as the red crosses. In some years, the monthly smoke $PM_{2.5}$ contributions over 100% are due to the overestimates of simulated smoke $PM_{2.5}$.