

S1. Data

In-situ Coverage

Owing to difficult operating conditions instrument coverage was variable across and within all flights. Of twenty flights undertaken during SAMBBA (b731 – b750) all but two (b733 and b747) were full science flights. Hence, ratios derived from different instruments are not always comparing the same sample points in time and space. Optimum bin sizes for the calculation of the median profiles were chosen based on containing approximately the same number of data points in each bin.

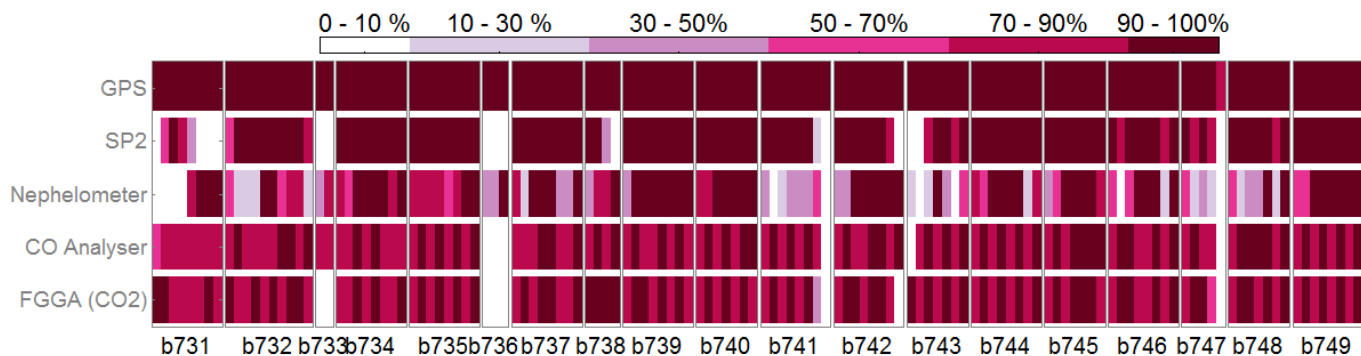


Figure S1. Schematic of percentage data coverage for key parameters, per 30 minute segment flight time

In-situ Acquisition

A comprehensive set of all variables were measured during SAMBBA on a 1 HZ base. Profile time series were averaged into 50-m altitude bins for the automated analysis presented here. Both datasets are available upon request.

Ancillary Products

The MODIS instrument, operating on-board the Aqua and Terra satellites, provided retrievals of daily Aerosol Optical Depth (AOD) at 550 nm on a 1° x 1° grid (specific products: MYD08_D3.061, MOD08_D3.061). Instruments on-board the Tropical Rainfall Measuring Missions (TRMM) satellite mission provided a measure of precipitation rate (in mm/hr) at a 3-hr interval on a 0.25° x 0.25° grid (specific product: 3B42.007) these products were obtained from the Giovanni online data system, developed and maintained by the NASA GES DISC (<http://disc.sci.gsfc.nasa.gov/giovanni>).

The fire data were produced by the University of Maryland and acquired from the online Fire Information for Resource Management System (FIRMS; <https://earthdata.nasa.gov/data/near-real-time-data/firms/abouts>; specific product: MCD14ML).

Model wind fields and soil moisture values were obtained from the ERA-Interim global atmospheric reanalysis product (Dee et al., (2011); <http://www.ecmwf.int/en/research/climate-reanalysis/era-interim>). All soil moistures reported in the manuscript are for soil water layer 1. When soil moistures are linked to profiles, the value used is the nearest grid box at the closest 6-hour time step.

Land cover data was provided at a 300 m resolution from the European Space Agency Land Cover Maps (v2.07), which are based on observations from the MEdium Resolution Imaging Spectrometer (MERIS), Advanced Very High Resolution Radiometer (AVHRR), SPOT-VGT, PROBA-V and PROBA-B satellites. The land cover data was re-gridded on to a 5' x 5' grid for presentation purposes in this manuscript. This product is based on the UN Land Cover Classification System (LCCS), which was simplified based on the approach of Chen et al., (2013) for MODIS IGBP (International Geosphere–Biosphere Programme) data and is displayed below:

LCSS Scheme		New Scheme	LCSS Scheme		New Scheme
Class	Class Name		Class	Class Name	
0	no_data		100	mosaic_tree_and_shrub	SAVN
10	cropland_rainfed	AGRI	110	mosaic_herbaceous	SAVN
11	cropland_rainfed_herbaceous_cover	AGRI	120	shrubland	SAVN
12	cropland_rainfed_tree_or_shrub_cover	AGRI	121	shrubland_evergreen	SAVN
20	cropland_irrigated	AGRI	122	shrubland_deciduous	SAVN
30	mosaic_cropland	AGRI	-126	grassland	SAVN
40	mosaic_natural_vegetation	AGRI	-116	lichens_and_mosses	
50	tree_broadleaved_evergreen_closed_to_open	EVGN	-106	sparse_vegetation	
60	tree_broadleaved_deciduous_closed_to_open	DECF	-104	sparse_shrub	
61	tree_broadleaved_deciduous_closed	DECF	-103	sparse_herbaceous	
62	tree_broadleaved_deciduous_open	DECF	-96	tree_cover_flooded_fresh_or_brakish_water	
70	tree_needleleaved_evergreen_closed_to_open	EVGN	-86	tree_cover_flooded_saline_water	
71	tree_needleleaved_evergreen_closed	EVGN	-76	shrub_or_herbaceous_cover_flooded	
72	tree_needleleaved_evergreen_open	EVGN	-66	urban	
80	tree_needleleaved_deciduous_closed_to_open	DECF	-56	bare_areas	
81	tree_needleleaved_deciduous_closed	DECF	-55	bare_areas_consolidated	
82	tree_needleleaved_deciduous_open	DECF	-54	bare_areas_unconsolidated	
90	tree_mixed	DECF	-46	water	
			-36	snow_and_ice	

S2. Data Processing

Individual Profile Operations

Individual profiles were analysed to determine common features to elucidate the drivers of the pollutant vertical distribution. As a precursor step, to ensure pollutant and thermodynamic parameters were comparable, profile data was resampled into 50 metre altitude bins - the value in any given bin being the median of all points in that range.

Convective Mixing Layer (CML): This was estimated from a manual inspection of each profile and the spread of these automated methods: i) location of the point $d\theta/dz \geq 2 \text{ K km}^{-1}$ (Fisch et al., 2004), following Seidel et al., (2010) the location of the minimal vertical gradient of ii) specific humidity, iii) relative humidity, iv) potential temperature and v) refractivity, plus vi) the Wang and Wang, (2014) approach of combining methods 2-5. The spread amongst these was often large for each profile, given the complex atmospheric structure encountered (e.g. multiple residual mixing layers), hence the necessity for a manual approach. For example, in Fig. 3, techniques 1-6 were in close agreement (100-150 m) given the sharp gradient in all parameters.

Thermodynamic Layers: The lifted condensation level (LCL), level of free convection (LFC) and limit of convection (LOC) were automatically calculated from the tephigrams in Fig. 3. These parameters were only calculated if data was present below 500 m. In profiles where the CML, LCL, LFC and LOC could not be calculated, e.g. a profile between 3-8 km with no near surface data, values from temporally and geographic neighbouring profiles were applied, if appropriate.

5 **Pollutants well mixed in the convective boundary layer:** Determined via comparison of the pollutant vertical gradient and pollutant variability within the CML. For the same definition thresholds to be used across all pollutants they were standardised via z-score normalisation. The gradient was calculated from fitting a straight line throughout the normalised data points within the CML region. The variability was represented via the standard deviation of this data. If the gradient was between +/- 0.0005 and the standard deviation less than 1.5 that pollutant was considered well mixed.

10 **Plumes:** Identified via first considering pollutants separately (steps 1-3) and then combined (step 4), via the following methodology:

1) A baseline was calculated as the moving 25th percentile over a 1500 m window to represent the underlying profile shape.
2) A value was then added to this baseline to create an upper threshold, any points above which are considered in excess of the background haze and represent a significant enhancement. The values chosen were the campaign 90th percentiles (above
15 background) within the convective boundary layer (< 4 km), based on aircraft straight and level runs: 331 ppb for CO, 18.6 ppm for CO₂, 1.32 µg sm⁻³ for rBC and 125 Mm⁻¹ for σ_{sp}.

3) Any given exceedance of this threshold was defined as a potential plume. The footprint of this was then identified by locating the above/below altitudes at which pollutant values returned to 'near baseline'. This was defined as the baseline plus the above background campaign 10th percentile of each pollutant within the lowest 4 km, calculated from the flight straight
20 and level runs.

4) The plumes identified from each pollutant profile were then collated. If any plume footprints overlapped, these were deemed co-incident and the plume footprint was redefined to encompass the region of these plumes.

To our knowledge this is the first attempt at identifying plumes from vertical profiles within an already polluted boundary layer. Reported pollutant values/ratios are calculated as the plume integrated value across the total plume footprint.

25 **Horizontal wind speed maximum:** Starting at the lowermost altitude, the horizontal wind speed was interrogated to determine if it was greater than the wind speed at the next two altitudes. At points above 100 m, the wind speed was also required to be greater than the two altitudes below. A horizontal wind speed maxima was defined if these conditions were satisfied and the wind speed was also greater than some threshold. The threshold was defined in a similar manner to that used for plume identification: the moving 25th percentile over a 1500 m window (the baseline) plus a value of 1 m s⁻¹. Once a maxima is defined, the identification
30 procedure begins again at the following altitude plus 100 m.

Horizontal positive wind shear region: Based on the assumption regions that a region of strong positive vertical wind shear will end in a local wind speed maxima. As such the method for deriving these was utilised (see above), although the threshold was greater (+ 2m s⁻¹) given the coarser scale of wind shear regions. When the differential of the wind speed profile was less than zero at three successive altitudes below the wind speed maximum altitude, the base of the wind shear region was identified.

35 **Horizontal wind speed maximum coincident with pollutant maximum:** Pollutant maximum in each profile were identified in a similar manner to steps 1-3 in identifying plumes, only the upper threshold was lower – calculated as the campaign 25th percentile of each pollutant within the lowest 4 km, calculated from the flight straight and level runs. Data points +50, 0, and -50 m above each of the horizontal wind speed maximum were interrogated to determine if they also contained a pollutant maximum.

Horizontal wind shear region coincident with pollutant reduction: If the pollutant profile in the wind shear regions identified
40 (see above) satisfied with the following three conditions, it was considered the shear was acting to reduce the pollutant concentrations: i) the differential of the pollutants was majoritively negative, ii) the pollutant median of the upper half of the shear

range was lesser than that of the lower half, and iii) The range covered in this reduction was greater than some threshold, defined here as the campaign 25th minus 5th percentile from straight and level runs. Capturing this feature programmatically is difficult, yielding occasional spurious results (Sect. S5). The results provide a broad and approximate insight into the prevalence of interactions between the horizontal wind speed and pollutant distribution rather than strict quantitative results.

- 5 **Residual Layer:** A pollutant residual layer was identified if burdens between the top of the mixing layer (or if absent, the surface) and 4 km exceeded those of unpolluted background conditions, defined as 0.1 ng cm⁻³ for rBC (Artaxo et al., 2013), 15 Mm⁻¹ for σ_{sp} (Rizzo et al., 2013) and 140 ppb for CO (Andreae et al., 2012). CO₂ not included in this analysis given the difficulty of determining unpolluted background concentrations.

- 10 **Enhanced pollutant concentrations at high altitude:** Identified if pollutant burdens above 4 km were in excess of the unpolluted background conditions provided for residual layer identification.

Miscellaneous

The scattering only aerosol optical depth (AOD) was calculated for each profile and plume therein as the integral of σ_{sp_amb} at ambient temperature and pressure.

- 15 The dropsonde altitude in metres was derived by applying the relationship between pressure and altitude from the aircraft to the dropsonde measured pressure.

Local time was defined as UTC minus 4 hours.

The relationships in Fig. 9 and trends in Fig. 5 are modelled by robust multilinear regressions using iteratively reweighted least squares with a bisquare weighting function.

- 20 The bulk of the analysis in this manuscript was undertaken with MATLAB 2012a. In the above instances, where the moving median is calculated, the function 'moving' was utilised – available at the MATLAB file exchange:
<http://uk.mathworks.com/matlabcentral/fileexchange/8251-moving-averages---moving-median-etc>.

S3. Tables

Feature	All	CO	CO ₂	rBC	σ _{sp_dry}
	% Profile frequency (# instances/ # profiles)				
Well mixed thermodynamic and pollutant mixing layer - AM	6 ^{2/34}	23 ^{7/31}	25 ^{6/24}	50 ^{15/30}	45 ^{15/33}
Well mixed thermodynamic and pollutant mixing layer - PM	22 ^{5/23}	43 ^{9/21}	50 ^{11/22}	41 ^{7/17}	50 ^{11/22}
Pollutant residual layer present above mixing layer	73 ^{52/71}	88 ^{60/68}		89 ^{55/62}	97 ^{67/69}
Pollutant residual layer in absence of mixing layer	62 ^{16/26}	76 ^{19/25}		88 ^{21/24}	100 ^{24/24}
Enhanced pollutant concentrations at high altitude	9 ^{8/94}	50 ^{45/90}		4 ^{3/74}	6 ^{5/84}
Coincident positive horizontal wind shear and pollutant reduction	16 ^{14/90}	59 ^{51/87}	48 ^{40/84}	41 ^{33/81}	41 ^{35/85}
Coincident horizontal wind speed and pollutant maximum	3 ^{3/102}	23 ^{23/98}	20 ^{19/96}	30 ^{28/94}	32 ^{31/98}
Plume	18 ^{12/65}	45 ^{28/62}	57 ^{35/61}	70 ^{40/57}	55 ^{36/65}
	% Total occurrence (# instances/ # all plumes)				
Plume	10 ^{12/122}	35 ^{41/118}	34 ^{39/114}	71 ^{75/105}	42 ^{50/119}

Table S1. Prevalence of interactions between atmospheric structure and biomass burning pollutants

	% Plume occurrence in layer (# layer instances)																							
	Surface → CML top		Surface → LCL		CML top → LCL		LCL → CML top		LCL → LFC		LFC → LOC		LOC → Ceiling		LFC → Ceiling									
	AM	PM	AM	PM	AM	PM	AM	PM	AM	PM	AM	PM	AM	PM	AM	PM								
CO	17	(30) 28	(18)	33	(15) 10	(10)	5	(37) 0	(21)	0	(5) 20	(5)	10	(51) 13	(32)	10	(20) 0	(10)	5	(19) 8	(13)	3	(37) 14	(22)
CO₂	44	(25) 17	(18)	71	(14) 10	(10)	20	(35) 0	(22)	25	(4) 0	(5)	4	(49) 9	(33)	0	(19) 0	(11)	0	(19) 0	(14)	0	(36) 0	(22)
rBC	11	(28) 27	(15)	33	(15) 0	(9)	14	(35) 17	(18)	20	(5) 20	(5)	27	(48) 17	(29)	17	(18) 0	(9)	21	(19) 9	(11)	23	(35) 10	(21)
σ_{sp_dry}	9	(32) 20	(20)	27	(15) 11	(9)	13	(38) 0	(23)	0	(5) 20	(5)	24	(50) 15	(34)	10	(20) 8	(13)	16	(19) 6	(16)	5	(37) 16	(19)

5 **Table S2. Prevalence of pollutant plumes in different thermodynamic layers, segregated by morning and afternoon. Bracketed values represent sample frequency of layers.**

S4. Figures

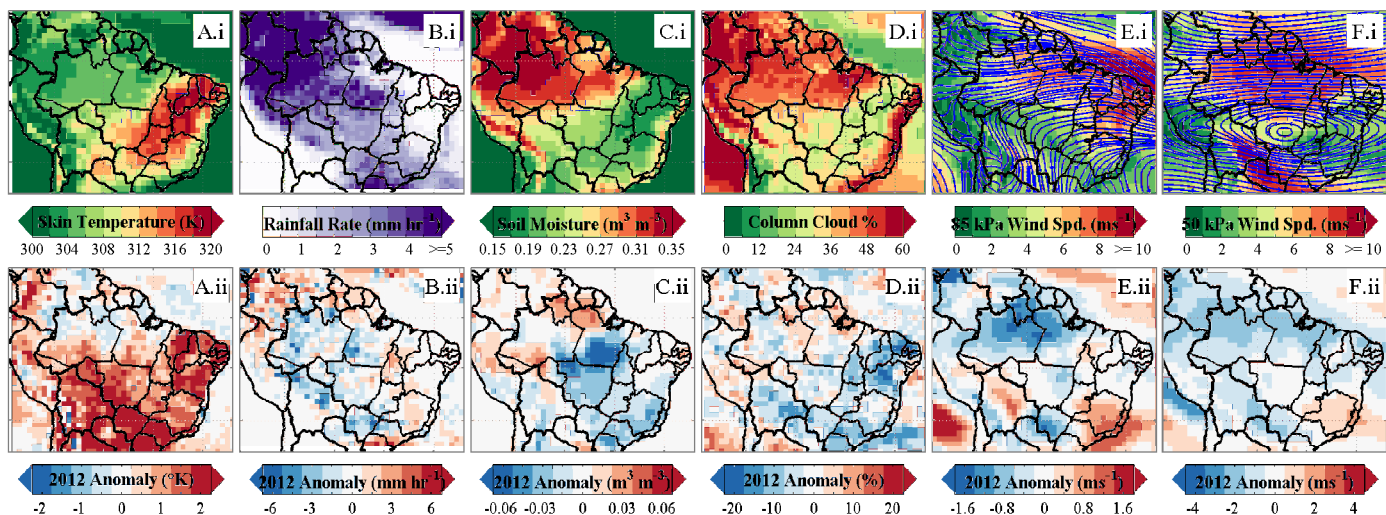


Figure S2. 2000-2014 climatology and 2012 anomaly of meteorological parameters. Skin temperature and cloud fraction are from the AIRS satellite instrument. Rainfall rates are from the TRMM satellite mission. Soil moisture and winds are from ECMWF reanalysis

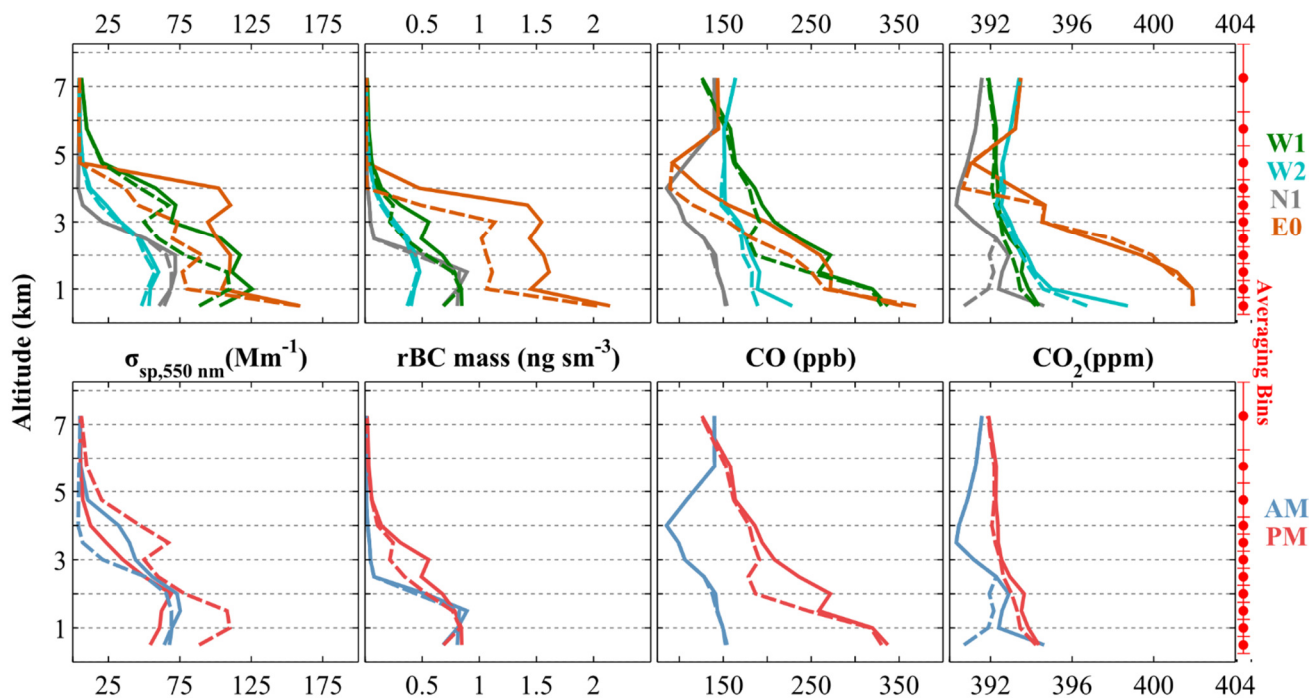


Figure S3. Pollutant profiles averaged over a range of sample scenarios with (dashed line) and without (solid line) plumes removed.

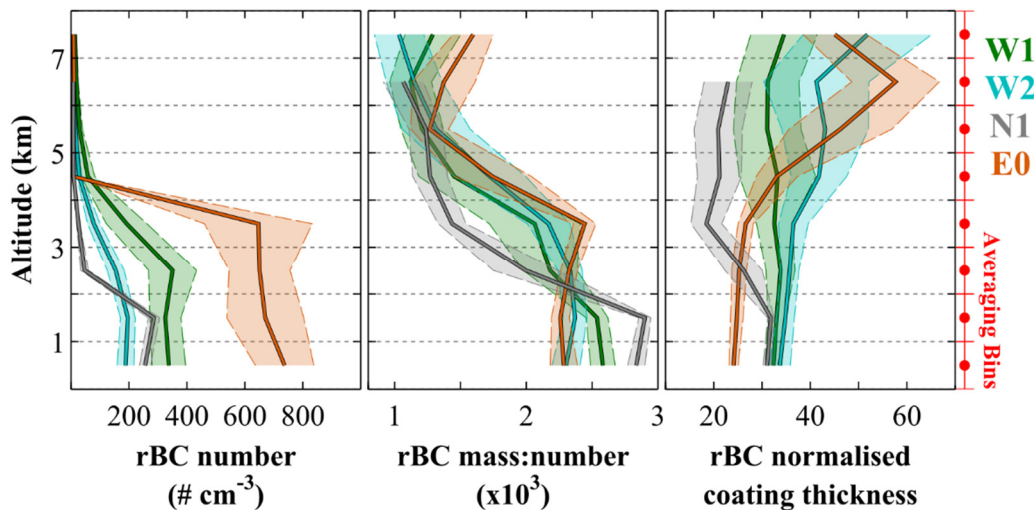


Figure. S4. Additional rBC properties by regime

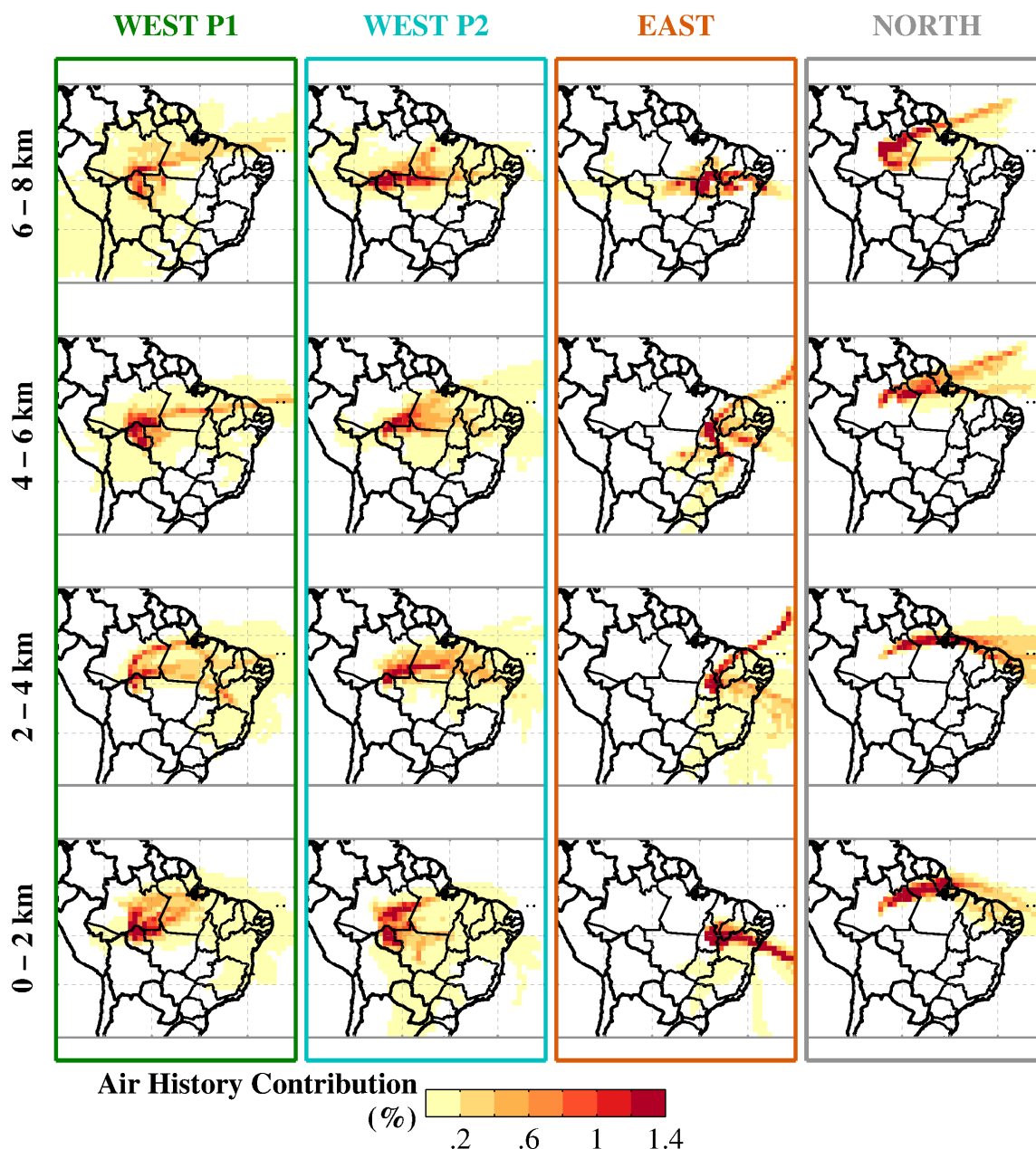


Figure. S5. History of air masses sampled during SAMBBA flights in each regime. Trajectories were released at 60 second intervals along each straight and level run and profile and calculated 120 hours backward. The occurrence in each 1° grid was calculated and normalised across the whole domain to produce % contributions. Trajectories were calculated using the HYSPLIT model (Stein et al., 2015) using GDAS (Global Data Assimilation System) meteorological fields

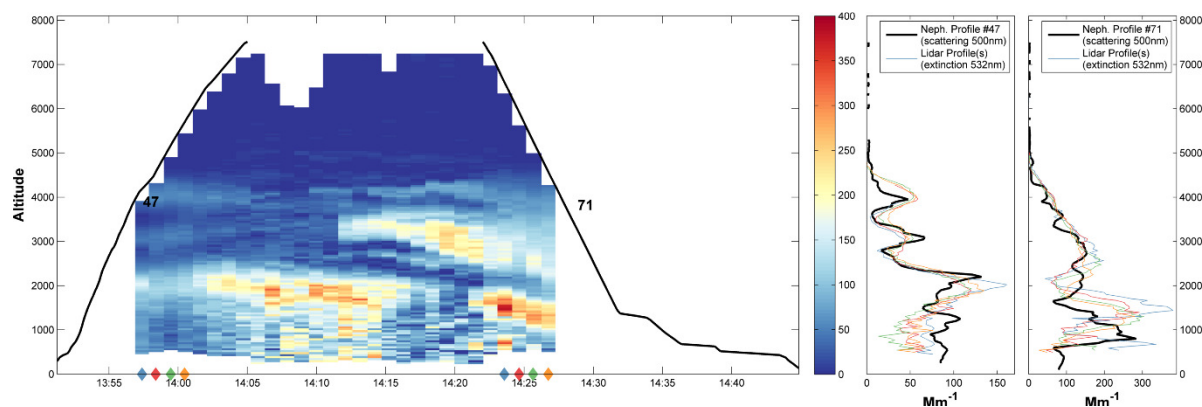
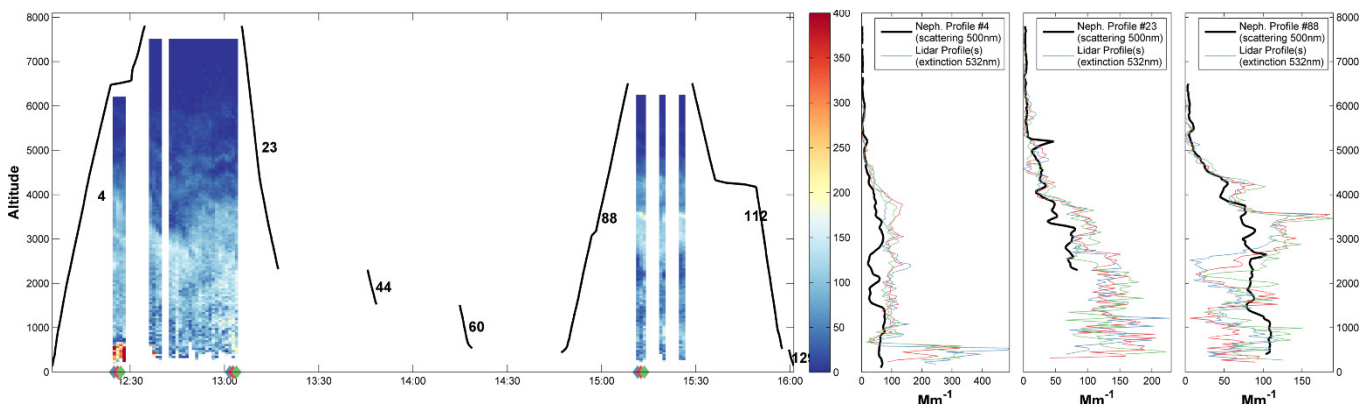


Figure S6.a. Lidar curtain (main panel) for flight b733. Side panel shows a comparison of in-situ σ_{sp_dry} to local Lidar extinction coefficients (at 532). σ_{sp_dry} is reported at ambient temperature and pressure, with the profile# corresponding to the library of profiles

(Sect S5). Colour of lidar profiles corresponds to the coloured markers at the base of the main panel. Lidar curtain reproduced with permission from (Marenco et al., 2016).



5 Figure S6.b. As (a) but flight b734

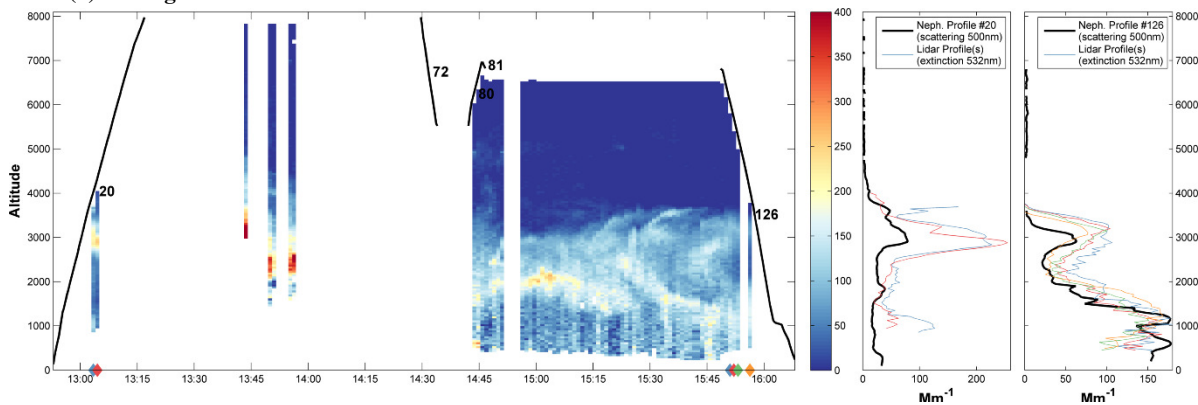
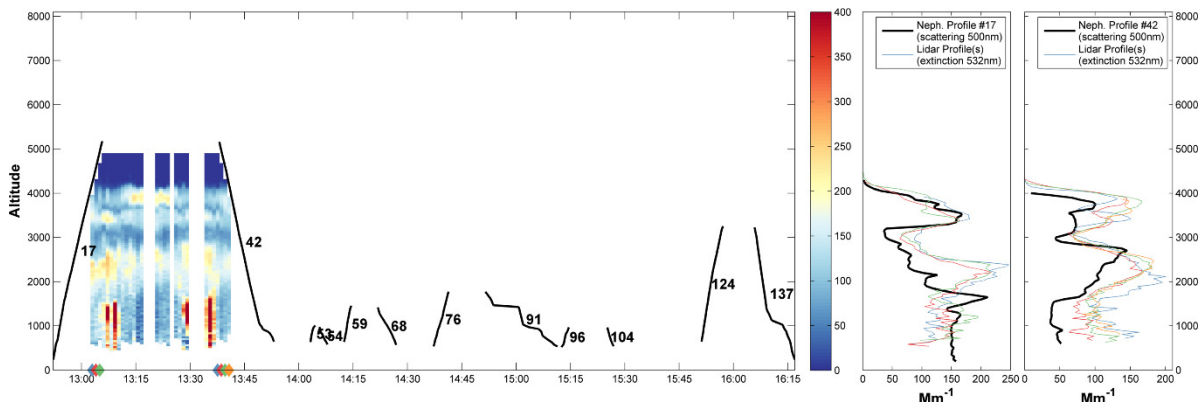


Figure S6.c. As (a) but flight b741



10 Figure S6.d. As (a) but flight b742

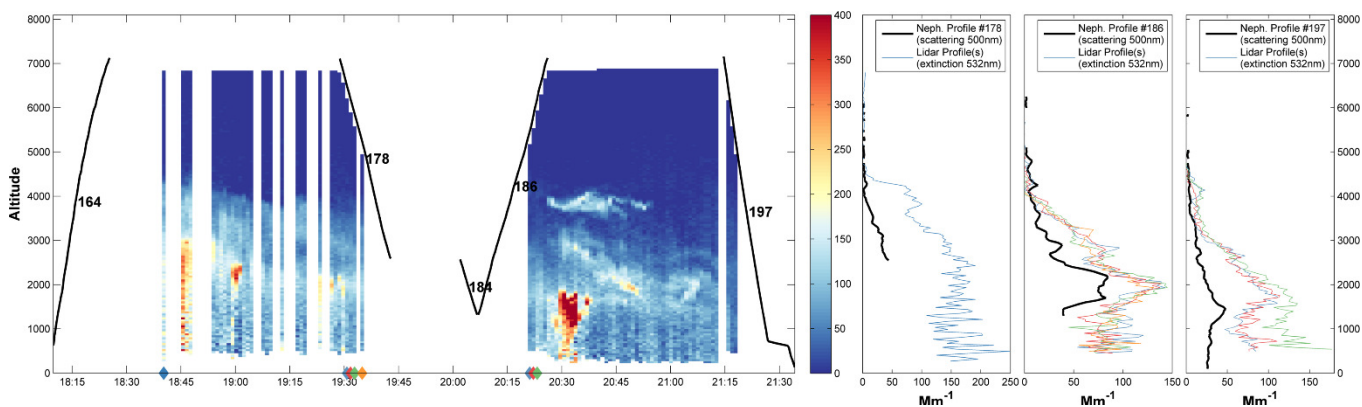


Figure S6.e. As (a) but flight b743

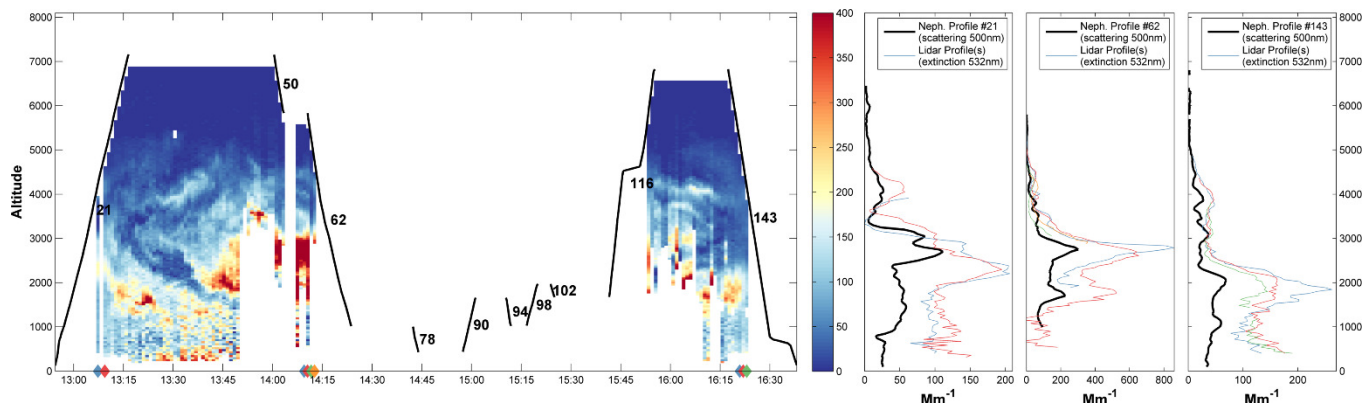
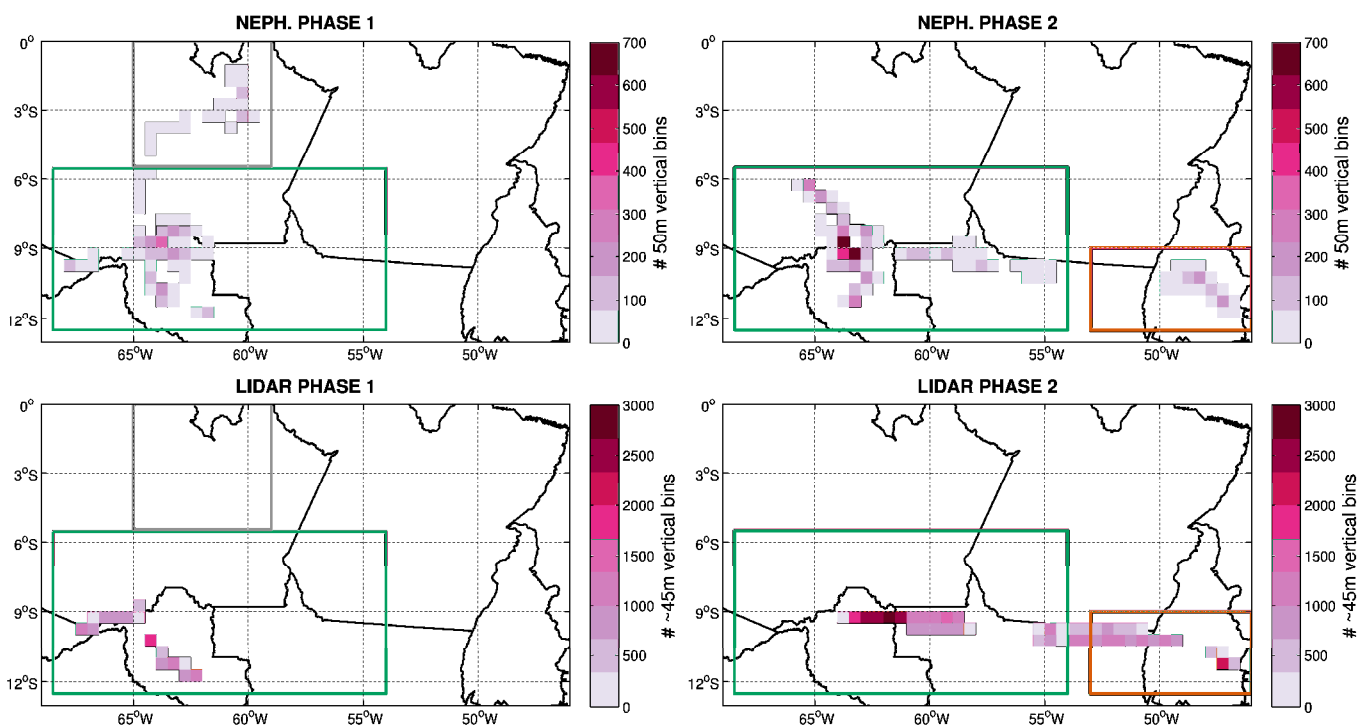


Figure S7.f. As (a) but flight b746



5 Figure S7. Geographic profile frequency plot on 1-degree grid during phase 1 (left) and phase 2 (right) for nephelometer (top) and lidar (bottom). East (orange), west (turquoise) and north (grey) domains displayed, the former two of which represent the regions averaged over for the profiles presented in Fig. 10.

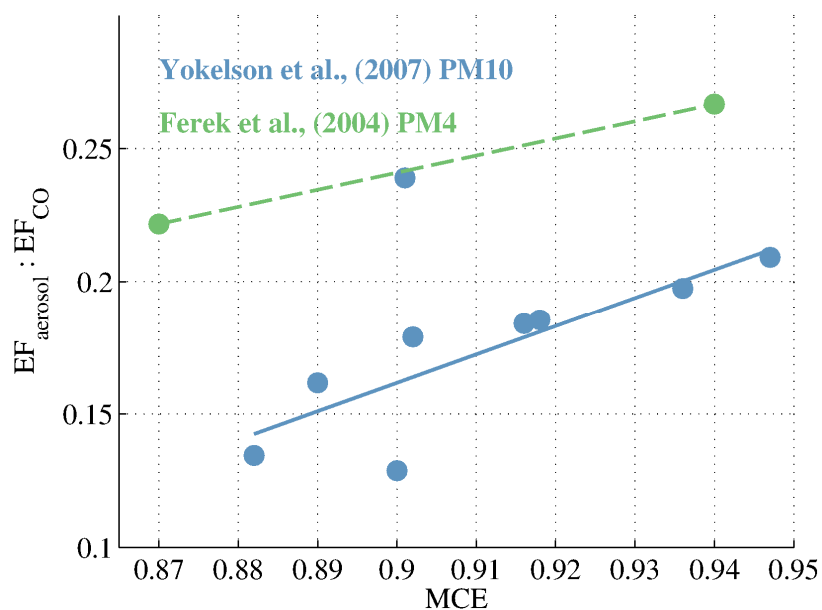


Figure S8. Emissions factors for aerosol:CO plotted against the MCE. Derived from past literature studies in tropical South America (Ferek et al., 1998; Yokelson et al., 2007).

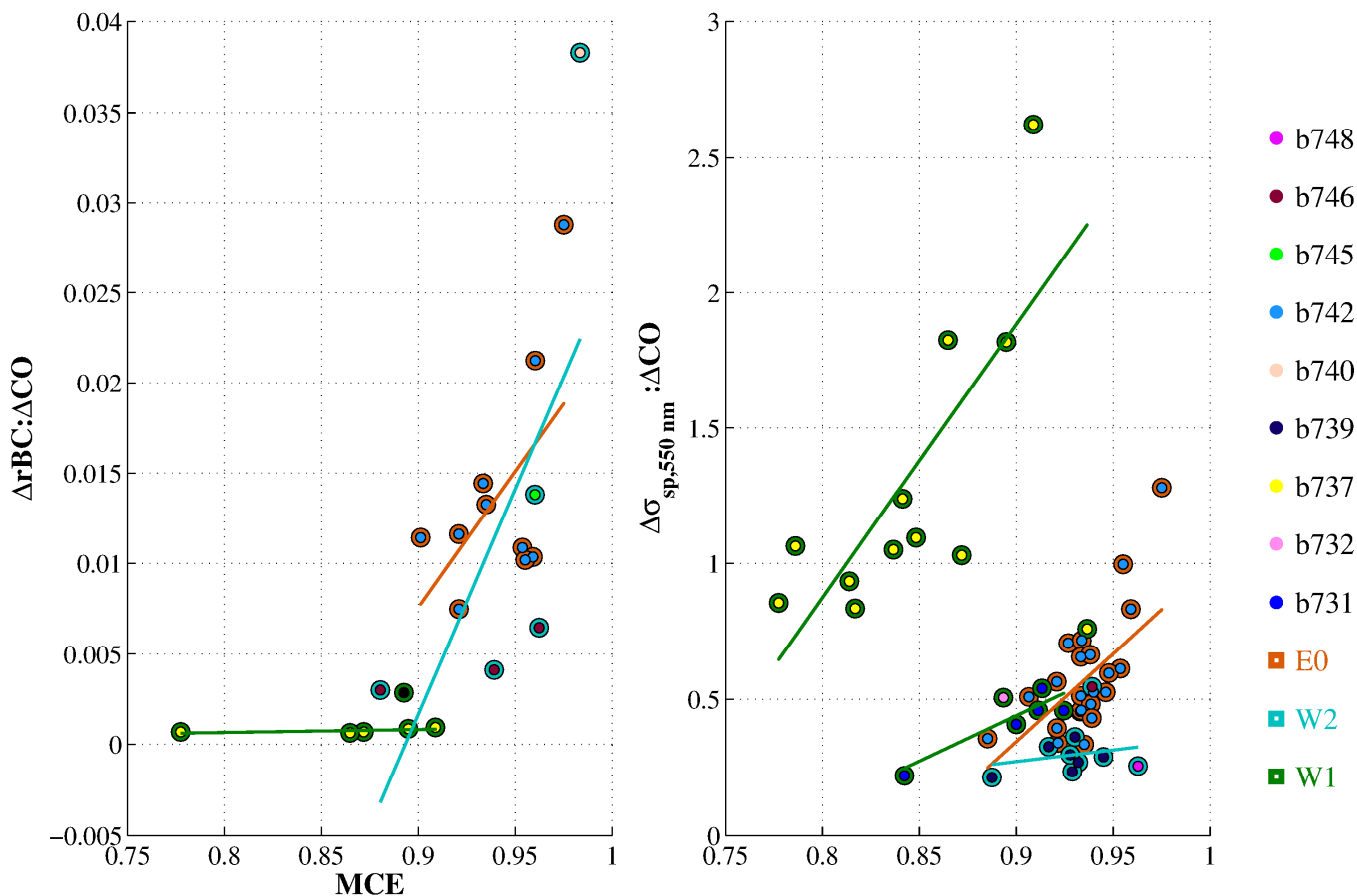


Figure S9. Enhancement factors calculated from straight and level plume intercepts during SAMBBA for a) rBC and b) $\sigma_{sp,dry}$ ratioed to CO against the modified combustion efficiency (MCE). Enhancement ratios and MCE calculated following the approach of (Hodgson et al., 2018). Relationships are modelled by robust multilinear regressions using iteratively reweighted least squares with a bisquare weighting function. These relationships are categorised by regime, although W1 is split into the contributions from flight b737 and other flights – this is because the single large deforestation wildfire sampled in b737 is likely unique for the region (Hodgson et al., 2018)

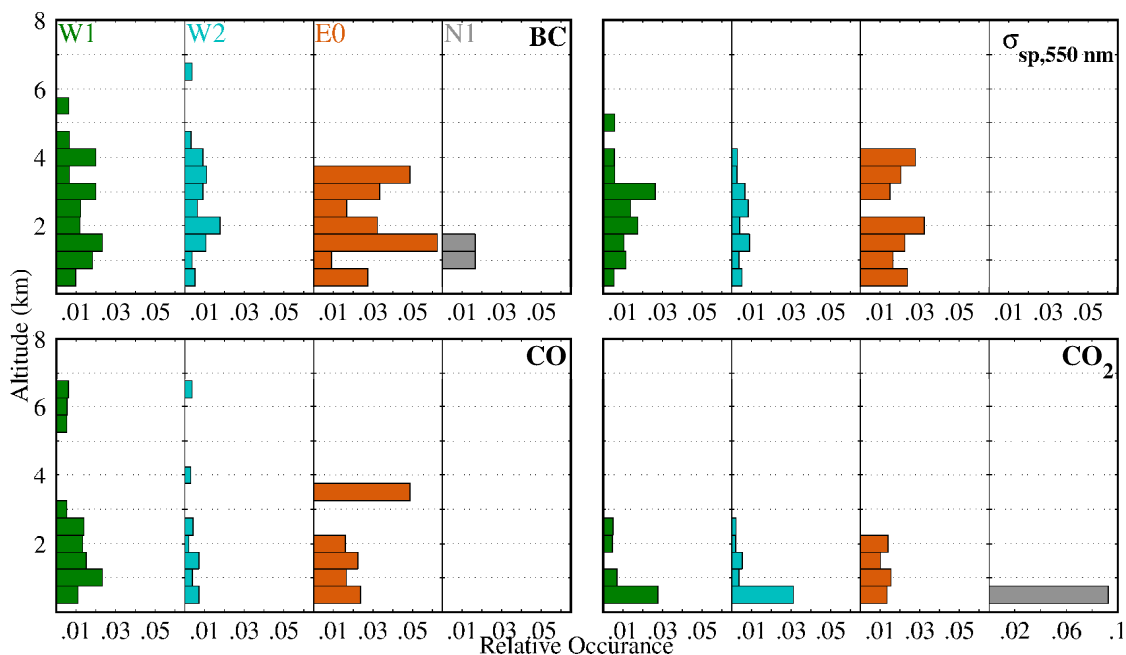


Figure. S10. Histograms of plume frequency by altitude, weighted by sample points

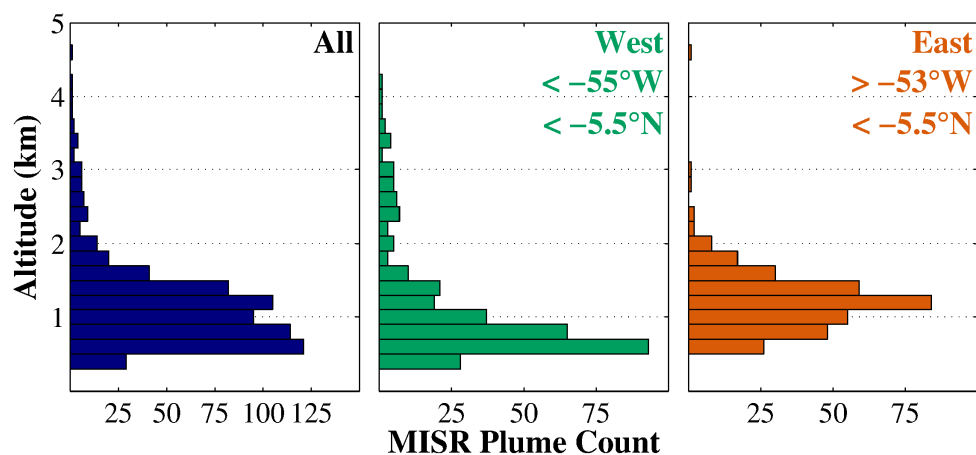


Figure S11. Plume injection height distribution across all the entire domain (left), west (middle) and east (right), from the MISR Plume Height Project for the Aug-Sept 2007 period (data acquired from <https://www-misr.jpl.nasa.gov/getData/accessData/MisrMinxPlumes/>)

References

- 5 Andreae, M. O., Artaxo, P., Beck, V., Bela, M., Freitas, S., Gerbig, C., Longo, K., Munger, J. W., Wiedemann, K. T. and Wofsy, S. C.: Carbon monoxide and related trace gases and aerosols over the Amazon Basin during the wet and dry seasons, *Atmos. Chem. Phys.*, 12(13), 6041–6065, doi:10.5194/acp-12-6041-2012, 2012.
- Artaxo, P., Rizzo, L. V., Brito, J. F., Barbosa, H. M. J., Arana, A., Sena, E. T., Cirino, G. G., Bastos, W., Martin, S. T. and Andreae, M. O.: Atmospheric aerosols in Amazonia and land use change: from natural biogenic to biomass burning conditions, *Faraday Discuss.*, 165(0), 203–235, doi:10.1039/C3FD00052D, 2013.
- 10 Chen, Y., Morton, D. C., Jin, Y., Collatz, G. J., Kasibhatla, P. S., van der Werf, G. R., DeFries, R. S. and Randerson, J. T.: Long-term trends and interannual variability of forest, savanna and agricultural fires in South America, *Carbon Manag.*, 4(6), 617–638, doi:10.4155/cmt.13.61, 2013.
- Dee, D. P., Uppala, S. M., Simmons, a. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. a., Balsamo, G., 15 Bauer, P., Bechtold, P., Beljaars, a. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, a. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N. and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, *Q. J. R. Meteorol. Soc.*, 137(656), 553–597, doi:10.1002/qj.828, 2011.
- 20 Ferek, R. J., Reid, J. S., Hobbs, P. V., Blake, D. R. and Liousse, C.: Emission factors of hydrocarbons, halocarbons, trace gases and particles from biomass burning in Brazil, *J. Geophys. Res. Atmos.*, 103(D24), 32107–32118, doi:10.1029/98JD00692, 1998.
- Fisch, G., Tota, J., Machado, L. a. T., Silva Dias, M. a. F., da F. Lyra, R. F., Nobre, C. a., Dolman, a. J. and Gash, J. H. C.: The 25 convective boundary layer over pasture and forest in Amazonia, *Theor. Appl. Climatol.*, 78(1–3), 47–59, doi:10.1007/s00704-004-0043-x, 2004.
- Hodgson, A. K., Morgan, W. T., O’Shea, S., Bauguitte, S., Allan, J. D., Darbyshire, E., Flynn, M. J., Liu, D., Lee, J., Johnson, B., Haywood, J. M., Longo, K. M., Artaxo, P. E. and Coe, H.: Near-field emission profiling of tropical forest and Cerrado fires in Brazil during SAMBBA 2012, *Atmos. Chem. Phys.*, 18(8), 5619–5638, doi:10.5194/acp-18-5619-2018, 2018.
- Marengo, F., Johnson, B., Langridge, J. M., Mulcahy, J., Benedetti, A., Remy, S., Jones, L., Szpek, K., Haywood, J., Longo, K. 30 and Artaxo, P.: On the vertical distribution of smoke in the Amazonian atmosphere during the dry season, *Atmos. Chem. Phys.*, 16(4), 2155–2174, doi:10.5194/acp-16-2155-2016, 2016.
- Rizzo, L. V., Artaxo, P., Müller, T., Wiedensohler, A., Paixão, M., Cirino, G. G., Arana, A., Swietlicki, E., Roldin, P., Fors, E. O., Wiedemann, K. T., Leal, L. S. M. and Kulmala, M.: Long term measurements of aerosol optical properties at a primary forest site in Amazonia, *Atmos. Chem. Phys.*, 13(5), 2391–2413, doi:10.5194/acp-13-2391-2013, 2013.
- 35 Seidel, D. J., Ao, C. O. and Li, K.: Estimating climatological planetary boundary layer heights from radiosonde observations: Comparison of methods and uncertainty analysis, *J. Geophys. Res.*, 115(D16), D16113, doi:10.1029/2009JD013680, 2010.
- Stein, A. F., Draxler, R. R., Rolph, G. D., Stunder, B. J. B., Cohen, M. D. and Ngan, F.: NOAA’s HYSPLIT Atmospheric Transport and Dispersion Modeling System, *Bull. Am. Meteorol. Soc.*, 96(12), 2059–2077, doi:10.1175/BAMS-D-14-00110.1, 2015.
- 40 Wang, X. Y. and Wang, K. C.: Estimation of atmospheric mixing layer height from radiosonde data, *Atmos. Meas. Tech.*, 7(6), 1701–1709, doi:10.5194/amt-7-1701-2014, 2014.

Yokelson, R. J., Karl, T., Artaxo, P., Blake, D. R., Christian, T. J., Griffith, D. W. T., Guenther, A. and Hao, W. M.: The Tropical Forest and Fire Emissions Experiment: overview and airborne fire emission factor measurements, *Atmos. Chem. Phys.*, 7(19), 5175–5196, doi:10.5194/acp-7-5175-2007, 2007.

5 **S5. Individual Profile Library**

Each profile is displayed in a similar format to Fig. 3, in order of time of day sampled.

