Ozone Trends and Mortality Risk: The Growing Need for Machine Learning Predictions in Bogotá, Colombia

Supplementary Material

Daniela Bustos[1], Diana Garcia[1], Nestor Y. Rojas[1], Ellie A. Lopez-Barrera[2], Carlos Peña[3], Alejandro Casallas[4,5]

[1] Universidad Nacional de Colombia, Bogotá, Colombia

[2] Universidad Sergio Arboleda, Instituto de Estudios y Servicios Ambientales - IDEASA, Bogotá, Colombia

[3] Universidad Sergio Arboleda, Escuela de Ciencias Exactas e Ingeniería, Bogotá, Colombia

[4] Institute of Science and Technology Austria, Klosterneuburg, Austria

[5] Max Planck Institute for Chemistry, Mainz, Germany

Stations	Pollutants						Meteorological variables				Time Resolution and Period
	O_3	NO	NO_2	PM_{10}	$PM_{2.5}$	CO	Wind speed	Temp	Solar rad	HR	
Bolivia	✓	✓	✓	✓	✓	✓					Hourly (2021-2023)
Bosa		✓	✓	✓	✓	✓		✓	✓		Hourly
Carvajal	✓	✓	✓	✓	✓	✓	✓	✓			Hourly (2013-2023)
Centro de Alto Rendimiento (CDAR)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	Hourly (2013-2023)
Colina	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	Hourly
Ciudad Bolívar	✓	✓	✓	✓	✓	✓	✓	✓		✓	Hourly (2020-2023)
Fontibón	✓	✓	✓	✓	✓	✓	✓	✓		✓	Hourly (2018-2023)
Guaymaral	✓	✓	✓	✓	✓	✓		✓	✓	✓	Hourly (2013-2023)
Jazmín	✓	✓	✓	✓	✓	✓	✓	✓	✓		Hourly (2020-2023)
Kennedy	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	Hourly (2014-2023)
Las Ferias	✓	✓	✓	✓	✓	✓	✓	✓		✓	Hourly (2013-2023)
Min Ambiente	✓			✓	✓	✓	✓				Hourly (2013-2023)
Móvil Fontibón	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	Hourly (2021-2023)
Puente Aranda	✓	✓	✓	✓	✓	✓		✓			Hourly (2013-2023)
San	✓	✓	✓	✓	✓		✓	✓	✓	✓	Hourly (2013-2023)

Cristóbal											
Suba	✓	✓	✓	✓	✓	✓	✓	✓			Hourly (2013-2023)
Tunal	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	Hourly (2013-2023)
Usaquén	✓	✓	✓	✓	✓	✓	✓	✓			Hourly (2013-2023)
Usme	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	Hourly (2020-2023)

Table SM1. Coverage of pollutant and meteorological variables by monitoring station in Bogotá (2013–2023).

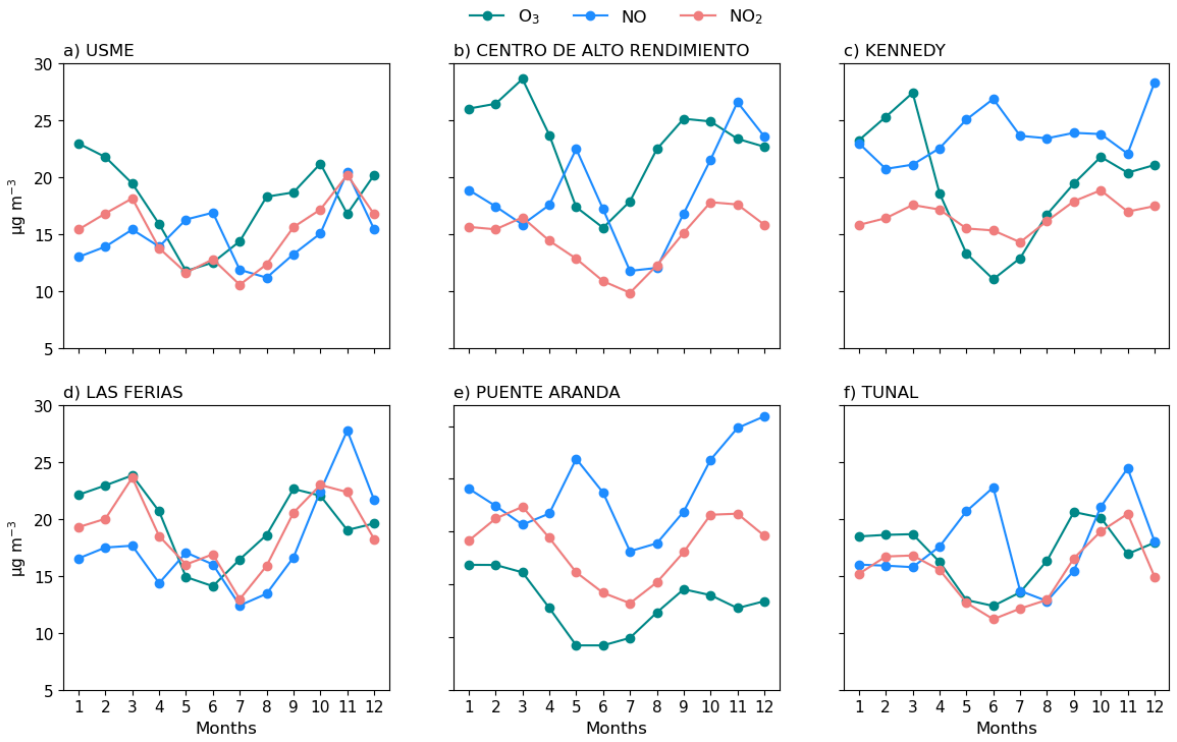

Figure SM1. Multi-year monthly averages of O_3, NO, and NO_2 behavior

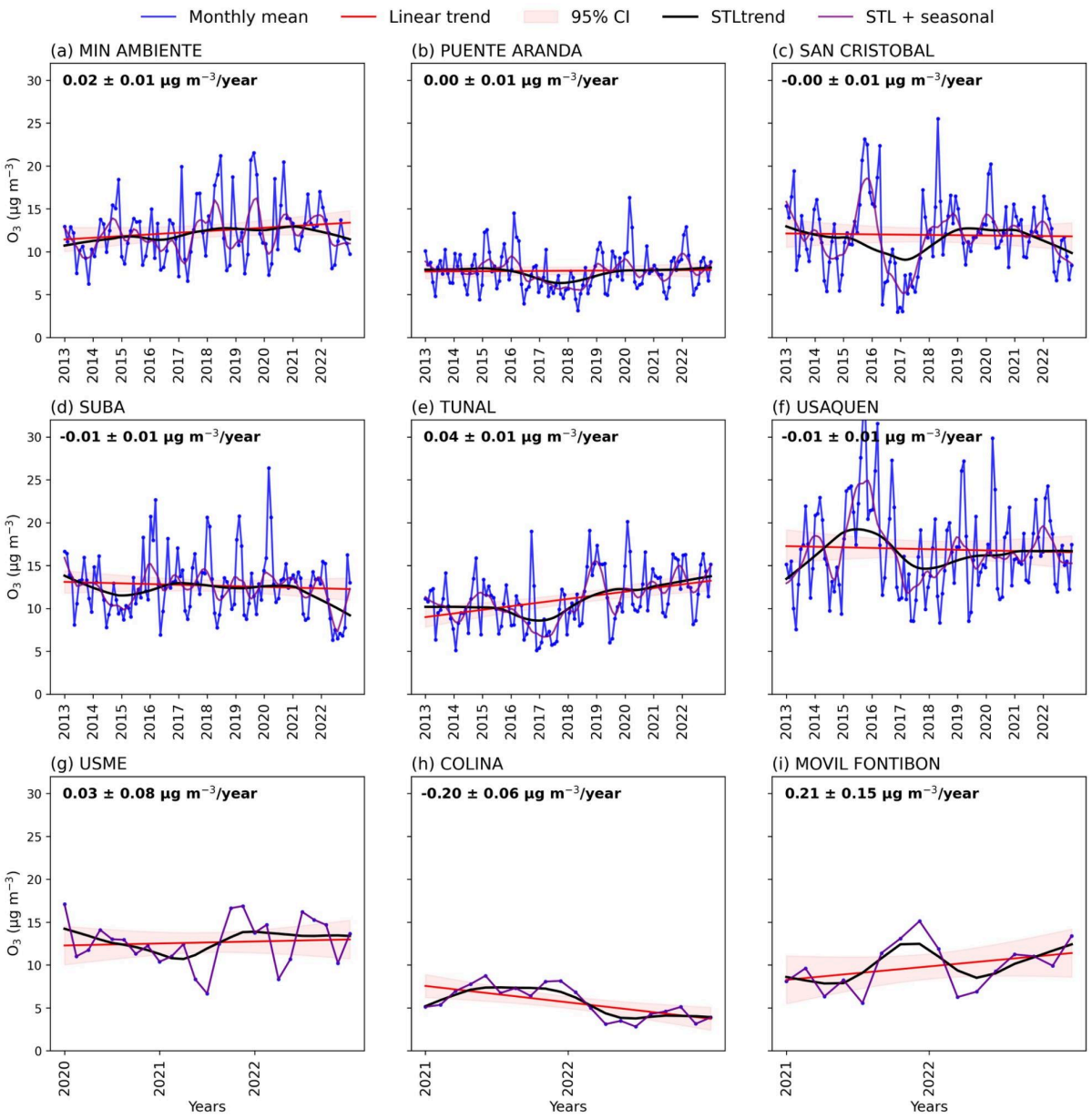

Figure SM2. Multi-annual trend of monthly ozone levels calculated from the OSMDA8 in Bogotá, obtained by STL decomposition and GLS regression. The purple line represents the seasonal component, the black line indicates the long-term trend, and the red line shows the GLS-adjusted slope reflecting the general linear trend of ozone over 2013-2023 (dependent on data availability), with the red shading depicting the 95% confidence interval (CI).

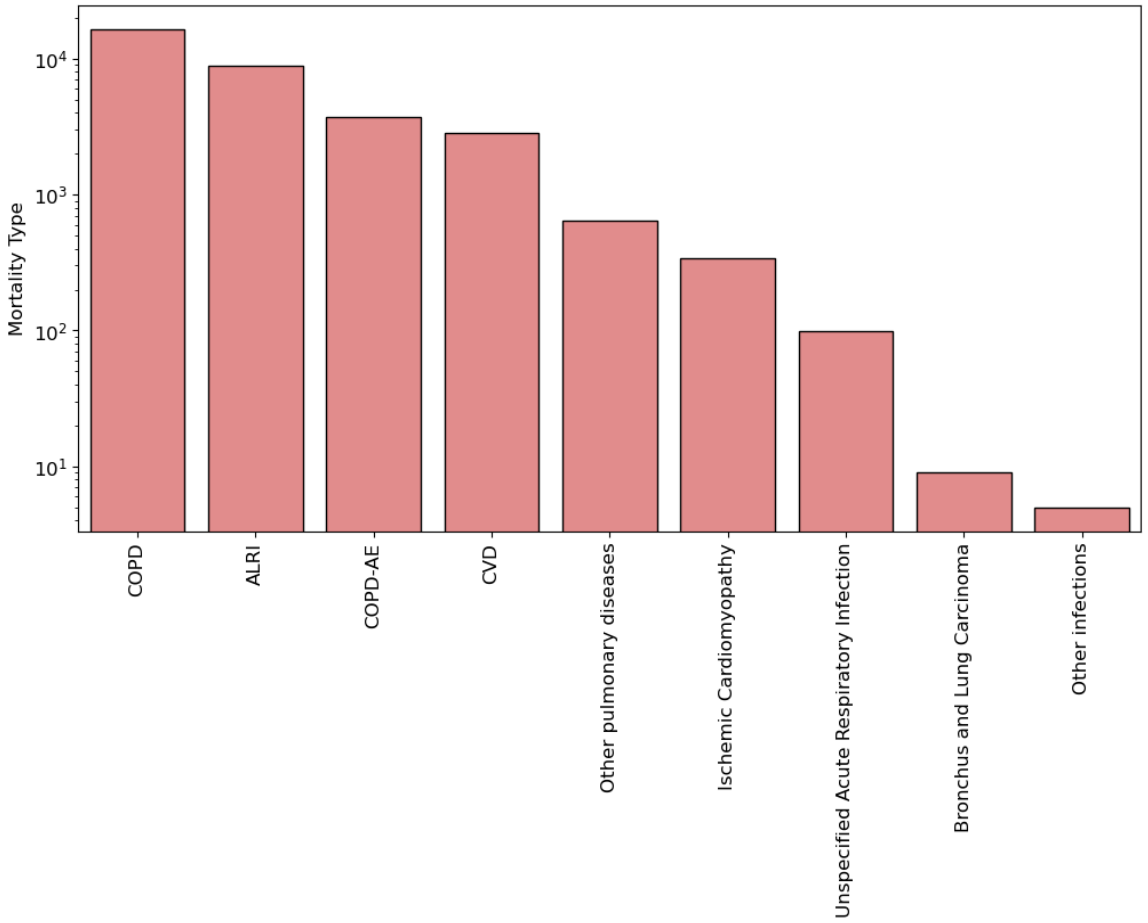

Figure SM3. Distribution of specific causes of mortality attributable to ozone in Bogotá. ARLI:Acute Lower Respiratory Infection, COPD: Chronic Obstructive Pulmonary Disease, COPD AE: acute exacerbation of COPD, CVD: Cardiovascular Disease Other pulmonary diseases , Ischemic Cardiomyopathy , Unspecified Acute Respiratory Infection, Bronchus and Lung Carcinoma , Other infections.

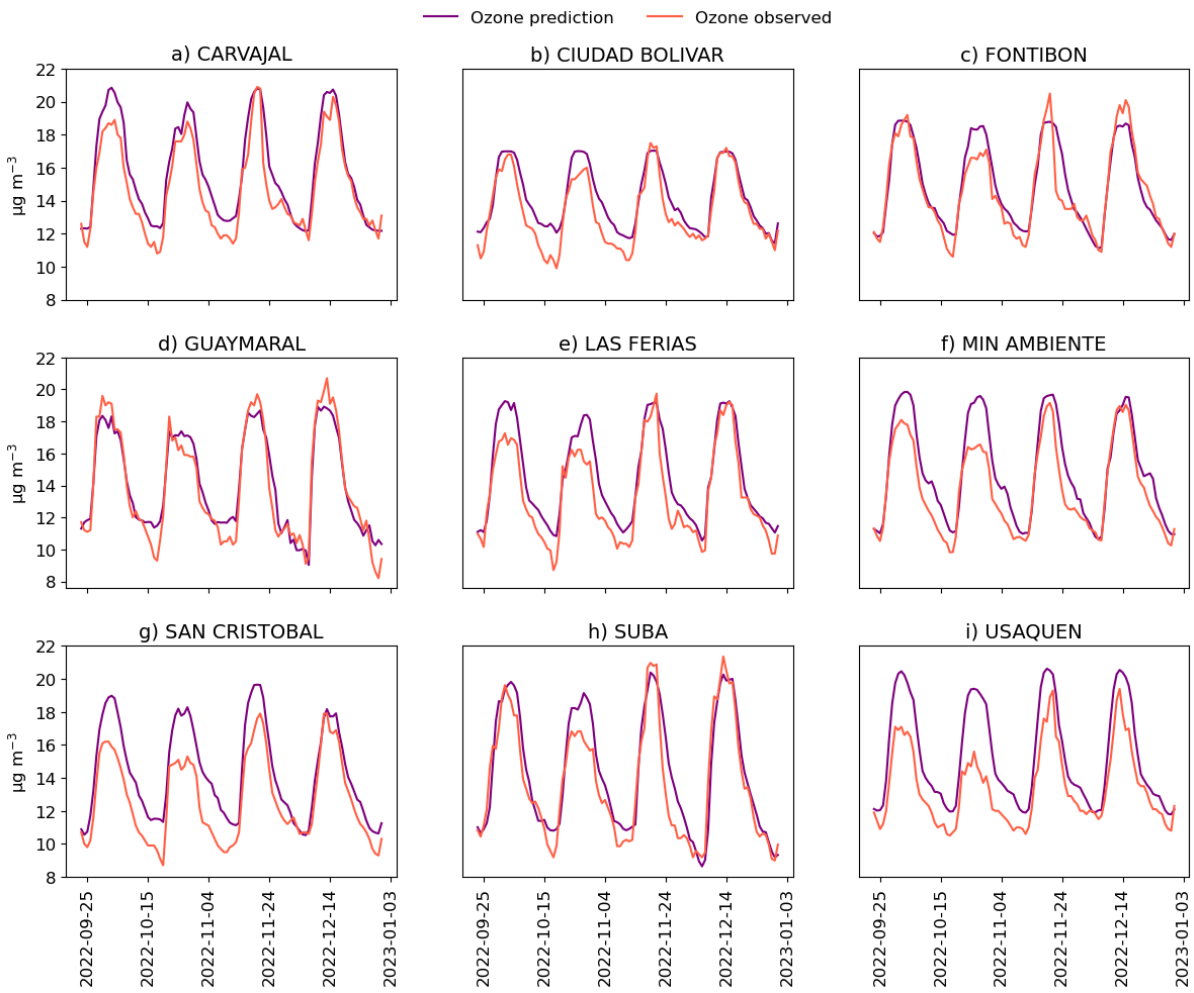

Figure SM4. Comparison between observed and predicted ozone concentration values at the other stations with the largest datasets in the RMCAB. The purple line represents predicted ozone, and the yellow line represents observed ozone.

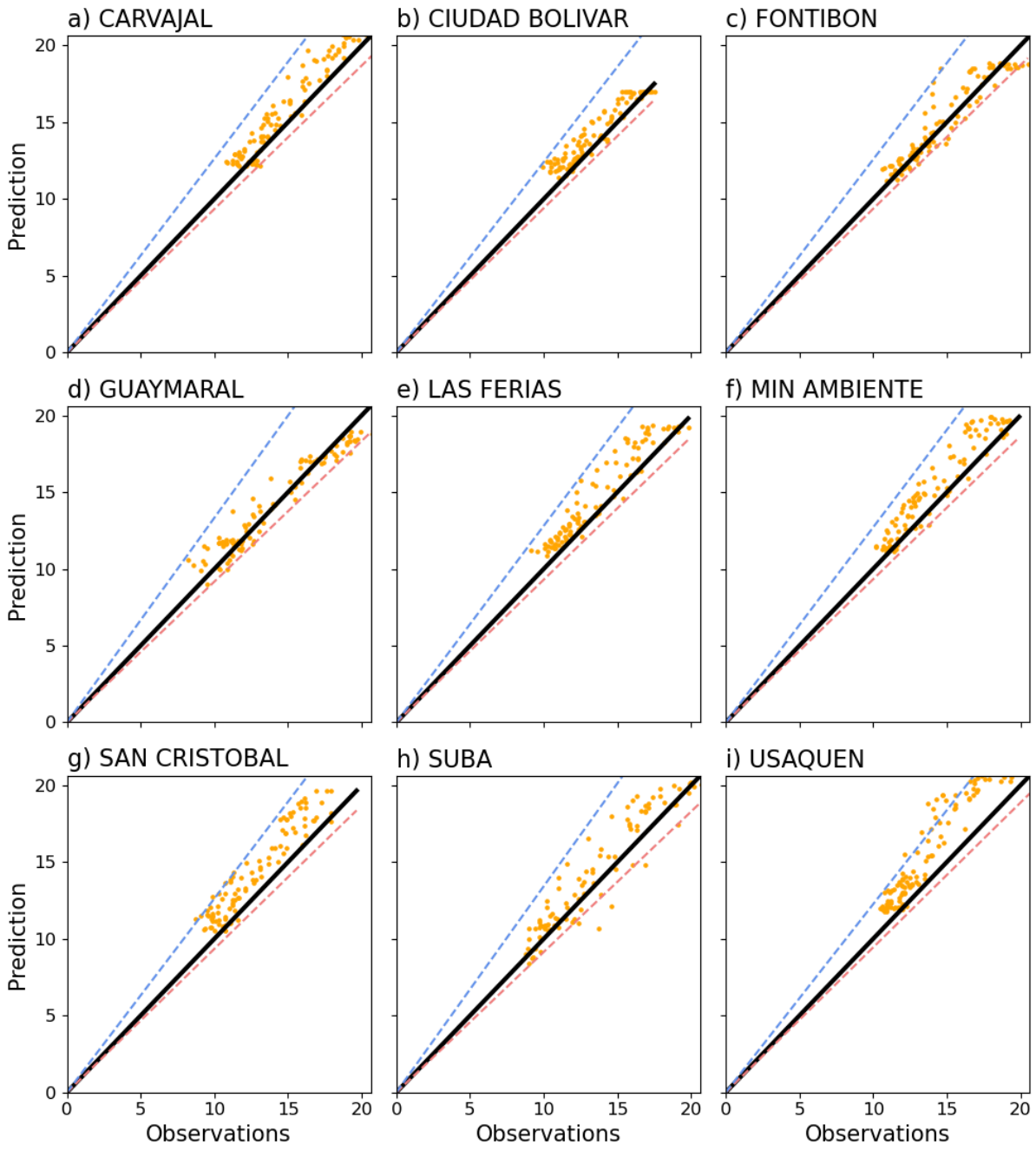

Figure SM5. Scatter plot of ozone observations versus predictions (units of µg/m³) from stations that do not measure all the variables required as inputs to the model. The black line corresponds to the 1:1 linear regression; the blue and red dotted lines correspond to 1:2 and 1:0.5, respectively.

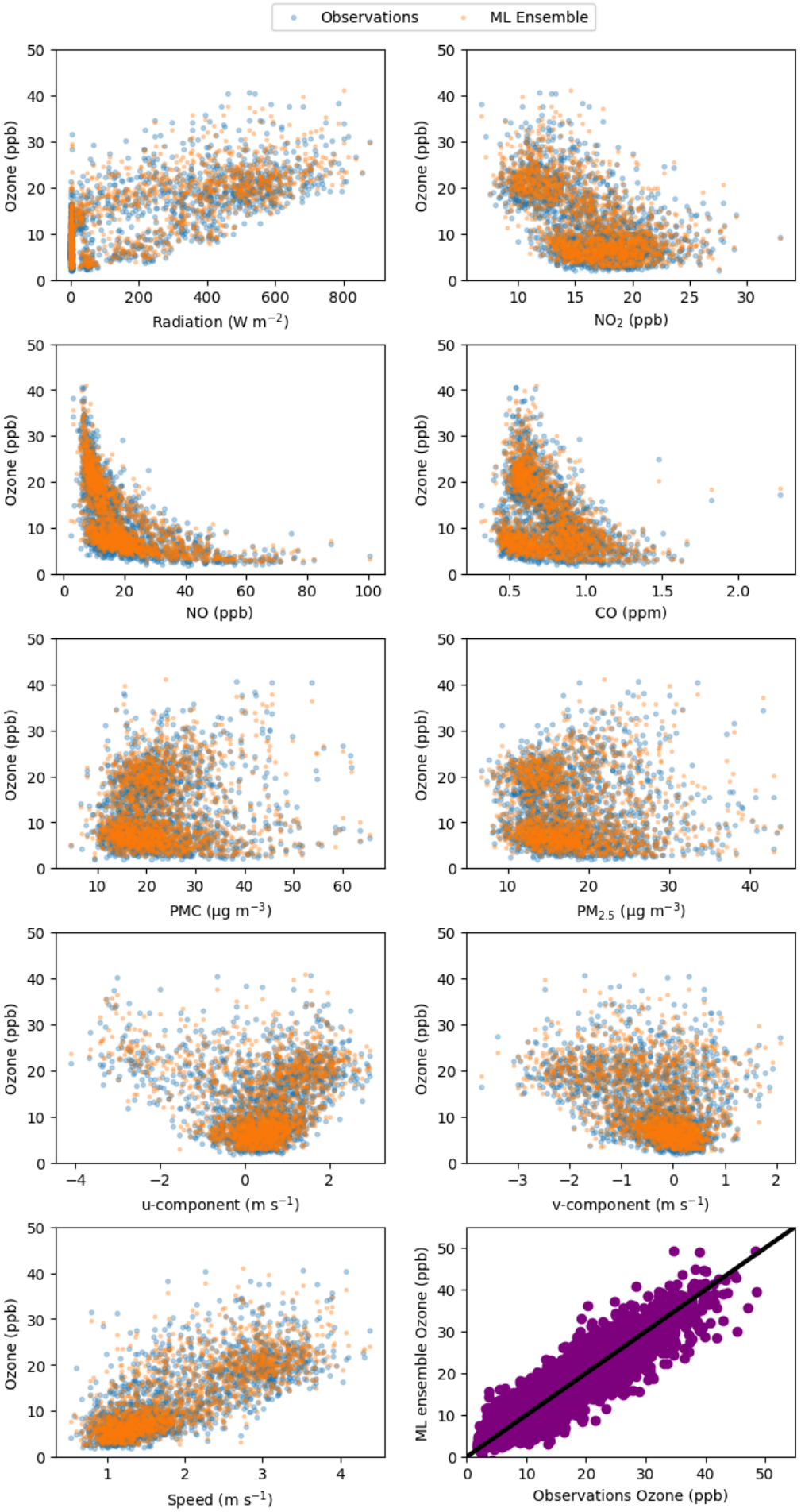
Observations
ML Ensemble
Ozone (ppb)
Radiation (W m^{-2})
NO_2 (ppb)
NO (ppb)
CO (ppm)
PMC (µg m^{-3})
$PM_{2.5}$ (µg m^{-3})
u-component (m s^{-1})
v-component (m s^{-1})
Speed (m s^{-1})
ML ensemble Ozone (ppb)
Observations Ozone (ppb)

Figure SM6. Data dispersion for surface O_3 in Bogotá based on the ML ensemble means for a subset of 2000 randomly subsampled data points from the validation dataset. O_3 is plotted as a function of radiation, NO_2, NO, CO, PMC, $PM_{2.5}$, wind components (u and v), and wind speed. The lower right panel plots the ozone observation versus the ML ensemble mean.