Supplementary Information
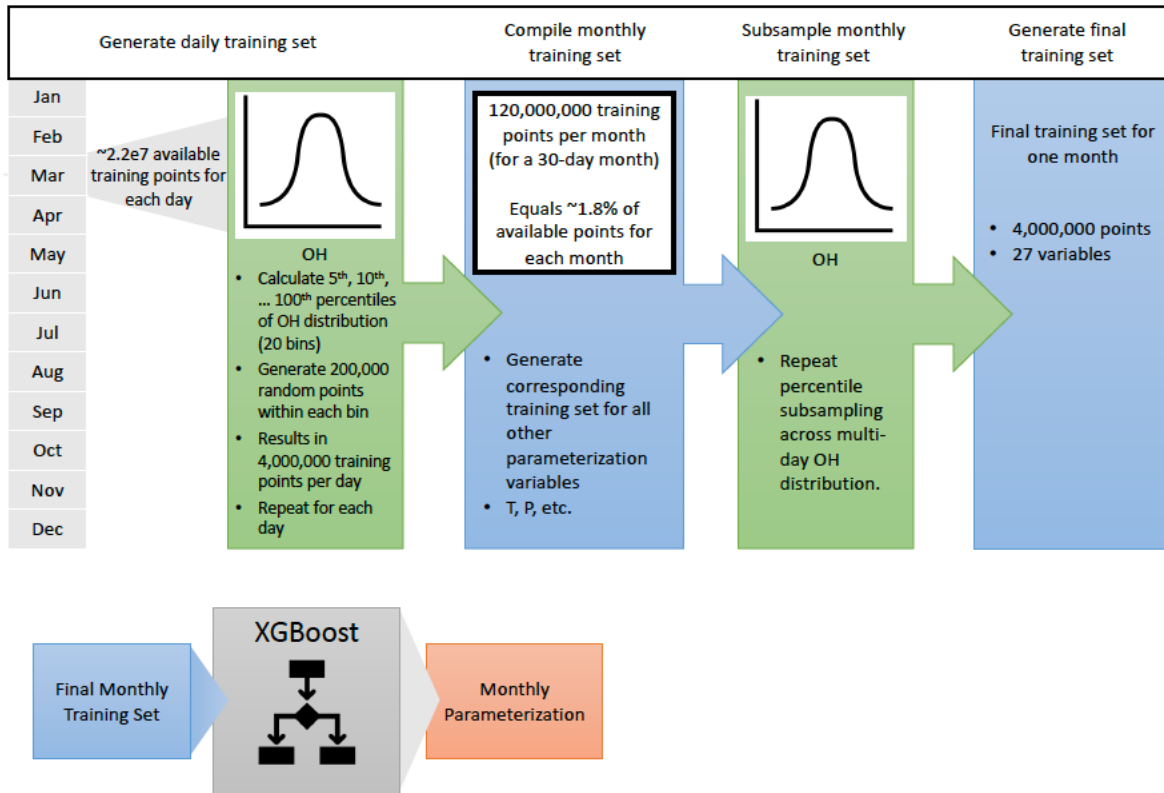
Figure S1: A schematic showing the process used to generate the training set for the parameterization of OH and the parameterization itself.
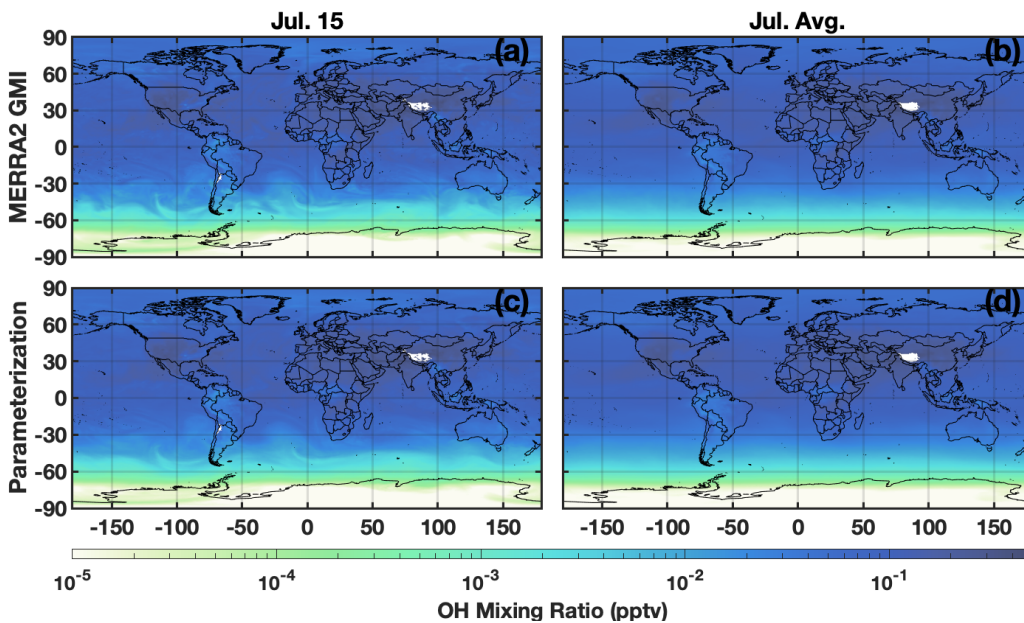
Figure S2: Mixing ratio of OH averaged over the LFT (top of the PBL to 500 hPa) for July 15th, 2005 (left) and averaged across all days in July 2005 (right). OH from the MERRA2 GMI simulation is shown on the top, and OH from the parameterization is shown on the bottom.
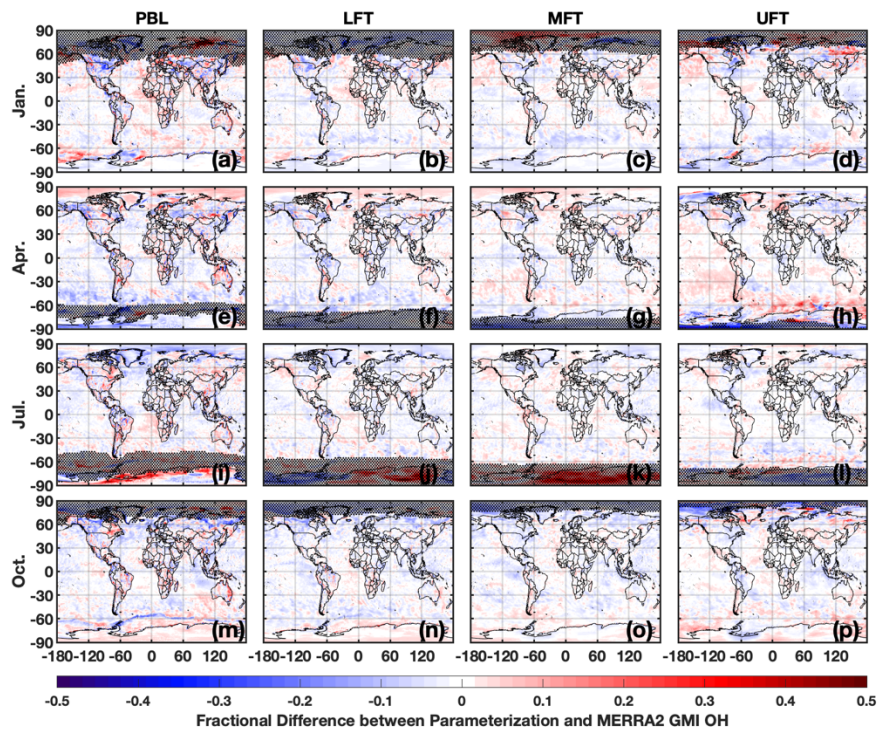
Figure S3: Fractional difference between the parameterized OH and OH from the MERRA2 GMI simulation averaged across all days for January, April, July, and October 2005 for the PBL, LFT, MFT, and UFT. Regions with low OH, defined as a mixing ratio of less than 0.005 pptv, are indicated with stippling.
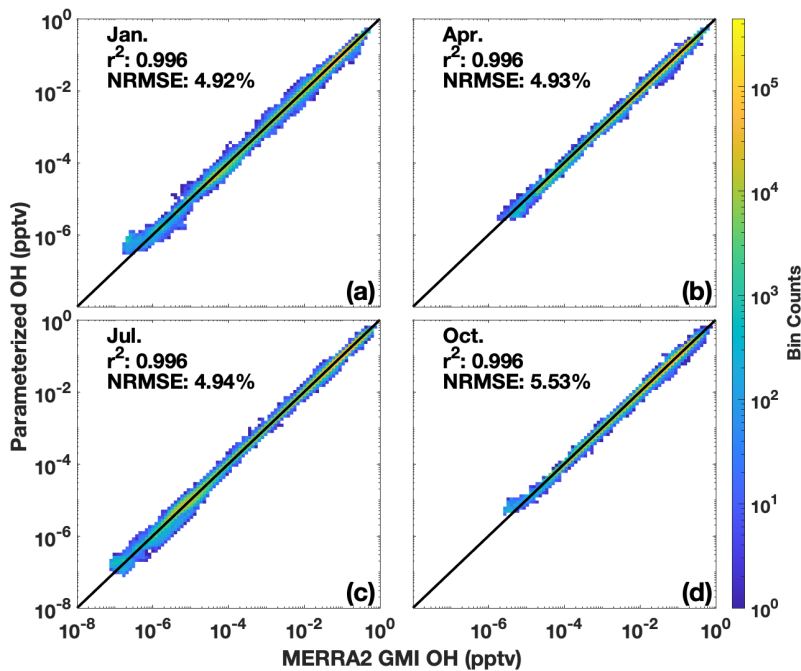
Figure S4: Scatter density plot of OH from the MERRA2 GMI simulation plotted against OH calculated by the parameterization for Jan. (a), Apr. (b), Jul. (c), and Oct. (d) 2005. Colors indicate the number of data points in each bin. The r^2 of a linear least squares regression and the NRMSE for each month are also indicated.
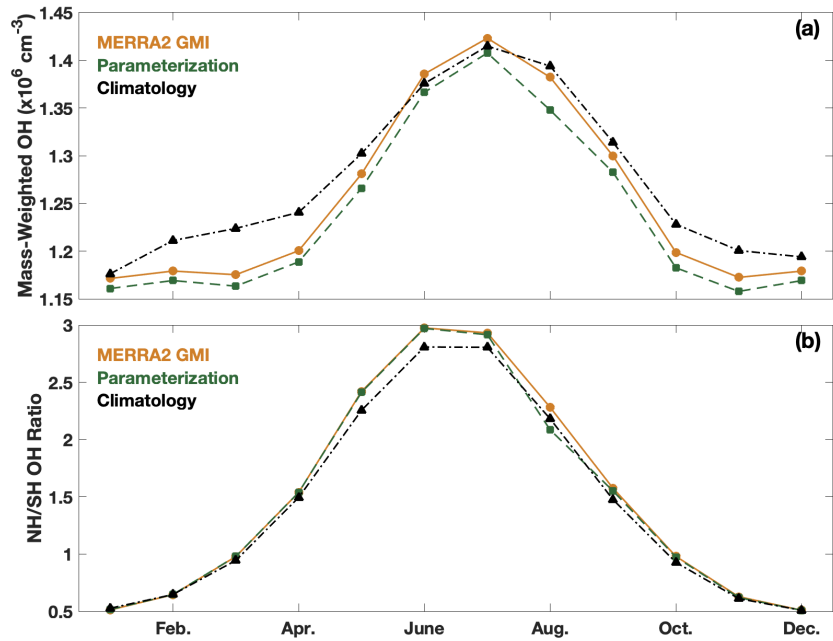
Figure S5: Global mean mass-weighted OH (a) and the ratio of Northern to Southern Hemispheric OH (b) as calculated by the parameterization (green squares) and MERRA2 GMI (orange circles) for 2005. The climatological mean calculated from MERRA2 GMI for 1980 – 2019 is also indicated by the black triangles.
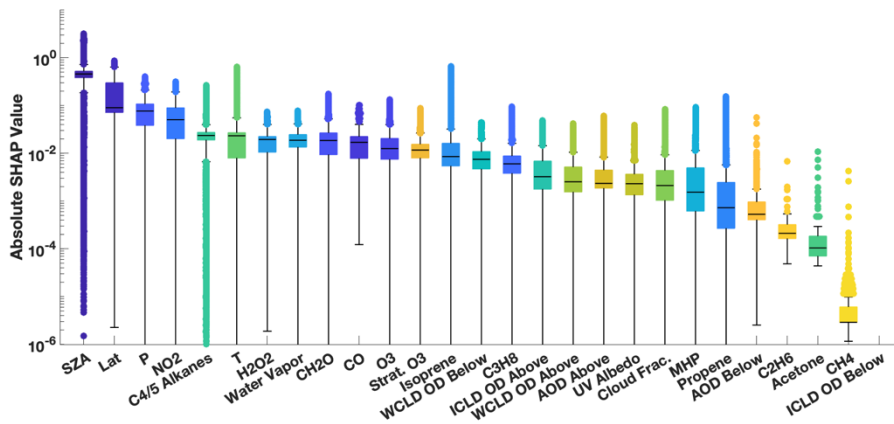
Figure S6: Distribution of the absolute SHAP value for each parameterization input for January from an untuned version of the parameterization. Species are sorted by order of relative importance. The median is indicated with the black line, edges of the box represent the interquartile range, and whiskers represent the 5th and 95th percentile. Values outside this range are indicated with circles.
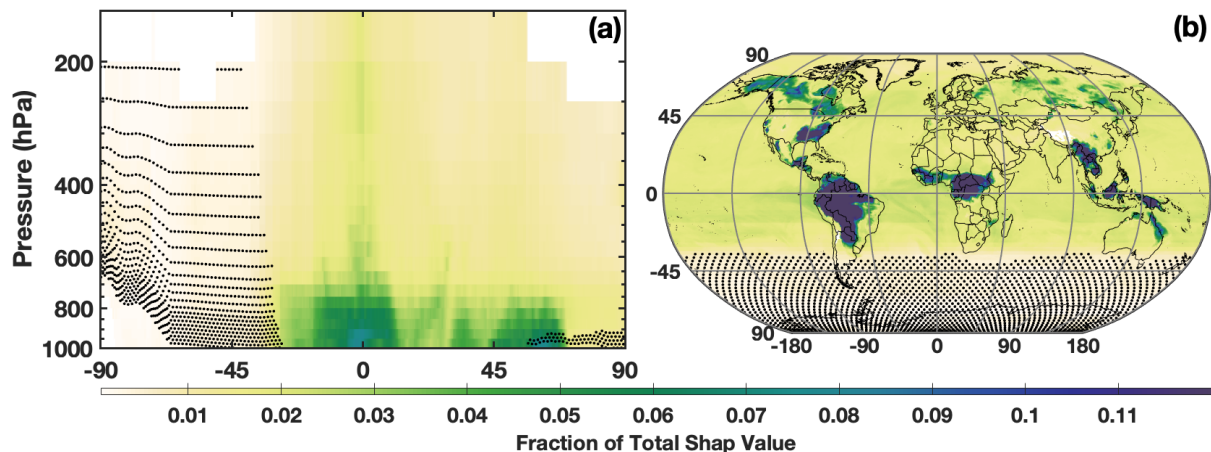
Figure S7: The fraction of the contribution of the Isoprene SHAP value to the sum of the absolute SHAP value of all inputs is shown for the zonal mean (a) and the LFT (b). Regions where mean OH mixing ratios are below 0.03 pptv, the point below which the untuned parameterization is unable to reasonably predict OH, are indicated by the hatching.
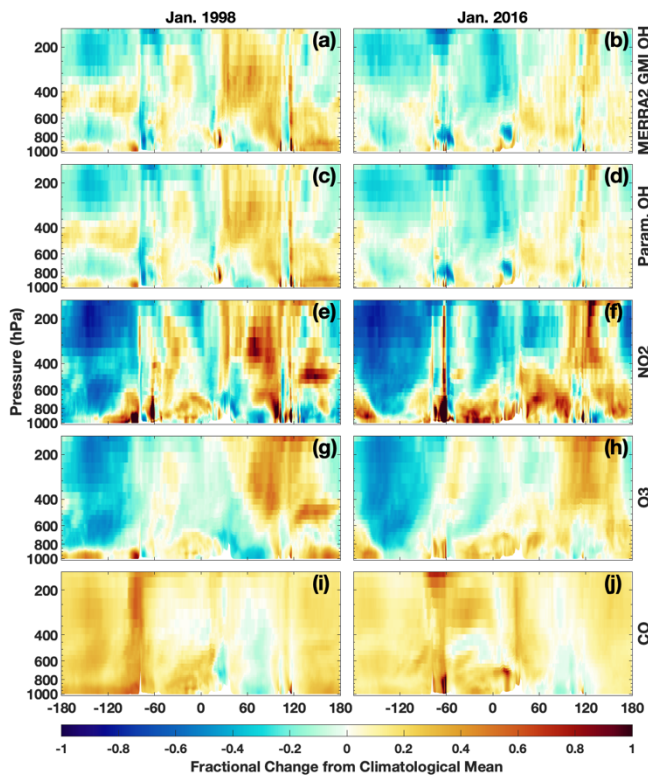
Figure S8: Fractional difference in the indicated variable between January 1998 (left) and the climatological mean (1980 – 2019) calculated from the MERRA2 GMI run for a zonal slice along the equator. Analogous plots for January 2016 are shown on the right. Species shown are OH from MERRA2 GMI (a,b), OH calculated from the parameterization (c,d), NO_2 (e,f), O_3 (g,h) and CO (i,j).
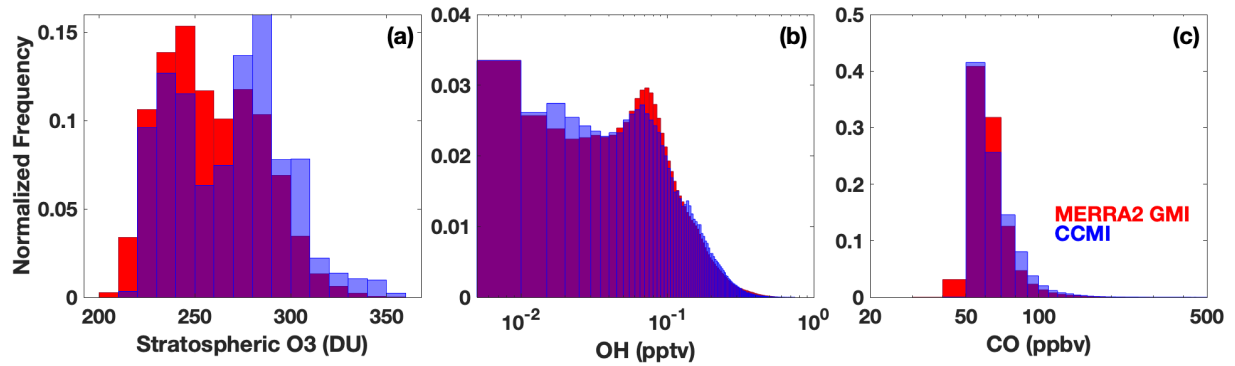
Figure S9: Histograms showing the distribution of stratospheric column O$_3$ (a), tropospheric OH (b) and tropospheric CO (c) from the MERRA2 GMI parameterization training dataset (red) and from the CCMI simulation for 2000 (blue). _Purple indicates areas of overlap between the two distributions._

Table S1: The NRMSE and τ_{CH4} from MERRA2 GMI and the parameterization from January and July for years omitted from the training dataset.

Year	January			July		
	NRMSE (%)	Parameterization τ_{CH4} (years)	M2GMI τ_{CH4} (years)	NRMSE (%)	Parameterization τ_{CH4} (years)	M2GMI τ_{CH4} (years)
1985	5.4	8.74	8.68	4.8	6.86	6.79
1995	4.8	8.55	8.47	5.6	6.62	6.50
2005	4.9	8.39	8.33	4.9	6.61	6.53
2015	4.9	8.38	8.34	5.2	6.82	6.78