S1. Intercomparison of NO_2 vertical column densities

We intercompare the magnitudes of tropospheric NO_2 vertical column densities (VCDs) from the NASA, the DOMINO and the QA4ECV retrievals in Figure S1. Air mass factors (AMFs) for each retrieval product are calculated using the GEOS-Chem vertical profiles and the scattering weights (w, for NASA product) or averaging kernels (A, for DOMINO and QA4ECV products, where $A = w/\text{AMF}$) from each product following equations 1 – 4 in Qu et al. [2017]. The VCDs in Figure S1 are calculated using the tropospheric NO_2 slant column densities from each product and the corresponding AMFs calculated above to minimize the impact of differences in prior simulations of these retrievals.

OMI NO_2 VCDs show the largest magnitude in the DOMINO product, and the smallest magnitude in the NASA product. The NO_2 VCDs from the GC-adj simulations are larger than the OMI NASA retrieval in eastern China, the eastern US, and parts of Europe. The differences between simulations and observations are used to drive the NO_x emission inversion, and they suggest reductions of emissions in these regions. The differences between the GC-adj simulations and the DOMINO VCDs are all negative, suggesting increases of emissions globally. The GC-adj simulations have larger NO_2 VCDs than the QA4ECV retrieval in North China Plain and France but have smaller VCDs in the rest of the world. The magnitudes of the negative differences using the QA4ECV retrieval are smaller than the magnitudes using the DOMINO retrieval, and they are larger than the magnitudes using the NASA retrieval. This is consistent with the top-down emission estimates, in which the magnitude of the QA4ECV posterior NO_x emissions lies between the NASA and the DOMINO posterior emissions.

S2. Evaluations with CNEMC ozone measurements

Compared with the CNEMC ozone measurements in China in 2015 (Figure S3), GC-adj simulations using the HTAP emissions in 2010 have the smallest NMB and NMSE in all ozone metrics. However, this better performance using prior simulations may only be a coincidence of using bottom-up NO_x emissions for a different year. As shown in Section 3.3, the annual budget of NO_x emissions in China decreased by 10.7% (NASA) and 3.5% (DOMINO) from 2010 to 2015, and their spatial distributions also change [Duncan et al., 2016]. GCv12 ozone simulations using the prior emissions (CEDS) show the smallest NMSE in all metrics in 2015, and show the smallest NMB in annual mean daytime ozone and annual mean 24-hour ozone concentrations. Ozone simulations using the NASA posterior emissions show the smallest NMB in daytime ozone (annual mean and NH summertime), 24-hour ozone (annual mean and NH summertime) and annual mean MDA8 ozone, whereas simulations using the DOMINO posteriors show the smallest NMB in MDA8 ozone.

S3. Trend of global surface ozone

As shown in Figure S5, the peak of NH summertime ozone in 2012 is captured by all model simulations and can be explained by the anomalously high temperature in that year [Fiore et al., 2015; Lin et al., 2017]. In both simulations, the impact of assimilation on improving biases in ozone over the 2006 – 2014 period also depends upon the ozone metric. GCv12 simulations using the 2010 anthropogenic emissions for all years have the best correlations with the trend of measured MDA8 ozone (R^2=0.91), reflecting the large impact of meteorology on the trend of MDA8 ozone. All annual mean ozone metrics from the measurements show the smallest values in 2009 and continue to increase through 2016. All simulations of annual daytime and MDA8 ozone have opposite trends to the increase in the measurements from 2012 to 2014 for reasons that remain unclear. All NH summertime ozone metrics from TOAR show decrease (9.1% for daytime, 9.3% for MDA8, 6.0% for 24 hour) from 2006 to 2009, and increase (5.7% for daytime, 5.0% for MDA8, 8.0% for 24 hour) from 2009 to 2014. This trend is not well captured in most of the simulations.

S4. Evaluations with ozone sonde in Trinidad Head in 2016

Ozone simulations from global models can be used as boundary conditions in regional models for further air quality studies. With this goal in mind, we also evaluate the vertical profile of ozone simulations at Trinidad Head in 2016 with the ozonesonde measurements in Figure S9. The choice of the year 2016 is a combined consideration of a high ozone year in North America and performing ozone evaluations at a more recent year than the base year for which we performed the inversions. In the spring, the GCv12 results are more consistent with measurements near the surface, whereas GCadj performs better in the free troposphere. The NMSEs in GC-adj simulations are one order of magnitude smaller than GCv12 simulations. The GCv12 simulations using the DOMINO posterior NO_x emissions have the smallest NMSE (0.015) and NMB (-10.6%). The GC-adj simulations using the HTAP 2010 emissions turn out to have the smallest NMSE (0.0018) and NMB (3.2%) for 2016 simulations. However, the differences in NMSE and NMB between the simulations using the HTAP inventory and the posterior NO_x emissions are small, different by 11.1% and 9.5%, respectively.

We also evaluate the time series of ozone simulations between 700 – 900 hPa at Trinidad Head in 2016 (Figure S10). The GC-adj ozone simulations using the HTAP inventory have the smallest NMSE (0.0018) and NMB (3.2%) (NASA posterior: NMSE = 0.0020, NMB = 3.5%; DOMINO posterior: NMSE = 0.0064, NMB = 7.8%). The GCv12 ozone simulations using the DOMINO posterior NO_x emissions have the smallest NMSE (0.026) and NMB (-11.5%) (in comparison, NASA posterior: NMSE = 0.096, NMB = -25.2%; CEDS: NMSE = 0.034, NMB = -13.9%). The improved simulations of ozone mean concentrations between 700 – 900 hPa using the posterior NO_x emissions in GCv12 reflect better characterization of ozone boundary conditions and timeseries for regional scale air quality studies.

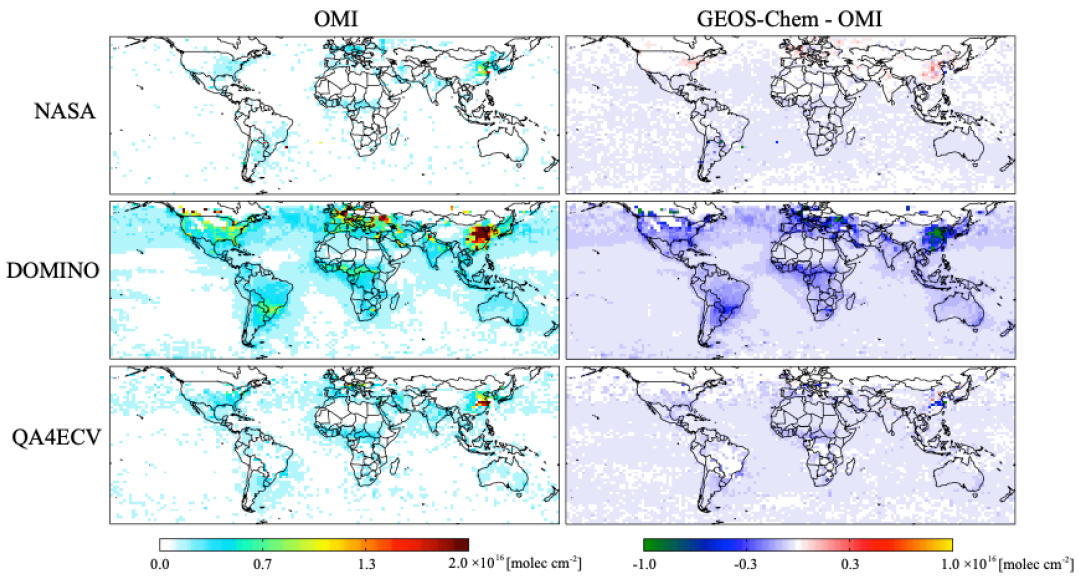

Figure S1. Tropospheric NO_2 VCDs in January 2010. The left column shows OMI VCDs calculated using vertical profiles from the GC-adj simulation and scattering weight / averaging kernels from the corresponding product. The right columns show the differences between NO_2 slant column densities in GC-adj simulations and OMI retrievals.

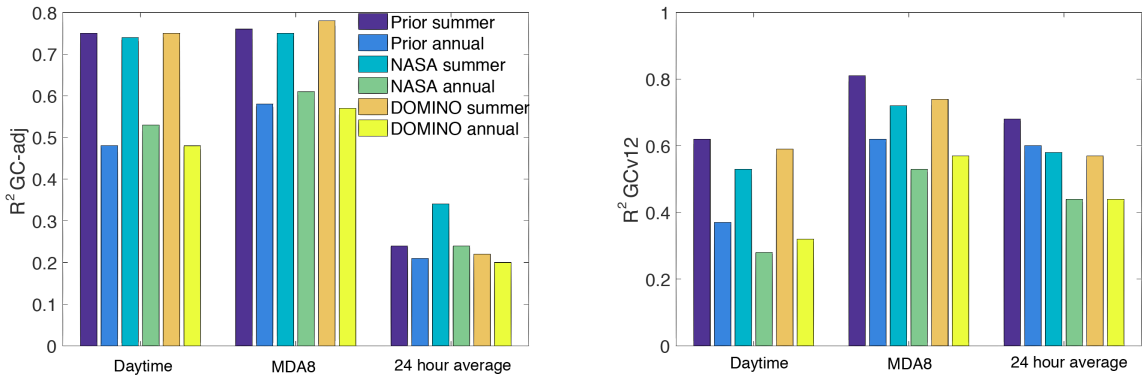

Figure S2. The coefficients of determination of the simulated and measured (from TOAR) surface ozone concentrations in 2010. The values are calculated by taking the annual or NH summer mean of surface measurements and GEOS-Chem simulations, which are sampled at each monitoring location, and calculated across different sites.

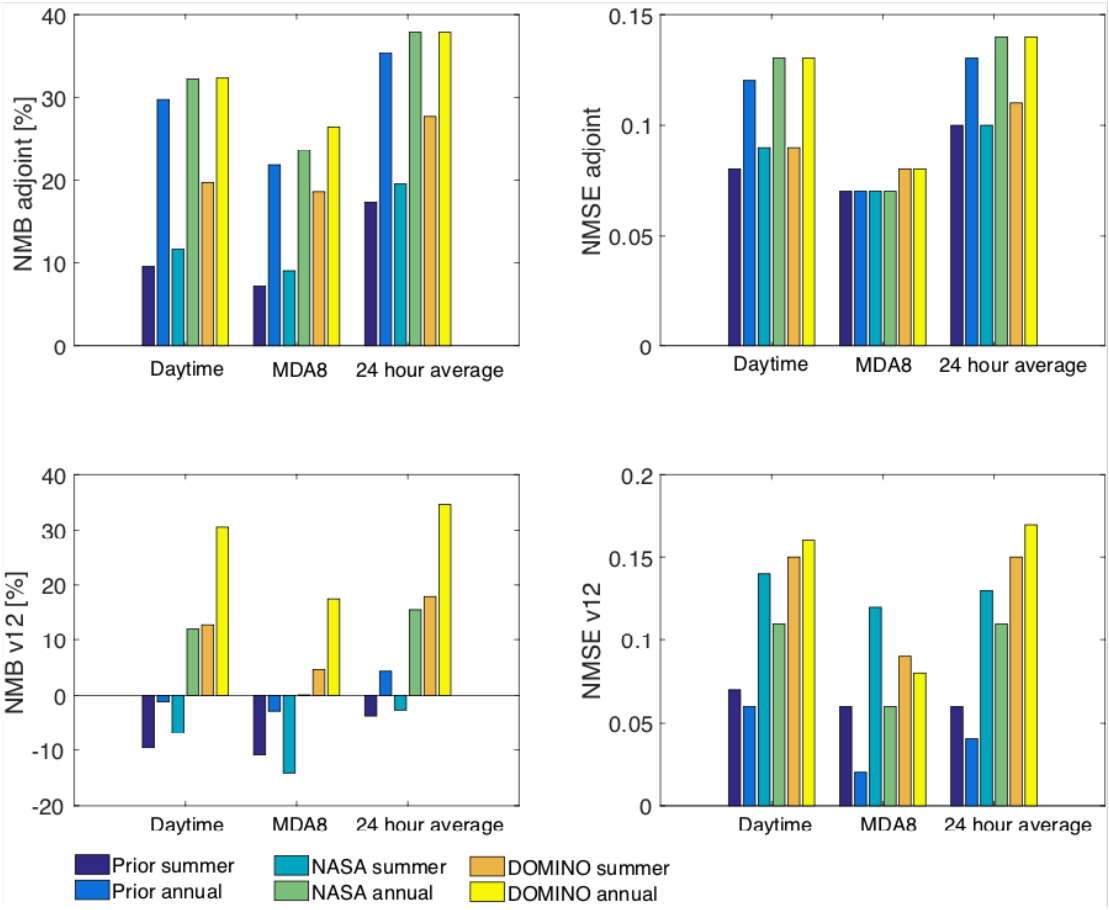

Figure S3. NMB and NMSE of annual mean and NH summertime surface ozone concentrations when comparing measurements from CNEMC in China in 2015 with GC-adj (top) and GCv12 (bottom) simulations. The simulations are input with three sets of NO_x emissions: CEDS bottom-up inventory (HTAP for GC-adj and CEDS for GCv12), posterior emissions constrained by NASA product, and posterior emissions constrained by DOMINO product.

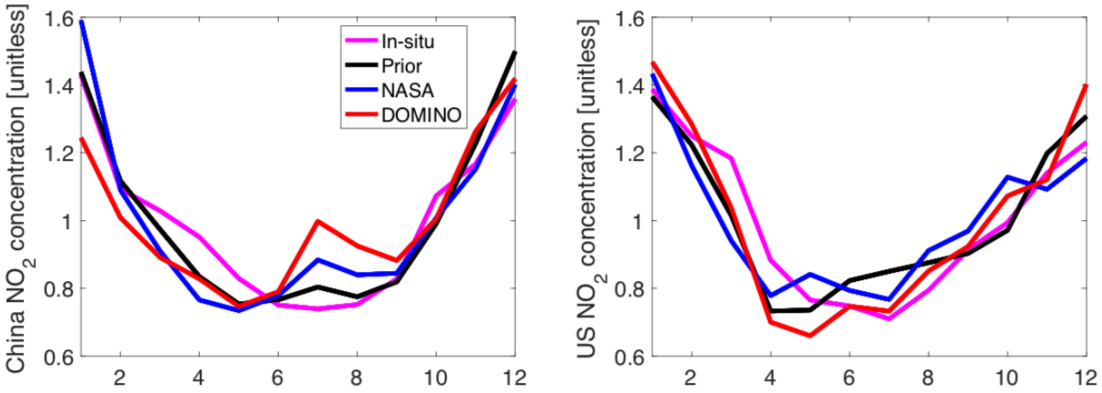

Figure S4. Monthly NO_2 concentrations in China (left) and the US (right) normalized to annual mean concentrations in 2015. The coefficients of determinant between simulations and measurements are 0.96 (prior), 0.82 (DOMINO posterior), and 0.92 (NASA posterior) in China and 0.92 (prior), 0.94 (DOMINO posterior), and 0.88 (NASA posterior).

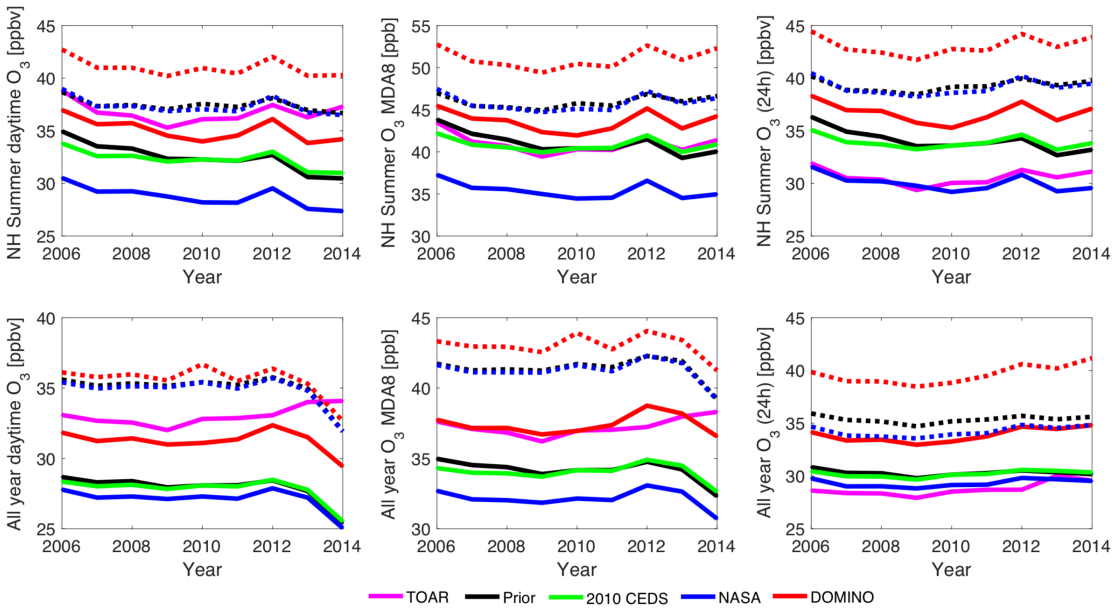

Figure S5. The trends of 2-meter ozone concentrations from 2006 to 2014. Surface measurements are from the TOAR database (magenta line). Simulations are from the GCv12 (solid lines) and the GC-adj (dotted lines). NO_x emissions are from bottom-up inventories (HTAP 2010 for GC-adj, CEDS in each corresponding year for GCv12, black lines), 2010 CEDS (green lines), NASA posterior (blue lines), and DOMINO posteriors (red line).

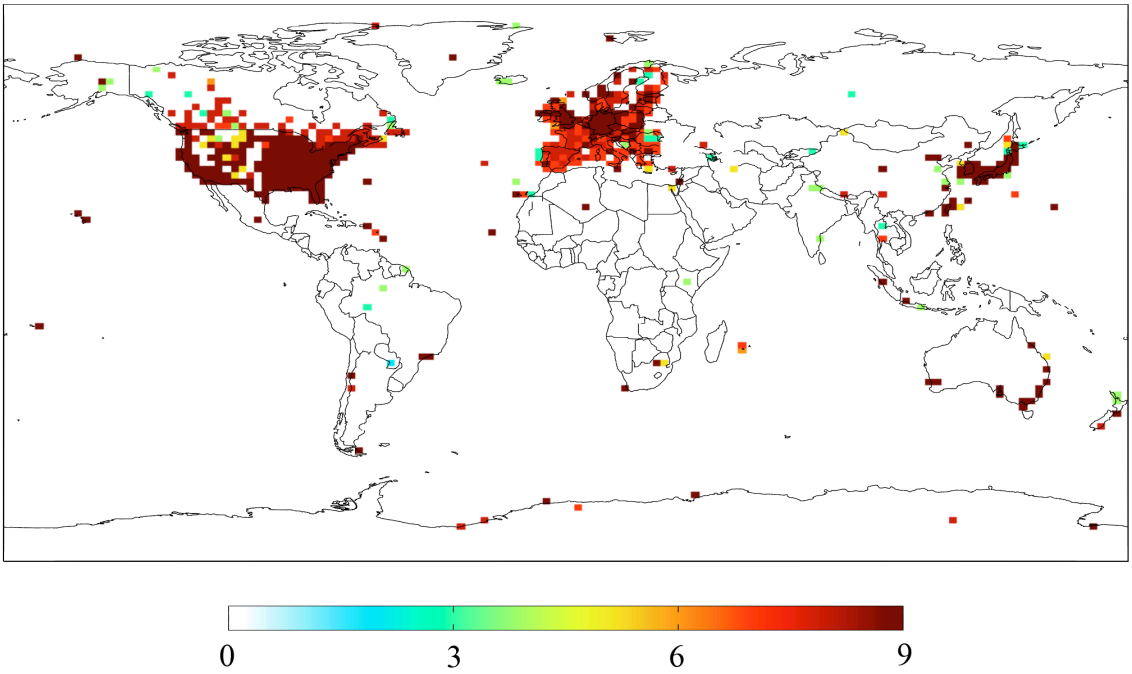

Figure S6. Number of years of ozone measurements in each grid cell.

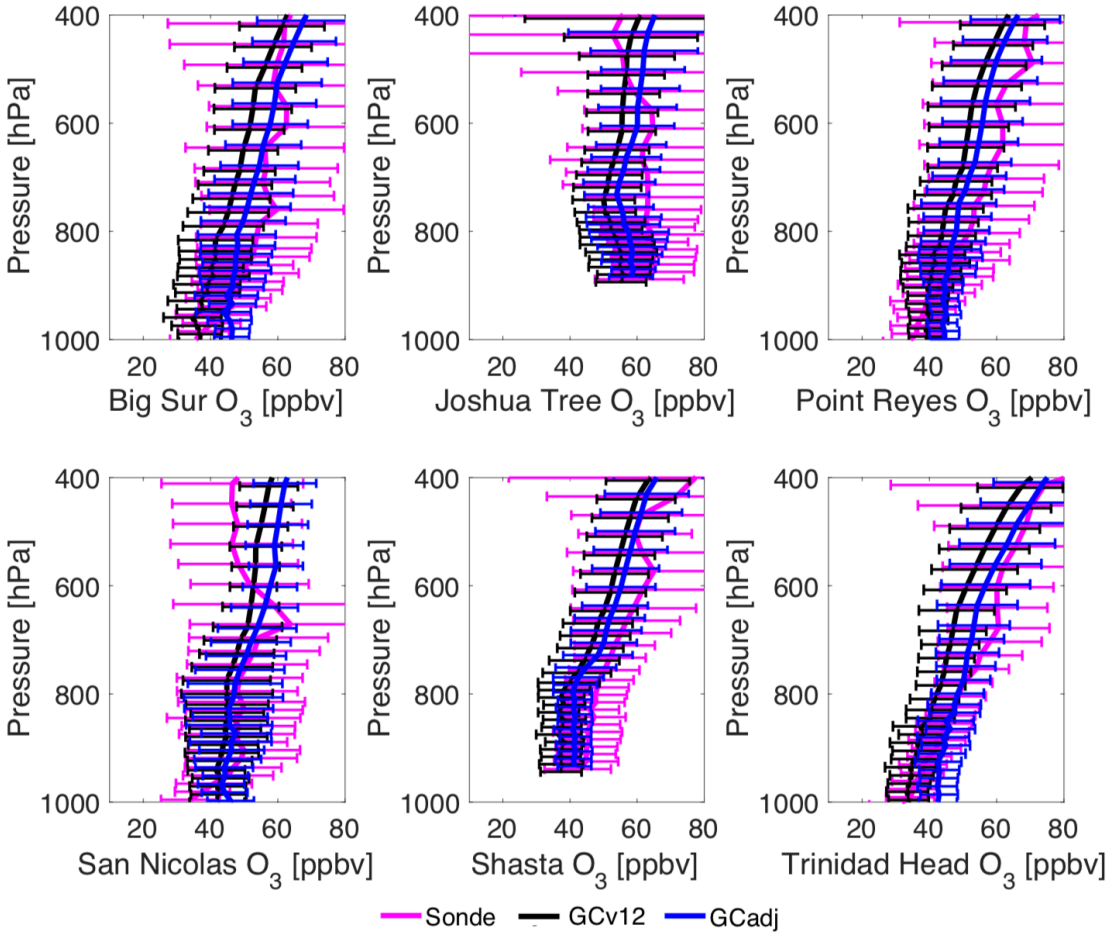

Figure S7. Ozone vertical profiles averaged over May and June of 2010 from 6 ozonesonde measurement sites from the IONS-2010 field experiment in California (magenta line). Black lines are from the GCv12 simulations. Blue lines are from the GC-adj simulations. The horizontal bars show standard deviations of measurements at each vertical layer.

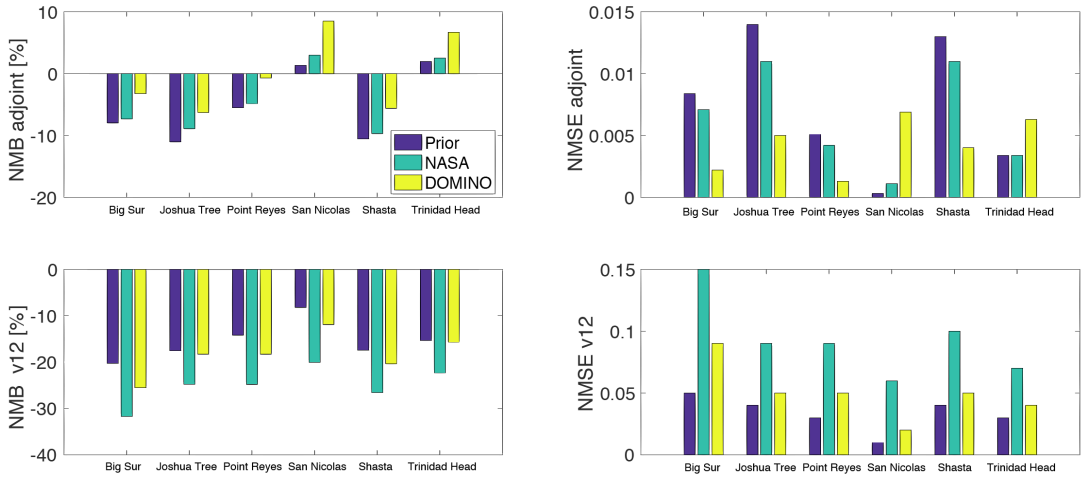

Figure S8. NMSE and NMB of GC-adj (top) and GCv12 (bottom) simulations at 700hPa to 900 hPa in 2010 evaluated with IONS-2010 field experiments. Three sets of NO_x emissions, i.e., bottom-up inventory (HTAP for GC-adj, CEDS for GCv12), DOMINO posterior, and NASA posterior, are input in each model.

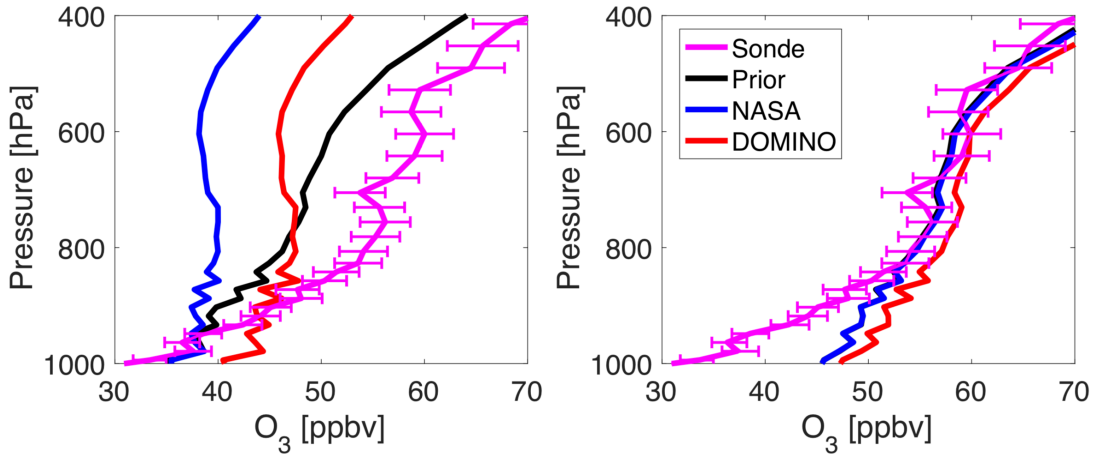

Figure S9. Period-mean ozone vertical profile from March to May, 2016. Simulations are from GCv12 (left) and GC-adj (right). The sonde measurements are available on Mar 2, 11, 17, 23, 31, April 6, 13, 19, 28, May 5, 10, 19, 25.

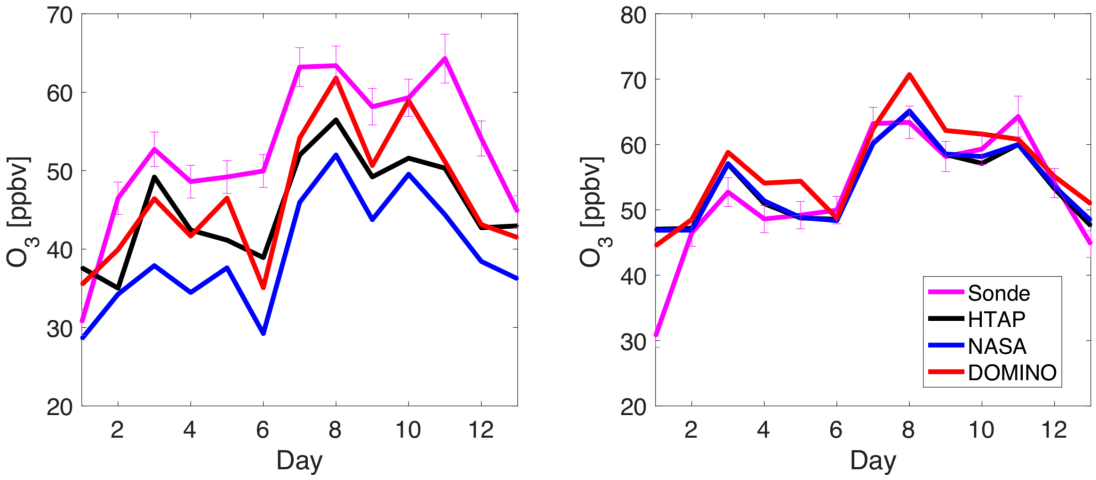

Figure S10. Timeseries of ozone mean concentration between 700 hPa to 900 hPa in all 13 days that sonde measurements are made from March to May, 2016. Simulations are from GCv12 (left) and GC-adj (right).

Table S1. Coefficients of determination between simulations and TOAR measurements of NH summertime ozone trend (2006-2014)

GCv12		Prior (CEDS)	2010 CEDS	NASA	DOMINO
Daytime	NH Summer	0.47	0.53	0.55	0.72
MDA8	NH Summer	0.81	0.91	0.85	0.89
24 hour average	NH Summer	0.55	0.83	0.69	0.88
GC-adj		Prior (HTAP)	-	NASA	DOMINO
Daytime	NH Summer	0.70	-	0.77	0.81
MDA8	NH Summer	0.80	-	0.88	0.86
24 hour average	NH Summer	0.91	-	0.95	0.96

Table S2. Coefficient of determination between ozone simulations and TOAR measurements at remote surface sites in

	GCv12			GC-adj		
	Prior	NASA posterior	DOMINO posterior	Prior	NASA posterior	DOMINO posterior
Mauna Loa (Night)	0.39	0.48	0.43	0.44	0.47	0.46
Mt Bachelor (Night)	0.35	0.32	0.38	0.13	0.16	0.24
Mt Bachelor	0.47	0.46	0.52	0.30	0.32	0.36
Lassen Volcanic	0.82	0.61	0.72	0.62	0.65	0.75
Great Basin	0.67	0.65	0.65	0.67	0.72	0.73
Sequoia	0.90	0.84	0.88	0.85	0.86	0.89