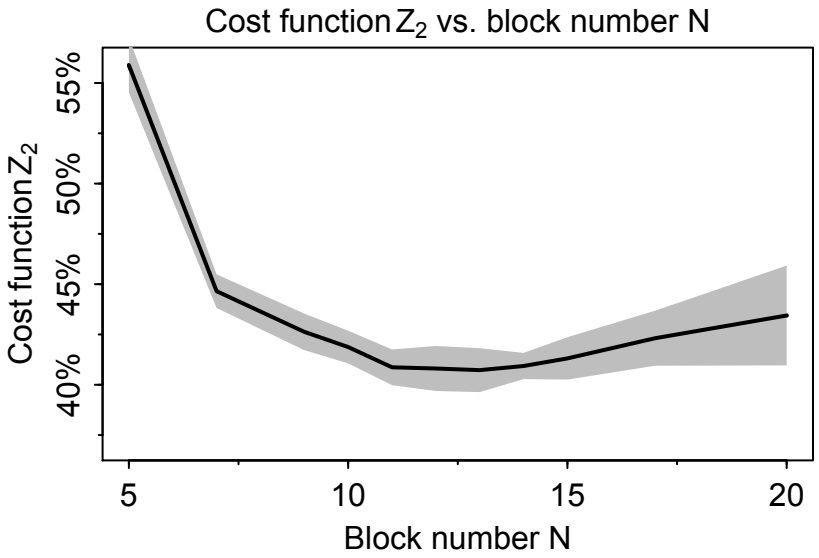

Figure S1. The cost function Z_2 with different block number N. Results are shown for the first day of February, April, August and November sampled every 6 hours. Shaded area denotes for 1 standard deviation.

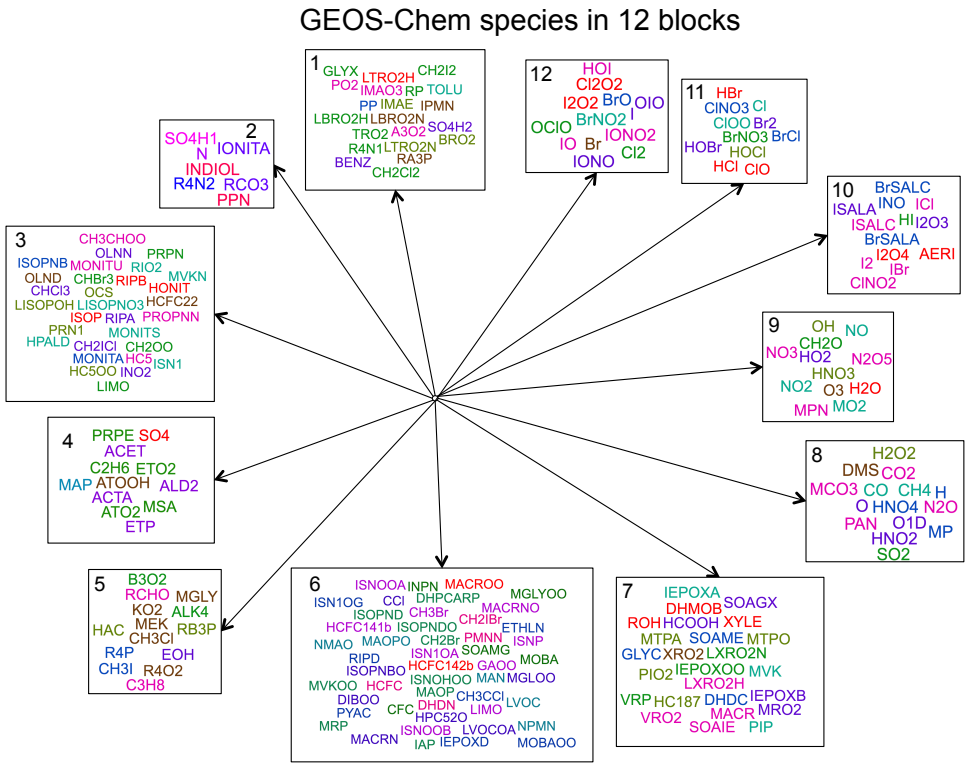

Figure S2. GEOS-Chem species in a 12-block system. The 228 species are split into 12 blocks based on the similarity of production and loss rates using the simulated annealing algorithm. More details can be found in text. The full names of these acronyms can be found at http://wiki.seas.harvard.edu/geos-chem/index.php/Species_in_GEOS-Chem.

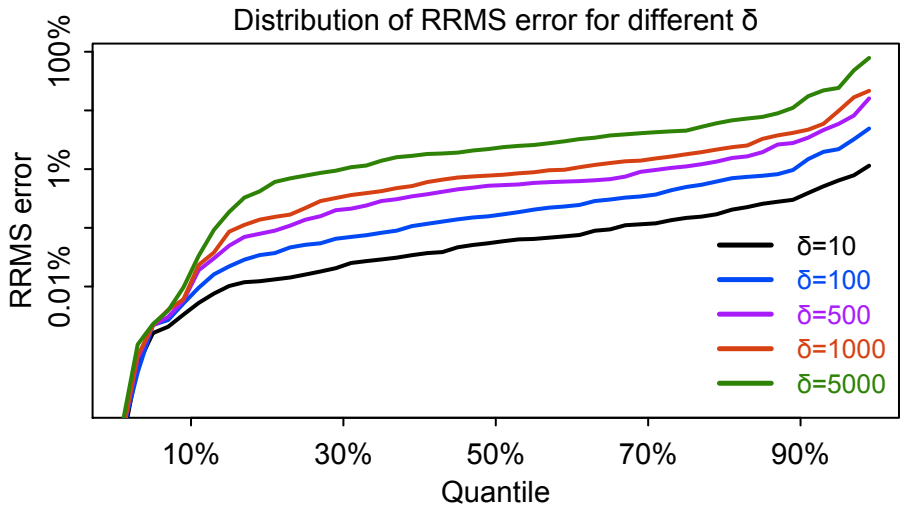

Figure S3. Distribution of RRMS error for different thresholds δ (*N*=12 and *M*=20). The y axis is on the logarithmic scale.

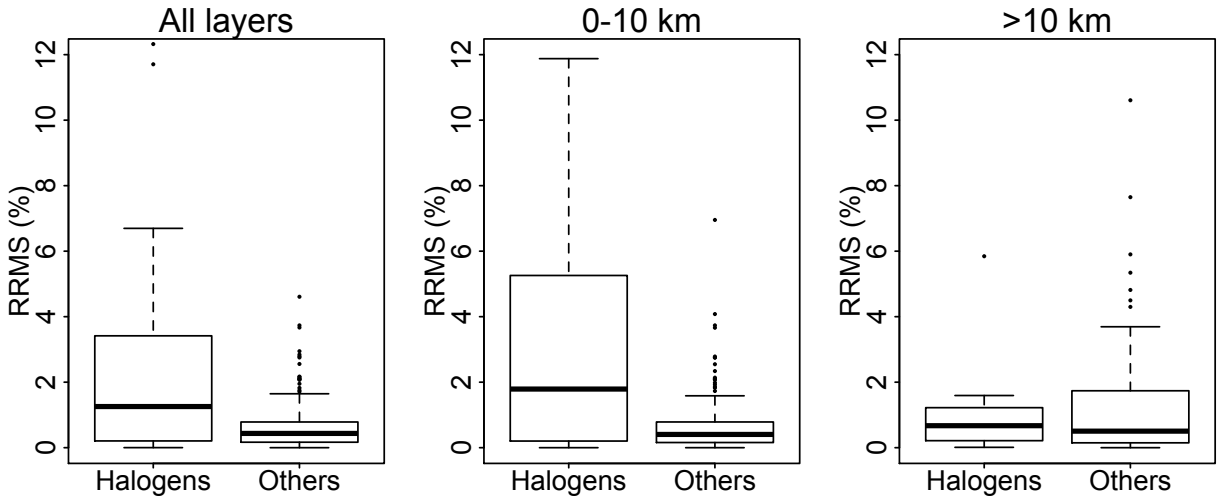

Figure S4. The RRMS error of halogen and non-halogen species for (a) all layers, (b) 0-10 km and (c) above 10 km. Results are shown for the daily-average concentrations of all species on the last day of the 2-year simulations, using 20 chemistry regimes and a threshold δ of 500 molecules cm^{-3} s^{-1}.

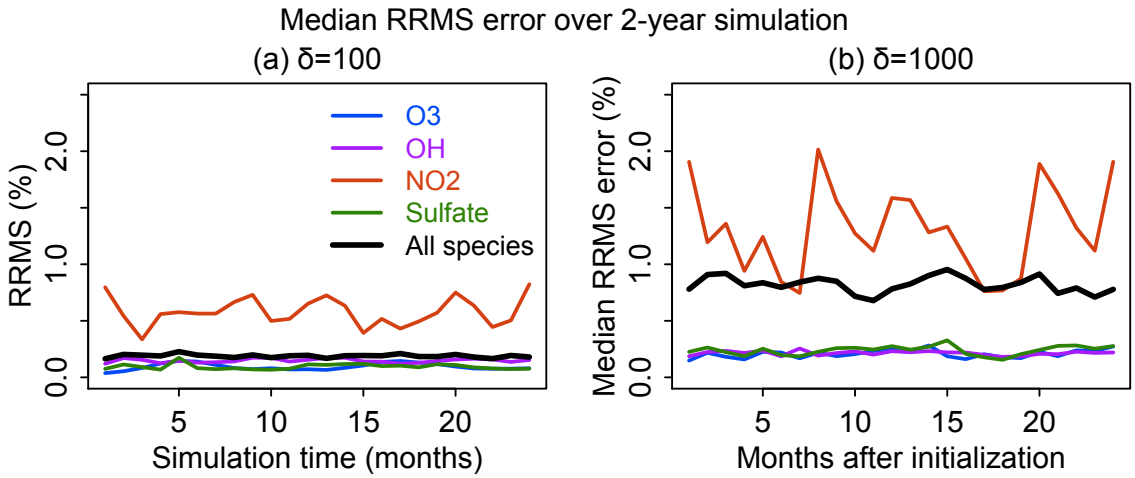

Figure S5. Same as Figure 4, but using a threshold of 100 and 1000 molecules cm^{-3} s^{-1} to partition the fast and slow species.