

Supporting Information for A Graph Theory-Based Algorithm for the Reduction of Atmospheric Chemical Mechanisms

Section S1: Species weightings used for isoprene error measurements.

For Figure 2, the error was calculated using six box model run conditions in Table 1 and the following species with weights: OH:1, HO₂: 1, NO: 1, NO₂: 1, O₃: 0.5, formaldehyde: 1, isoprene: 0.5, IEPOX: 1, methylglyoxal: 0.5, glyoxal: 0.5, acyl-peroxy radical: 0.8, methacrolein: 0.5, methyl-vinyl-ketone: 0.5, isoprene tetra-functional nitrates: 1, other isoprene tetra-functional compounds: 1, and isoprene hydroxy nitrates: 1. These weighting represent the proportional contribution of this species to the aggregate error measure. The error is calculated using equation 14, with weightings represented by the variable ω .

Figure S1: Isoprene concentrations for 109-species, 74-species, and 15-species AMORE 2.0 reduced isoprene mechanisms compared to the full Caltech Isoprene mechanism, Caltech Reduced Plus isoprene mechanism, Caltech Mini isoprene mechanism, and AMORE v1.2 isoprene mechanism for six conditions (Table 1).

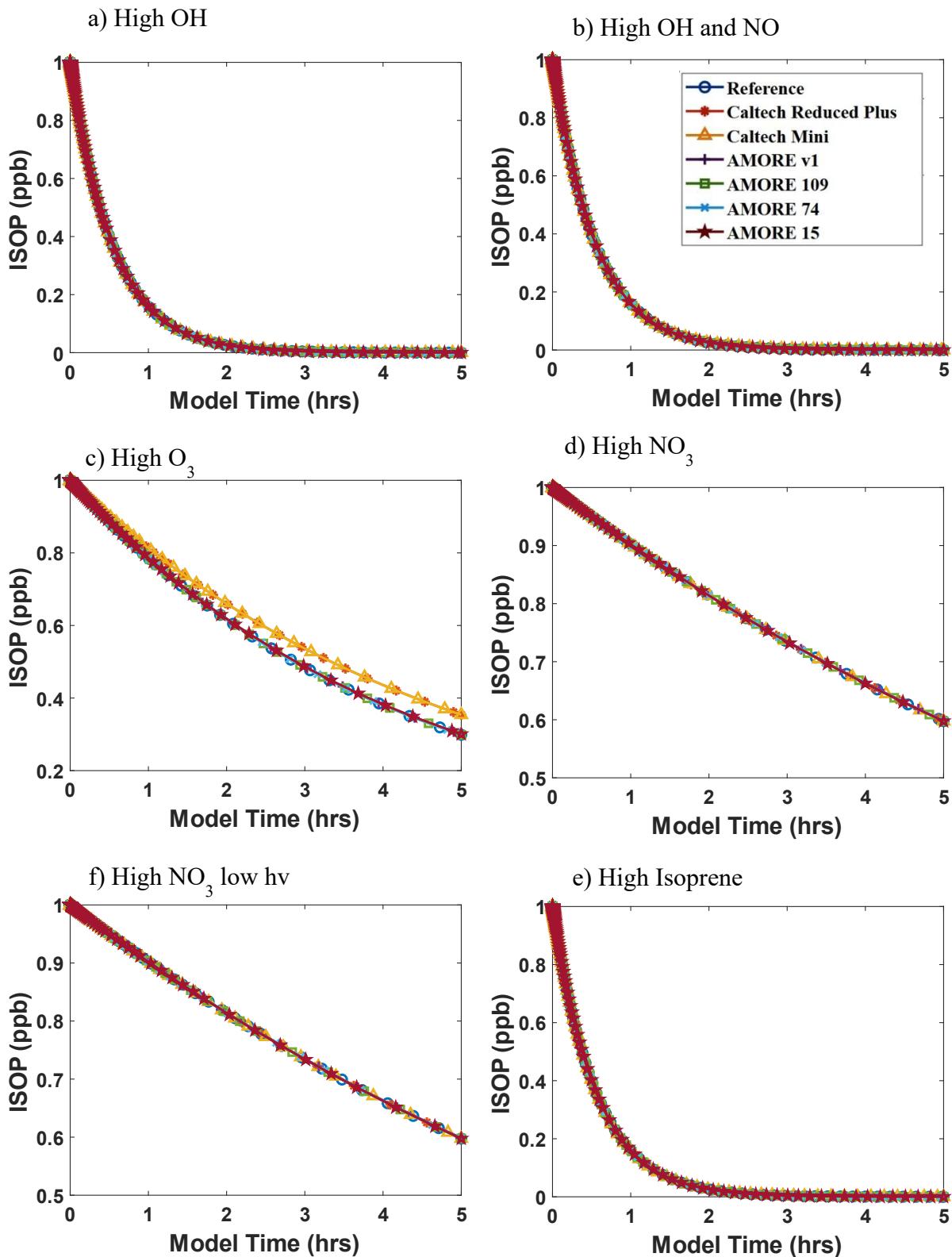
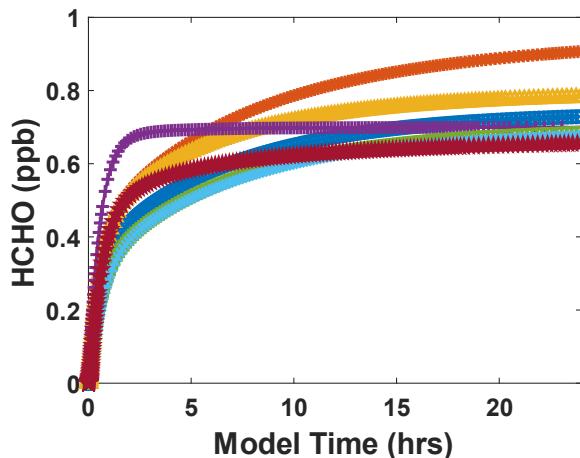
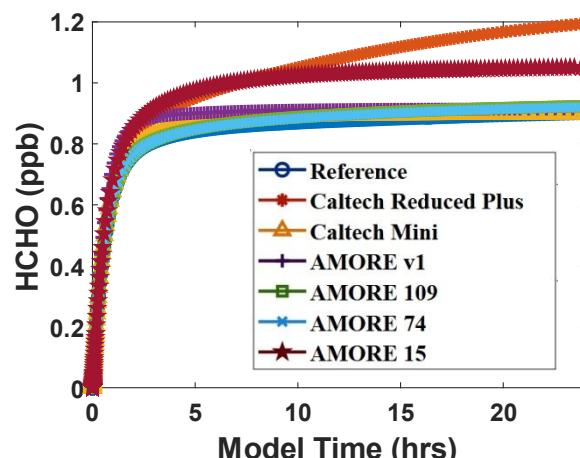


Figure S2: Formaldehyde (HCHO) concentrations for 109-species, 74-species, and 15-species AMORE 2.0 reduced isoprene mechanisms compared to the full Caltech Isoprene mechanism, Caltech Reduced Plus isoprene mechanism, Caltech Mini isoprene mechanism, and AMORE v1.2 isoprene mechanism for six conditions (Table 1).

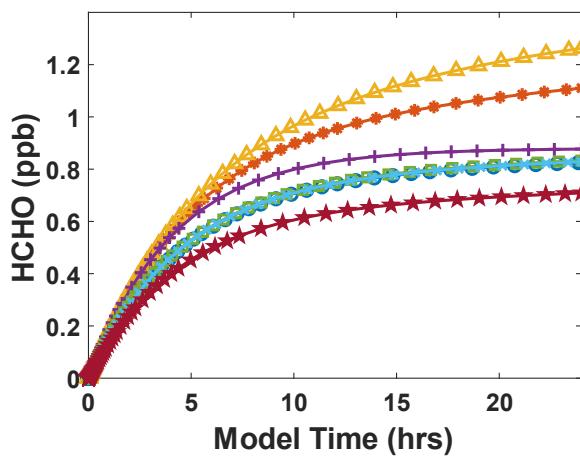
a) High OH



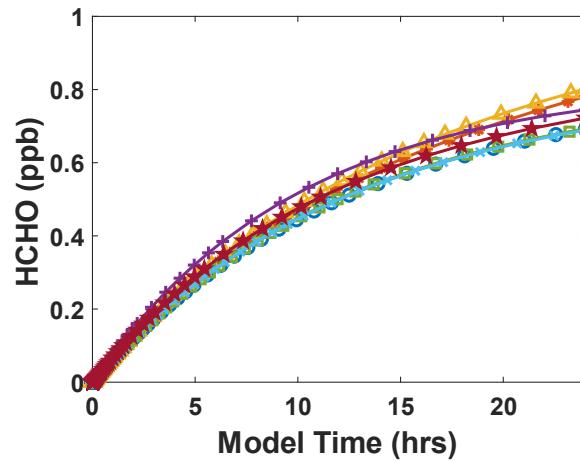
b) High OH and NO



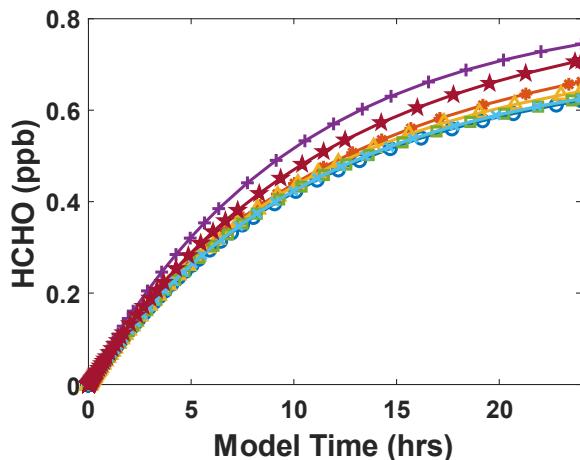
c) High O₃



d) High NO₃



f) High NO₃ low hv



e) High Isoprene

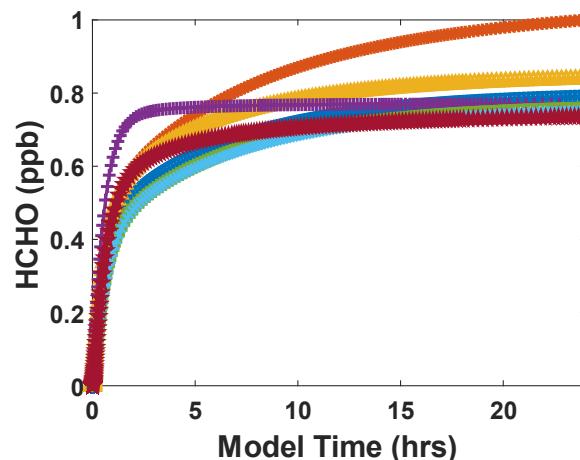
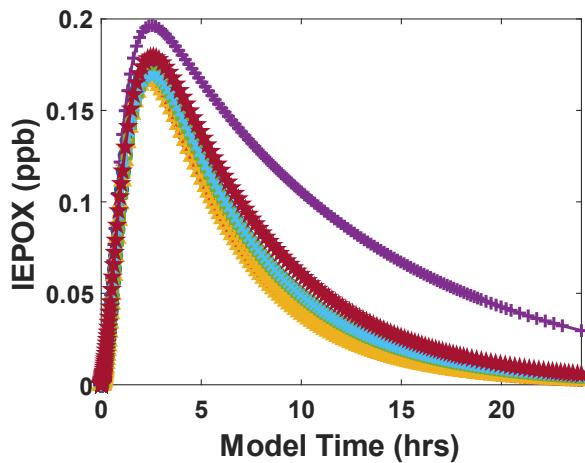
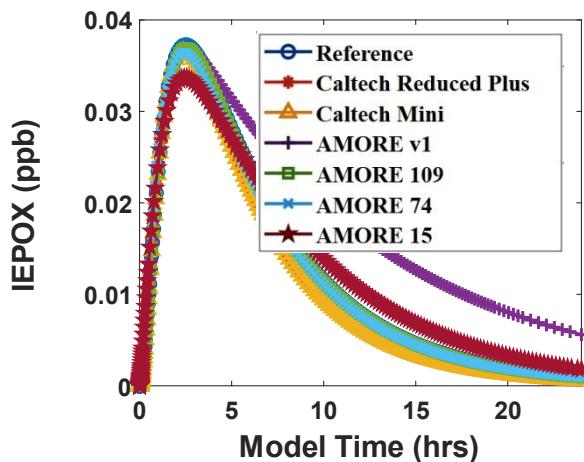


Figure S3: Isoprene epoxy-diol (IEPOX) concentrations for 109-species, 74-species, and 15-species AMORE 2.0 reduced isoprene mechanisms compared to the full Caltech Isoprene mechanism, Caltech Reduced Plus isoprene mechanism, Caltech Mini isoprene mechanism, and AMORE v1.2 isoprene mechanism for six conditions (Table 1).

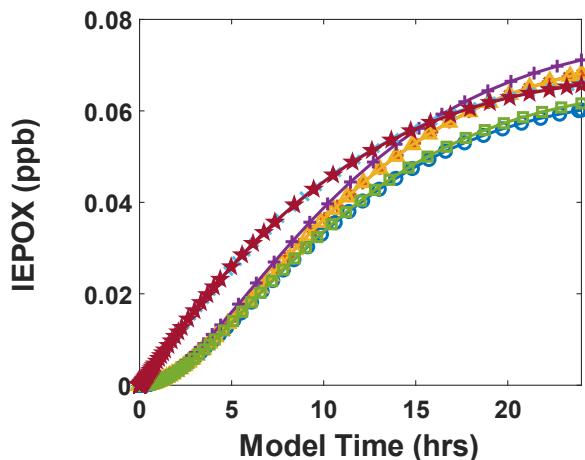
a) High OH



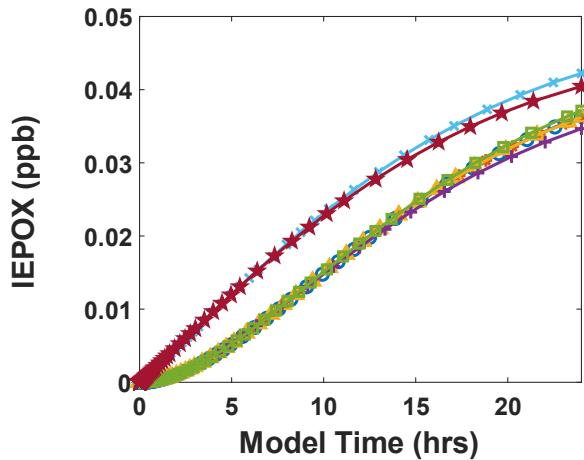
b) High OH and NO



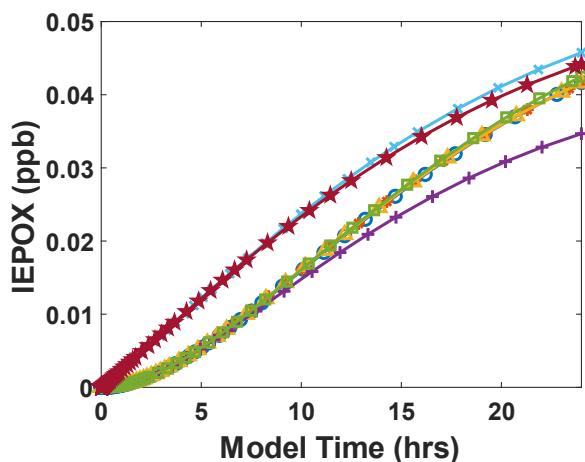
c) High O₃



d) High NO₃



f) High NO₃ low hv



e) High Isoprene

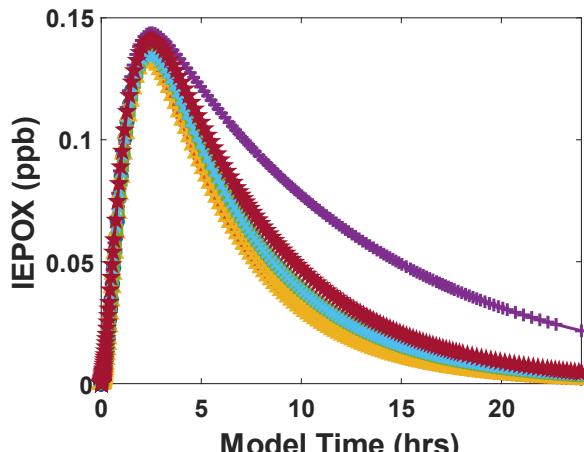
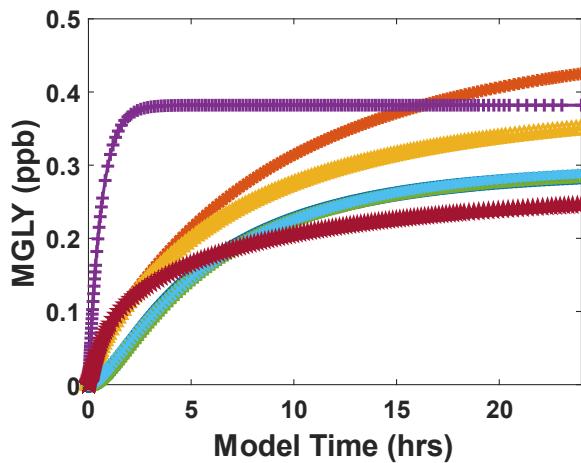
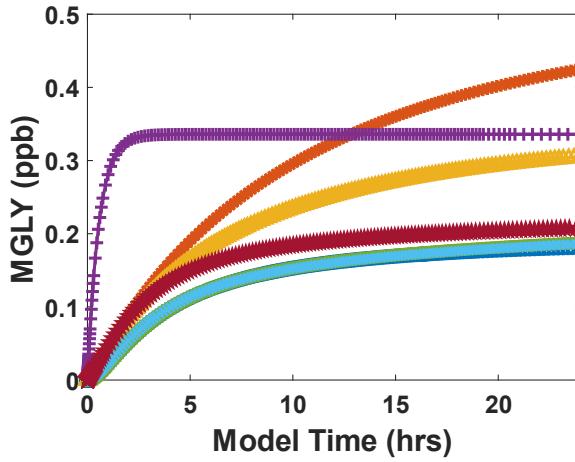


Figure S4: Methyl Glyoxal (MGLY) concentrations for 109-species, 74-species, and 15-species AMORE 2.0 reduced isoprene mechanisms compared to the full Caltech Isoprene mechanism, Caltech Reduced Plus isoprene mechanism, Caltech Mini isoprene mechanism, and AMORE v1.2 isoprene mechanism for six conditions (Table 1).

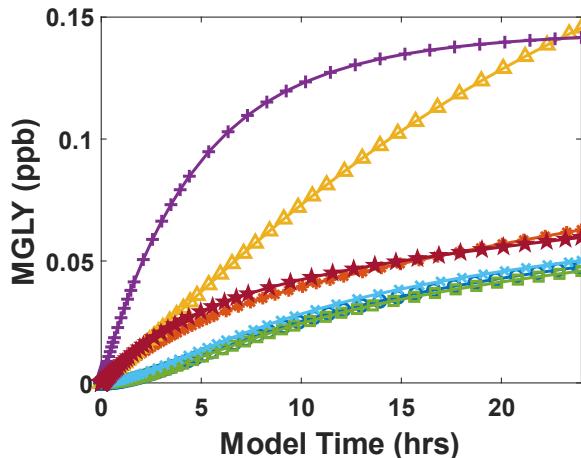
a) High OH



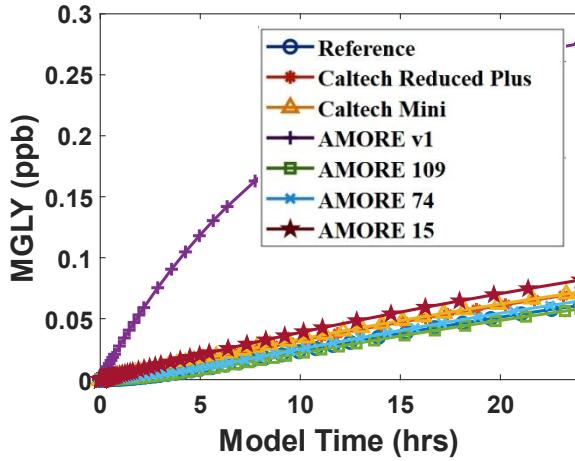
b) High OH and NO



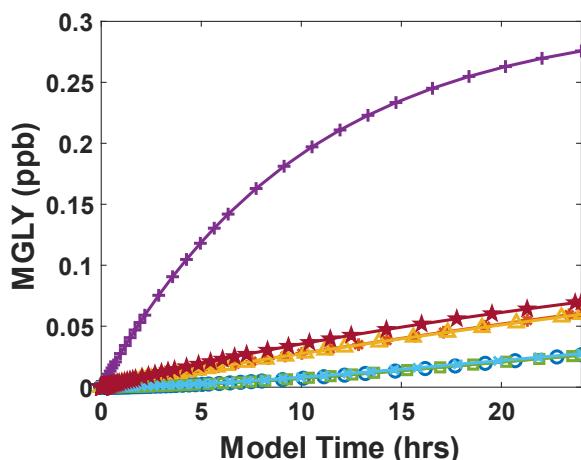
c) High O₃



d) High NO_x



f) High NO_x low hv



e) High Isoprene

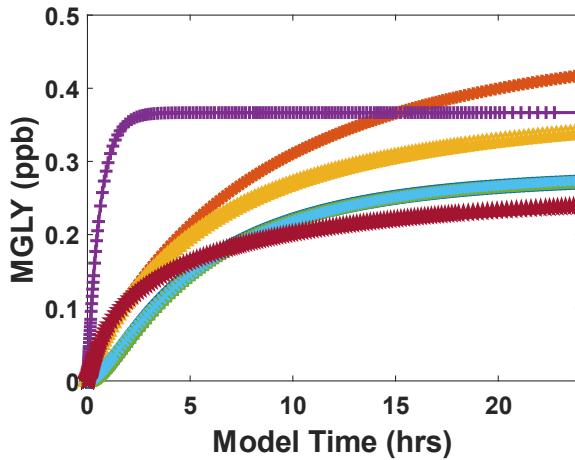
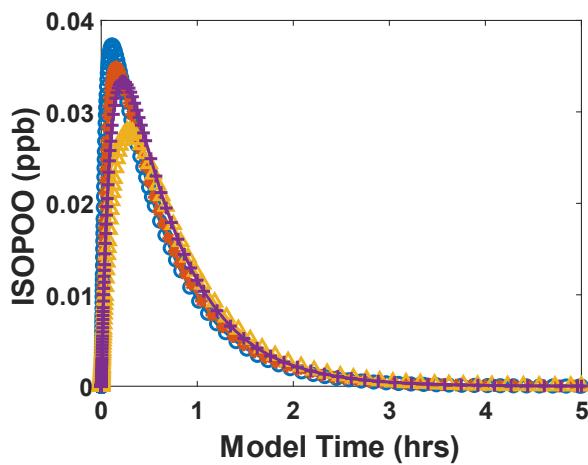
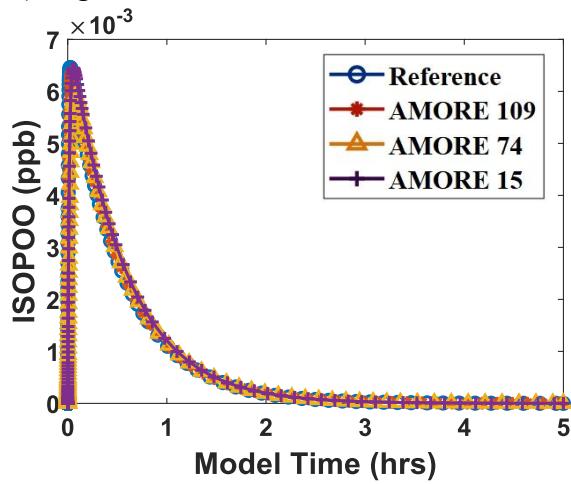


Figure S5: ISOPOO concentrations for 109-species, 74-species, and 15-species AMORE 2.0 reduced isoprene mechanisms compared to the full Caltech Isoprene mechanism for six conditions (Table 1).

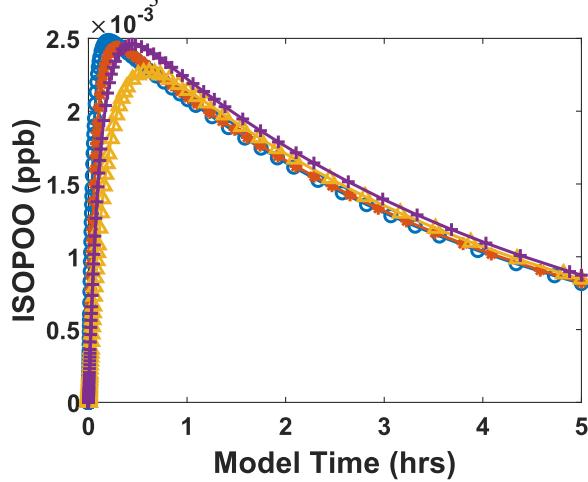
a) High OH



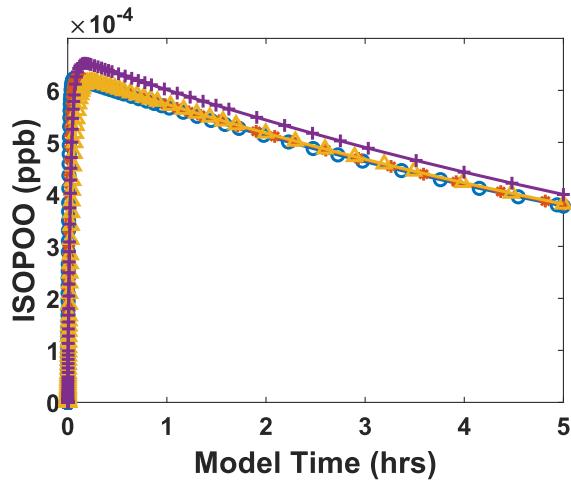
b) High OH and NO



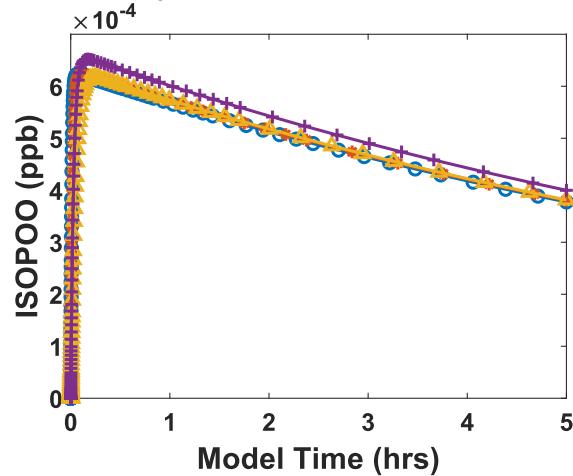
c) High O₃



d) High NO₃



f) High NO₃ low hν



e) High Isoprene

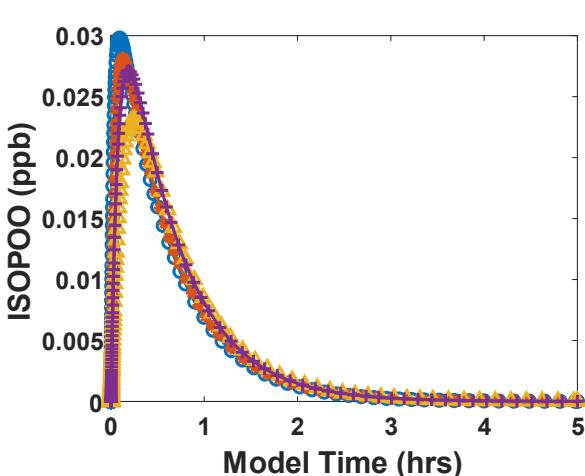
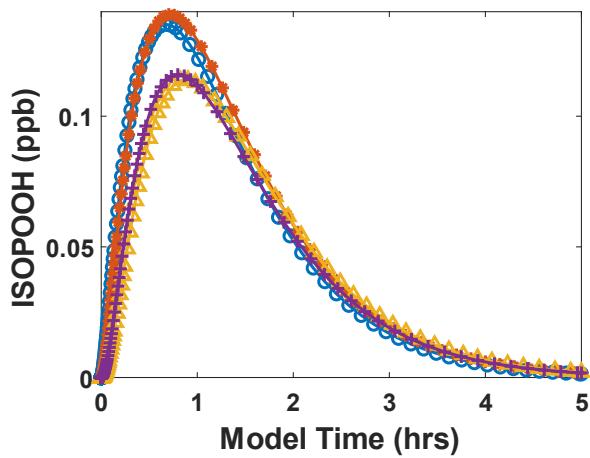
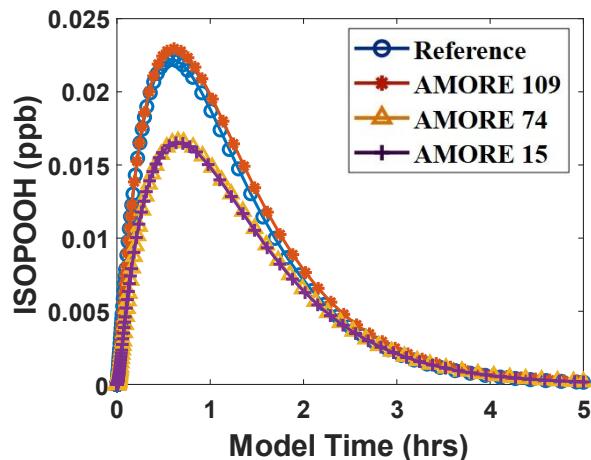


Figure S6: ISOOPOOH concentrations for 109-species, 74-species, and 15-species AMORE 2.0 reduced isoprene mechanisms compared to the full Caltech Isoprene mechanism for six conditions (Table 1).

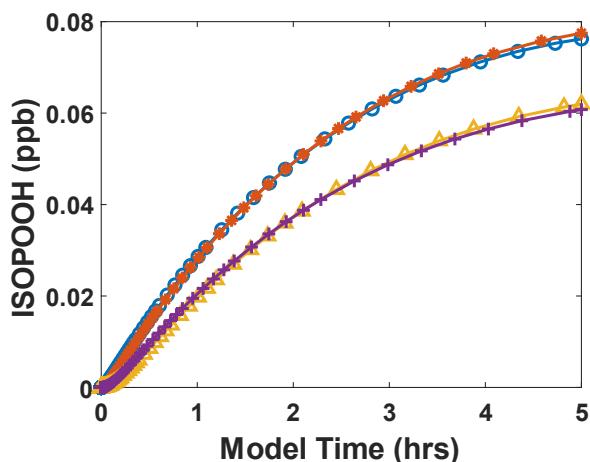
a) High OH



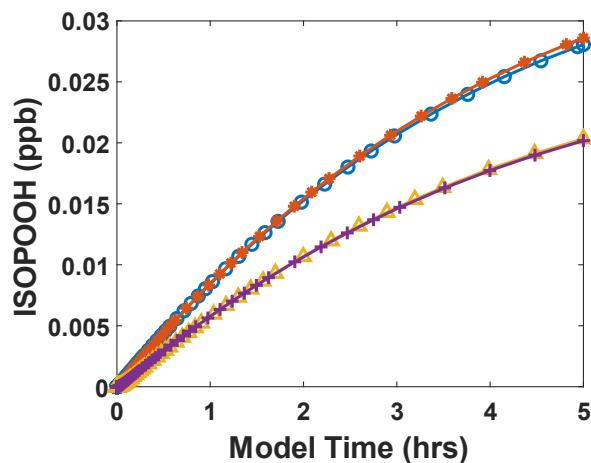
b) High OH and NO



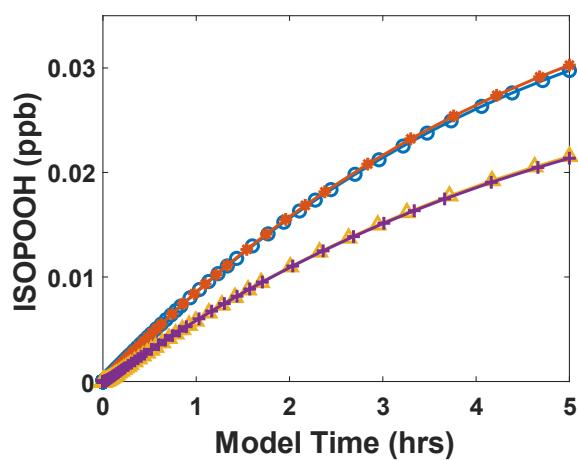
c) High O₃



d) High NO₃



f) High NO₃ low hv



e) High Isoprene

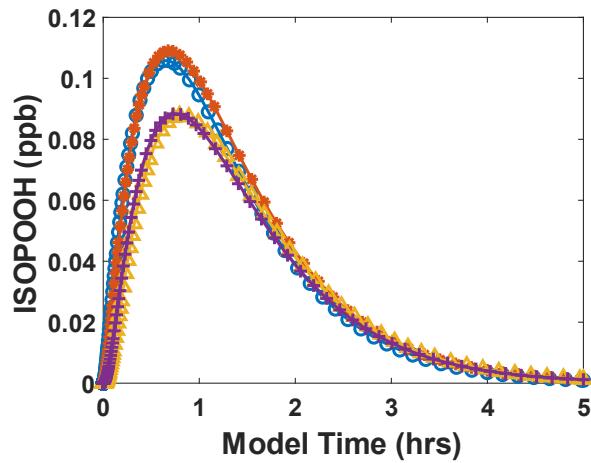


Figure S7: Bifunctional Nitrate concentrations for 109-species, 74-species, and 15-species AMORE 2.0 reduced isoprene mechanisms compared to the full Caltech Isoprene mechanism for six conditions (Table 1).

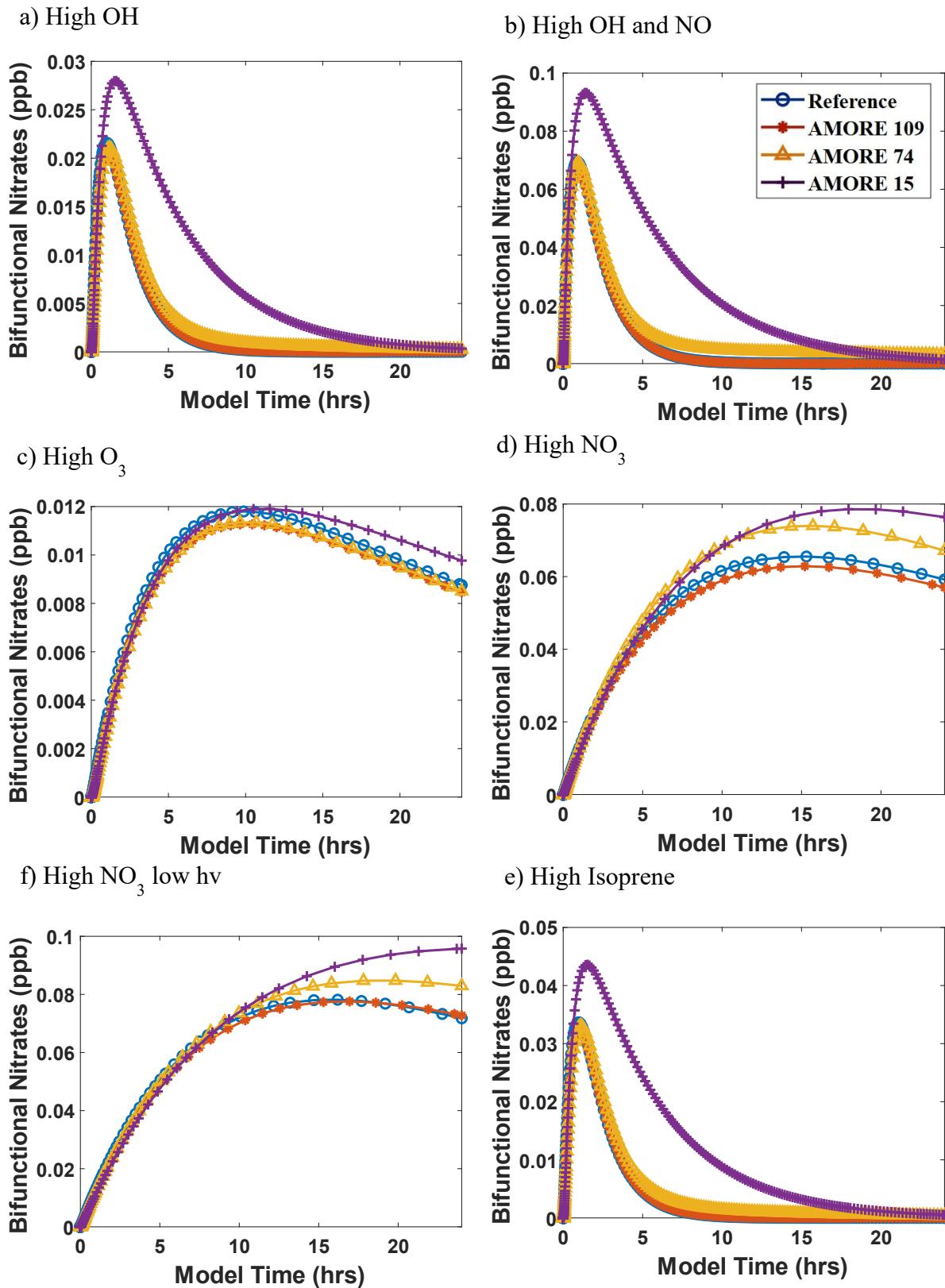


Figure S8: Multifunctional Isoprene Nitrate concentrations for 109-species, 74-species, and 15-species AMORE 2.0 reduced isoprene mechanisms compared to the full Caltech Isoprene mechanism for six conditions (Table 1).

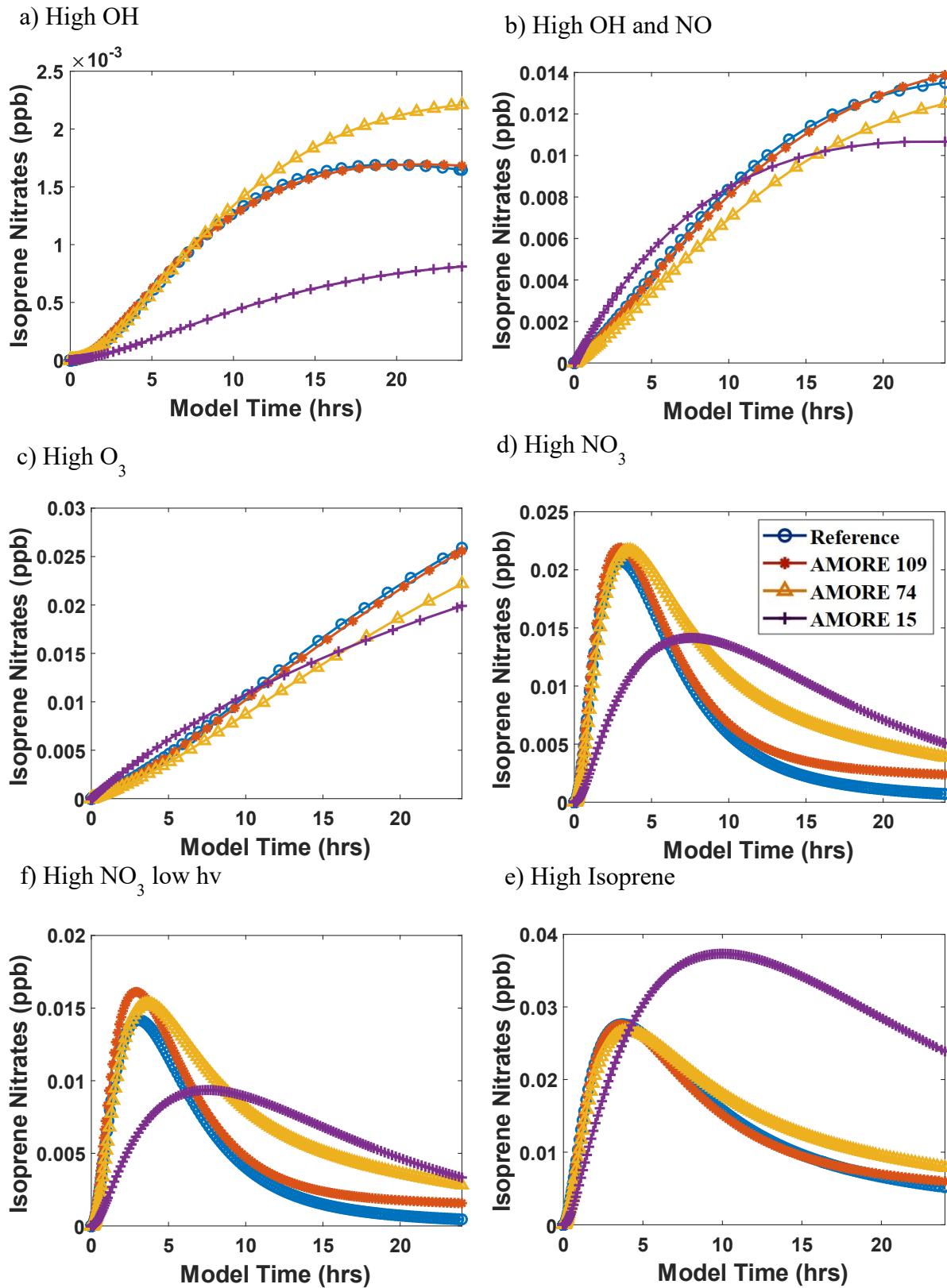


Figure S9: Isoprene Tetrafunctional compound concentrations for 109-species, 74-species, and 15-species AMORE 2.0 reduced isoprene mechanisms compared to the full Caltech Isoprene mechanism for six conditions (Table 1).

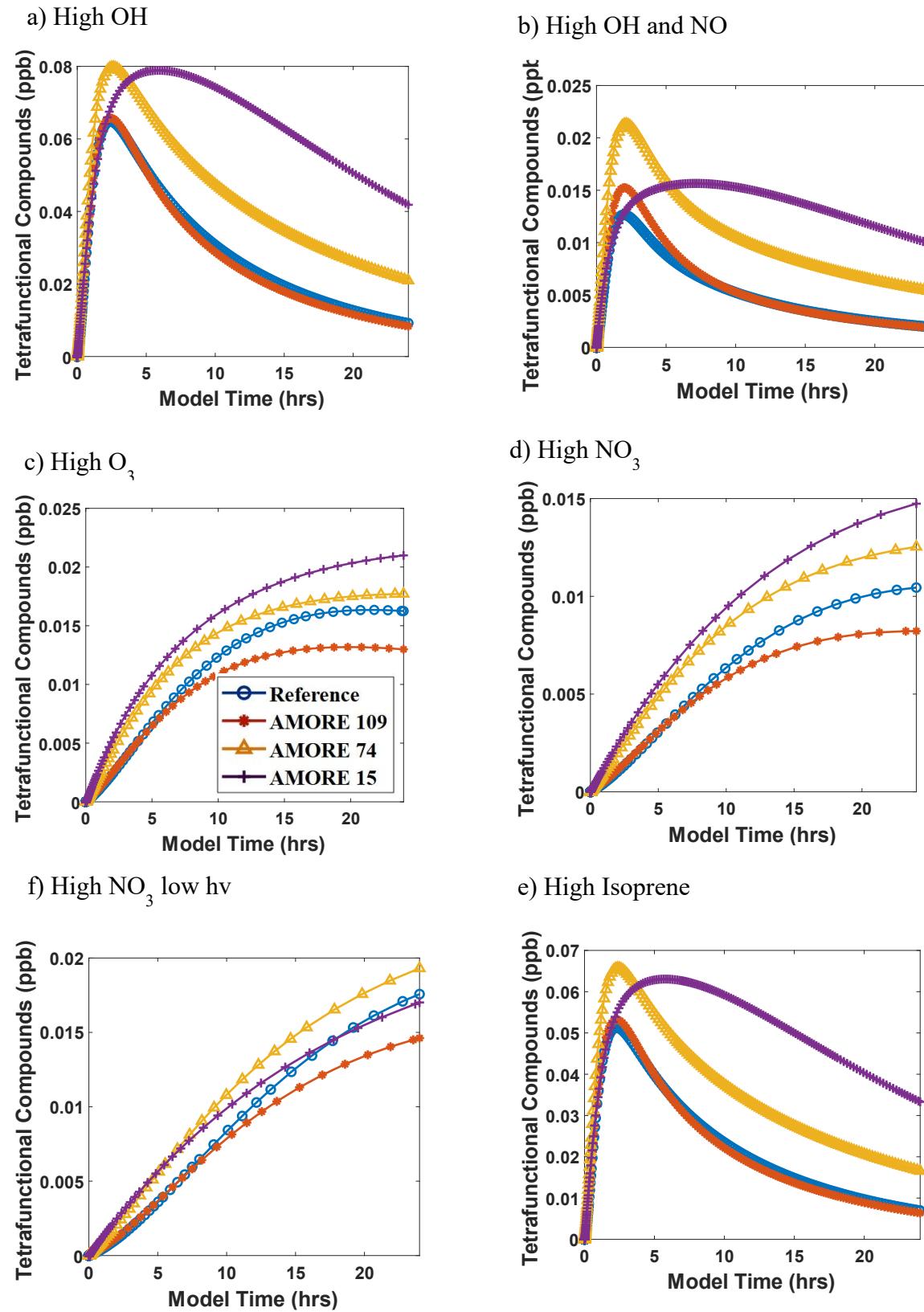
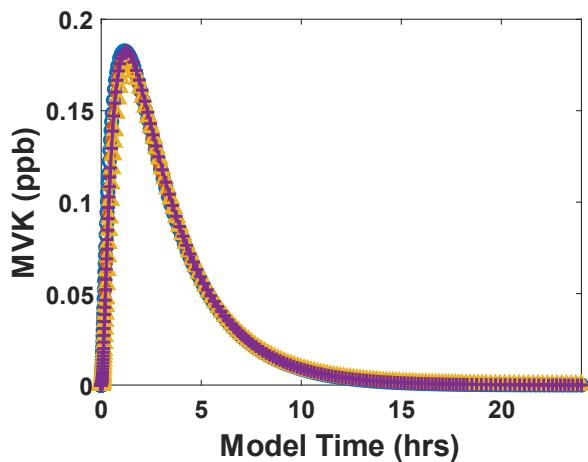
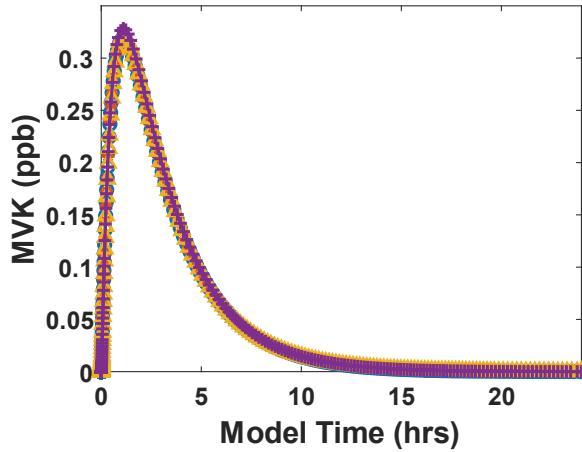


Figure S10: Methyl vinyl ketone (MVK) concentrations for 109-species, 74-species, and 15-species AMORE 2.0 reduced isoprene mechanisms compared to the full Caltech Isoprene mechanism for six conditions (Table 1).

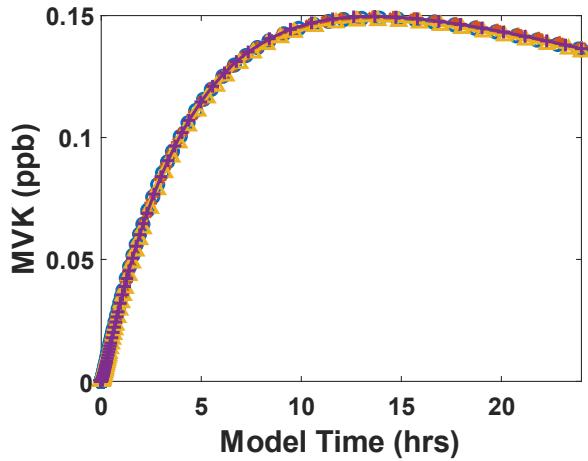
a) High OH



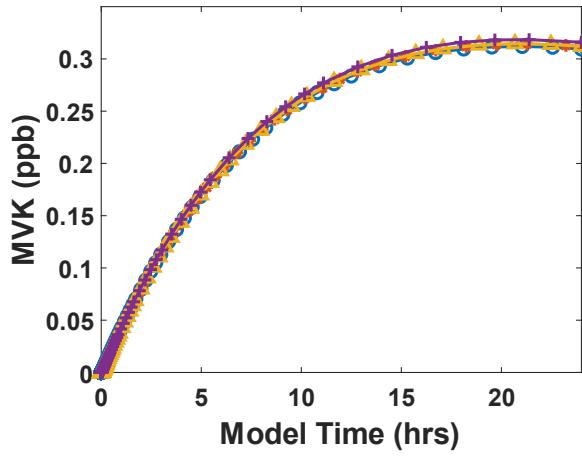
b) High OH and NO



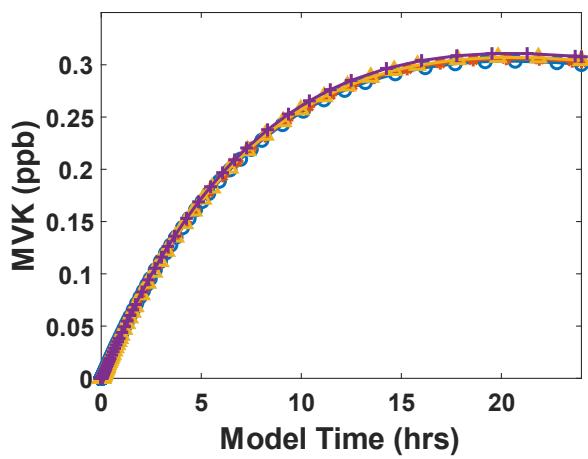
c) High O₃



d) High NO₃



f) High NO₃ low hν



e) High Isoprene

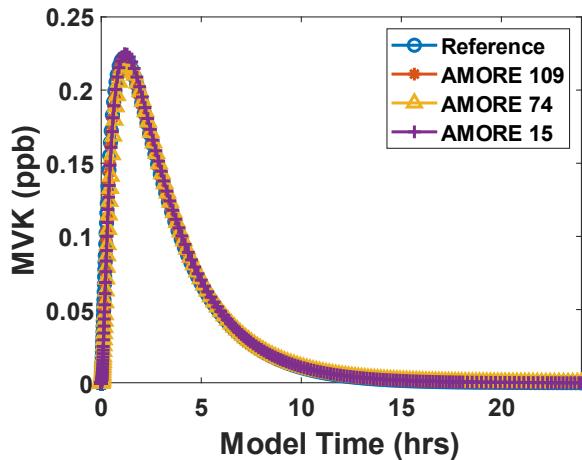
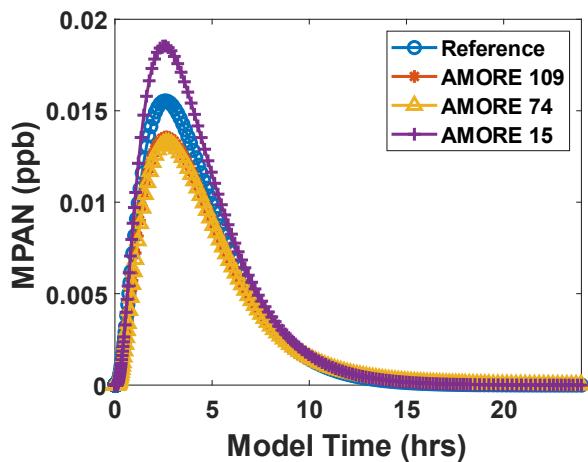
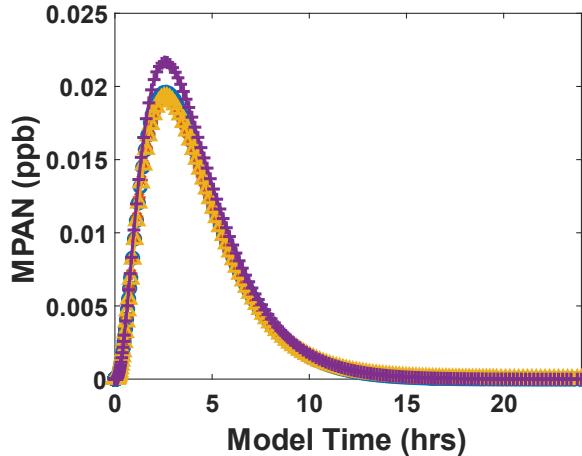


Figure S11: Methyl Peroxy Acetyl Nitrate (MPAN) concentrations for 109-species, 74-species, and 15-species AMORE 2.0 reduced isoprene mechanisms compared to the full Caltech Isoprene mechanism for six conditions (Table 1).

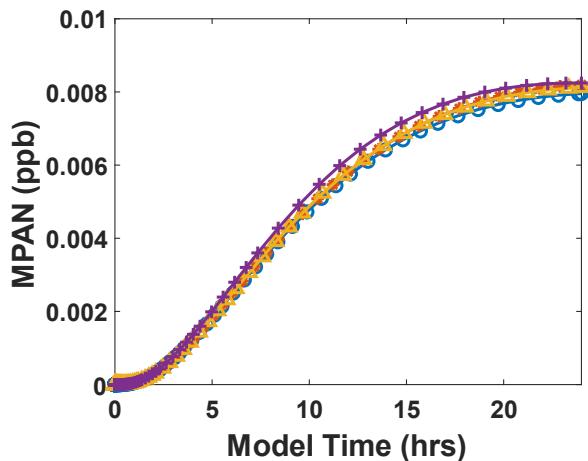
a) High OH



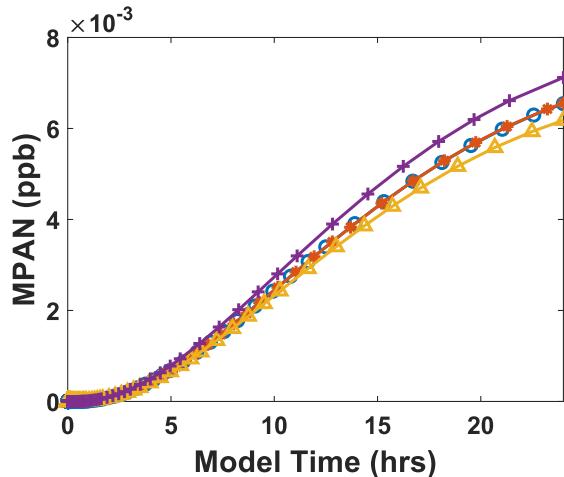
b) High OH and NO



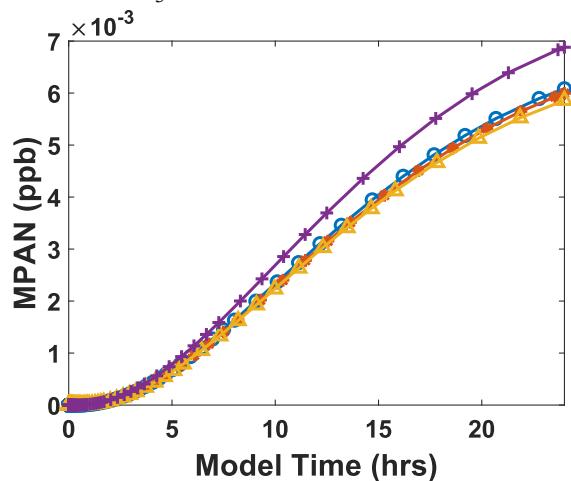
c) High O₃



d) High NO₃



f) High NO₃ low hν



e) High Isoprene

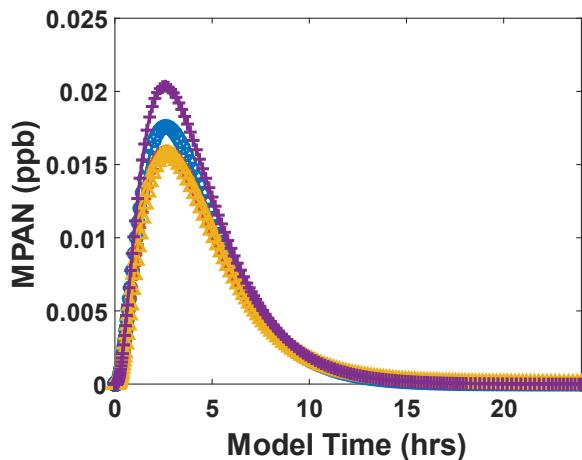


Figure S12: Methacrolein (MACR) concentrations for 109-species, 74-species, and 15-species AMORE 2.0 reduced isoprene mechanisms compared to the full Caltech Isoprene mechanism for six conditions (Table 1).

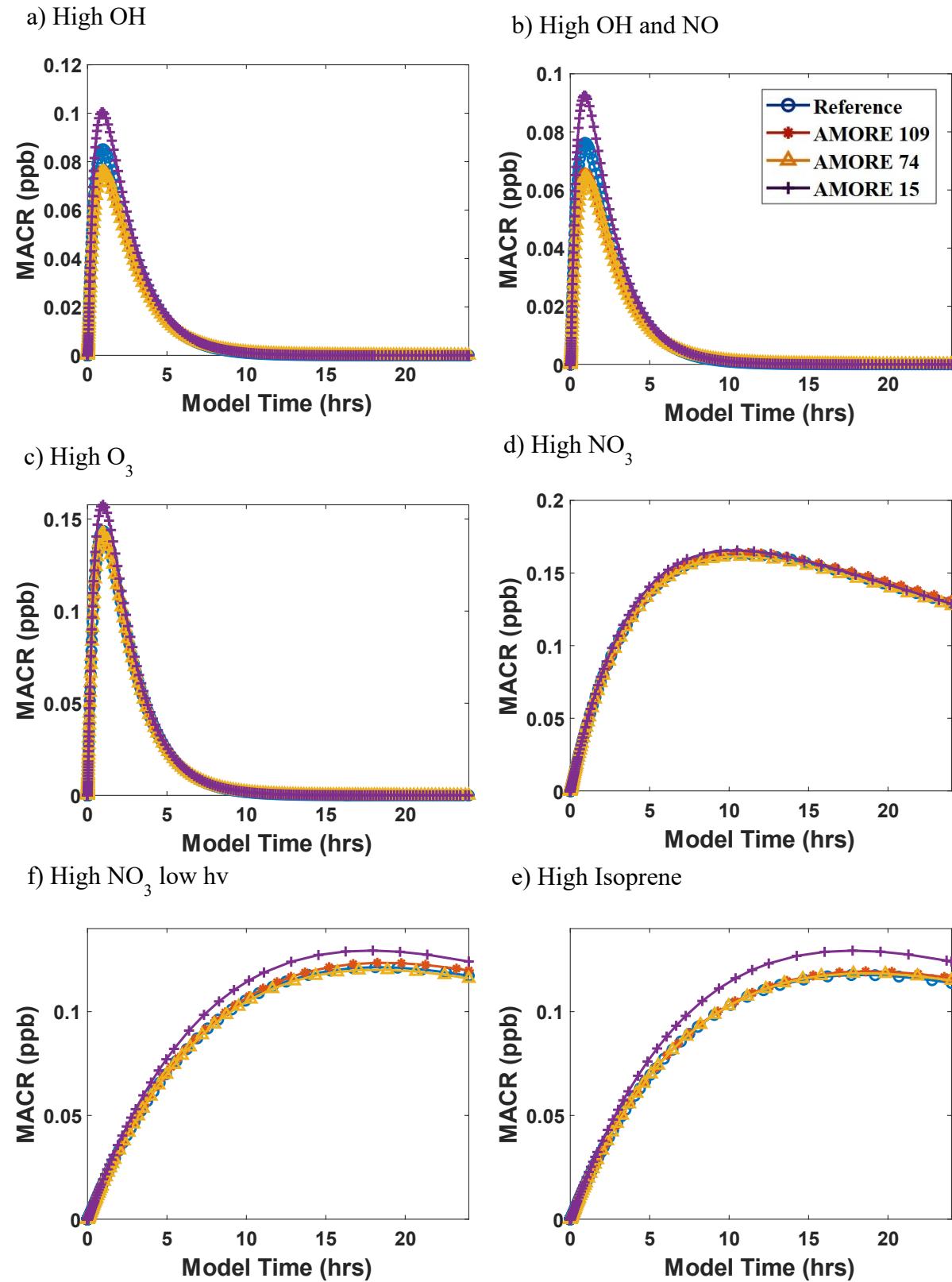


Figure S13: Methacrolein (MACR) concentrations for 109-species, 74-species, and 15-species AMORE 2.0 reduced isoprene mechanisms compared to the full Caltech Isoprene mechanism for six conditions (Table 1).

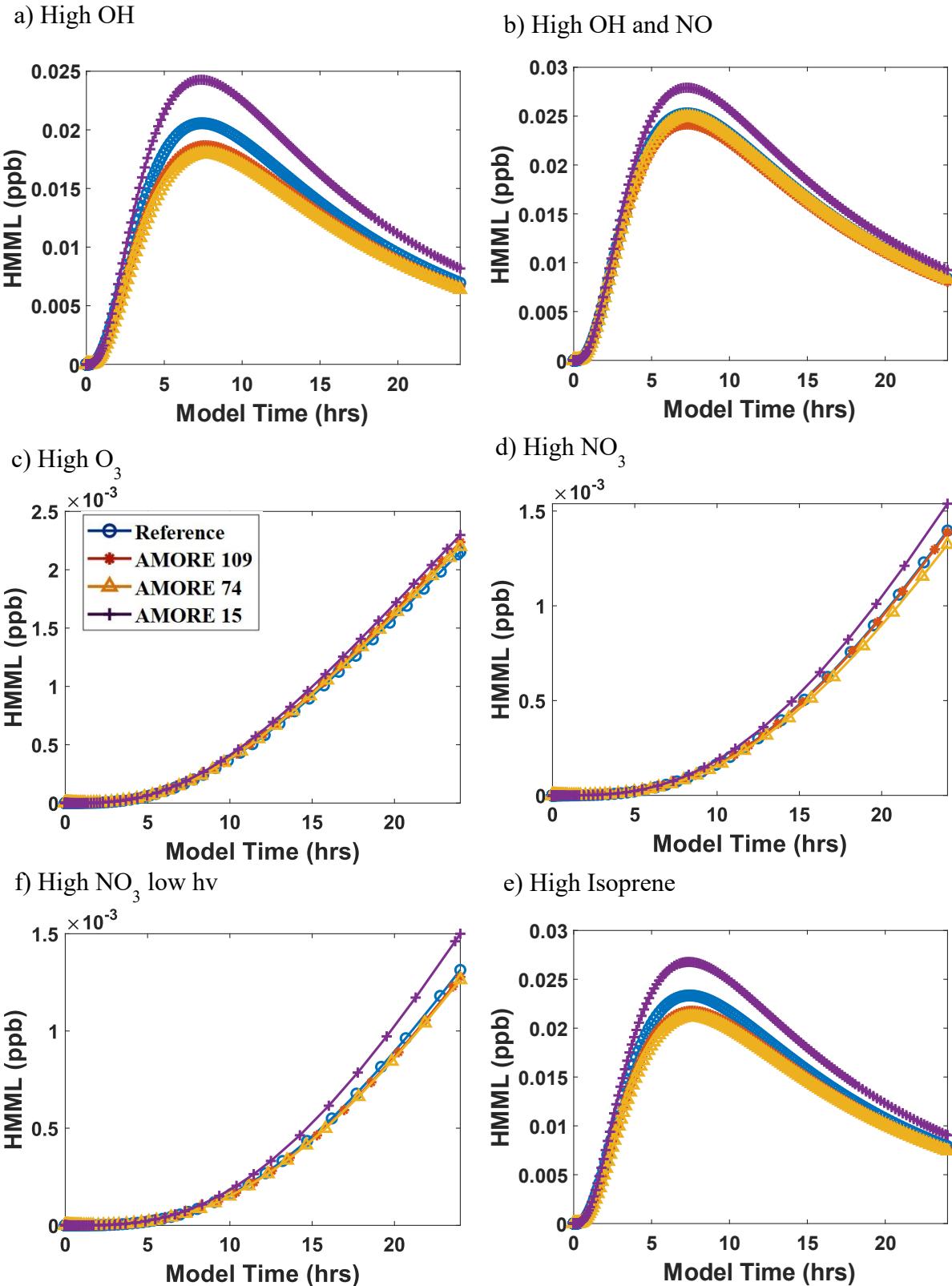
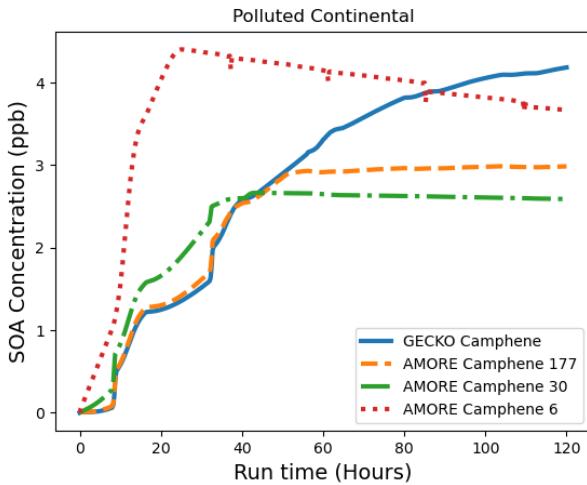
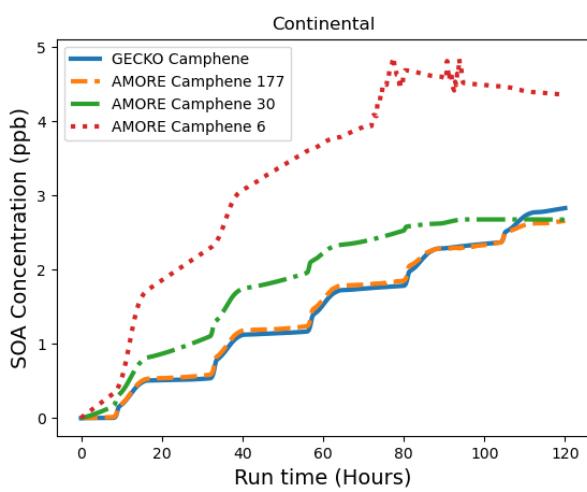


Figure S14: SOA concentration of 177-species, 30-species, and 6-species AMORE 2.0 reduced camphene mechanisms compared to the full GECKO-A camphene mechanism for 9 conditions (Table S1) over 120 hours.

a)



b)



c)

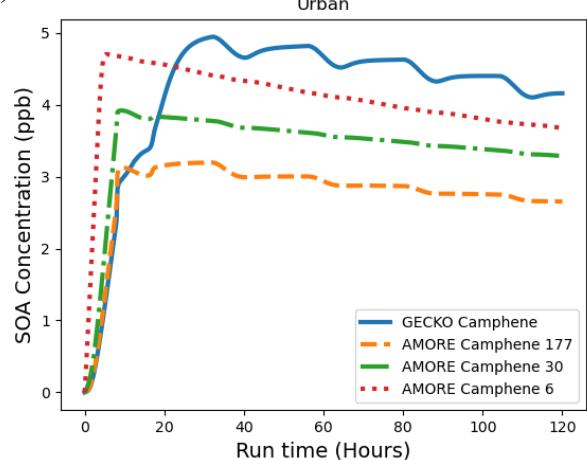
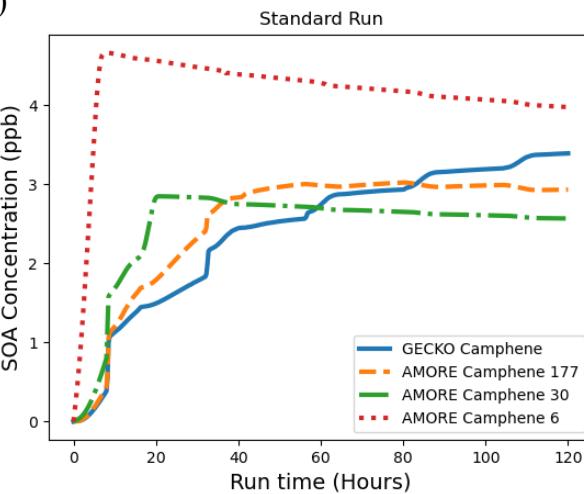
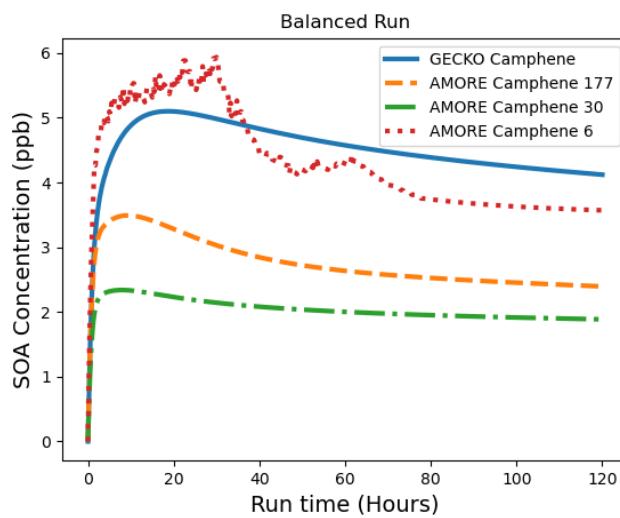


Figure S14 continued:

d)



e)



f)

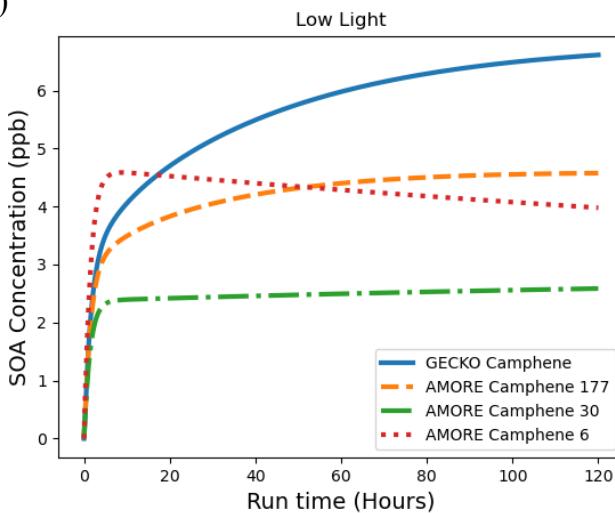


Figure S14 continued:

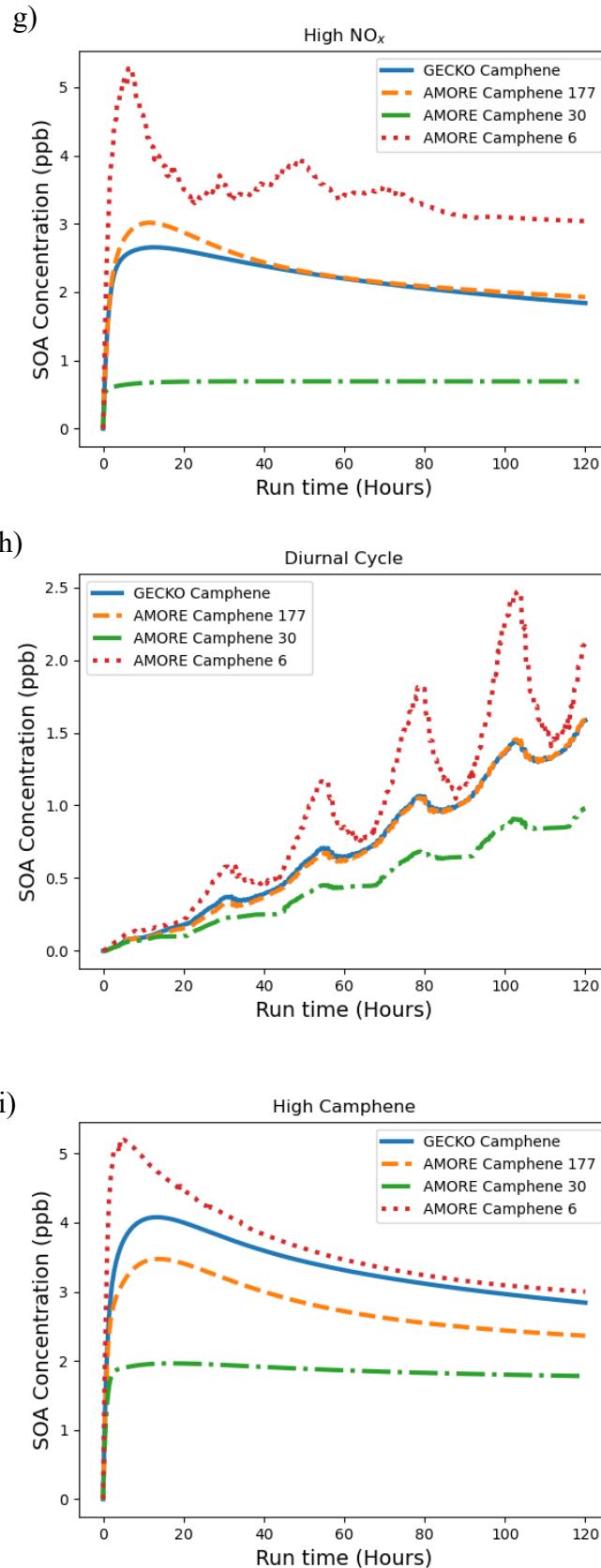
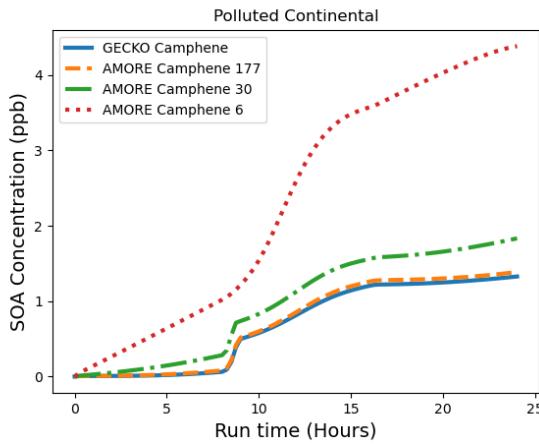
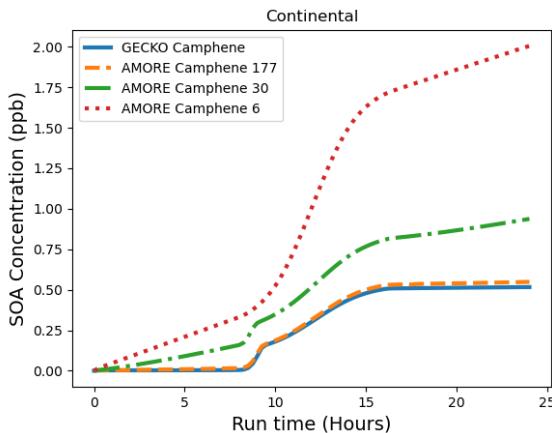


Figure S15: SOA concentration of 177-species, 30-species, and 6-species AMORE 2.0 reduced camphene mechanisms compared to the full GECKO-A camphene mechanism for 9 conditions (Table S1) over 25 Hours.

a)



b)



c)

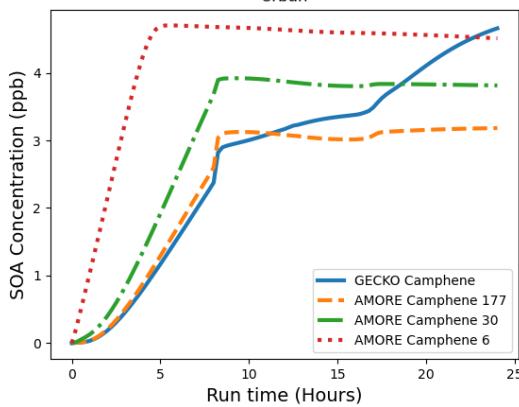
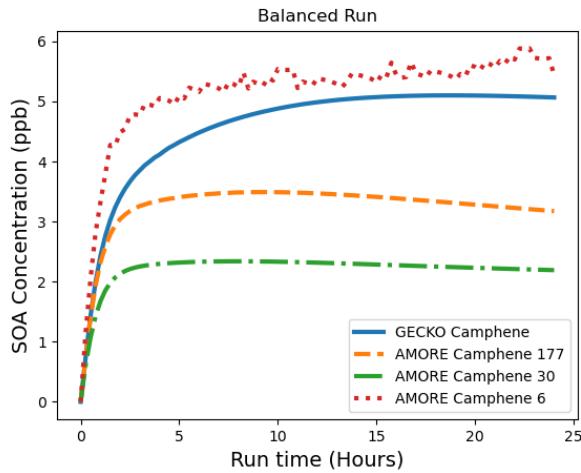
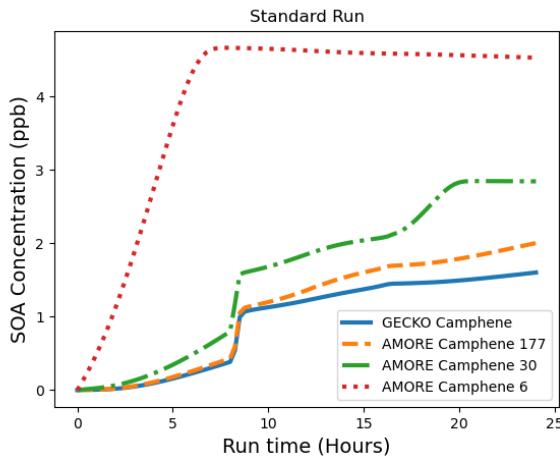


Figure S15 continued:

d)



e)



f)

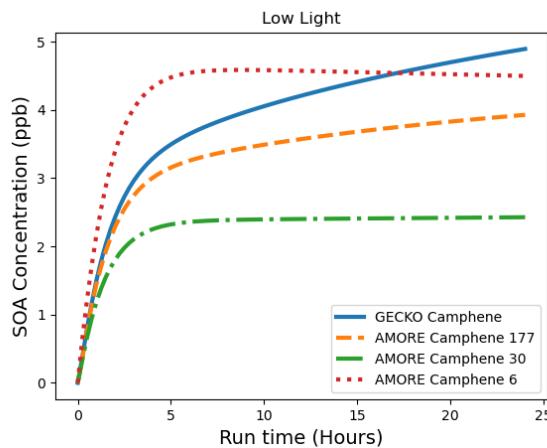
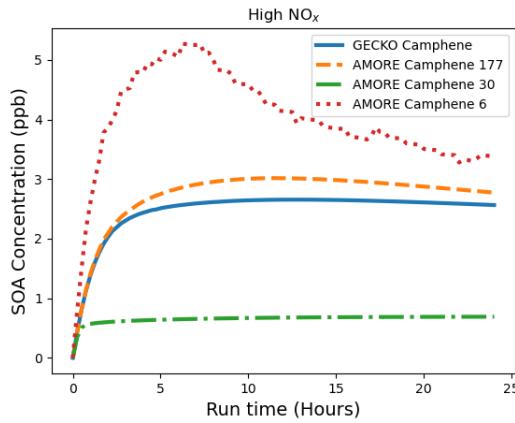
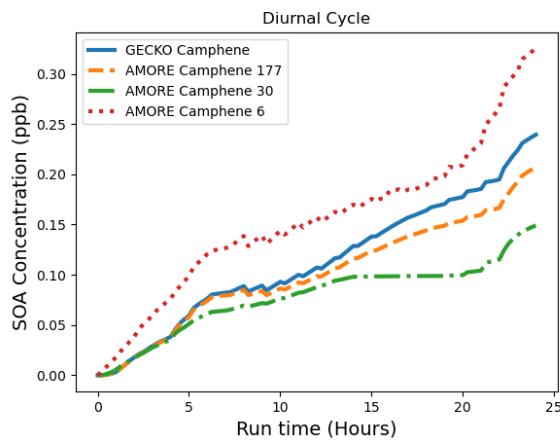


Figure S15 continued:

g)



h)



i)

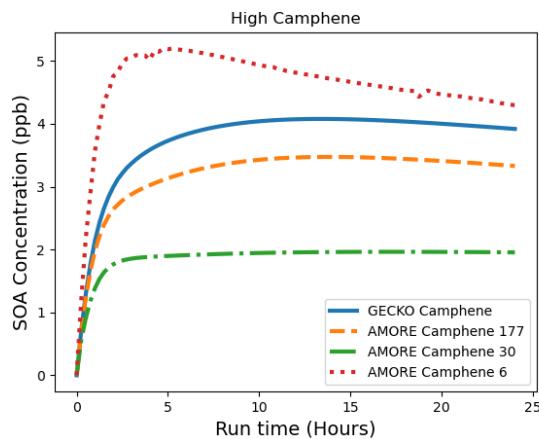
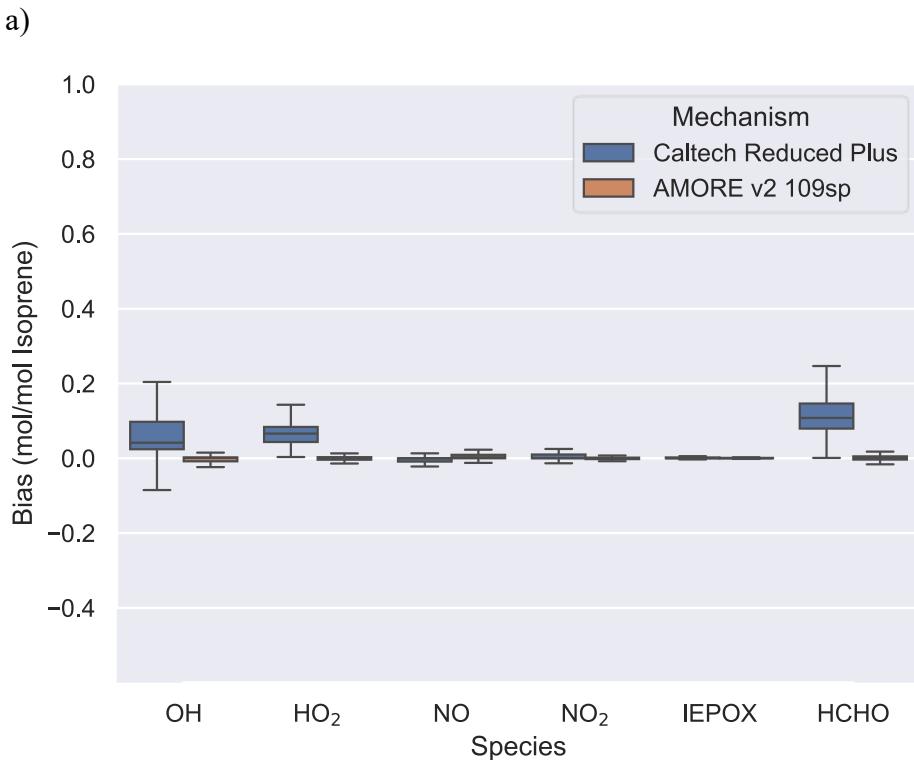


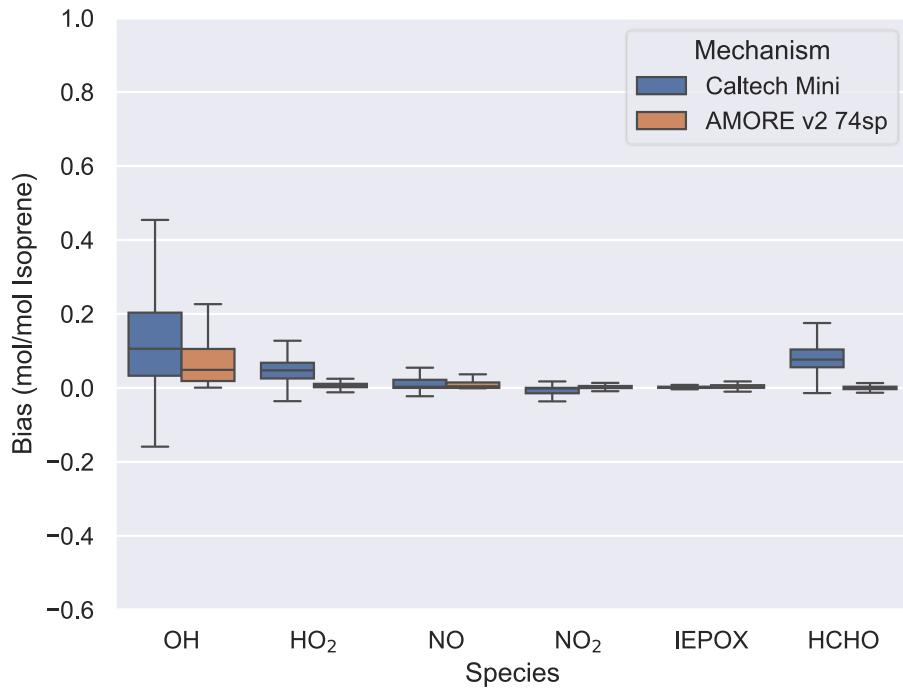
Table S1: Conditions used in the GECKO-A box model for the Camphene mechanism plots. Fixed values are bolded.

Condition	Camphene	HO	HO ₂	NO	O ₃	NO ₃	SZA
Continental	10	0.0001	0.03	0.1	41	0.00005	Diurnal
Polluted Continental	10	0.0002	0.03	0.4	41	0.0001	Diurnal
Urban	10	0.0002	0.003	6	41	0.0001	45
Standard	10	0	0	10	41	0	Diurnal
Balanced	10	0.0002	0.007	0.08	20	0.007	0
Low hν	10	0.00005	0.003	0.08	20	0.007	90
High NO _x	10	0.0002	0.007	2	0	0	0
High Camphene	100	0.0002	0.003	2	50	0.007	45
Diurnal	0.5	0.00001	0	0.005	21	0	Diurnal

Figure S16: Complementary plots to Figure 3 for 6-hour runtimes.



b)



c)

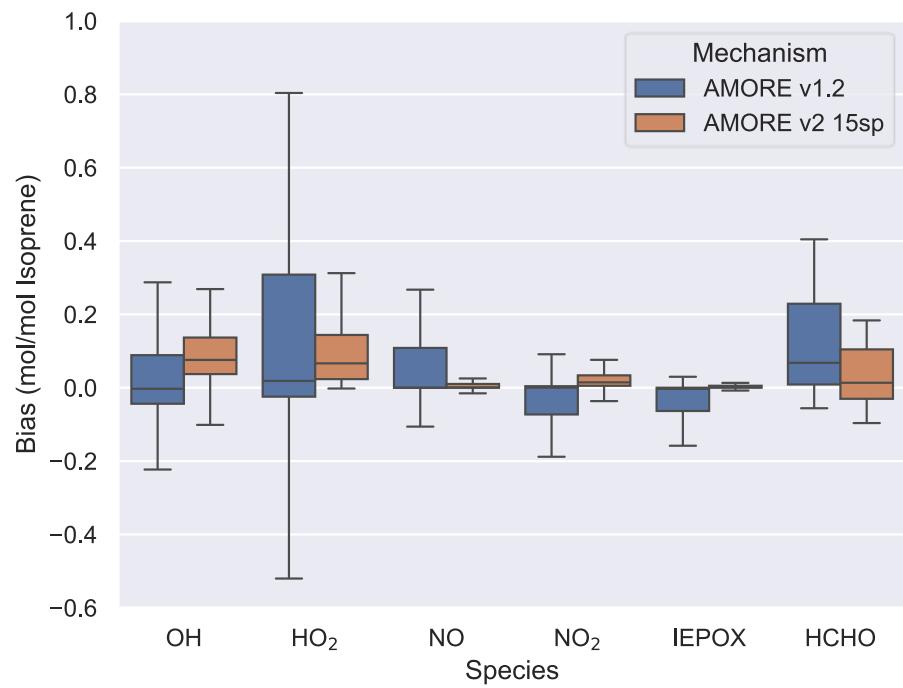
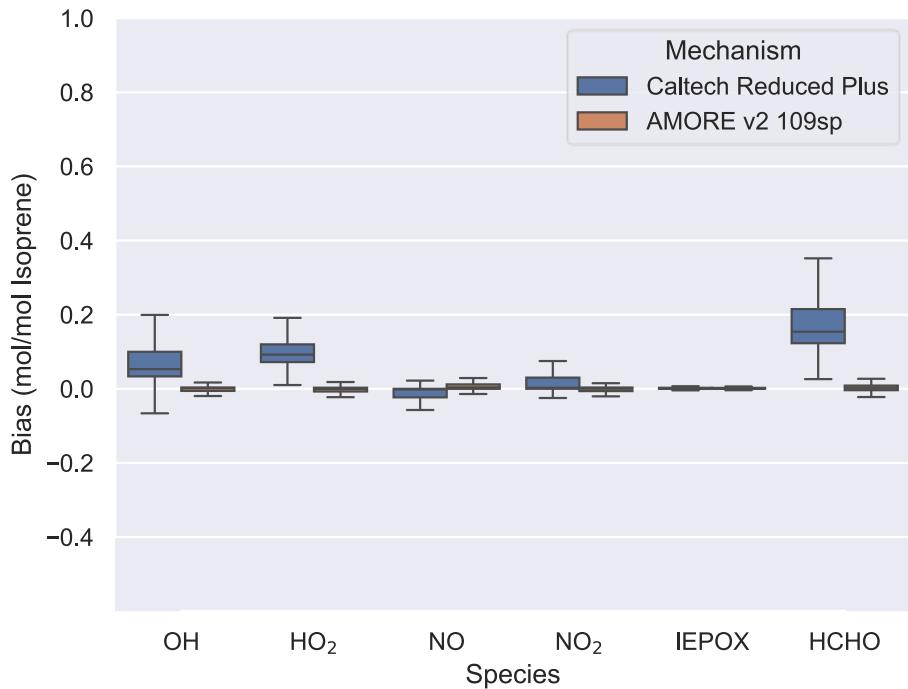
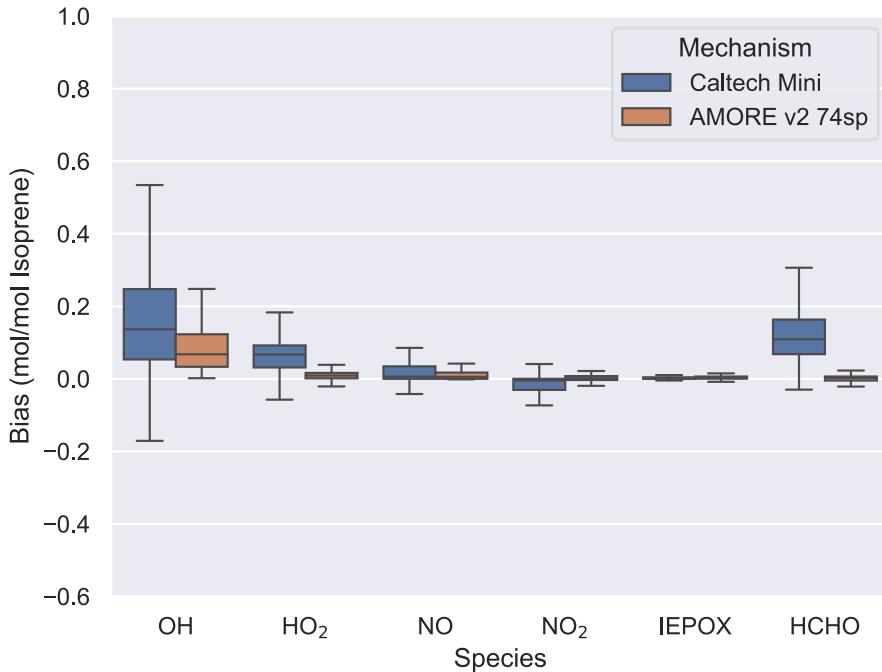


Figure S17: Complementary plots to Figure 3 for 12-hour runtimes.

a)



b)



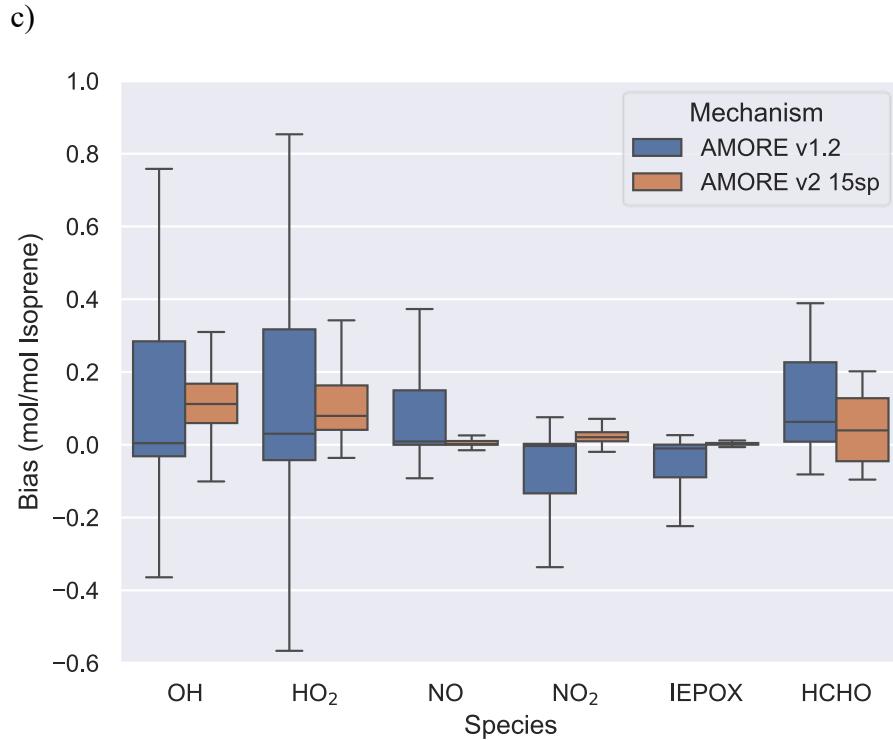
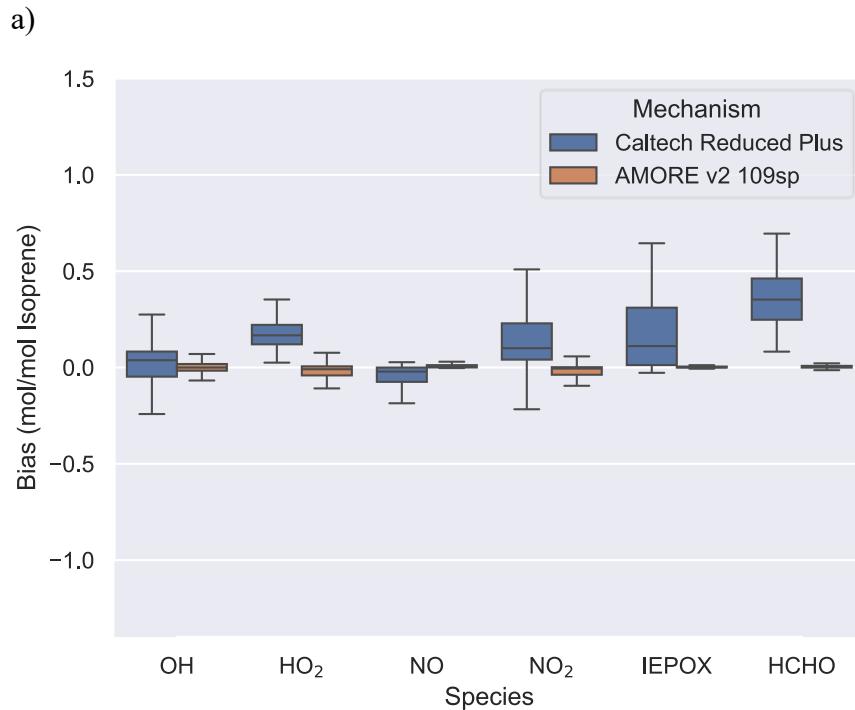
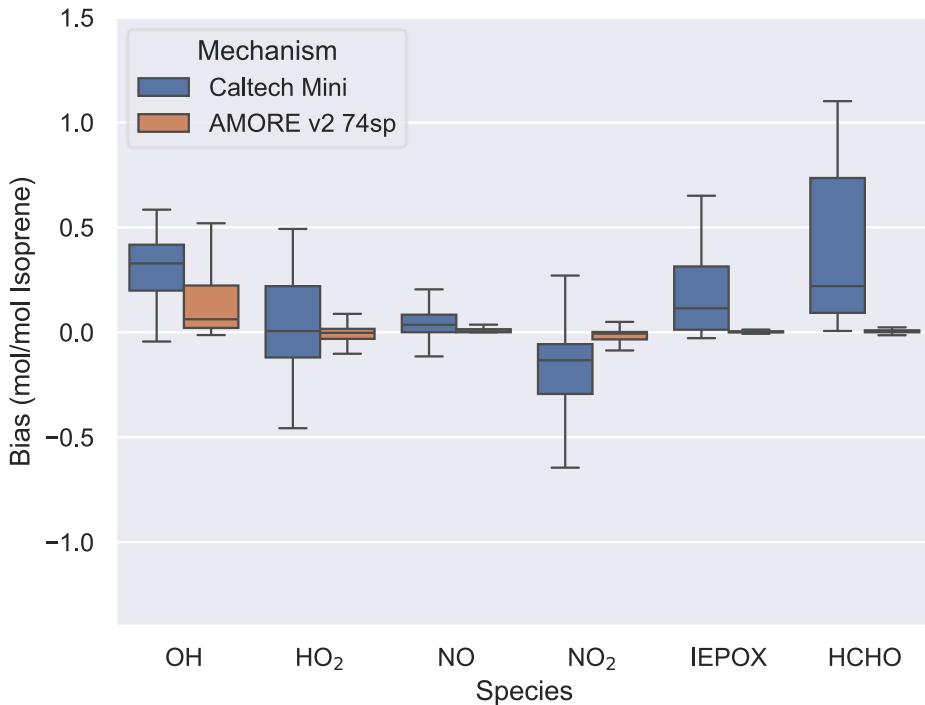


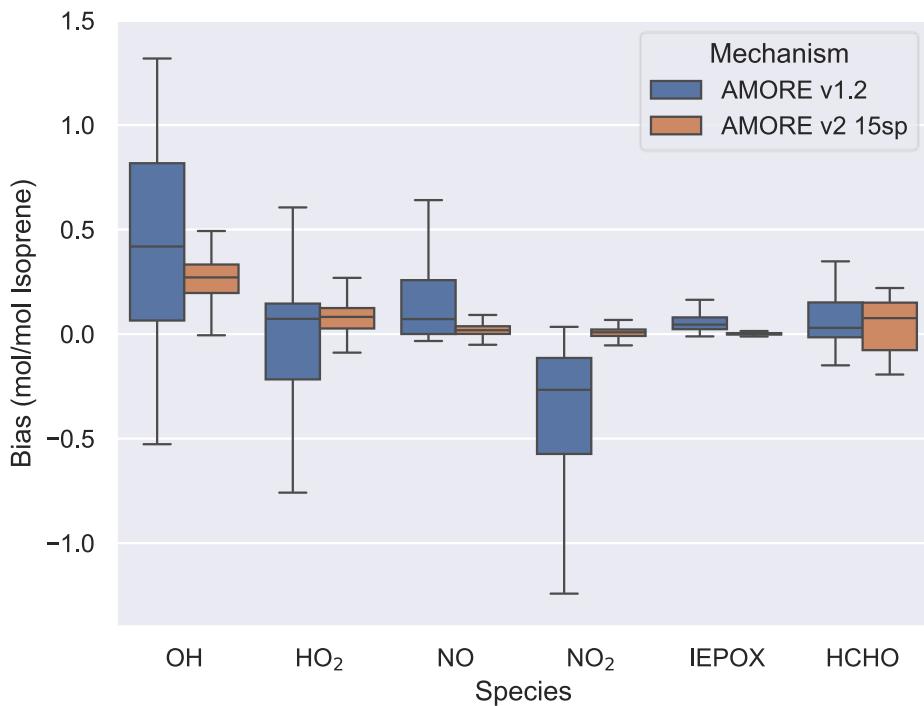
Figure S18: Complementary plots to Figure 3 for 200-hour runtimes.



b)



c)



Section S2: Discussion of condition selection for the AMORE 2.0 algorithm.

One of the key inputs of the AMORE 2.0 algorithm is the selection of atmospheric conditions as inputs to the algorithm. These conditions include all parameters needed to calculate the relative rate of the mechanism reactions, which includes meteorological parameters such as temperature, pressure, and solar intensity, and also the concentration of secondary reactants. The relative rate is defined as the rate of reaction per unit concentration of the primary reactant. This relies on being able to evaluate the rate constant and having a concentration for the second reactant of the reaction if present.

The necessary species concentrations provided depend on the full mechanism. Ideally, a concentration (in ppb) will be provided for all secondary reactants. The user must provide a set of species to be considered as secondary reactants by default, which generally includes OH, HO₂, NO, NO₂, NO₃, O₃, the methyl radical, and the acyl-peroxy radical. If these species are present in a reaction, then they will be considered the secondary reactant, regardless of their position in the reaction reactant list (1st or 2nd). If there are two reactants and neither reactant is in this secondary reactant set, then the 2nd species in the reactant list will be assumed to be the secondary reactant, and a concentration should be provided for it. If there is no concentration provided for a secondary reactant, a low default concentration will be assigned to calculate the relative rate of reaction.

As such, concentrations of the most common secondary reactants (listed above) is essential for accurate reduced mechanisms. In addition, most mechanisms contain reactions with organic radicals as the second reactant. In the GECKO-A box model, these species are binned into groups, termed PERO1, PERO2, etc. In the isoprene mechanism, these species are listed explicitly. We recommend running a box model of the full mechanism and taking ½ the maximum concentration of any organic radical which is a secondary reactant. These concentration values should be used as input conditions for the algorithm.

Secondary reactant concentrations are generally not constant. For example, NO_x concentrations vary significantly with location and time of day. To calculate the relative rate, one concentration must be selected. To work around this limitation, the AMORE 2.0 algorithm accepts multiple input conditions, which will lead to the calculation of multiple relative rates. These multiple rates will be used to construct multiple graphs which are then averaged together to create a representative graph. So, by providing multiple conditions, the reduced mechanism will be representative of a broader set of conditions.

A simple example involving the species NO and OH would be to provide a high and low NO condition. The two conditions would be: {OH: 10⁻⁴ ppb, NO: 0.05 ppb}, {OH: 5*10⁻⁵, NO: 2 ppb}. The first stages of the AMORE 2.0 algorithm, including the calculation of relative rates,

species yields, and graph construction, would proceed for both of these conditions. Then, the graph and yields would be averaged over the two conditions and the remainder of the algorithm would proceed with the average value.

For the mechanisms presented in the paper, we chose to use six input conditions. This was a balance between having broader condition representation and algorithm efficiency, as more conditions increase the run time. The conditions used are shown in Table 1 in the paper.

There are two approaches to condition selection. The first is to prioritize one condition per mechanism and create multiple mechanisms for multiple conditions. If this method is chosen, then only one condition should be used for each reduced mechanism. The second approach is to create a broad scope reduced mechanism from multiple conditions. We have conducted additional investigations into the impact of selecting different numbers of input conditions on the accuracy of the reduced mechanism output.

To do this, we again utilize the 1000 data points from a July 2016 single day global GEOS-Chem run. This data set was created by selecting data points from this run where isoprene concentrations were higher than 0.1 ppb. This data set is available in the supplementary files and is representative of the conditions that would be expected for a reduced isoprene mechanism in GEOS-Chem. This data set covers a wide range of conditions, shown in Table S2.

Table S2: Minimum, maximum, and average values for the parameters in the GEOS-Chem dataset. All species have units of parts per billion (ppb). Sun is a unitless fraction of full solar intensity.

	Isop	OH	HO2	NO	NO2	O3	NO3	Sun	Temp (K)	Air Den (kg/m ³)
Minimum	0.10	2.5*10 ⁻⁷	8.0*10 ⁻⁵	0	0.0012	4.9	1.9*10 ⁻⁸	0	227	0.34
Maximum	111	9.3*10 ⁻⁴	3.6*10 ⁻²	5.7	18	85	1.7*10 ⁻³	0.39	315	1.3
Mean	2.6	8.5*10 ⁻⁵	7.3*10 ⁻³	0.056	0.21	28	3.3*10 ⁻⁵	0.011	286	1.0

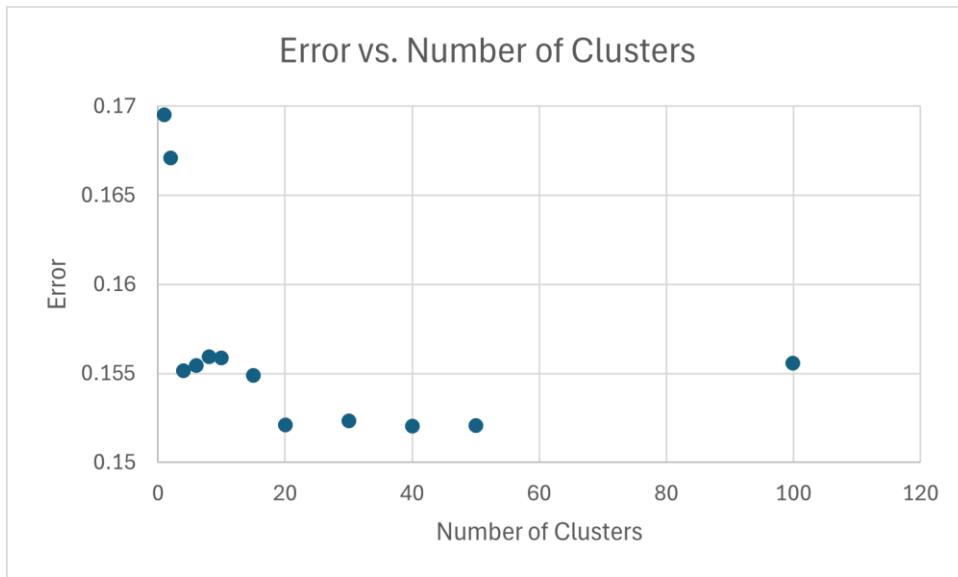
From this dataset, we utilized k-means clustering to divide the dataset into bins. This was done by normalizing the dataset and then running the sci-kit learn k-means clustering python package. Each cluster contains a subset of the data points which are relatively close to each other relative to the rest of the data. The cluster centers obtained from this method were used as input conditions to the AMORE 2.0 algorithm. We repeated this process for 1, 2, 4, 6, 10, 20, 30, 40, 50 and 100 clusters to determine the impact of increasing the number of clusters on the performance of the resulting reduced mechanisms.

For each set of clusters, we ran the AMORE 2.0 algorithm to 53, 43, 33, 28, 22, and 14 species. We measured the performance of each of these reduced mechanisms against the full mechanism under all of the 1000 original data points using the rapid yield algorithm with Equation 14 and comparing the following species: OH, HO2, NO, NO2, NO3, O3, methyl radical, acyl-peroxy

radical, isoprene, IEPOX, MVK, formaldehyde, HAC, PAN, GLYC, PYRAC, MGLY, HCOOH, MACR, MPAN, GLYX, HMML.

The average performance of all of the reduced mechanisms across all 1000 data points was used to determine the performance of the selected clusters. Figure S19 shows the performance of the different numbers of clusters.

Figure S19: Performance (error) of AMORE 2.0 reduced isoprene mechanisms across a dataset of 1000 GEOS-Chem derived datapoints reducing using varying number of clusters from that dataset as inputs to the algorithm.

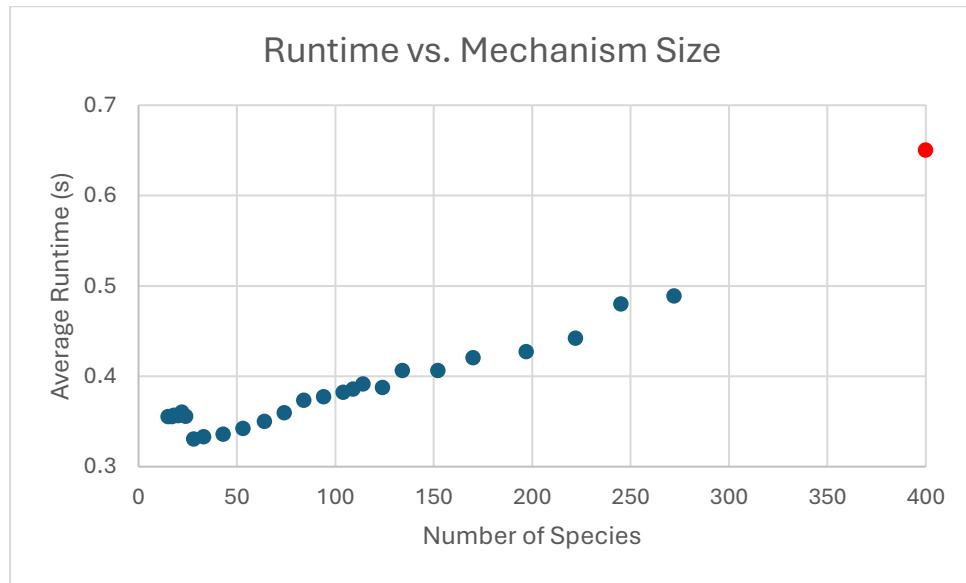


Very few (1 or 2) clusters did not perform well in this test, suggesting that at least 4 input conditions are necessary for accurate reduced mechanisms for a large dataset using this method. However, beyond 4 clusters, there is no clear correlation between number of clusters and performance. While 20, 30, 40, and 50 clusters collectively performed the best, 100 clusters was no better than 4. Given the inconclusive results, we suggest that the user use at least 4 input conditions if using the k-means method for a large representative dataset. Furthermore, we suggest that any reduced mechanisms are tested in a box model to confirm performance against the desired conditions.

Section S3: Mechanism runtimes in the F0AM box model.

The ultimate goal of mechanism reduction is create mechanisms that are more computationally efficient to run in large scale models. Previous work (Yang et al from the main paper) has demonstrated that the AMORE v1.1 mechanism was able to reduce GEOS-Chem model runtimes by up to 10%, and speed up chemistry by up to 25% due to the replacement of a much larger isoprene model (Caltech Mini). While we did not conduct any transport model runs for this work, we have done preliminary testing of AMORE reduced model runtimes in the F0AM box model. These runtimes are not expected to carry over proportionally to a chemical transport model, but do demonstrate the general decrease in runtimes attributable to mechanism reduction. Figure S20 shows the runtime of several AMORE 2.0 reduced isoprene mechanism with the full isoprene mechanism runtime for reference. These runtimes are averaged over six different run conditions in the F0AM box model (Table 1).

Figure S20: Mechanism average runtime vs. number of species over six conditions (Table 1) in the F0AM box model for different AMORE 2.0 isoprene mechanisms with the Caltech full mechanism (red) for reference.



The runtime scales roughly linearly with the number of species, with the smallest mechanisms having roughly half the runtime as the full mechanism. One interesting feature is that the smallest mechanisms (<28 species) have slightly higher runtimes than slightly larger mechanisms. It is unclear whether this is an artifact of uncertainty in the measurement. Either way, we do not expect this data to carry over precisely to CTMs due to the significant differences in the mechanics of CTMs compared to box models. This result does demonstrate that the AMORE 2.0 reduced mechanisms do succeed at reducing the model runtime.