Supporting Information for AMORE-Isoprene v1.0: A new reduced mechanism for gas-phase isoprene oxidation

1. Full Isoprene Mechanism Update

The Caltech full Isoprene mechanism (Bates and Wennberg 2018) was updated for this work. The original mechanism did not contain complete oxidation pathways for all of the species present in the mechanism. The reasoning behind this was that several species did not have published or known oxidation schemes. Since this mechanism was not designed for use in 3D models, the incomplete chemistry was not an issue.

However, for this work, complete chemistry was needed, as we were attempting to utilize the accuracy of the full mechanism for the reduction process. To do so, we needed to update the full mechanism to contain oxidation pathways for all species. Most of the work for this process was already done by Bates et al. in the preparation of their reduced plus isoprene scheme. However, they completed oxidation pathways only after lumping many multifunctional isoprene species together. They used SAR and existing mechanisms (MCM) to complete the chemistry for these species.

We utilized the information from their reduced plus isoprene scheme to create a complete full mechanism. To do so, we first created a correspondence between lumped species in the reduced plus mechanism and species in the full mechanism without an oxidation pathway. From there, the reactions of the lumped species were replicated for the un-lumped species. If a lumped species was a reactant, then a new reaction was created for each species that it represented. If a lumped species was a product, then the reaction coefficient was divided evenly into the represented species set, thus conserving carbon flux between the two mechanisms.

In addition to incomplete isoprene chemistry, the Caltech full mechanism did not complete oxidation pathways for species considered outside of the isoprene scheme. The oxidation chemistry for these additional species was taken from the MCM scheme.

The updated Caltech full isoprene mechanism is ideal for box model simulations for the purpose of mechanism reduction, as it can be considered more accurate for dynamic oxidant concentrations and common oxidation products such as formaldehyde.

We have done some preliminary testing of this mechanism in comparison to experimental data, but it was difficult to attain meaningful results. Thus, further assessment of this updated mechanism is required.

In the supplementary files, there is a folder entitled “Caltech Full Mechanism (AMORE update)” which contains all of the files related to this mechanism. The “READ_ME” file gives a brief description of each file in the folder.

2. Priority Species List

Below is a list of priority species which were included in the final mechanism and tested for accuracy under multiple conditions.

Table S.1

Name	Name in Mechanism
Isoprene	ISO
Isoprene epoxydiols (lumped)	IEPOX
Isoprene nitrates (lumped)	ISON
Formaldehyde	HCHO
Methyl vinyl ketone	MVK
Methyl glyoxal	MGLY
Glyoxal	GLY
Methacrolein	MACR
Peroxyacetyl nitrate	PAN
Peroxyacetyl radical	ACO3
Methyl radical	MO2
Hydroxyl radical	OH
Hydroperoxyl radical	HO2
Nitric oxide	NO
Nitrogen dioxide	NO2
Nitrate radical	NO3
Ozone	O3

3. Directed Relation Graph method discussion

The Directed Relation Graph (DRG) method has been employed in the past to reduce chemical mechanisms. This method was initially trailed for AMORE and it was determined that the method was unsuitable for use in this project. The reasoning behind this is discussed below.

The DRG method starts by representing the full mechanism as a graph. This aspect of the method was retained in the AMORE algorithm. In the DRG method, all species are represented as nodes within the mechanism graph. Edges are created between species in which the first

species contributes directly to the production of the second species. The edge is directed from the first species to the second species.

Each edge is assigned a weighting based on the fraction of production of the end species that is contributed by the start species. This weighting quantifies the relationship between the two species and serves as a natural metric of which edges are more important.

The DRG reduction method works by removing edges from lowest weight onward until a desired mechanism size is reached. Initially, the removal of edges leads to the severing of ties between two species, usually by removing the reactions or components of reactions that they share. However, once a species has been isolated from all other species, it is effectively removed as well. Thus, edge removal leads to the removal of both reactions and species. There are several methods that elaborate on this process, by complexifying the edge weighting scheme. Hothever, the primary issue with DRG for the application to the isoprene mechanism is not the weighting scheme, but rather the method of removal.

In the DRG method, model reduction only occurs through the removal of species and reactions. Thus, this method assumes that there is a sufficient number of unimportant species and reactions that can be identified and removed without loss of accuracy. The implication for the isoprene mechanism, where our target mechanism is one to two orders of magnitude smaller than our original mechanism, is that upwards of 90% of the species and reactions can be removed without significant impact to the species we are trying to retain. Clearly, this assumption is not valid for this application. Additionally, given that important species are divided amongst multiple longer paths, the minimum achievable mechanism size that still maintained a connection between isoprene and the species listed in SI section 2 using the DRG method was still larger than the desired mechanism size of <10 intermediate species.

Given this issue, there was no apparent way to implement the DRG method in creating the AMORE-isoprene mechanism. The AMORE algorithm differs from the DRG method by attempting to summarize the core elements of the entire mechanism first by creating a much smaller representative structural graph, and then filling in the coefficients by probing the full mechanism. This is a categorically different approach because it does not rely on removing unimportant parts of the mechanism.

4.1 Cycle to node reduction algorithm

One of the key components of the AMORE-isoprene algorithm is the Yield Estimation Algorithm (YEA) (section 2.3.1). The YEA is used to estimate the time independent yields of any isoprene derived species assuming all reactions have gone to completion and all oxidant concentrations are held constant. The primary source of difficulty and inaccuracy in this method is the treatment of cycles. Cycles are instances in which directed edges form a loop between species, meaning that carbon can return to its starting point within the cycle.

In a non-cyclical section of the graph, carbon moves in one direction, and can be tracked simply by measuring the branching ratio of the set of products from a given reaction. Within cycles,

carbon may move within the cycle multiple times and the distribution of carbon to products connected to the cycle is not obvious. However, the distribution of carbon is uniquely determined by the input conditions (under constant oxidant concentrations) and thus should be able to be incorporated into a model.

The goal of the cycle sub algorithm is to accurately estimate the yield of species that are sinks for a cycle from a source species for the cycle. Source species are defined as those with directed edges that enter the cyclical section. Sink species are defined as those with directed edges that go from the cycle to the sink species. Source and sink species are outside of the cycle, and thus their carbon is not cycled.

The first step in the cycle sub algorithm is to treat the entire cycle as a single entity that behaves like a normal species, with carbon passing through to other species. Thus, the complicated dynamics that occur within the cycle must be estimated. The equation below shows the y_a of species A that is in or connected to the cycle.

$$\mathrm{k}_{r,net} = k_r \times [reactant] \text{ if reactant} = \text{input oxidant}, k_r \text{ else}$$
$$\mathrm{f}_{a,b} = \sum_{r=1}^{R} k_{r,net} \times c_{b,r}/c_{a,r} \text{ if a} = \text{reactant, b} = \text{product}$$
$$\mathrm{y}_a = \sum_{n=1}^{N} f_{n,a} \text{ if } \mathrm{f}_{a,n} = 0, (f_{n,a}/f_{a,n})\alpha \text{ if } \mathrm{f}_{a,n} > 0$$

R is the number of reactions, r is an individual reaction, k_r is the rate constant for reaction r, k_{rnet} is the effective rate constant under constant oxidant concentration, N is the number of species, n is an individual species, f_{a,b} is the production of b from a, c_{a,r} is the stoichiometric coefficient of a in reaction r, and y_a is the yield of a species within a cycle. Here, \alpha is a tunable parameter which modulates how much is contained within the cycle versus distributed outward. When this parameter is large, flux tends to spread evenly through the cycle before being distributed outward, leading to a more even distribution outward. When this parameter is small, flux is quickly distributed outside of the cycle, leading to an uneven distribution within. This parameter was tuned to most accurately replicate the yields from the F0AM simulation, and a value of 0.0000001 was used, signifying a tendency to quickly distribute flux outward.

4.2 YEA Performance

The performance of the YEA was measured in comparison to the simulated yields from the F0AM 0-D box model. The YEA was used in place of the box model to reduce computation time. Thus, comparison to the box model was used to determine whether the reduction time was worth the accompanied loss in accuracy.

Times were measured on a dell Inspiron 15 with a Ryzen 5000 series processor. The YEA algorithm was able to calculate yields for 10 species under 128 different conditions (1280 yield values) in 4.27 seconds. Running the equivalent F0AM simulations took 132 seconds for the ODE solver, 387 seconds to calculate the yields from the concentration data, and 909 seconds to calculate the yield of isoprene nitrates which is a large grouping of species. This leads to a total run time of 1428 seconds to calculate the equivalent yields in F0AM. The F0AM yield

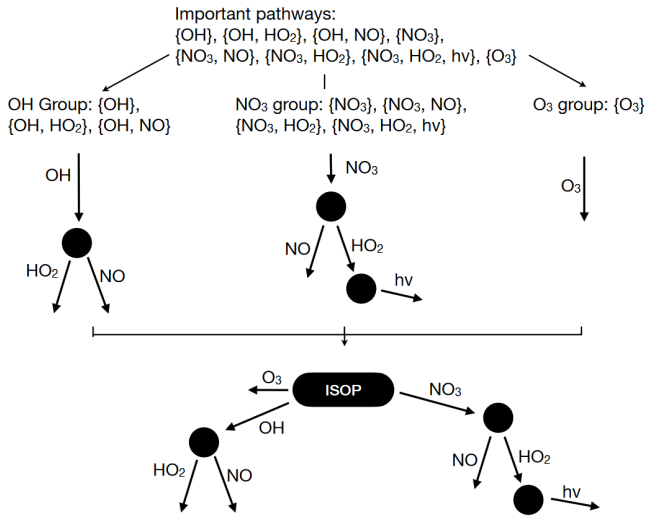
calculation was not optimized for time performance, so they may have significant room for improvement. The YEA was between 30 and 300 times faster than the equivalent F0AM simulation.

The accuracy of the YEA was measured by direct comparison to the F0AM yields. Between all species and conditions, the YEA was off by a factor of 2 on average, with a maximum discrepancy of a factor of 11. Given that yields can span many orders of magnitude, and the AMORE algorithm functions based on changes in yields, this is strong performance for the use case and decrease in time cost. For each measured species, a linear regression analysis was run comparing the F0AM yields to the YEA. The slopes of all regressions were positive as expected, and the R^2 values ranged from 0.01 (MVK) to 0.7 (IEPOXd). Species with higher R^2 values tended to have higher variation in yields. A csv file containing the yields of F0AM, YEA, and the input conditions is given in the supporting files (YEA_analysis_data).

5. Recombination of paths in the algorithm process:

The figure below outlines the process of the pathway combination algorithm. After the important pathways are identified, they are grouped based on shared oxidants, and then paths are created within these groupings which minimize the number of necessary intermediate steps.

Figure S.1



6. Algorithmically generated mechanism

Below is the algorithmically generated mechanism, with species names assigned manually for clarity. Mechanism structure was determined using the pathway importance algorithm and reaction stoichiometric coefficients were determined using the yield estimation algorithm. Rate laws were not algorithmically determined.

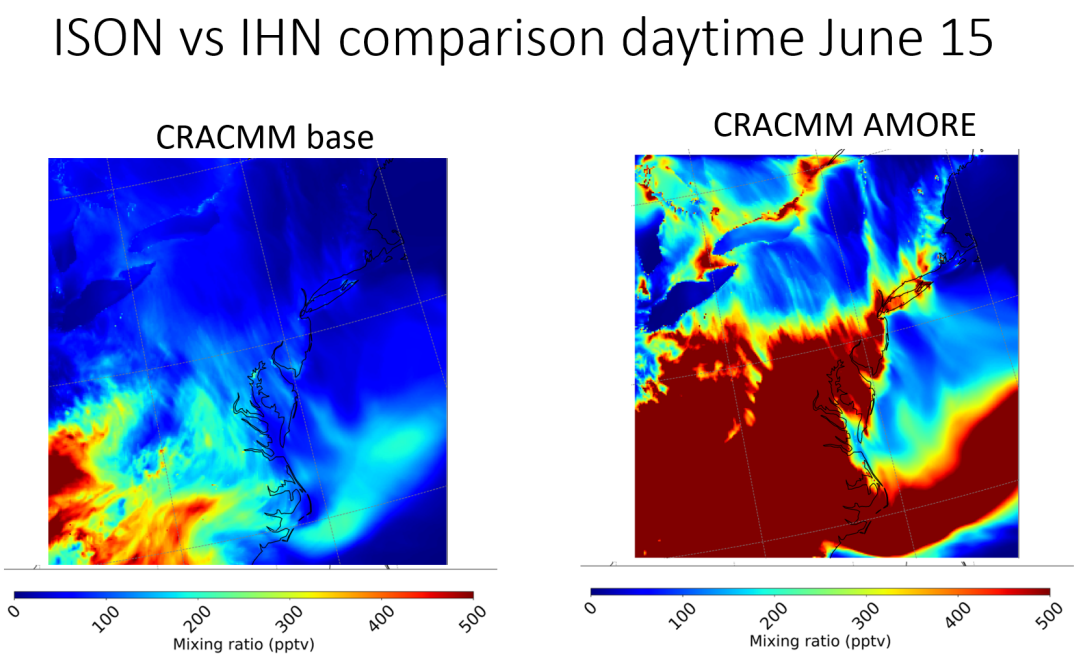
Table S.2

Number	Reaction
1	ISOP + OH --> ISOPOO
2	ISOPOO --> 0.022 IEPOX + 0.040 ISOPN + 0.277 MVK + 0.037 GLY + 0.224 MACR + 0.867 MGLY + 1.12 HCHO
3	ISOPOO + NO --> 0.049 IEPOX + 0.344 ISOPN + 0.405 MVK + 0.027 GLY + 0.321 MACR + 0.196 MGLY + 1.468 HCHO
4	ISOPOO + HO2 --> 0.072 IEPOX + 0.041 ISOPN + 0.362 MVK + 0.091 GLY + 0.281 MACR + 0.711 MGLY + 1.286 HCHO
5	ISOP + NO3 --> INO2
6	INO2 --> 2.3 ISOPN + 0.166 MVK + 0.334 GLY + 0.022 MACR + 0.288 MGLY + 0.44 HCHO
7	INO2 + HO2 --> IPN
8	IPN --> 0.014 IEPOX + 1.94 ISOPN + 0.163 MVK + 0.582 GLY + 0.023 MACR + 0.323 MGLY + 0.415 HCHO
9	IPN + hv --> 1.6 ISOPN + 0.166 MVK + 0.264 GLY + 0.026 MACR + 0.115 MGLY + 1.22 HCHO
10	INO2 + NO --> 2.127 ISOPN + 0.199 MVK + 0.064 GLYX + 0.029 MACR + 0.084 MGLY + 0.788 HCHO
11	ISOP + O3 --> 0.046 IEPOX + 0.135 ISOPN + 0.094 MVK + 0.023 GLY + 0.192 MACR + 0.2 MGLY + 0.974 HCHO

7. Discussion of isoprene hydroxy nitrates (IHN)

The first version of the optimized AMORE mechanism was run in CMAQ for initial evaluation. The results were generally favorable, however it was noted that the species IHN had very high concentrations, as shown in the figure below, which compares IHN concentrations in AMORE-isoprene to ISON concentrations, a comparable species in CRACMM base. The concentrations of IHN were much higher than the ISON concentrations, and upon further investigation, it was found that IHN in AMORE-isoprene was also much higher than in the Caltech full mechanism. Although IHN was not initially considered an important species, newer research has suggested that it is worth retaining IHN accuracy for NOx cycling (Vazquez et al 2020).

Figure S.2



8. Table of rate law descriptions

The table below describes the rate laws for each reaction and how they were determined.

Table S.3

#	Reaction	Rate Law	Source
1	ISO + O3 = 0.07 MACR + 0.189 MVK + 0.58HCHO + 0.25 HO + 0.25 HO2 + 0.58 HCHO + 0.08 MO2 + 0.1 ACO3 + 0.09 H2O2 + 0.1 MACP + 0.461 MACR + 0.14 CO + 0.28 ORA1 + 0.15 OLT	1.58E-14 exp(-2000/T)	Caltech, sum of isoprene + O3 rates
2	ISO + NO3 = INO2 + 0.3 HCHO + 0.3 NO2 + 0.3 ISON	2.95E-12 exp(-450/T)	Caltech, sum of isoprene + NO3 rates

3	ISO + HO = ISOP + 0.02 MO2	2.69E-11 exp(390/T)	Caltech, sum of isoprene + OH rates
4	ISOP + HO2 = ISHP + 0.6 HO2 + 0.15 HCHO	4.5E-13 exp(1300/T)	Standard form of many HO2 reactions in Caltech full, A constant calibrated for best match
5	ISOP + NO = 0.14 IHN + 0.7 HCHO + 0.44 MVK + 0.88 HO2 + 0.78 NO2 + 0.28 MACR + 0.021 GLY	2.7E-12 exp(350/T)	Standard form of many NO reactions in Caltech full, A constant calibrated for best match
6	ISHP + HO = ISOP	4.6E-12 exp(200/T)	Modeled after reaction from Caltech full: ISOP1OH2OOH + OH = ISOP1OH2OO : 4.6E-12*EXP(200./TEMP);
7	INO2 + HO2 = IPN + HO	3.14E-14 exp(580/T)	Standard form of many HO2 reactions in Caltech full, A constant calibrated for best match
8	INO2 + NO = 0.2 ISON + 0.9 HCHO + 0.5 MGLY+ 0.8 MVK+ 0.5 NO2+ HO2+ 0.1 MO2	9.42E-16 exp(580/T)	Calibrated for INO2 pathway balancing
9	IPN + HO2 = 0.2 ISON + 0.8 NO2 + 0.4 HCHO + 0.05 GLY + 0.1 MGLY + 0.4 MACR + HO2 + 0.94 MVK + 0.1 MO2	3.4E-11 exp(390/T)	Calibrated for HCHO production rate
10	IHN + HO = ISON + HO + 0.2 IEPOX	2.4E-7 exp(580/T)	Calibrated for IHN concentration accuracy
11	ISHP + HO = 0.05 IPC + 0.15 HCHO + 0.05 MGLY + 0.15 MACR + 0.02 GLY + 0.2 MVK + 0.4 NO2 + 0.58 IEPOX + 0.8 HO	2.97E-11 exp(390/T)	Standard form of many OH reactions in Caltech full, A constant calibrated

			for best match
12	ISHP = 0.4 HCHO + 0.1 MGLY + 0.06 ACO3	Photol(HCHO_RAD_RACM2)	Chosen from RACM2 photolysis options to best match rate in box model
13	IPC + NO = 0.35 NO2 + 0.8 NO	1e-10	Calibrated for NOx cycling
14	ISON + HO = CO + 0.12 NO2	5e-11	Calibrated for ISOPN accuracy
15	ISON + NO3 = CO	2e-14	Calibrated for ISOPN accuracy
16	IHN = HNO3	2.3e-5	Published IHN degradation rate
17	IEPOX + HO = HO	5E-11 exp(-400/T)	Match IEPOX degradation rates
18	ISOP + MO2 = HO2 + 1.31 HCHO + 0.159 MACR + 0.250 MVK + 0.250 MOH + 0.250 ROH + 0.023 ALD + 0.018 GLY+ 0.016 HKET	3.4E-14exp(221/T)	RACM2
19	ISOP + ACO3 = 0.5 HO2 + 0.5 MO2 + 1.048 HCHO + 0.219 MACR + 0.305 MVK + 0.5 ORA2	8.4E-14 exp(221/T)	RACM2
20	ISOP + APIP2 = 0.96 HOM + 0.48 ROH + 0.48 HCHO + 0.48 MVK + 0.48 HO + 0.48 HO2 + 0.04 ELHOM	1e-10	RACM2
21	ISOP + APINP2 = 0.96 HOM + 0.48 ROH + 0.48 HCHO + 0.48 MVK + 0.48 NO2 + 0.48 HO2 + 0.04 ELHOM	1e-10	RACM2
22	ISOP + LIMNP2 = 0.96 HOM + 0.48 ROH + 0.48 HCHO + 0.48 MVK + 0.48 NO2 + 0.48 HO2 + 0.04 ELHOM	1e-10	RACM2

9. Species naming convention

The species in the AMORE-Isoprene mechanism were named using common naming conventions. All species with direct analogues were named the same as in other mechanisms. This includes isoprene epoxy diols (IEPOX) and isoprene hydroxy nitrates (IHN). Other species were named in a similar manner to the Caltech reduced mechanism, with I serving as the signifier for the isoprene base, and letters H, N, C, P standing for hydroxyl, nitrates, carbonyl, and hydroperoxy groups respectively. The RACM2 naming convention is more abbreviated than that of Caltech reduced, and that convention was retained. Thus, names in the F0AM files do not all match those in the CMAQ formatted files. The table below lists each species, the representative chemical formula, and the names given in both CMAQ and F0AM format. Representative chemical structures are included in the supplementary files.

Table S.4

Species	Chemical Formula	CMAQ	F0AM	Smiles String
Isoprene	C_5H_8	ISO	ISOP	C=CC(=C)C
isoprene epoxy diols	$C_5H_{10}O_3$	IEPOX	IEPOX	CC(O)(CO)C1CO1
Isoprene hydroperoxy radical	$C_5H_9O_3$	ISOP	ISOPOO	[O]OC/C=C(\CO)/C
Isoprene hydroperoxide	$C_5H_{10}O_3$	ISHP	ISOPOOH	OOCC=C(C)CO
Isoprene nitrates (lumped)	$\sim C_5H_8N_2O_6$	ISON	ISOPN	-
Isoprene hydroxy nitrate	$C_5H_9NO_4$	IHN	IHN	C/C(=C/CON(=O)=O)CO
Isoprene hydroperoxy carbonyl	$C_5H_8O_3$	IPC	IPC	C/C(C=O)=C/COO
Isoprene nitrate peroxy radical	$C_5H_8NO_5$	INO2	INO2	C/C(=C\CO[O])CON(=O)=O
Isoprene hydroperoxy nitrate	$C_5H_9NO_5$	IPN	IPN	C/C(=C\COO)CON(=O)=O
Peroxy radicals formed from MACR + OH	MW = 101	MACP	MACP	
Formic Acid	MW = 46	ORA1	ORA1	

Terminal alkenes	MW = 42	OLT	OLT	
C3 and higher alcohols	M2 = 60	ROH	ROH	
methanol	MW = 47	MOH		
C3 and higher aldehydes	MW = 58	ALD		
Hydroxy ketone	MW = 74	HKET		
Acetic acid and higher acids	MW = 60	ORA2		
		HOM		
		ELHOM		
		APINP2		
Peroxy radicals from d-limonene and other cyclic diene-terpenes	MW = 185	LIMNP2		
Peroxy radicals from alpha-pinenes and other cyclic terpenes with one double bond	MW = 185	APIP2		

10. Multi mechanism box model error tables

Errors (as defined in section 2.4.3) were measured for AMORE-Isoprene, Caltech Reduced Plus, Carbon Bond 3, and CRACMM in the F0AM box model under the six conditions described in table 3. A table containing species by species errors for each mechanism under the six conditions is in the Supporting Information file folder. The file is named "Multi_mechanism_box_model_error_tables".

11. Error Metric Behavior

Plot shows behavior of single species error metric for different sample concentration profiles compared to a reference profile. The error metric ranges from 0 to 1.

Figure S.3

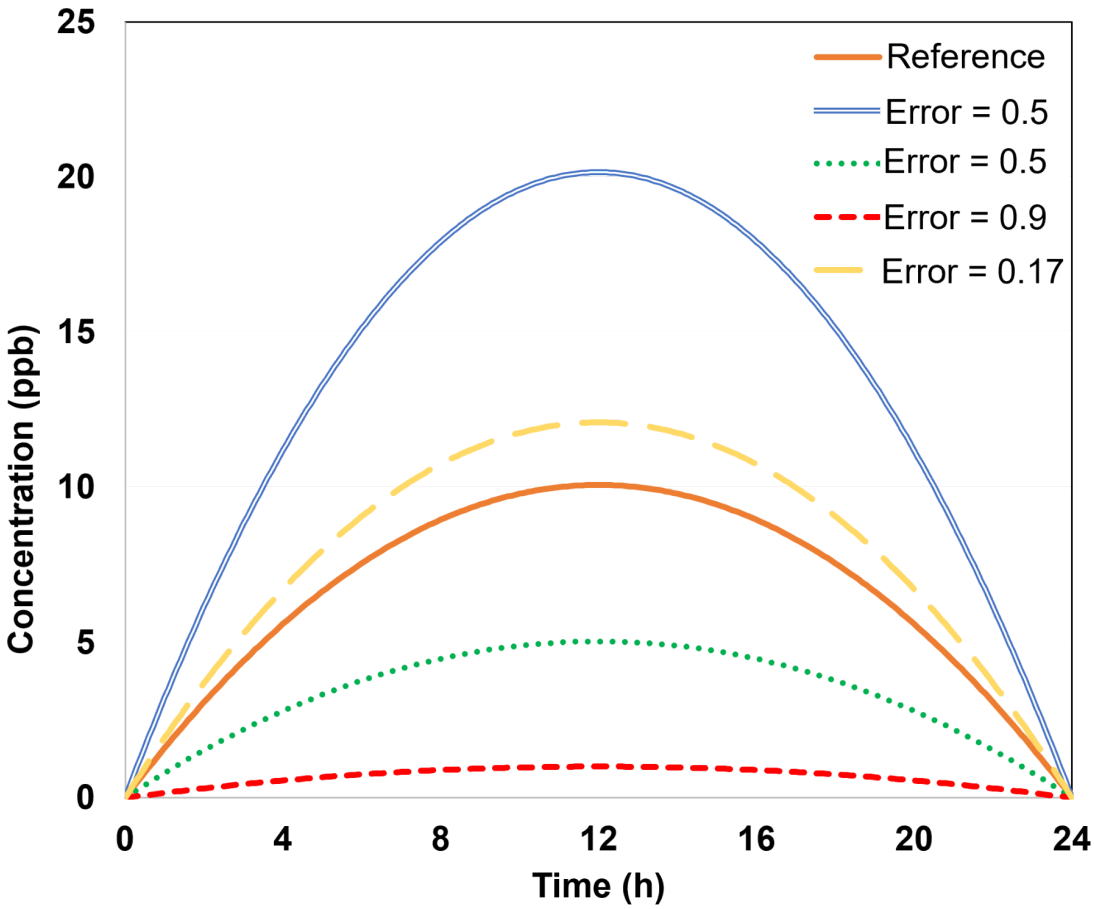
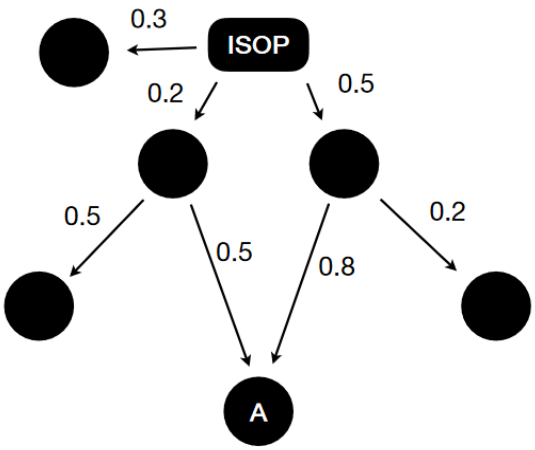


Figure S.4

Visual demonstration of yield estimation algorithm for a simple directed acyclic graph. The yield is shown for species A from isoprene.



$Y_A = 0.2*0.5 + 0.5*0.8 = 0.5$