Authors' Response to Reviews of

GENerator of reduced Organic Aerosol mechanism (GENOA v1.0): An automatic generation tool of semi-explicit mechanisms

Zhizhao Wang, Florian Couvidat, Karine Sartelet

Geoscientific Model Development, `https://doi.org/10.5194/egusphere-2022-245`

RC: ***Reviewers' Comment***, AR: Authors' Response, ☐ Manuscript Text

Reviewer # 1

1. General comments

RC: *This paper discusses procedures developed to reduce the size of large gas/aerosol mechanisms to greatly reduced mechanisms that give predictions of secondary organic aerosol (SOA) that agree with those of the larger mechanism in a selected set of environmental conditions to within a specified tolerance. Because of the complexity of the atmospheric reactions or organic compounds, it is necessary to use reduced mechanisms in practical airshed model applications, but most reduced mechanisms used in airshed model applications were developed primarily focused on accurate ozone predictions. However, SOA predictions are much more affected by the chemical complexity of organic reactions than predictions of ozone, and developing methods to reduce mechanisms for without significantly affecting SOA predictions is an important research priority in atmospheric chemical mechanism development.*

In fact, work with GECKO suggest that multi-generation mechanisms may be necessary for reliable SOA predictions, and even the MCM, which greatly lumps reactions of 2nd and higher generation products, may be too reduced for this application. However, the use of MCM as an example is sufficient to illustrate the method, and as discussed below MCM may be about the largest mechanism that could be reduced using the method discussed in this work, given current computer capabilities. In any case, without some way to reduce these huge mechanisms, we have no choice to continue to rely on the empirical and parameterized SOA models that are adjusted to fit SOA yields measured environmental chamber data without consideration of the actual chemistry and how different chemical conditions in the atmosphere affect SOA yields. If we had suitable reduction methods for SOA predictions, then use of SOA models that are based on actual chemistry might become practical.

AR: The authors would like to first thank Dr. William Carter for his detailed comments and insightful suggestions. Many of these suggestions are constructive and have been taken into account for the further development of the GENOA algorithm.

We have carefully considered the comments and revised the manuscript accordingly. Please note that the line and section numbers mentioned in the response correspond to the version of the manuscript before revision.

RC: *The method discussed here involves use of a 3D grid model representing a large continental domain and various seasons, with SOA calculated using the large mechanism to be reduced as the starting point, and then uses sets of 0D scenarios derived from selected grid cells and times during the grid model simulation in an algorithm to develop the reduced mechanism. This has the disadvantage that it requires a full 3D calculation* *with the full mechanism, which may be possible for MCM, but not for huge GECKO-like mechanisms that are probably what are really required. It may be possible to revise or supplement this method so that 0D calculations with selected scenarios may be sufficient to serve as the standard, but that is not discussed in this work. However, the method discussed here is a useful starting point, and is worth publishing for this reason.*

AR: In the paper, 3-D simulations were conducted using implicit gas-phase mechanisms rather than explicit gas-phase mechanisms. The 3-D simulation results are used to provide a realistic range of conditions (for ozone and radical concentrations and meteorological conditions) to train semi-explicit SOA mechanisms with GENOA. As the 3-D results have been evaluated using measurement data ([Couvidat et al., 2018]), implicit gas-phase mechanisms can be utilized to simulate the concentrations of oxidants and major air pollutants (e.g., ozone, radicals, other inorganic pollutants), as well as environmental parameters (e.g., relative humidity and temperature). Also, as the reviewer pointed out, 3-D simulations with the full mechanism are too time-consuming. Therefore, 3-D simulations with implicit gas-phase mechanisms are sufficient to obtain near-realistic conditions on which to train the semi-explicit mechanisms.

The paper has been updated to include details about implicit gas-phase mechanisms of the 3-D CHIMERE model and rephrased how 3-D results are used for 0-D simulations:

> *Section 2.2, line 193:*
>
> The version of CHIMERE and its configuration is described in Lanzafame et al. (2022). The ~~monthly average~~ 3-D CHIMERE simulations were conducted with the implicit gas-phase MELCHIOR2 mechanism (Derognat et al., 2003), which contains 120 reactions and less than 80 lumped species. The MELCHIOR2 mechanism describes the degradation of sesquiterpenes by three oxidant initiated reactions (HUMULE reacts with OH, O3, and NO3, respectively), where the the species HUMULE represents the lumped class of all sesquiterpenes.
>
> The monthly diurnal profiles of hourly meteorological data (e.g., temperature, relative humidity), and hourly concentrations of ~~oxidants and~~ oxidant, radical, and other inorganic species were extracted from each location. That information is required in the 0-D simulations with SSH-aerosol (see section 2.4) to reproduce SOA concentrations and compositions under near-realistic conditions. Since the reduced SOA mechanism focuses only on SOA formation, the meteorological data and the concentrations of oxidants, radicals and inorganics are assumed to remain intact during the 0-D SOA simulation. The coordinates and time of each condition are also provided to calculate the solar zenith angle. Because the reduction focuses on the impact on SOA variation, and because no inorganic reactions are considered in the reduced chemical mechanism, the oxidant, radical and inorganic concentrations~~are fixed as the hourly background~~, as well as the environmental parameters, are fixed to the diurnal profiles obtained from the CHIMERE data in 0-D SOA simulations. The concentration of HUMULE (denoted C_{SQT} as the CHIMERE surrogate for sesquiterpene~~(denoted C_{SQT}~~) is used to estimate the SQT concentration.

RC: *The focus of this paper is a specific software package developed for mechanism reduction, which they call GENOA, though the main interest of this paper from a scientific perspective is the method itself. The name of the software is somewhat misleading because it is not actually a mechanism generator, but instead is a method to reduce existing mechanisms, potentially including those developed by actual mechanism generators such as GECKO.*

AR: The role of GENOA is more like a "generator of reduced mechanisms" than as a "generator of mechanisms". In our opinion, the term "generator" might be better than other terms such as "producer" or "reducer". However, it is true that the term "generator" could still be misleading. Thus, GENOA is always referred to as a "generator of reduced mechanisms," rather than a "generator" alone, as GENOA is designed to generate mechanisms from reduction, not from scratch.

RC: *This could potentially be a useful tool for the atmospheric chemistry research community, if suitably documented and made publicly available. The fact that they include a users manual along with the Supplementary Information suggests that this is the intention of the authors. If it has sufficient flexibility, it could possibly be used for other criteria besides SOA predictions and perhaps even for other applications besides atmospheric modeling, such as, for example, combustion modeling or reducing large liquid-phase mechanisms. However, the users manual does not contain sufficient information to actually run the model for general applications, other than duplicating the results given in this paper. In particular, it would need information on how to interface this with output of existing 3D air quality models.*

AR: Thank you for the positive comments. Certainly, one of the goals of GENOA is to build the reduction algorithm that can be used for other modellers and even for other applications besides SOA prediction.

As explained previously, the monthly diurnal profiles of hourly meteorological data (e.g., temperature, relative humidity), and hourly concentrations of oxidant, radical, and other inorganic species were extracted from 3-D results. The user can, therefore, simply compute these profiles from 3-D results for GENOA reduction.

We have added the following explanation to the manual:

> *GENOA user's manual, section 4, page 15:*
>
> For example, to run GENOA over a specific domain and using specific 3D model results, the user has to construct the files described in section 2.2.2 to provide the monthly profiles for different variables.

RC: *It is worth pointing out that the reduced mechanisms developed using this are strictly speaking reliable to give predictions to within the desired tolerances only for the airshed conditions used in its development. They developed this example using the conditions of all of Europe for a whole year, but it may be more practical to use this to optimize mechanisms for specific urban scenarios, for use in regulatory modeling. To be useful for this, the users manual would need to be improved so it can be used with other 3D models and modeling scenarios.*
Although this paper is reasonably well written, it does have areas where improvements are needed before it is accepted for publication, and I have some suggestions. These are given below in approximate order of importance.

AR: In the example of BCARY reduction, the reduced "Rdc." mechanism was trained under training conditions that covered both high and low NOx regimes. The training condition "ADD2", for instance, received high NO concentrations from Milan, and thus is considered a typical urban condition. Consequently, the "Rdc." mechanism could be adequate for use over Europe or in specific urban scenarios, as it was developed under both low and high-NOx conditions and efficiently reproduces the SOA concentrations in most metropolises (e.g., Paris, Berlin).

Compared to the "Rdc." mechanism, undoubtedly, more reduction could be achieved if the SOA mechanism was trained specifically for urban scenarios using an urban-specific training dataset. In other words, as pointed out by the reviewer, training with fewer and more targeted conditions will optimize the resulting SOA mechanism (more accurate and smaller), which is intended for a small domain or specific scenarios.

2. Improvements need to be done

RC: *It looks relative changes in SOA concentrations are used as the criteria to test a reduction approach. This means that 5% error in a grid cell where almost no SOA is formed is given equal weights to 5% error in a grid cell with high SOA, where the model prediction is relatively more important. Wouldn't absolute error be a better criterion, or at least among the criteria employed? Shouldn't there be a cutoff to remove cells with very low SOA, or was this incorporated implicitly by the choice of testing scenarios?*

AR: Thank you for such an insightful suggestion. We used relative errors in the evaluation of the reduction, because they are sensitive to small changes and it is easy to set universal criteria for conditions with a wide range of SOA concentrations. Because sesquiterpene has high SOA yields, all 0-D simulations in the paper resulted in SOA concentrations consistently exceeding 1 μg/m^3 (with an initial BCARY concentration of 5 μg/m^3). Therefore, for BCARY reduction, there may be no need for a cut-off concentration.

We have added the information to the paper:

> *Section 2.3, line 267:*
>
> The initial BCARY concentration is ~~taken equal to 5~~ set to five μg m^{-3} in order to ensure high SOA production the (SOA concentration is always greater than one μg m^{-3} at all evaluated conditions at all conditions).

However, for other precursors that have SOA yields lower than sesquiterpene, a cut-off concentration may be required for conditions with low SOA production, where large relative errors can occur as a result of low SOA concentrations.

The user manual has been updated to inform the user that there are several options that can be taken to avoid this issue:

> *GENOA user's manual, section 4, page 15:*
>
> It should be noted that reduced mechanisms may cause large errors when there is a low SOA concentration. The reason might not be directly linked to reduction of the mechanism performance, but to the evaluation criterion, based on relative errors, which naturally has large variations when the absolute value is small. There are several solutions to resolve this uncertainty:
>
> - The user can set high initial concentrations of the studied SOA precursor in order to ensure high SOA production.
> - The user may specify a threshold SOA concentration for condition selection in the training and pre-testing datasets.
> - The user may also evaluate reductions using a different type of error.

RC: *The 3D model simulations of continental Europe employing the chemistry-transport model CHIMERE was used as the standard against which the reductions were compared. However, CHIMERE mechanism is not exactly the same as MCM, and I could not find an indication of whether the "CHIMERE surrogate for sesquiterpene" (line 198) is exactly the same as the MCM b-caryophyllene mechanism as used in this work. If that is the case, it should be stated explicitly when CHIMERE is first mentioned, since I couldn't find such a* *statement in Lanzafame et al (2022). If it is not the case, then justification needs to be given as to whether this is an appropriate standard against which to test the reductions.*

AR: As detailed in reply to one of the general comments, we have not used 3-D but 0-D SOA simulations as a reference to evaluate the reductions. In the 0-D simulation, the concentration profiles of ozone, radicals, and inorganics, as well as environmental parameters, were derived from the 3-D CHIMERE simulations.

In CHIMERE v2020r1, there are several gas-phase chemical mechanisms embedded. We have used the MELCHIOR2 mechanism [Derognat et al., 2003], which contains 120 reactions and less than 80 lumped species. Evidently, it is an implicit gas-phase mechanism that is highly simplified compared to the explicit MCM mechanism.

The species HUMULE in CHIMERE (the MELCHIOR2 mechanism) does not refer to a specific sesquiterpene, but to a lumped class that includes all sesquiterpene species. As we use "HUMULE" concentrations only to determine whether sesquiterpenes are present at a specific location and select several conditions for varying sesquiterpene concentrations in the pre-testing dataset, there is no consistency issue.

Relevant modifications have been done to *section 2.2 line 19* of the paper, as mentioned previously.

RC: *It is not clear to me whether the concentrations of the inorganic species such as OH, O3, HO2, etc were constrained to be exactly what was calculated by CHIMERE in the 0D models for the selected locations, or if they were calculated using the mechanisms using boundary or initial conditions somehow obtained from CHIMERE. If the former (which I presume to be the case) this should be stated explicitly, and if the latter than more detail about the inputs to the 0D models need to be provided, if only in an Appendix.*

AR: Sorry for the ambiguity. As mentioned in reply to the previous comment, the concentrations of certain species (i.e., ozone, radicals, and other inorganics) are constrained to be exactly what was calculated in the 3-D CHIMERE simulations for the selected conditions (locations + month). The concentrations of other species resulting from sesquiterpene degradation are simulated using the 0-D SSH-aerosol model. As the CHIMERE concentrations represent 24-hour monthly averages, the diurnal profiles are repeated for five days in 0-D simulations (the simulation time is five days to consider adequately SOA formation and aging processes).

We have added the explanation of how the coupling is generally resolved between VOC species and oxidants in modeling:

> *Section 1, line 36:*
>
> ~~Along with~~ To complete implicit gas-phase mechanisms, implicit SOA mechanisms have been developed to model specifically the SOA formation without modifying the concentrations of ozone and major radicals (Kim et al., 2011) . In 3-D modeling, implicit SOA mechanisms or parameterizations are usually added to implicit gas-phase mechanisms, conserving the oxidant chemistry of the implicit gas-phase mechanism.
>
> Implicit SOA mechanisms are often established based on experimental data from smog chamber experiments to represent the formation and evolution of SOA, such as the two-product empirical SOA model (Odum et al., 1996) and the volatility basis set (VBS) that splits VOC oxidation products into a uniform set of volatility "bins" Donahue et al., 2006).

And the focus of our semi-explicit SOA mechanisms:

> *Section 2, line 80:*
>
> The generated semi-explicit mechanisms are designed to preserve the accuracy of explicit mechanisms for SOA formation, while keeping the number of reactions/species low enough to be suitable for large-scale modeling, particularly ~~3D~~ in 3-D AQMs. The focus of the semi-explicit mechanism is solely on the accurate modelling of SOA. Because ozone, major radicals, and other inorganics are also affected by inorganic and other VOC chemistry, their concentrations are not tracked with the semi-explicit mechanism. Instead, they are simulated using existing implicit gas-phase chemical mechanisms.

RC: *The computation of stoichiometric coefficients and rate constants for new lumped species (as shown in Figure 2) based on concentrations and lifetimes need to be discussed. The use of "average produced concentrations from five-day 0D simulations" needs more discussion than as a footnote in the table. Presumably the [HO2], [NO], etc. concentrations used to compute the lifetimes are also averaged? What are the 5 0D simulations they use to compute the average? The later discussion indicates they use 8 scenarios for initial training, and more than that when they are close to being finished.*

AR: Sorry for the ambiguity. As mentioned in the previous reply, all 0-D simulations are conducted for five days in order to consider adequately SOA formation and aging processes. In the reduction by lumping, the weighting ratio is necessary to weigh the proportion of lumped molecules in the new molecule, as well as to compute the stoichiometric coefficients and rate constants in the new reaction. In order to calculate the weighting ratio, we tried different methods and eventually decided to compute it based on average concentration over the lifespan of the particle under training conditions. A five-day period is chosen as a compromise between a shorter period that may not reflect aging, and a longer period that may less adequately address the SOA formation. Considering the limited computational capacity, only one weighting ratio calculation method is utilized in the BCARY reduction.

To ensure coherency, all reduction parameters are calculated based on five-day 0-D simulation results of training conditions. For example, in the BCARY reduction, the lifetime of inorganics (e.g., HO2, NO) is calculated as the average of 16 simulations (two five-day simulations starting at 0h/12h for eight training conditions). Considering that inorganic concentrations repeat the diurnal profiles of 3-D simulations, the five-day average is equivalent to the daily average. This is different from the concentrations of sesquiterpene oxidation products, where a five-day average could take into account both formation and aging processes.

For clarity, the following explanation has been added to the paper:

> *Section 2.1.1, line 126:*
>
> As detailed in Table 3, f_w is computed as a function of chemical lifetime τ following the computation of Seinfeld and Pandis (2016), and the reference concentrations C_r that are the arithmetic mean concentrations ~~of a set from five-day 0D simulations of~~ calculated from 0-D simulations using the explicit VOC mechanism~~under~~. Both τ and C_r are based on averages of simulations across all training conditions.
>
> *Section 2.3, line 268:*
>
> Unless stated otherwise, ~~a 5-day simulation is performed~~ two simulations are performed for each condition starting at midnight (0 h) and noon (12 h)~~for each condition, focusing on~~, taking into account both the daytime and nighttime chemistry. All 0-D simulations are run for five days in order

to consider adequately SOA formation and aging processes.

RC: *The maps indicate that the reduction errors are the greatest in Southern Europe than in the North. Is that because there is more secondary SOA predicted for the South, which should be more photochemically reactive? It might be useful to show a relationship between SOA level predicted and the reduction error for the various grid cells or scenarios.*

AR: In order to illustrate the relationship between testing errors and SOA levels, we have added the SOA concentration map (Fig. 1) to the main paper, which corresponds to the results in the error map (Figure 3 of the paper):

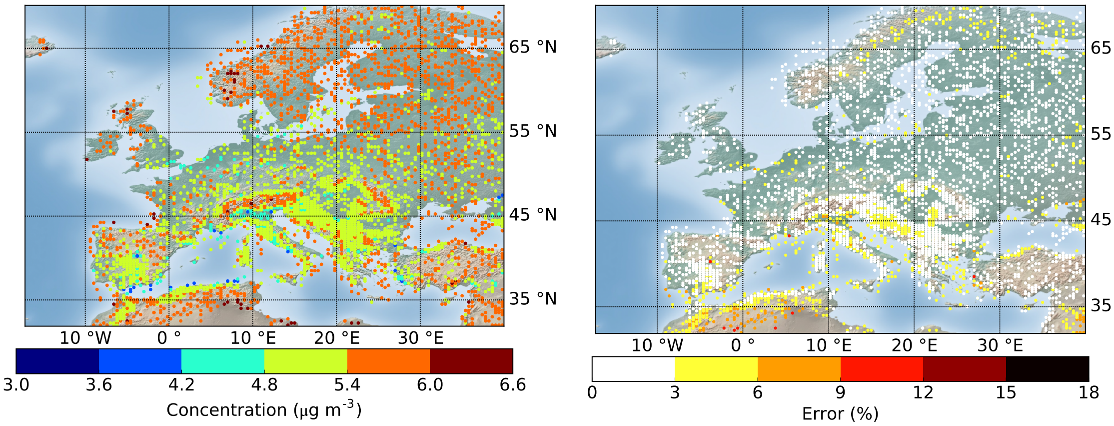


Figure 1: SOA concentration (left panel) and testing error (right panel) maps of the testing results of the "Rdc." mechanism in summer (July and August).

We can see that SOA concentrations are generally higher under Nordic and North African conditions (The same initial BCARY concentration is set ($5\ \mu$ g/m^3 for all conditions). As only one photolysis reaction is preserved in the "Rdc." mechanism (248 are preserved in the original MCM), the "Rdc." mechanism is probably not be very sensitive to photochemical reactivity. As can be seen from the bar plot in Fig. 5 in the paper, the testing errors in July and August over Southern Europe are actually close to 3 %, although they seem to be higher (placed in the error category between 3 % and 6 %) than others. In northern Africa, the high error might be partially attributed to the photochemical reactivity, but the main cause is still the high temperatures and dry climate.

It should also be noted that although some conditions on the map appear to have large errors (e.g., in northern Africa), the results may still be acceptable as long as they adhere to the pre-testing criteria (average error $\leq$ 3 %, and maximum error $\leq$ 20 %). As the average SOA concentration map of all testing results may be too complex to interpret, it is not included in the main paper but in Appendix C for reference (Fig. 2):

RC: *On line 169 they state that "jumping" is restricted to cases where the intermediate to be removed results in the formation of only a single compound. However, this is not the case for the examples they show on Table 4, where each intermediate forms more than one compound. (Reactions 14 and 15 in Table 5 reflect merged competing reactions, and are not explicit.) In fact, any rapidly reacting intermediates that have several reactions that form different products, can be "jumped" without affecting simulation results as long as their reactions are either unimolecular or with O2, so lifetimes would not vary with conditions. This is*

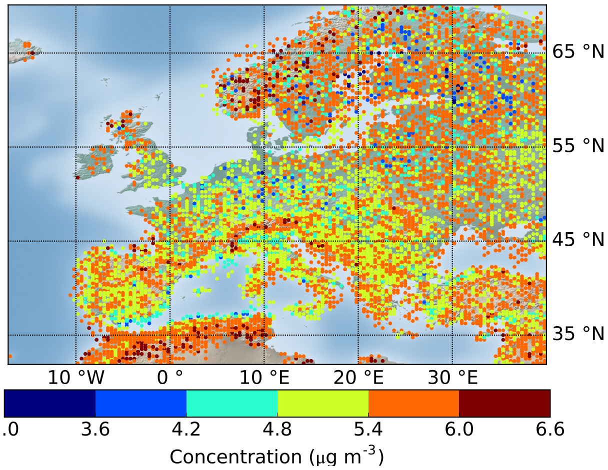


Figure 2: The average SOA concentration map of the testing results of the "Rdc." mechanism on all 12 159 testing conditions.

the case almost all of the alkoxy radicals in explicit atmospheric mechanisms. This should be pointed out in the discussion. The main approximation with "jumping" compounds with several reactions would that it would not deal with temperature effects if some of the competing reactions have different temperature dependences in the full model. Whether this is acceptable for their continental model could be determined by their algorithm.

AR: The jumping example shown in Table 4 of the paper (line 169) is indeed a mistake, as we mixed up the concept of jumping with a reduction process "pre-reduction" that are applied to the original explicit mechanism. We have added one section describing the pre-reduction (with an example of Table 1) to the main paper:

> *Section 2, line 103, before Section "Reduction strategies":*
>
> A pre-reduction process is conducted on the original MCM mechanism before it is used as the reference mechanism for the reduction. This process skips very fast reactions (i.e. reaction rate of 1.0×10^6 corresponding to a lifetime of 1 μs) to avoid numerical problems. For computational efficiency, it also combines elementary reactions with the same reactants into combined reactions with non-integer stoichiometric coefficients.
>
> An example is shown in Table 1, where the original MCM reactions No. 1 to 7 have first been merged into the combined reactions No. 8 to 10. The pre-reduction compacts the reaction list (from 1 626 to 1 242 reactions), improving the reduction efficiency. The pre-reduction also skips two biradicals (i.e., BCALOOA and CH2OOF), that are extremely reactive and disintegrate instantaneously with a kinetic rate coefficient of 10^6 s^{-1}. As a result, the reactions No. 8 to 10 can then be repented by one reaction No. 11, whose kinetic rate coefficient corresponds to that of the reaction producing the skipped species (in this case, the ozonolysis of BCAL, reaction No. 9).

We have also rephrased the jumping section with an new example (Table 2):

Table 1: Reactions before and after pre-reduction , where MCM species BCALOOA and CH2OOF are skipped over by their degradation products.

No.	Reactions	Kinetic rate coefficient
1	BCAL + O3 → BCALOOA + HCHO	$1.1 \times 10^{-16} \times 0.670$
2	BCAL + O3 → BCLKET + CH2OOF	$1.1 \times 10^{-16} \times 0.330$
3	BCALOOA → BCALOO	$1.0 \times 10^{6} \times 0.5$
4	BCALOOA → C146O2 + OH	$1.0 \times 10^{6} \times 0.5$
5	CH2OOF → CH2OO	$1.0 \times 10^{6} \times 0.370$
6	CH2OOF → CO	$1.0 \times 10^{6} \times 0.500$
7	CH2OOF → HO2 + CO + OH	$1.0 \times 10^{6} \times 0.130$
8	BCAL + O3 → BCALOOA + CH2OOF + BCLKET + HCHO	1.1×10^{-16}
9	BCALOOA → 0.5 BCALOO + 0.5 C146O2 + OH	1.0×10^{6}
10	CH2OOF → 0.37 CH2OO + 0.13 HO$_2$ + 0.63 CO + 0.13 OH	1.0×10^{6}
11	BCAL + O3 → 0.5 BCALOO + 0.5 C146O2 + 0.37 CH2OO + BCLKET + HCHO + 0.13 HO$_2$ + 0.63 CO + 1.13 OH	1.1×10^{-16}

Section 2.1.3, line 164:

~~As shown in Table 4, the ozonolysis of BCAL results in the formation of the excited Criegee biradical compounds BCALOOA and CH2OOF. These two biradicals are extremely reactive and disintegrate instantaneously with a kinetic rate coefficient of 10^6 s^{-1}. As a result of the jumping strategy, three MCM reactions (reactions Nos. 13, 14, and 15 in Table 4) are reduced to one reaction R5, whose kinetic rate coefficient corresponds to that of the reaction producing the skipped species (in this case, the ozonolysis of BCAL, reaction No.13).~~

As shown in Table 4, the alkoxy radicals BCALOO formed from ozonolysis of BCAL (reaction No. 11 in Table 1) is jumped over to its only destruction product BCLKET. Consequently, the reactions No. 12 to 16 are removed, and reaction No. 11 is updated to reaction No. 17. Currently, the jumping strategy is considered when the destruction of a single compound (to be jumped) results in the production of a single compound (jumping). The difference in carbon numbers between reduced species can not exceed three in order to prevent significant differences in organic mass before and after jumping.

Table 2: Reactions before and after the jumping strategy, where MCM species BCALOO is jumped over by its degradation product BCLKET.

No.	Reactions	Kinetic rate coefficient
12	BCALOO + CO → BCLKET	1.2×10^{-15}
13	BCALOO + NO → BCLKET + NO2	1.0×10^{-14}
14	BCALOO + NO2 → BCLKET + NO3	1.0×10^{-15}
15	BCALOO + SO2 → BCLKET + SO3	7.0×10^{-14}
16	BCALOO → BCLKET + H2O2	$1.4 \times 10^{-17} \times [H_2O]$
17	BCAL + O3 → 1.5 BCLKET + 0.5 C146O2 + 0.37 CH2OO + HCHO + 0.13 HO_2 + 0.63 CO + 1.13 OH	1.1×10^{-16}

Currently, only jumping to a single compound was done, as doing jumping to multiple compounds may not affect significantly the reduction of BCARY-SOA mechanism . Furthermore, compounds involved in several reactions with different temperature effects cannot be jumped over if they lead to multiple compounds without neglecting the temperature effect. After removing some reactions or some species, some compounds can become a candidate for "single" jumping. At the end of the reduction, only two compounds remain that could be candidate for jumping to multiple compounds. However, these two compounds (C133O and BCBOO) compounds are involved in several reactions with different temperature effects and therefore cannot be jumped over.

RC: *The example of lumping they present in Table 2 gives a rather trivial reduction example that should not really require an algorithm. The chemical lifetimes of the 3 peroxy radical species being lumped are the same for the major loss processes, with the only difference being the RO2+RO2 reactions that make almost no contribution (see Figure 2), so the "weighting ratios" should actually be very close the stoichiometric coefficients in the explicit mechanism (reaction "0" in Table 1), and the simulation results should also be the very close. An additional example where the lumping may not be as obvious, and which may potentially be more dependent* *on environmental conditions, should also be given. This would aid the discussion of the lumping approach and give the reader a better understanding of the lumping approach and its general utility and potential drawbacks. Or is the kind of peroxy radical lumping illustrated in Table 2 the only type of lumping they actually did? If so, this should be stated.*

AR: Thank you for the suggestion. Typically, lumping occurs between compounds sharing similar chemical pathways or molecular structures. While the example of lumping in Table 2 is trivial, it serves to illustrate how lumping merges reactions and build new surrogates.

Taking into account the reviewer's comments, two additional examples from the BCARY reduction have been added to appendix:

In Appendix C:

Besides the example shown in Sect. 2.2.3, two additional examples have been added from the BCARY reduction: one illustrates the lumping of two similar compounds formed by different reactions, and the other illustrates the lumping of two more distinct compounds. The first example is the MCM species C1313NO3 and C152NO3 (Fig. 3). These two species come from different reactions. The molecular structures of both compounds are similar (they contain organic nitrates, aldehydes, and alcohols), but C152NO3 contains an additional carboxylic acid where C1313NO3 contains an aldehyde. The corresponding reactions before and after lumping are summarized in Table 3, where the new surrogate "mC1313NO3" is built from C1313NO3 with a weighting ratio of 83% and C152NO3 with a weighting ratio of 17%. As a result of this lumping, the average error increase under training conditions is 0.001 % (the tolerance is 0.01 %).

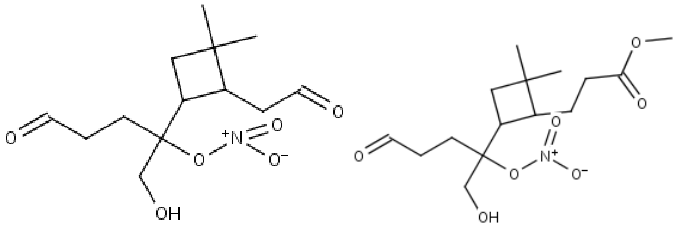


Figure 3: Molecular structures of the MCM species C1313NO3 (left) and C152NO3 (right).

Another example of lumping is the MCM species BCALBOC and C1310OH (Fig. 4). Unlike the previous example, these two species are more distinct. According to MCM, BCALBOC are generated through O_3-initiated reactions, while C1310OH are generated through high-generation oxidations. There is less similarity in the structures or chemical reactions of the two molecules. MCM contains the OH reaction of BCALBOC, and the O_3 and OH reactions of C1310OH. However, this reduction was accepted since lumping them only increased the average error by 0.01 % under training conditions (the tolerance was 1 %). The new surrogate "mBCALBOC" is constructed from BCALBOC with a weighting ratio of 98 % and C1310OH with a weighting ratio of 2 %.

As C1310OH has a low weighting ratio, the lumping would be substituted by replacing (a special case of lumping), where the weighting ratio of BCALBOC is set to 100 % and of C1310OH is set to 0 %. In that case, instead of forming a new surrogate, C1310OH is replaced by BCALBOC. In BCARY reduction, this type of replacing was not used, but it can be activated by the user by setting the weighting ratio threshold.

RC: *I don't understand the difference between "removing species" and "removing elementary-like reactions". Strictly speaking, all species that are removed should be done by renormalizing yields from competing*

Table 3: Reactions related to the reduction of MCM species C1313NO3 and C152NO3 via lumping.[a]

Reactions before lumping	Kinetic coefficient
Production	
C1313O2 + NO → C1313NO3	KRO2NO×0.134
C152O2 + NO → C152NO3	KRO2NO×0.136
Destruction	
C1313NO3 + OH → C116CHO + HCHO + NO2	5.59×10^{-11}
C152NO3 + OH → BCLKBOC + HCHO + NO2	1.58×10^{-11}
Reactions after lumping	
Production	
C1313O2 + NO → 0.134 mC1313NO3 + 0.866 C1313O + 0.866 NO2	KRO2NO
C152O2 + NO → 0.136 mC1313NO3 + 0.864 C152O + 0.864 NO2	KRO2NO
Destruction	
mC1313NO3 + OH → C116CHO + HCHO + NO2	$5.59\times10^{-11} \times 0.82945$
mC1313NO3 + OH → BCLKBOC + HCHO + NO2	$5.59\times10^{-11} \times 0.17055$

[a] 0.82945 is the exact weighting ratio of C1313NO3 and 0.17055 is the exact weighting ratio of C152NO3.

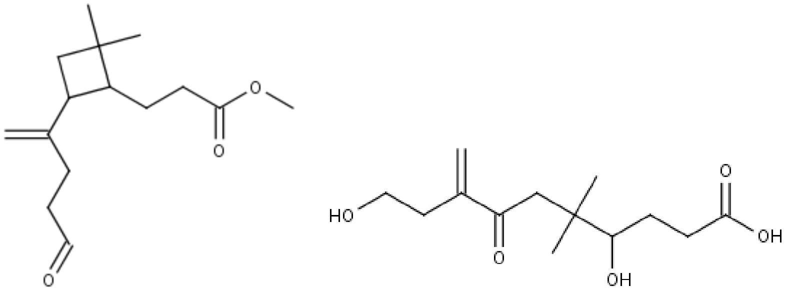


Figure 4: Molecular structures of the MCM species BCALBOC (left) and C1310OH (right).

Table 4: Reactions related to the reduction of MCM species BCALBOC and C1310OH via lumping.[a]

Reactions before lumping	Kinetic coefficient
Production	
BCOOA → BCALBOC	$1.0\times 10^{6}\times 0.15$
C1310O2 → C1310OH	$2.5\times 10^{-13}\times$[RO2]$\times 0.2$
Destruction	
BCALBOC + O3→ BCBOOA + HCHO	$1.1\times 10^{-16}\times 0.670$
BCALBOC + O3→ BCLKBOC + CH2OOF	$1.1\times 10^{-16}\times 0.330$
BCALBOC + OH→ C152O2	6.98×10^{-11}
C1310OH + OH → C1310CO + HO2	6.2×10^{-11}
Reactions after lumping	
Production	
BCOOA → mBCALBOC	$1.0\times 10^{6}\times 0.15$
C1310O2 → mBCALBOC	$2.5\times 10^{-13}\times$[RO2]$\times 0.2$
Destruction	
mBCALBOC + O3 → BCBOOA + HCHO	$1.10\times 10^{-16}\times 0.670\times 0.97675$
mBCALBOC + O3 → BCLKBOC + CH2OOF	$1.10\times 10^{-16}\times 0.330\times 0.97675$
mBCALBOC + OH → C1310CO + HO2	$6.20\times 10^{-11}\times 0.97675$
mBCALBOC + OH → C152O2	$6.98\times 10^{-11}\times 0.023251$

[a] 0.97675 is the exact weighting ratio of BCALBOC and 0.023251 is the exact weighting ratio of C1310OH.

pathways if the reactions are lumped, which is the same as converting them to "elementary-like reactions", and then removing the less important reactions. Without renormalization of the other branching ratios, the removal would cause an unnecessary loss of mass or, worse, an artificial radical sink. An artificial radical sink may not affect SOA predictions if the concentrations of the inorganics are constrained in the test simulations (and thus would pass the tests in the reduction algorithm discussed here if that were the case), but the reduced mechanism would be unsuitable for predicting radical levels if used in a full model. Even small artificial radical sinks may have non-negligible effects on the radical chain reactions involved in many photooxidation mechanisms.

AR: It is true that the reduction with the current removing strategies may have an impact on the formation of inorganic radicals. However, as indicated in several replies previously, we develop GENOA primarily for SOA concentrations and not for the concentrations of inorganic radicals. Since inorganic radical formation and destruction are still simulated with implicit gas-phase chemical mechanisms, the removing strategies used in BCARY reduction are acceptable, as long as SOA concentrations are reproduced.

RC: *Section 2.2.3 describes the 12,159 conditions used for the testing datasets. It should be stated in that section that Figure D1 has a map of these locations. What are the 4,717 conditions mapped on Figure 6?*

AR: These 4,717 conditions (out of the total 12 159 testing conditions) are the testing conditions during the summer period from July to August. As the map of all testing conditions is difficult to read, we present only the testing results of the two months in the main paper. The testing results of all conditions can be found in the Appendix B for reference.

The locations of testing conditions have also been added in the main paper:

> *Section 2.2.3, line 253:*
>
> Fig. B1 indicates the locations the testing dataset as well as the testing results for BCARY reduction.

RC: *Are any of the lumping criteria listed around Lines 134-143 options that can be varied by the user? Define "extremely short lifetimes". Do they allow lumping compounds that react with different atmospheric species or types of reaction (e.g., with OH, O3, NO, etc, by photolysis, or unimolecular), but happen to have comparable lifetimes for most scenarios?*

AR: Yes, the lumping criteria listed around lines 134-143 can be modified by the user. We have summarized the user-chosen criteria into Table 5, which will be available in an Excel file in the supplementary material.

For intermediate compounds, such as biradicals (e.g, MCM species BCAOO), which degrade rapidly to closed shell molecules, jumping is considered to be more appropriate than lumping. In BCARY reduction, we checked the species type and ruled biradical them out of lumping as they have a low lifetime (approximately a few microseconds). In the paper, the term "extremely short lifetimes" is used to indicate that intermediate compounds are better suited for jumping since they have a short lifetime.

Table 5: The user-chosen reduction options and parameters applied to BCARY reduction.

Category	Meaning (and notion if mentioned in the paper)	Threshold
reduction evaluation	ϵ_{ref}, the maximum error compared to the reference mechanism under training conditions	$\leq$ 1% to 10%
reduction evaluation	ϵ_{pre}, the maximum error compared to the previous validated mechanism under training conditions	$\leq$ 1% to 10%
reduction evaluation	$\epsilon^{ave}_{pre-testing}$, the maximum error compared to the reference mechanism under pre-testing conditions	$\leq$ 3%
reduction evaluation	$\epsilon^{max}_{pre-testing}$, the maximum error compared to the reference mechanism under pre-testing conditions	$\leq$ 25%
reduction evaluation	reduction error calculation method	see paper
reduction strategy	searching order in species/reaction lists	high-to-low generation
reduction strategy	type & order of applied reduction strategies	see paper
removing reactions	B_{rm}, Maximum hourly branching ratio of the reaction	$\leq$ 5% to 100%
jumping	carbon number difference	≤ 3
jumping	minimum P_{sat} of species considered to be jumped	$\geq 1{,}0 \times 10^{-6}$ atm
lumping	order of species list: find lumping for condensables	order by P_{sat}
lumping	weighting ratio f_w calculation method	see paper
lumping	functional group check I: restrict to species with same groups	7 Types, see paper
lumping	functional group check II: do not participate lumping	ROO species
lumping	mass difference	≤ 100 g mol^{-1}
lumping	lifetime difference	$\leq$ 10-fold
lumping	carbon number difference	≤ 2
lumping	weighting ratio threshold by which a species is considered to be replaced by another	≤ 0 %
replacing	Minimum mass of species participating in replacing	> 100
replacing	carbon number difference	≤ 3
Aerosol-oriented treatment	number of condensables to activate aerosol-oriented treatments	≤ 20
0-D simulation	simulation starting times	0 h & 12 h

The term "extremely short lifetimes" has been rephrased to avoid confusion:

> *Section 2.1.1, line 141:*
>
> - ~~No lumping for intermediate compounds with extremely short lifetimes (e.g.,~~ Lumping is not considered for biradicals (ROO) ~~), as they are candidate for jumping instead~~that degrade rapidly into closed shell molecules, as jumping is considered to be more appropriate for these compounds.

RC: *It probably would be appropriate to point out that the MCM mechanism they use as the starting point has already been reduced to some extent. It already "jumps" carboncentered radicals by replacing them by the peroxy radical formed when they react. It also combines elementary reactions with the same reactants into combined reactions with noninteger stoichiometric coefficients. Removing species from such reactions should not be done without renormalizing coefficients representing the competing processes, as discussed above.*

AR: As previously explained, the rationale behind the "removing species" strategy is justified by the fact that we do not intend to track the destruction or production of inorganic radicals. The section entitled "Pre-reduction" has been added to the main paper (in *section 2, line 103*, as detailed previously), which describes the treatments on the original MCM mechanism before training.

RC: *Figure 2 and Table A1 use "RRNO3" to indicate the fraction of peroxy radicals reacting with NO3. The term "RNO3" should be used instead, to be consistent with the terminology for the other peroxy loss processes. I found this confusing when I first saw this.*

AR: The term "R_{RNO3}" was adopted, because the other term "R_{NO3}" had already been used for the reacting ratio of NO_3 with SOA precursor. To avoid confusion, we have changed "R_{RNO3}" to "$R_{NO3-RO2}$", along with other reacting ratios showing the reactivity with peroxy radicals (RO_2):

> *Section 2.2.1, line 222:*
>
> - The reacting ratios of RO_2 species with NO (R~~$_{NO}$~~$_{RO_2-NO}$), HO_2 (R~~$_{HO_2}$~~$_{RO_2-HO_2}$), NO_3 (R~~$_{RNO_3}$~~$_{RO_2-NO_3}$), and other RO_2 species (R~~$_{RO_2}$~~$_{RO_2-RO_2}$), whose sum equals 1, indicate the relative reactivity of successive reactions with RO_2 species.

3. Suggestions

RC: *It would be better if the four items in the reduction strategy were discussed in the order they are actually applied in the algorithm. "Lumping" and "replacing" is discussed first, but actually "removing" and "jumping" is done first, which is appropriate.*

AR: Thank you for the helpful suggestion. Reduction strategies have been reordered in section 2.1 as follows: "removing", "jumping", "lumping" and "replacing".

RC: *It would be useful to include a table of options or parameters that control the reduction process, indicating which can be changed by the user, and the value(s) of the parameters used in the examples in this work. This would be particularly relevant if the intention is to make this software available for use by other researchers.*

AR: Thank you for the suggestion. A table of reduction options and parameters applied to BCARY reduction has been added to an Excel file (Table 5), as part of the supplementary material.

RC: *The geographical and meteorological conditions for the 8 0D scenarios used for the training dataset are summarized on Table 5, and a map of their locations is shown on Figure C1 in an Appendix. The map showing the locations is useful and should be in the main paper, and if the points are shown along with the "errors" on Figure 6, it would take no added journal space and also indicate whether there is an relationship between the errors and the locations of the training scenarios. It also would be helpful for Table 5 to have columns showing the amounts of SOA calculated, the major daytime loss process for b-caryophyllene, and at least the day and night fractions that peroxy radicals react with NO. These are the most important chemical differences in the scenarios, and including this on the table would give this information in a more compact and perhaps clearer manner than on Figure 2. Figure 2 is also useful, but it is difficult to read and could be formatted better.*

AR: Thank you for the helpful suggestions. We have put together the locations of the training and pre-testing datasets in one figure (Fig. 5), and have attached it to the main paper (The figure number corresponds to the new numbering in the revised paper.):

> *Section 2.2.2, line 243:*
>
> The locations of the training and pre-testing conditions are presented in the Fig. 3.

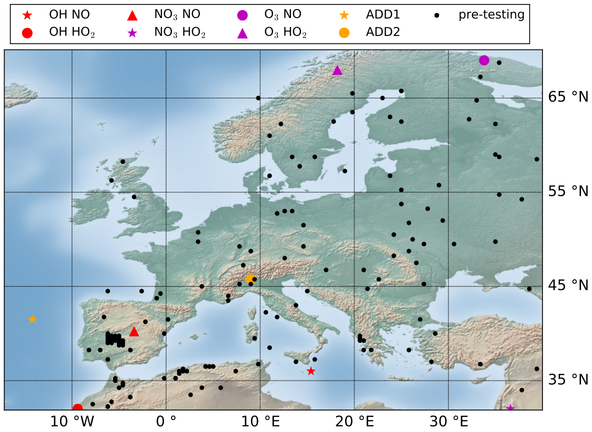


Figure 5: Locations of the training (see legends) and pre-testing (white scattered dots) datasets used in the BCARY reduction.

We have also added two columns to the training dataset information table (Table 5 of the paper), including average simulated SOA concentrations and average NO reactivity ratios with proxy radicals (R$_{RO_2-NO}$, indicating the chemical regimes (high/low NOx)). Considering that figure 2 of the paper may be sufficient to illustrate the major daytime loss process for BCARY (the reactivity of BCARY with O3, OH, or NO3 at 12h shown in the left panel), we decided to not repeat this information in the table. On the right half panel of Figure 2, the R$_{RO_2-NO}$ ratio at 0h and 12h indicates the fractions of NO that react with peroxy radicals at day and at night, respectively. Since they are consistent with the day and night fractions that peroxy radicals react with NO, the additional information is not added to the table (daily average NO ratio would be enough). For clarity, the digits in Table 5 of the paper have also been changed for temperature and relative humidity.

Meanwhile, Table 5 of the paper has been revised as follows (Table 6):

Table 6: Geographic and meteorological conditions of the training dataset

Condition Name[a]	Lat	Lon	Time	TEMP	RH	R_{NO}[b]	SOA[c]
	°N	°E	month	K	%	%	$\mu g/m^3$
OH NO	36.0	15.4	Jul.	~~299.4~~299	~~78.6~~79	60	4.1
OH HO_2	32.0	-9.4	Jul.	~~295.9~~296	~~76.7~~77	20	6.1
NO_3 NO	40.25	-3.4	Jul.	~~302.4~~302	~~27.9~~28	69	4.4
NO_3 HO_2	32.0	36.6	Aug.	~~302.2~~302	~~38.7~~38	29	5.7
O_3 NO	69.0	33.8	Jan.	~~260.7~~261	~~84.2~~84	99	5.2
O_3 HO_2	68.0	18.2	Dec.	~~265.5~~266	~~88.7~~89	25	4.6
ADD1	41.5	-14.2	Dec.	~~288.6~~289	~~75.8~~76	20	5.5
ADD2	45.75	9.0	Dec.	~~279.1~~279	~~84.5~~85	100.	4.4

[a] from left to right: name, latitude, longitude, time period, average temperature, average relatively humidity, daily average NO reacting ratio, simulated total SOA concentration of the training conditions.
[b] the daily average NO reacting ratio is calculated out of the RO_2 reactivity of NO, HO_2, NO_3, and RO_2. Conditions with high R_{NO} ratio are considered as in high NO_x regime. [c] the initial concentration of BCARY is 5 $\mu g/m^3$.

RC: *It would also be useful to include a map showing locations where the 150 testing scenarios were derived. This may cause too much clutter if added to Figure 6, but should at least be in the Appendix. There should also be a map showing the SOA levels calculated by the CHIMERE model for the European domain. Perhaps it could be combined with the figure in the Appendix showing the locations of the testing scenarios, though it may be useful enough to be in the main text.*

AR: Thank you for the suggestions. As explained in one previous reply, we have combined the locations of the pre-testing dataset and the training dataset in one figure to the main paper. Meanwhile, the maps showing the average SOA concentrations of the testing conditions in summer (July and August) and all times (Fig. 1 and Fig. 2) have been added to the main paper and to Appendix B, respectively.

RC: *Figure 4 would be more useful if it included some brief indication of what kinds of compounds the model species refer to.*

AR: Thank you for the suggestion. We have added Figure 6 to Appendix C, which includes the molecular structures of MCM species kept in the final reduced mechanism. Since merged species (names starting with "m", nine species in total) contain complex functional groups and can be surrogates for multiple MCM compounds, they are difficult to classify. Therefore, we have added Table 7 to Appendix B, which lists the names of the "Rdc." species with corresponding lumped species in original MCM. Particularly for condensables, an Excel file (Fig. 7) has also been added to the supplementary materials, which contains the molecular structures of "Rdc." condensables in UNIFAC structural group format.

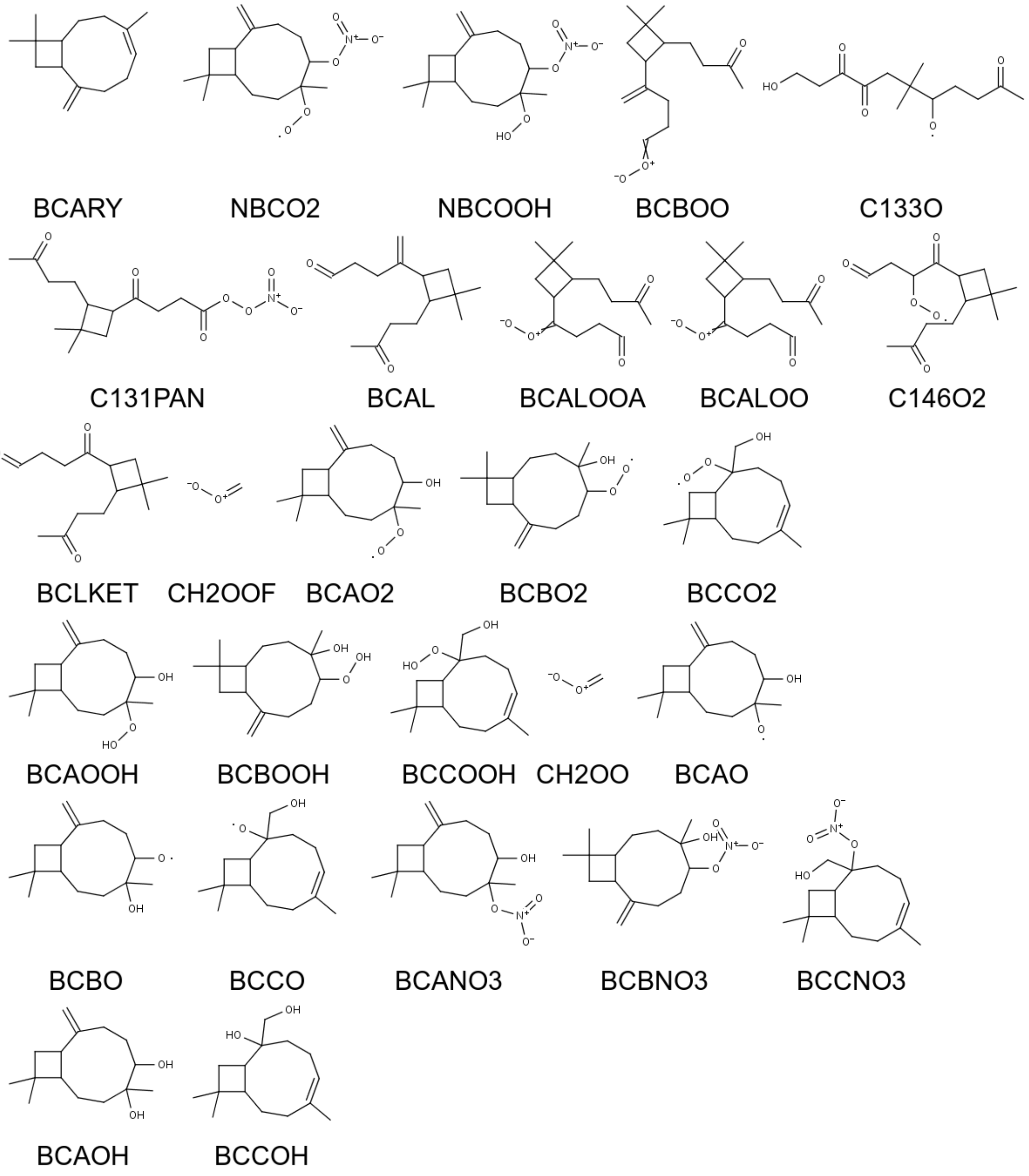
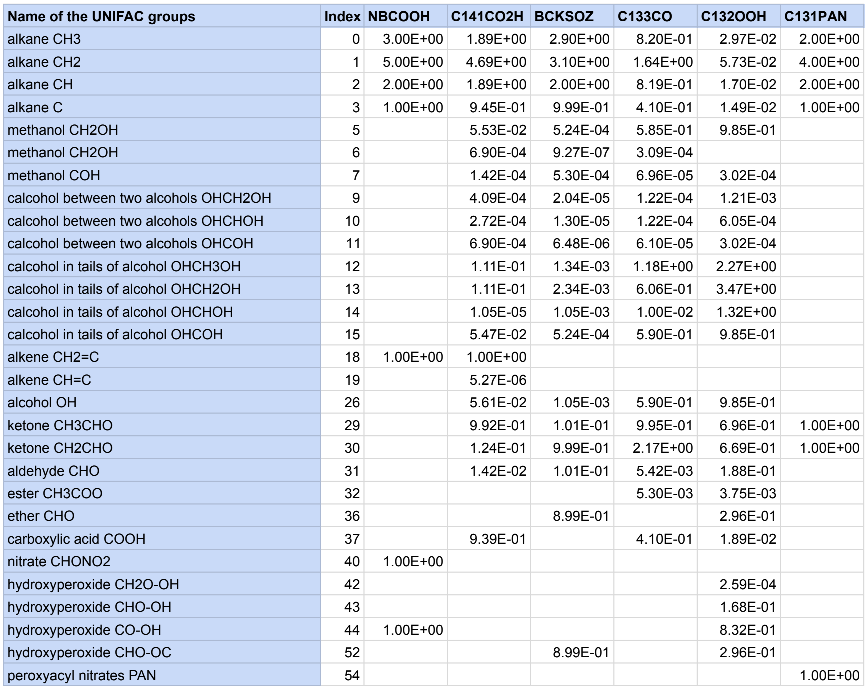


Figure 6: The molecular structures of the MCM species that are mentioned in this paper.



Name of the UNIFAC groups	Index	NBCOOH	C141CO2H	BCKSOZ	C133CO	C132OOH	C131PAN
alkane CH3	0	3.00E+00	1.89E+00	2.90E+00	8.20E-01	2.97E-02	2.00E+00
alkane CH2	1	5.00E+00	4.69E+00	3.10E+00	1.64E+00	5.73E-02	4.00E+00
alkane CH	2	2.00E+00	1.89E+00	2.00E+00	8.19E-01	1.70E-02	2.00E+00
alkane C	3	1.00E+00	9.45E-01	9.99E-01	4.10E-01	1.49E-02	1.00E+00
methanol CH2OH	5		5.53E-02	5.24E-04	5.85E-01	9.85E-01	
methanol CH2OH	6		6.90E-04	9.27E-07	3.09E-04		
methanol COH	7		1.42E-04	5.30E-04	6.96E-05	3.02E-04	
calcohol between two alcohols OHCH2OH	9		4.09E-04	2.04E-05	1.22E-04	1.21E-03	
calcohol between two alcohols OHCHOH	10		2.72E-04	1.30E-05	1.22E-04	6.05E-04	
calcohol between two alcohols OHCOH	11		6.90E-04	6.48E-06	6.10E-05	3.02E-04	
calcohol in tails of alcohol OHCH3OH	12		1.11E-01	1.34E-03	1.18E+00	2.27E+00	
calcohol in tails of alcohol OHCH2OH	13		1.11E-01	2.34E-03	6.06E-01	3.47E+00	
calcohol in tails of alcohol OHCHOH	14		1.05E-05	1.05E-03	1.00E-02	1.32E+00	
calcohol in tails of alcohol OHCOH	15		5.47E-02	5.24E-04	5.90E-01	9.85E-01	
alkene CH2=C	18	1.00E+00	1.00E+00				
alkene CH=C	19		5.27E-06				
alcohol OH	26		5.61E-02	1.05E-03	5.90E-01	9.85E-01	
ketone CH3CHO	29		9.92E-01	1.01E-01	9.95E-01	6.96E-01	1.00E+00
ketone CH2CHO	30		1.24E-01	9.99E-01	2.17E+00	6.69E-01	1.00E+00
aldehyde CHO	31		1.42E-02	1.01E-01	5.42E-03	1.88E-01	
ester CH3COO	32				5.30E-03	3.75E-03	
ether CHO	36			8.99E-01		2.96E-01	
carboxylic acid COOH	37		9.39E-01		4.10E-01	1.89E-02	
nitrate CHONO2	40	1.00E+00					
hydroxyperoxide CH2O-OH	42					2.59E-04	
hydroxyperoxide CHO-OH	43					1.68E-01	
hydroxyperoxide CO-OH	44	1.00E+00				8.32E-01	
hydroxyperoxide CHO-OC	52			8.99E-01		2.96E-01	
peroxyacyl nitrates PAN	54						1.00E+00

Figure 7: UNIFAC functional group decomposition for the condensable species in the "Rdc." mechanism.

Table 7: The new surrogate in the "Rdc." mechanism and the corresponding lumped species in the original MCM mechanism. Noted that the "Rdc." surrogates may also go through other reductions (e.g., jumping) that does not affect their structure.

Rdc. surrogate	lumped MCM species
mBCSOZ	BCSOZ, BCAL, BCKET
mC141CO2H	C141CO2H, C143CO, C1310CO, BCALCCO, C143OH, BCCOH, BCAOH
mBCALO2	BCALO2, C146O2, C142O2, BCKAO2, C147O2
mBCKSOZ	BCKSOZ, BCLKET, BCALOH, BCKBCO, BCKAOH, BCSOZOH
mC131CO3	C131CO3, C141CO3, C1211CO3, C137CO3
mC131O2	C131O2, C144O2, C143O2, BCLKAO2, C152O2, BCLKCO2
mC132OOH	C132OOH, BCSOZOOH, C133OOH, C146OOH, C147OOH C1313OOH, BCLKBOOH, BCLKAOOH, C152OOH, C145OOH C148OOH, C144OOH, BCALOOH, BCKBOOH, C151OOH
mC133O2	C133O2, C1313O2
mC133CO	C133CO, C131CO2H, C148CO, C145OH, C1313OH, BCLKBOH, BCLKAOH C152OH, C151OH, C147OH, BCLKACO, C148OH, C1211CO2H

References

[Couvidat et al., 2018] Couvidat, F., Bessagnet, B., Garcia-Vivanco, M., Real, E., Menut, L., and Colette, A. (2018). Development of an inorganic and organic aerosol model (chimere 2017β v1. 0): seasonal and spatial evaluation over europe. *Geosci. Model Dev*, 11(1):165–194.

[Derognat et al., 2003] Derognat, C., Beekmann, M., Baeumle, M., Martin, D., and Schmidt, H. (2003). Effect of biogenic volatile organic compound emissions on tropospheric chemistry during the atmospheric pollution over the paris area (esquif) campaign in the ile-de-france region. *J. Geophys. Res.-Atmos.*, 108(D17).