https://doi.org/10.5194/egusphere-2022-245
Preprint. Discussion started: 10 June 2022

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

Zhizhao Wang[1,2], Florian Couvidat[2], and Karine Sartelet[1]

[1]CEREA, École des Ponts ParisTech, EDF R&D, 77 455 Marne-la-Vallée, France.
[2]INERIS, Institut National de l'Environnement Industriel et des Risques, Verneuil-en-Halatte, France

Correspondence: zhizhao.wang@enpc.fr

Abstract. This paper describes the GENerator of Reduced Organic Aerosol Mechanisms (GENOA) that produces semi-explicit mechanisms for simulating the formation and evolution of secondary organic aerosol (SOA) in air-quality models. Using a series of predefined reduction strategies and evaluation criteria, GENOA trains and reduces SOA mechanisms from explicit chemical mechanisms (e.g., the master chemical mechanism (MCM)) under representative atmospheric conditions. As a consequence, these trained SOA mechanisms can preserve the accuracy of explicit VOC mechanisms on SOA formation (e.g., molecular structures of crucial compounds, effect of non-ideality and hydrophilic/hydrophobic partitioning of aerosols), with a size (in terms of reaction and species numbers) that is manageable for three-dimensional aerosol modeling (e.g., regional chemical transport models). Applied to the degradation of a sesquiterpene (β-caryophyllene) from MCM, GENOA builds a concise SOA mechanism (2% of the MCM size), consisting of 23 reactions and 15 species, six of them being condensable. The generated SOA mechanism has been evaluated for its ability to reproduce SOA concentrations under varying atmospheric conditions encountered over Europe, with an average error lower than 3%.

1 Introduction

Atmospheric aerosols have garnered considerable attention due to their undesirable effects on human health (Breysse et al., 2013) and climate (Seinfeld et al., 2016; McNeill, 2017), and as such, they must be well represented in air quality models (AQMs). Besides being directly emitted, aerosols can be formed in the atmosphere as secondary aerosols. Secondary organic aerosol (SOA) represents a significant fraction of aerosols (Hallquist et al., 2009), and is mainly formed by the condensation of the oxidation products from the degradation of volatile organic compounds (VOC) (Kanakidou et al., 2005).

The chemistry of VOC degradation is comprehensively described in explicit atmospheric chemical mechanisms, such as the near-explicit Master Chemical Mechanism (MCM) (Jenkin et al., 1997; Saunders et al., 2003), which details gas-phase chemical processes related to VOC oxidation, or the Generator for Explicit Chemistry and Kinetics of Organics in the Atmosphere (GECKO-A) (Aumont et al., 2005), which uses a prescribed protocol to assign complete reactions pathways and kinetic data to the degradation of VOCs. Explicit mechanisms represent the state of knowledge of atmospheric chemistry, including information about reaction pathways, kinetics data, and chemical structures (which may be used to deduce thermodynamic properties based on structure-activity relationships). Explicit mechanisms are mainly used in the box model for solving the multigener-

https://doi.org/10.5194/egusphere-2022-245
Preprint. Discussion started: 10 June 2022

ation oxidation of multitudinous VOCs (e.g., Yuan et al., 2018). While the near-explicit MCM mechanism has been applied to three-dimensional models (e.g., Ying and Li, 2011; Li et al., 2015), explicit mechanisms are generally too computationally intensive to be used in 3D air quality model because of the large number of species involved. For the sake of computational efficiency, AQMs generally use implicit gas-phase chemical mechanisms. Two major approaches are frequently adopted in order to build implicit chemical mechanisms:

- The lumped-species approach, which gathers into one surrogate compounds with analogous formulas and properties (e.g. SAPRC-07 Carter (2010), RACM2 Goliff et al. (2013))
- The carbon-bound or lumped-structure approach, which assumes that organic molecules have chemical behaviors equivalent to those of their decomposed functional groups (e.g. CB05 Sarwar et al. (2008))

These mechanisms were primarily developed for ozone simulation and reflect only the most significant chemical phenomena occurring in the atmosphere, where VOCs are represented by a limited number of model species.

Along with implicit gas-phase mechanisms, implicit SOA mechanisms have been 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). In the VBS approach, the successive evolution of oxidation products by aging is determined regardless of the chemical composition and structure of the species. Another implicit approach is based on the molecular surrogate approach (e.g., Griffin et al., 2003; Pun et al., 2006; Couvidat et al., 2012). Similarly to the gas-phase chemistry lumped-species approach, the VOC oxidation products are represented via the formation of a few SOA surrogates that are attached to a molecular structure (assumed to be representative of a myriad of semi-volatile compounds). By attaching a molecular structure to the surrogate, several processes otherwise not accounted for (like non-ideality, hygroscopicity, condensation on the aqueous phase of particles) can be represented in this approach. However, the choice of adequate molecular structures is crucial but it is highly uncertain, necessitating a precise estimation.

The computation of thermodynamic properties of aerosol (e.g., hydrophilicity, hydrophobicity, viscosity), additionally, requires knowing the molecular composition to take into account the whole complexity of the gas-particle partitioning (Kim et al., 2019). Therefore, tracking the whole complexity of the formation and aging of SOA with implicit chemical mechanisms can be problematic as it may not account for (or may oversimplify) some processes, such as non-ideality. These processes may be especially important to account for the non-linear relationship between the emissions of pollutants and the formation of aerosols (Huang et al., 2020).

The development of semi-explicit mechanisms is a compromise between the high computational time of explicit mechanisms and the lack of accuracy in the representation of chemical phenomena in the implicit mechanisms. This type of mechanism is being developed by reducing explicit mechanisms to a level of complexity suiting the computational constraints of AQMs. Recent developments of reduced mechanisms include the Common Representative Intermediates (CRI) mechanism (Jenkin et al., 2008; Watson et al., 2008) from an MCM reduction (Szopa et al., 2005), and the volatility basis set – Generator for Explicit Chemistry and Kinetics of Organics in the Atmosphere (VBS-GECKO) (Lannuque et al., 2018) from a GECKO-A

https://doi.org/10.5194/egusphere-2022-245
Preprint. Discussion started: 10 June 2022

reduction. However, the reduced mechanisms mentioned above do not track the detailed molecular structure of surrogates, but only consider specific properties of surrogates:

- CRI characterizes surrogates by their number of carbon-carbon and carbon-hydrogen bonds, which are reactive in the NO-to-NO_2 conversions with respect to ozone formation.
- VBS-GECKO groups organic surrogates by their volatility, as in the VBS approach (Donahue et al., 2006).

This study presents the development of the first version of the GENerator of reduced Organic Aerosol mechanism (GENOA) that generates customized semi-explicit chemical mechanisms appropriate for multi-scale AQM from explicit mechanisms, using surrogates assigned to molecular structures. As described in Sect.2, with effectiveness and efficiency, the new reduced mechanisms are designed to reproduce the complexity of the gas-phase oxidation by training under various atmospheric conditions and the non-ideality of gas-particle partitioning by using a molecular structure-preserving approach. GENOA also provides practical user-defined options, enabling users to specify the required reduction scale or accuracy. For gas/particle partitioning, a 0D box model SSH-aerosol (Sartelet et al., 2020) is modified and coupled with GENOA to simulate aerosol concentrations with the β-caryophyllene MCM mechanism. With SSH-aerosol, the effects of mass transfer between gas-phase and organic/aqueous phases, hygroscopicity, and non-ideality are taken into account in the reduction. The application of GENOA to the MCM degradation scheme of β-caryophyllene (BCARY) (Jenkin et al., 2012) is described in Sect.3. The β-caryophyllene species is selected because it widely serves as a benchmark for modeling the tropospheric chemistry and SOA formation relevant to sesquiterpene (SQT) (e.g., Li et al., 2015; Xavier et al., 2019), and its degradation has been evaluated in chamber simulations (Jenkin et al., 2012). The BCARY scheme in the STOCHEM-CRI mechanism (Khan et al., 2017) and the experiment data of Tasoglou and Pandis (2015) are also used for evaluation of the newly developed reduced mechanism.

2 Model development

The GENerator of Reduced Organic Aerosol Mechanisms (GENOA) is an algorithm that generates semi-explicit chemical mechanisms focusing on SOA formation. 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 AQMs.

As illustrated in Fig. 1, the processes in GENOA can be divided into two main sections: training, and testing. The training section, as detailed in Fig. 1, can be divided into two parts:

- Parameter selection, where the parameters to be used in the reduction cycle are selected automatically by GENOA from user-specified or preset values.
- Reduction cycle, where the actual reduction of the mechanism occurs.

https://doi.org/10.5194/egusphere-2022-245
Preprint. Discussion started: 10 June 2022

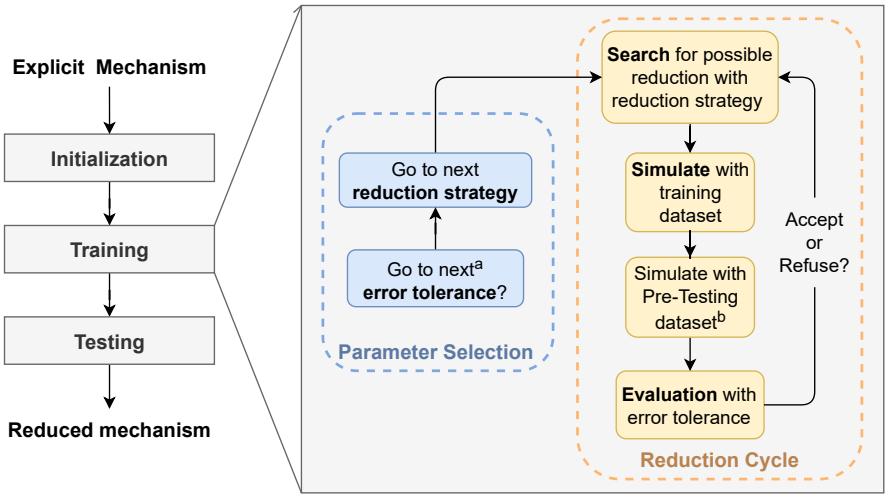

Figure 1. Flow chart indicating the three major procedures in GENOA and illustrating the main execution of the training section.

[a] GENOA uses the first value of the targeted variables for initialization, and passes to the next values for subsequent parameter updates.

[b] Activated under certain circumstances.

In the parameter selection, GENOA first assigns the error tolerance defined as the largest acceptable error induced by each change in the mechanism (see Sect.2.4), and then employs one of the reduction strategies along with its required parameters (see Sect.2.1).

Afterward, in the reduction cycle, GENOA searches for potential reductions according to the selected reduction strategy. The new mechanism with the first found reduction is then simulated over the conditions from the training dataset (a limited set of conditions used through all the reduction processes, Sect.2.2.1) or from the pre-testing dataset (a more extensive set of conditions used only at the end of the reduction process see Sect.2.2.2). The simulated total SOA concentrations are then compared to those simulated with the reference mechanism, where the differences are used to evaluate the potential reduction (see Sect. 2.4). In case the SOA differences are under the pre-defined error tolerances, the mechanism with the current reduction is accepted and serves as the basis for the next search for reduction. If the reduction is refused, the following reduction attempt starts with the previously validated mechanism. Once no more reduction is found, the current reduction cycle ends. The next step is either selecting the subsequent error tolerance and/or reduction strategy in the next parameter selection or terminating the GENOA training section. Finally, the performance evaluation of the final reduced mechanism is evaluated under a variety of environmental conditions denoted as the testing dataset (see Sect. 2.2.3). The 0D aerosol model SSH-aerosol is used to simulate SOA concentration and composition, which is required in all the GENOA sections (e.g., the initialization of reduction parameters, the evaluation of the reduced mechanism).

2.1 Reduction strategies

GENOA supports four types of reduction strategies:

https://doi.org/10.5194/egusphere-2022-245
Preprint. Discussion started: 10 June 2022

- Lumping: similar compounds are combined to form a new compound
- Replacing: one compound is replaced by another existing compound
- Jumping: an unstable compound with a short lifetime is assumed to degrade instantaneously into a stable compound (compound with significantly longer lifetime), as if the compound is "jumped over" in the mechanism.
- Removing: reactions, species, or gas-particle partitioning with negligible effects on SOA formation are eliminated from the mechanism.

For the BCARY reduction, the reduction strategies are employed in the following order: removing reactions, jumping, lumping, replacing, removing species, and finally removing gas-particle partitioning.

Table 1. Explicit reactions of MCM species BCAO2, BCBO2, BCCO2 in the degradation scheme of β-caryophyllene (BCARY)[a]

No.	Reactions	Kinetic rate coefficient[b]
0[c]	BCARY + OH → 0.408 BCAO2 + 0.222 BCBO2 + 0.37 BCCO2	1.97×10^{-10}
1	BCAO2 + HO_2 → BCAOOH	KAHO2 = KRO2HO2*0.975
2	BCAO2 + NO → 0.753 BCAO + 0.753 NO_2 + 0.247 BCANO3	KANO = KRO2NO
3	BCAO2 + NO_3 → BCAO + NO_2	KANO3 = KRO2NO3
4	BCAO2 + RO2 → 0.7 BCAO + 0.3 BCAOH	KARO2 = 9.2×10^{-14}
5	BCBO2 + HO_2 → BCBOOH	KBHO2 = KRO2HO2*0.975
6	BCBO2 + NO → 0.753 BCBO + 0.753 NO_2 + 0.247 BCBNO3	KBNO = KRO2NO
7	BCBO2 + NO_3 → BCBO + NO_2	KBNO3 = KRO2NO3
8	BCBO2 + RO2 → 0.8 BCBO + 0.2 BCAOH + 0.2 BCBCO	KBRO2 = 8.8×10^{-13}
9	BCCO2 + HO_2 → BCCOOH	KCHO2 = KRO2HO2*0.975
10	BCCO2 + NO → 0.753 BCCO + 0.753 NO_2 + 0.247 BCCNO3	KCNO = KRO2NO
11	BCCO2 + NO_3 → BCCO + NO_2	KCNO3 = KRO2NO3
12	BCCO2 + RO2 → 0.7 BCCO + 0.3 BCCOH	KCRO2 = 9.2×10^{-14}

[a] MCM v3.3.1. Species RO_2 represents the sum of all peroxy radicals.
[b] The same symbols is used to demonstrate the reduction strategies shown in Table 2,3. The precise values of kinetic rate coefficients can be found on the MCM website, in the unit $molecule^{-1}\cdot cm^3\cdot s^{-1}$.
[c] No.0 shows the production of BCAO2, BCBO2, and BCCO2, while the other reactions depicts their destruction processes.

2.1.1 Lumping strategy

The lumping strategy (consisting of lumping different compounds into a single surrogate compound) assumes that organic compounds with similar chemical structures may exhibit similar properties and undergo similar physico-chemical processes and may therefore be lumped together. With lumping, both the number of species and reactions decrease.

https://doi.org/10.5194/egusphere-2022-245
Preprint. Discussion started: 10 June 2022

Table 2. Reduced reactions of Table 1 and the computation of the weighting ratios, in case of lumping BCAO2, BCBO2, BCCO2 into a new surrogate mBCAO2.[a]

No.[b]	lumped[c]	Reactions[d]	Kinetic rate coefficient
R0	0	BCARY + OH → mBCAO2	1.97×10^{-10}
R1	1,5,9	mBCAO2 + HO_2 → $f_{w,a}$BCAOOH + $f_{w,b}$BCBOOH + $f_{w,c}$ BCCOOH	$f_{w,a}$* KAHO2 + $f_{w,b}$* KBHO2 + $f_{w,c}$ * KCHO2
R2	2,6,10	mBCAO2 + NO → 0.753 * ($f_{w,a}$BCAO + $f_{w,b}$ BCBO + $f_{w,c}$BCCO) + 0.753 * ($f_{w,a}$BCANO3 + $f_{w,b}$BCBNO3 + $f_{w,c}$BCCNO3) + 0.247 * ($f_{w,a}$+ $f_{w,b}$+ $f_{w,c}$) NO_2	$f_{w,a}$* KANO + $f_{w,b}$* KBNO + $f_{w,c}$* KCNO
R3	3,7,11	mBCAO2 + NO_3 → $f_{w,a}$BCAO + $f_{w,b}$BCBO + $f_{w,c}$ BCCO + ($f_{w,a}$+ $f_{w,b}$+ $f_{w,c}$) NO_2	$f_{w,a}$* KANO3 + $f_{w,b}$* KBNO3 + $f_{w,c}$ * KCNO3
R4	4,8,12	mBCAO2 + RO_2 → 0.7*$f_{w,a}$BCAO + 0.8*$f_{w,b}$BCBO + 0.7*$f_{w,c}$ BCCO + 0.2*$f_{w,b}$BCBCO + 0.3*$f_{w,a}$BCAOH + 0.2*$f_{w,b}$BCBOH + 0.3*$f_{w,c}$ BCCOH	$f_{w,a}$* KARO2 + $f_{w,b}$* KBRO2 + $f_{w,c}$ * KCRO2

Symbol[d]	Meaning[e]	Computation
$C_{r,a}$	reference concentration of BCAO2	average produced concentrations from five-day 0D simulations
τ_a	chemical lifetime of BCAO2	1/(KAHO2 [HO_2] + KANO [NO] + KANO3 [NO_3] + KARO2 [RO2])
τ_b	chemical lifetime of BCBO2	1/(KBHO2 [HO_2] + KBNO [NO] + KBNO3 [NO_3] + KBRO2 [RO2])
τ_c	chemical lifetime of BCCO2	1/(KCHO2 [HO_2] + KCNO [NO] + KCNO3 [NO_3] + KCRO2 [RO2])
$f_{w,a}$	weighting ratio of BCAO2	$\tau_a C_{r,a}/(\tau_a C_{r,a} + \tau_b C_{r,b} + \tau_c C_{r,c})$
$f_{w,b}$	weighting ratio of BCBO2	$\tau_b C_{r,b}/(\tau_a C_{r,a} + \tau_b C_{r,b} + \tau_c C_{r,c})$
$f_{w,c}$	weighting ratio of BCCO2	$\tau_c C_{r,c}/(\tau_a C_{r,a} + \tau_b C_{r,b} + \tau_c C_{r,c})$

[a] Name of new surrogate contains the letter "m" revealing lumping and the name of the relatively dominant lumped species. This notation of lumping is used hereafter.
[b] reaction number after lumping, where reactions R0 to R4 preserve the destruction of BCAO2, BCBO2, and BCBO2, and reaction R0 presenting the production.
[c] reaction numbers before lumping as presented in Table 1.
[d] subscript a, b, and c stands for BCAO2, BCBO2, and BCCO2, respectively.
[e] [X] the reference concentration of inorganic species, where X is HO_2, NO, NO_3, or RO_2 in this case. For radicals derived from the SOA precursor, the reference concentration is the produced concentration without considering their rapid destruction.

https://doi.org/10.5194/egusphere-2022-245
Preprint. Discussion started: 10 June 2022

The lumping strategy is illustrated by the comparison of Table 1 (reactions before lumping) and Table 2 (reactions after lumping). In this example, a total of 12 chemical reactions involving three organic compounds are reduced to four reactions with one new surrogate.

As demonstrated in the tables, the organic compounds BCAO2, BCBO2, and BCCO2 from the original MCM scheme are the peroxy radicals formed from the OH-initiated oxidation of β-caryophyllene (Table 1). Because these isomers undergo similar reactions with HO_2, NO, NO_3, they may conform to the lumping standard. When applying the lumping strategy, BCAO2, BCBO2, and BCCO2 are merged into a new surrogate named "mBCAO2" (Table 2).

The key parameter that drives the reduction accuracy is the weighting ratio f_w of lumping, corresponding to the weight of the original species in the new surrogate compound. As detailed in Table 2, 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 the explicit VOC mechanism under training conditions. The properties of the new surrogate compound (molecular structure, saturation vapor pressure, molar mass, kinetic of degradation, etc.) are estimated by weighting the properties of the initial compounds, while the stoichiometric coefficients and the kinetic rate coefficient of the new reaction are obtained by weighting those of the initial reactions.

In practice, GENOA attempts to lump only two species in a single reduction in order to ensure accuracy and effectiveness. Lumping is subject to certain restrictions:

- No lumping between a compound and its oxidation products.
- Due to specific chemical behavior, compounds with the following functional groups can only participate in lumping with the compounds sharing the same group: peroxyacetyl nitrates (PAN), organic nitrates (RONO2), organic radicals (R), oxy radicals (RO), peroxy radicals (RO_2), carboxylic acids (RC(O)OH), percarboxylic acids (RC(O)OOH).
- The difference in the molecular weight should be negligible (i.e. smaller than 100 g mol^{-1}).
- The difference in the carbon number should be no more than two.
- The difference in the chemical lifetime should be less than 10-fold.
- No lumping for intermediate compounds with extremely short lifetimes (e.g., biradicals (ROO)), as they are candidate for jumping instead.

2.1.2 Replacing strategy

The replacing strategy assumes that a compound with a negligible contribution to SOA formation can be substituted by a compound having a similar structure or undergoing the same reactions. In comparison to lumping, the replacing strategy reduces the number of reactions/species without creating new surrogate species.

Table 3 illustrates a reduction occurring via the replacing strategy (to be compared to the original mechanism in Table 1), assuming that BCAO2 is predominant in SOA formation. By substituting both BCBO2 and BCCO2 with BCAO2, the OH

https://doi.org/10.5194/egusphere-2022-245
Preprint. Discussion started: 10 June 2022

Table 3. Reduced reactions of Table 1, in case of replacing BCBO2 and BCCO2 by one existing species BCAO2.

No.[a]	replaced	Reactions	Kinetic rate coefficient
R0'	0	BCARY + OH → BCAO2	1.97×10^{-10}
R1'	1,5,9	BCAO2 + HO_2 → BCAOOH	KAHO2
R2'	2,6,10	BCAO2 + NO → 0.753 BCAO + 0.753 NO_2 + 0.247 BCANO3	KANO
R3'	3,7,11	BCAO2 + NO_3 → BCAO + NO_2	KANO3
R4'	4,8,12	BCAO2 + RO2 → 0.7 BCAO + 0.3 BCAOH	KARO2

[a] uses symbol ' to distinguish from the lumping reaction number in Table 2.

reaction of BCARY only leads to the production of BCAO2. The replacing strategy (Table 3) is expected to reduce more computational time than the lumping one (Table 2), since all reactions originating from the replaced species are removed from the mechanism. Hence, it does not require the computation of weighting ratios and new surrogates. However, as a compromise, replacing could be less accurate than lumping, because replacing may discard some compounds and part of the mechanism, and therefore, lead to more error.

Thus, in efforts to prioritize the accuracy of reduction, GENOA currently employs replacing only after lumping and exclusively on species from the same reaction. In this way, species that were not lumped (because the lumping was rejected or because they do not respect the lumping restriction) can be reduced by replacing. During the training of BCARY reduction, a restriction is applied on small organic compounds with molar mass less than 100 g mol^{-1}, which are excluded from replacing.

2.1.3 Jumping strategy

The jumping strategy relies on the assumption that compounds can be skipped in successive reactions, as long as it does not adversely impact the SOA concentration. In other words, the predecessor of an organic compound may directly form its destruction products. The jumping strategy is perfectly suited to intermediate compounds whose fast degradation may cause numerical stiffness, commonly including radicals such as oxy radicals (RO) or alkoxy radicals (ROO), as well as Criegee intermediates.

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).

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.

https://doi.org/10.5194/egusphere-2022-245
Preprint. Discussion started: 10 June 2022

Table 4. Reactions before and after the jumping strategy, where MCM species BCALOOA and CH2OOF are jumped over by their degradation products.

No.	Reactions	Kinetic rate coefficient
13	BCAL + O3 → BCALOOA + CH2OOF + BCLKET + HCHO	1.1×10^{-16}
14	BCALOOA → 0.5 BCALOO + 0.5 C146O2 + OH	1.0×10^{6}
15	CH2OOF → 0.37 CH2OO + 0.13 HO_2 + 0.63 CO + 0.13 OH	1.0×10^{6}
R5	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}

2.1.4 Removing strategy

The removing strategy assumes that chemical reactions and/or species with a low probability of contributing to the formation and evolution of SOA can be eliminated from the mechanism. In general, three types of removing are applied depending on the removed subject: (1) removing reactions, (2) removing compounds in both the gaseous and particle phases (completely removing a species from the scheme), and (3) removing the gas-particle partitioning of semi-volatile compounds (consider the semi-volatile compounds as VOCs without possibility to partition onto the particle phase, but preserve its gas-phase chemistry).

There is no particular restriction to exclude species from the reduction attempt via the strategy of removing compounds or removing gas-particle partitioning. However, for removing reaction, a threshold for its maximum hourly branching ratio under training conditions (B_{rm}) is applied to the reduction. To avoid over-reduction, a small B_{rm} is applied at the beginning of reduction. After going through the reductions for all reduction strategies, the value of B_{rm} is then incremented. In the reduction of BCARY, an ascending list of B_{rm} values equal to 5 %, 10 %, 50 % is employed, which is changed to 10 %, 50 %, and 100 % at the late stage (explained in Sect. 2.4). When B_{rm} equals to 100 %, GENOA evaluates the removal of each reaction.

Moreover, the searches for viable reductions via removing are conducted in reverse order of the reaction/species list, beginning with the species/reactions of the highest generation. The same reverse sequence is followed for other strategies, where the rank of saturation vapor pressure for condensable compounds is used exclusively for lumping. At each reduction, GENOA attempts to reduce only one species/reaction via removing, or one pair of compounds via lumping/ replacing/ jumping. This restriction allows for tracking exhaustively every detailed modification and its effect on SOA concentrations.

2.2 Datasets of atmospheric conditions applied to reduction

All the atmospheric conditions applied to the reduction are extracted from a 3D simulation spanning the latitudes from 32 °N to 79 °N and the longitudes 17 °W to 39.8 °E over continental Europe in a one-year period (2015) using the chemistry-transport model CHIMERE. The version of CHIMERE and its configuration is described in Lanzafame et al. (2022). The

https://doi.org/10.5194/egusphere-2022-245
Preprint. Discussion started: 10 June 2022

monthly average of hourly meteorological data (e.g., temperature, relative humidity), and hourly concentrations of oxidants and inorganic species were extracted from each location. 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 profiles obtained from the CHIMERE data. The concentration of the CHIMERE surrogate for sesquiterpene (denoted C_{SQT}) is used to estimate the SQT concentration. For the purpose of calculating reduction parameters (e.g., weighting ratio f_w, branching ratio B) and evaluating the reduced mechanisms, a dataset of representative physio-chemical conditions extracted from CHIMERE simulation results is employed in GENOA. Depending on their usages, three groups of conditions are defined: the training dataset, the pre-testing dataset, and the testing dataset.

2.2.1 Training dataset

The training dataset is the set of conditions used to initialize the reduction parameters, estimate the reference concentrations, as well as to evaluate the mechanism at each potential reduction. For a mechanism containing over 1000 reactions and 500 species, a complete reduction may require more than 10 000 SOA simulations to evaluate all the reduction attempts. To reduce the number of simulations and the computational cost, a limited number of conditions can be evaluated at each reduction attempt.

For the reduction of BCARY degradation, a training dataset of eight conditions is selected, which contains six chemistry-relevant conditions and two additional meteorological conditions. The geographic and meteorological information of each condition is described in Table 5, where the conditions cover a broad range in time (summer and winter conditions), temperatures ranging from 260 K to 302 K, and relative humidity from 39 % to 89 %.

Table 5. Geographic and meteorological conditions of the training dataset

Condition Name[a]	Lat °N	Lon °E	Time month	TEMP K	RH %
OH NO	36.0	15.4	Jul.	299.4	78.6
OH HO_2	32.0	-9.4	Jul.	295.9	76.7
NO_3 NO	40.25	-3.4	Jul.	302.4	27.9
NO_3 HO_2	32.0	36.6	Aug.	302.2	38.7
O_3 NO	69.0	33.8	Jan.	260.7	84.2
O_3 HO_2	68.0	18.2	Dec.	265.5	88.7
ADD1	41.5	-14.2	Dec.	288.6	75.8
ADD2	45.75	9.0	Dec.	279.1	84.5

[a] from left to right: name, latitude, longitude, time period, average temperature, average relatively humidity of the training conditions.

https://doi.org/10.5194/egusphere-2022-245
Preprint. Discussion started: 10 June 2022

The six chemistry-relevant conditions, which are named after the dominant oxidants (OH, O_3 and NO_3), focus on the influences of chemical regimes on SOA formation under either a high NO_X regime (represented by high NO concentrations) or a low NO_X regime (represented by high HO_2 concentrations). The two additional conditions included on the training dataset to improve the reduction, are referred to as ADD1 and ADD2.

The chemical regimes of the different conditions can by illustrated by seven competitive reacting ratios (equations are listed in Appendix Table A1):

- The reacting ratios of the precursor with the oxidants O_3 (R_{O_3}), OH (R_{OH}), and NO_3 (R_{NO_3}), whose sum equals 1, indicate the relative reactivity of the first-generation oxidation pathways that lead to the formation of distinct kinds of RO_2 species.
- The reacting ratios of RO_2 species with NO (R_{NO}), HO_2 (R_{HO_2}), NO_3 (R_{RNO_3}), and other RO_2 species (R_{RO_2}), whose sum equals 1, indicate the relative reactivity of successive reactions with RO_2 species.

These ratios indicate the competition between autoxidation and bimolecular reactions that result in different SOA types. A combination of these seven reacting ratios determines the chemical regime and favorable reaction pathways under a given atmospheric condition.

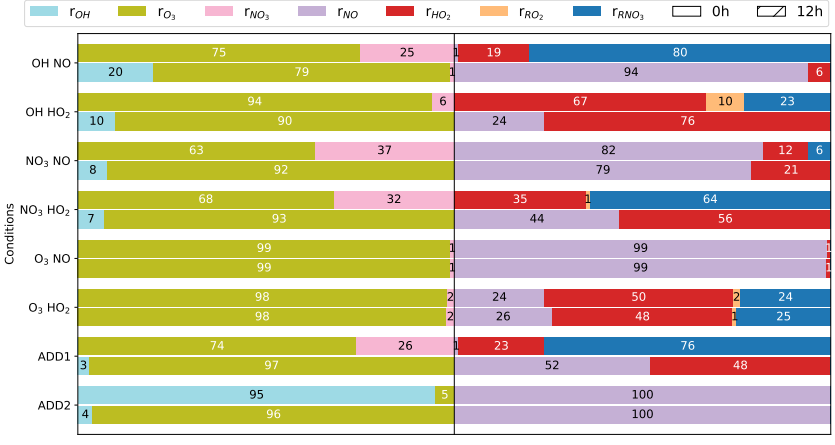

Figure 2. A bar plot showing the occupancy of seven reacting ratios in BCARY initiation reactions and RO_2 reactions, under the training conditions at midnight (0 h) and noon (12 h).

Fig. 2 describes the reacting ratios at midnight (0 h) and noon (12 h) for the training conditions. Under the majority of atmospheric conditions, O_3 is the dominant oxidant of BCARY due to the carbon-carbon double bonds that are subject to ozonolysis. The high O_3 training conditions have R_{O_3} exceeding 98 % at both noon and midnight. The bimolecular reactions with NO and HO_2 dominate RO_2 reactions in the MCM mechanism. Due to the low kinetic rate constants and low concentrations, the ratios of OH and NO_3 reacting with BCARY are relatively low (under 40 %). The high OH conditions are determined by the OH

https://doi.org/10.5194/egusphere-2022-245
Preprint. Discussion started: 10 June 2022

ratio at noon, while the high NO_3 conditions are determined by R_{RNO_3} at midnight. One specific exception is the additional condition ADD2, in which a high R_{OH} of 95 % at midnight is not due to an abundance of OH, but rather to extremely low concentrations of O_3 (2.9×10^{-4} ppb) and NO_3 (1.1×10^{-9} ppb) that leads to an absence of nighttime reactivity.

2.2.2 Pre-testing dataset

The pre-testing dataset contains a greater number of conditions than the training dataset, providing a more accurate estimation of the reduction mechanism on SOA formation. After the mechanism has been significantly reduced, the pre-testing dataset is included along with the training dataset in order to evaluate the reduction attempts at the late-stage reduction. At this point of reduction, a slight change in the mechanism significantly impacts the SOA concentrations; therefore, merely evaluating reduction based on the training dataset may not be adequate. Meanwhile, the size of the mechanism has already been significantly reduced, which makes the evaluation of each reduction attempt on the pre-testing dataset less computationally expensive. For the application to BCARY, a pre-testing dataset with 150 atmospheric conditions is selected from the testing dataset, among which 50 conditions for each level (low, medium, and high) of SQT emissions.

2.2.3 Testing dataset

The final reduced mechanism, obtained from training, is eventually evaluated with a large number of atmospheric conditions in the testing section. This set of conditions for the final evaluation is referred to as the testing dataset. Among all datasets, the results on the testing dataset are most likely to reflect the performance of the reduced mechanism for 3D modelling.

In the BCARY reduction, the testing dataset is selected based on the concentrations of the CHIMERE sesquiterpene surrogate. Its maximum hourly concentration C_{SQT} in ppb is used to exclude conditions with negligible SQT concentration. A testing dataset within a total of 12 159 conditions is applied (see Sect. 3.2), including all conditions (2 159 conditions) with high SQT concentration ($C_{SQT} \geq 0.1$ ppb), 5 000 random-select ones with medium SQT concentration (C_{SQT} ϵ between 0.01 and 0.1 ppb) and 5 000 random-select ones with low SQT concentration (C_{SQT} ϵ (0.001, 0.01]). The conditions with extremely low SQT concentration ($C_{SQT} < 0.001$ ppb) are not included on the testing dataset.

2.3 Settings for SOA simulations

The chemical composition and time variation of SOA due to gas-phase chemistry and condensation/evaporation are simulated using the 0D aerosol module SSH-aerosol (Sartelet et al., 2020). As detailed in Couvidat and Sartelet (2015), the gas/particle partitioning is estimated with Raoult's law (for the partitioning between the gas phase and the organic phase) and Henry's law (for the partitioning between the gas phase and the aqueous phase). Therefore, some properties of the condensable compounds, such as the saturation vapor pressure P_{sat} and the decomposition in functional groups, are crucial for the modelling. For BCARY derived organics, P_{sat} is calculated using UManSysProp (Topping et al., 2016) with the method of Myrdal and Yalkowsky (1997) combined with the boiling point estimation from Joback and Reid (1987), and the vapor pressure is computed using the method of Myrdal and Yalkowsky (1997). This functional group method was selected since it provides the best

https://doi.org/10.5194/egusphere-2022-245
Preprint. Discussion started: 10 June 2022

performance when compared with the chamber experiment data of Chen et al. (2012); Tasoglou and Pandis (2015), as discussed in Appendix A. Furthermore, the activity coefficient γ is calculated with the UNIQUAC Functional-group Activity Coefficients (UNIFAC) thermodynamic model (Fredenslund et al., 1975) for short-range interactions and the Aerosol Inorganic–Organic Mixtures Functional groups Activity Coefficients (AIOMFAC) model for medium-range and long-range interactions (Zuend et al., 2008). The initial BCARY concentration is taken equal to 5 μg m^{-3}.

Unless stated otherwise, a 5-day simulation is performed starting at midnight (0 h) and noon (12 h) for each condition, focusing on both the daytime and nighttime chemistry. For optimal computational efficiency, the gas-particle partitioning is assumed to be at thermodynamic equilibrium.

2.4 Settings for evaluation

For the different datasets, the performance of the reduced mechanism on SOA concentrations is evaluated using the fractional mean error (FME) computed with Eq. 1, where $C_{val,i}$ and $C_{ref,i}$ denote the SOA mass concentration at time step i simulated with the reduced and the reference mechanisms, respectively. To address the difference of performance of the reduced mechanisms at the early stage of the simulations (first day) and at the later stage, the error of one simulation is defined as the larger FME between the FME of the first simulation day 1 and the FME from day 2 to 5, and it is compared to the error tolerance specified in training (Eq. 1). For the evaluation on the training dataset, two errors are estimated compared to the previously verified reduced mechanism with a tolerance denoted ϵ_{pre}, and the MCM mechanism with a tolerance denoted ϵ_{ref}. As for the evaluation on the pre-testing dataset, one error compared to the MCM mechanism is calculated, with an error tolerance denoted $\epsilon_{pre-testing}$.

$$FME_{i1,i2} = \frac{2\sum_n^{i=1} abs(C_{val,i} - C_{ref,i})}{n\sum_n^{i=1}(C_{val,i} + C_{ref,i})}. \quad (1)$$

In order to begin with a conservative BCARY reduction, the initial values of ϵ_{pre} and ϵ_{ref} are both set to 1 %. The values of these error tolerances are then increased to larger values, reflecting the looser criteria used throughout the reducing. For every 1 % increase in ϵ_{ref}, ϵ_{pre} is stepped up by 1 % from 1 % to the value of ϵ_{ref}. The maximum values for both ϵ_{ref} and ϵ_{pre} are set to 10 %. When ϵ_{ref} reaches 3 %, the mechanism is expected to be largely reduced. From then, the evaluation under the pre-testing dataset is considered to be added to the reduction. This means that all subsequent reductions are evaluated using both the training and pre-testing datasets. The average/ maximum errors $\epsilon_{pre-testing}$ are restricted to be lower than 3 %/ 20 %, respectively. As a result of the above error tolerances, a reduced SQT-SOA mechanism with an average inaccuracy on SOA formation lower than 3 % (maximum 20 %) is expected.

Additionally, another error factor noted as the fractional bias (FB, computed as detailed in Eq. 2) is used to visualize the temporal performance of the reduced mechanism at each simulation time step. As examples, Fig. 7 and Fig. 9 show the average FB at each time step for the pre-testing conditions.

$$FB_i = 2\frac{C_{val,i} - C_{ref,i}}{C_{val,i} + C_{ref,i}} \quad (2)$$

https://doi.org/10.5194/egusphere-2022-245
Preprint. Discussion started: 10 June 2022

When trying to remove reactions, GENOA first removes reactions with low hourly branching ratios ($B_{rm} \leq 5$ %), since the removing reaction with B_{rm} is likely to have a minimal effect on SOA formation. After no reduction is accepted by all applied reduction strategies under the defined error tolerance, the value of B_{rm} is increased to 10 % and then 50 %.

2.5 Settings for aerosol-oriented treatments

In the late-stage training, an intense competition between different potential reductions is observed, and a minor modification may induce significant uncertainty into the mechanism and prevent further reduction. Besides, because the formation of aerosols costs more CPU time than gas-phase chemistry, certain specific treatments are employed in the late stage of training to reduce preferentially the number of condensables. These treatments are done when the size of the mechanism is below a certain threshold (20 for BCARY reduction). They are aerosol-oriented, aiming to encourage the reduction via the removing of condensable species. These treatments consist in:

- Restricting the reduction of the number of reactions. Thus, strategies that reduces the number of aerosols are favored to result in fewer condensable species.
- Bypassing the evaluation of aerosol-oriented reductions on the training dataset, when applied to lumping, replacing, and jumping. As a result, the aerosol-oriented reduction is evaluated only on the pre-testing dataset to avoid being rejected under some of the extreme conditions in the training dataset (which are less representative of average atmospheric conditions than the conditions of the pre-testing dataset).
- Applying an additional type of removing: removing elementary-like reaction.

The additional reduction strategy of removing elementary-like reaction is targeted to reaction with multiple products. After rewriting the reaction into a set of elementary-like reactions, each with one oxidation product, GENOA investigates the possibility to remove the elementary-like reactions one by one. In practice, removing elementary-like reaction is inserted after the strategy of removing reaction and before jumping, when no further reduction that reduces condensable species can be found with the current parameters.

3 Application to β-caryophyllene mechanism

GENOA is applied to the SQT degradation mechanism of the Master Chemical Mechanism v3.3.1 (Jenkin et al., 2012). Here β-caryophyllene (BCARY) is considered as a surrogate for SQT primary VOCs. The degradation of β-caryophyllene in MCM consists of 1 626 reactions and 579 species (223 radicals and 356 stable species). For the purpose of avoiding the numerical stiffness caused by solving the extremely rapid degradation of certain radicals, the kinetics of radicals that degrade with a kinetic rate coefficient larger than 1 s^{-1} are not considered and these radicals are ignored, as they are assumed to undergo instantaneous oxidation. This mechanism without fast radicals, hereafter referred to as MCM, contains 1 241 reactions and 493 species (137 radicals and 356 stable species). It is employed as the starting point and the reference for the reduction.

https://doi.org/10.5194/egusphere-2022-245
Preprint. Discussion started: 10 June 2022

Moreover, at the beginning of the GENOA training, all the stable species are assumed to be condensable (referred to as condensable species or condensables), and their saturation vapor pressures and activity coefficients are calculated based on their molecular structures (as detailed in Sect. 2). Applying the effective partitioning coefficients (K_p) described by Seinfeld and Pandis (2016), condensable species can be classified into: semi-volatile organic compound (SVOC, K_p between 10^{-2} and 10^{1} m^3 μg^{-1}), low volatile organic compound (LVOC, K_p between 10^{1} and 10^{4} m^3 μg^{-1}), and extremely low volatile organic compound (ELVOC, K_p larger than 10^{4} m^3 μg^{-1}).

The semi-explicit SQT-SOA mechanism "Rdc." presented in this section is trained from MCM with GENOA. Detailed descriptions of the building process and its chemical scheme are provided in Sect. 3.1. By the end of the training, "Rdc." is reduced from MCM to only 23 reactions and 15 species (reaction/species lists see Appendix B). The size of the "Rdc." mechanism is of the same order of magnitude as the additional BCARY SOA mechanism (28 reactions and 15 species) developed by Khan et al. (2017) for global modeling. As presented in Sect. 3.2, the "Rdc." mechanism accurately reproduces the SOA concentration and composition simulated by MCM with only six condensable species.

3.1 Building of the reduced SOA mechanism

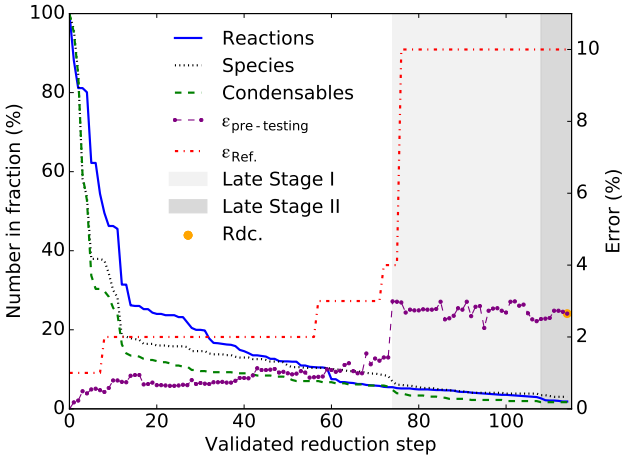

Figure 3. Reduction process of the "Rdc." mechanism showing the decrease of the number of reactions/species/condensables, and the evolution of the average error on pre-testing dataset ($\epsilon_{pre-testing}$), along with the error tolerance compared to MCM (ϵ_{ref}).

As shown in Fig. 3, the "Rdc." mechanism is built from 113 validated reduction steps. In GENOA, a reduction step refers to all reduction attempts based on the performed reduction strategy and reduction parameters, while a validated reduction step indicates at least one reduction attempt has been accepted at this step. The entire building process can be divided into three stages:

https://doi.org/10.5194/egusphere-2022-245
Preprint. Discussion started: 10 June 2022

- Early stage, from the first to the 74 th reduction step, where the number of condensable species drops to 20. The early-stage reduction is trained only on the training dataset with the seven pre-described reduction strategies. After ϵ_{ref} reaching 3 %, the list of B_{rm} is changed from [0.05, 0.10, 0.50] to [0.10, 0.50, 1.0].

- Late stage I, from the 75 th to the 107 th reduction step, where the number of condensables decrease to 7 and no more reduction can be found within $\epsilon_{ref} \leq 10$ % and $\epsilon_{pre} \leq 10$ %. On the 75 th reduction step, the size of the mechanism drops to 68 reactions and 41 species. From there, the reduction is trained on the pre-testing dataset if the condensable species are removed with lumping, replacing, or jumping. For reduction with other types of reduction strategies, it is first trained on the training dataset and then on the pre-testing datasets. From all reduced mechanisms with seven condensable species, GENOA selected the one with the minimum average errors on pre-testing dataset (2.44 %) to start the next stage.

- Late stage II, from the 108 th to the 113 rd reduction step. At this stage, the reduction strategy of removing elementary-like reactions is applied to the training. All reductions that reduce the condensables are evacuated exclusively on the pre-testing dataset. From the 108 th reduction step to the end of the training, the size of the reduced mechanism was reduced from 37 reactions and 19 species to 23 reactions and 15 species, among which the number of condensable species is reduced from 7 to 6. The average (maximum) error of the final reduced mechanism "Rdc." is 2.65 % (17.00 %) under the pre-testing dataset compared to MCM.

Table 6. Reduction accomplished per each reduction strategy during the building process of the "Rdc." mechanism.

Strategy	No. (fraction[a] %)		
	Reaction[b]	Species	Condensables
Removing reactions	594(48)	38(8)	26(7)
Removing elementary-like reaction[c]	8(1)	0(0)	0(0)
Lumping	0(0)	171(35)	110(31)
Replacing	25(2)	39(8)	31(9)
Jumping	138(11)	79(16)	43(12)
Removing species	453(37)	151(31)	108(30)
Removing partitioning	0(0)	0(0)	32(9)
Removed in total	1218(98)	478(97)	350(98)

[a] fraction of the original number (of reactions or species) that is reduced by the strategy.
[b] columns from left to right are the number (and fraction) of reduced chemical reactions, reduced total gas-phase species, and reduced gas-phase species that can condense on the particle phase, compared to MCM with 1 241 reactions and 493 species (356 condensables).
[c] only applied in the reduction at late stage II.

The extent of the reduction due to each strategy is summarized in Table 6. Compared to MCM, up to 99 % of reactions and 97 % species are reduced in "Rdc.". As expected, the reduction strategy of removing reaction contributes the most to the decrease

https://doi.org/10.5194/egusphere-2022-245
Preprint. Discussion started: 10 June 2022

in the number of reactions (48 %), followed by the strategy of removing species with a contribution of 37 % and lumping with a contribution of 31 %. Meanwhile, both lumping and removing species are significant in the reduction of species, by 35 % and 31 %, respectively. The number of condensable species decreases in proportion to the number of species, except for the strategy of removing partitioning. In that case, the gas-particle partitioning is removed and the species remains in the gas phase with no changes in the chemical mechanism.

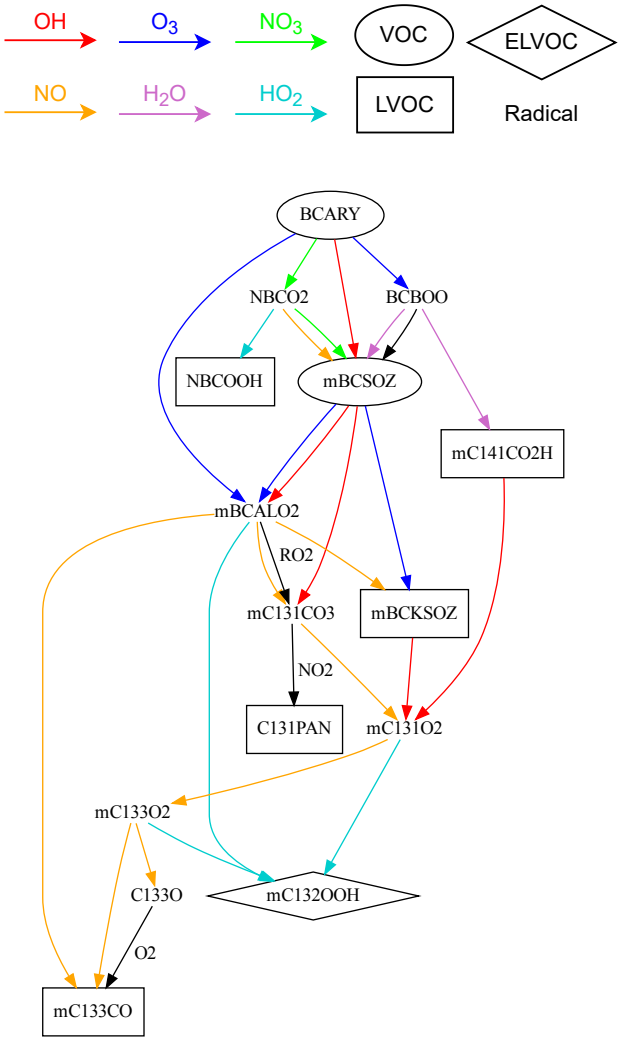

Figure 4. Representation of the chemical scheme of the "Rdc." mechanism. Species of types VOC, LVOC, ELVOC are presented in ellipse, square and diamond boxes, respectively. Radicals are written in plain text, without boxes. Reaction with OH, O_3, NO_3, NO, and H_2O, HO_2 are shown by arrows in red, blue, green, orange, purple, and sky-blue, respectively. The complete species and reaction lists of the "Rdc." mechanism are in Appendix Table B1 and C1, respectively.

https://doi.org/10.5194/egusphere-2022-245
Preprint. Discussion started: 10 June 2022

As shown in Fig.4, which describes the chemical scheme of the "Rdc." mechanism, the three oxidants (i.e., O_3, OH, and NO_3) initiate reactions, leading to common oxidation products (e.g., mBCSOZ, mBCALO2) that dominates the successive oxidations. The different reaction pathways under high or low NO_X regimes are presented in "Rdc." with the reactions with NO or HO_2, respectively, which results in the formation of different types of SOA: mBCKSOZ, mC133O, and C131PAN (in the presence of NO_2) under high NO_X conditions, and mC132OOH under low NO_X conditions. Other pathways, for example, the bimolecular reactions of the Criegee intermediate BCBOO with water vapour, and the RO_2 reaction of mBCALO2, are also preserved in the "Rdc." mechanism. The six condensable species in "Rdc." can be categorized into one SVOC, four LVOCs, and one ELVOC, according to the effective partitioning coefficient calculated on pre-testing dataset. The SOA concentration per volatility class is discussed in Sect. 3.2.

Compared to MCM, "Rdc." simplifies a considerable number of reactions that have small impacts on SOA formation (e.g., photolysis reactions) under the majority of atmospheric conditions, and merges a large number of compounds with similar chemical properties. The main oxidation products from the first two generation oxidation of MCM are preserved mainly through the "Rdc." species mBCSOZ, which is a lumped surrogate of several MCM representative BCARY derived oxidation products: BSCOZ (the major secondary ozonize with a molar yield $\geq$ 65 % reported by Jenkin et al. (2012)), BCAL (the primary product formed from both OH and O_3-initiated chemistry), and BCKET (from OH-initiated reactions).

3.2 Evaluation of the reduced SOA mechanism

3.2.1 Reproduction of the SOA concentrations

During the testing procedure, the "Rdc." mechanism is evaluated at 12 159 locations, with two different starting times (0 h and 12 h). Compared to MCM, "Rdc." presents a high level of accuracy with an average error of 2.66 % and a maximum error of 17.29 %. The monthly distribution of the number of the testing conditions as well as the testing errors are described in Fig. 5. The error is lower than 10 % for more than 99 % of the simulations. The summer conditions, between June and September, covering more than half of the testing conditions (63 %, 15 294 conditions), result in an average error of 2.37 % and a 3^{rd} quartile error of 2.85 %. As compared to summer conditions, testing results in winter conditions from October to January (19 % of the testing dataset, 4 570 conditions) display slightly higher uncertainty, with an average error of 3.79 % and a 3^{rd} quartile error of 5.36 %.

An error map of testing conditions in July and August is displayed in Fig. 6. It reveals more detailed information, especially on the outliers during this period (An error map of all testing conditions is shown in the Appendix B.). It shows that the "Rdc." mechanism induces low errors, lower than 6 %, for most of the testing conditions. The conditions with errors over 6 % are mainly concentrated in northern Africa near the Atlas Mountains and in the Eastern Mediterranean, where the conditions most likely correspond to a dry Mediterranean climate with low RH and high temperature. Other conditions with errors above 6 % are dispersed in the Pô valley of Northern Italy and along the coasts of southern Spain. More accurate results could be obtained with stricter parameters for reduction (e.g., lower error tolerance), or with updating the conditions (e.g., training and pre-testing datasets) covering more extreme conditions in the training process.

https://doi.org/10.5194/egusphere-2022-245
Preprint. Discussion started: 10 June 2022

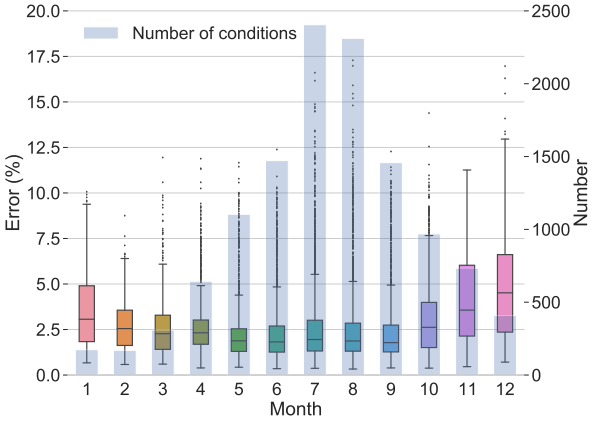

Figure 5. Monthly distribution of the testing results (errors compared to MCM) of the "Rdc." mechanism in box plot, and number of testing conditions in histogram.

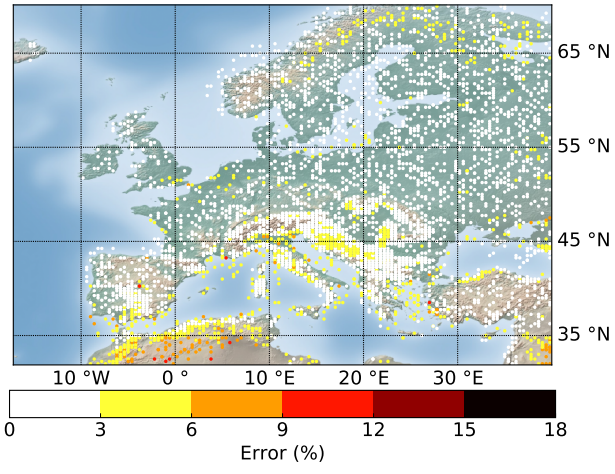

Figure 6. Error map that describes the geographic distribution of the testing results of the "Rdc." mechanism in July and August. The total number of conditions displayed is 4 717.

https://doi.org/10.5194/egusphere-2022-245
Preprint. Discussion started: 10 June 2022

3.2.2 Reproduction of the SOA composition

The SOA concentrations and chemical composition simulated with the "Rdc." mechanism and with MCM are compared in this section. The temporal profiles of the total SOA concentrations on an average of the pre-testing dataset and non-ideal conditions are displayed in Fig. 7. Throughout the entire 5-day simulation period, there is excellent agreement between hourly SOA concentrations simulated with MCM and those obtained from the "Rdc." mechanism. The SOA concentration builds up rapidly in the first few hours, where the results of the "Rdc." mechanism present relatively larger fluctuations (The maximum FB of 3.74 % is observed at 1 h on the average pre-testing results).

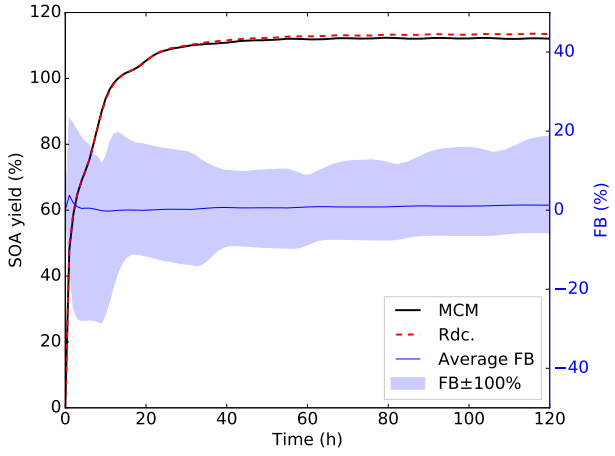

Figure 7. Temporal variation of total SOA concentration simulated with the pre-testing dataset and MCM (red, dotted), and the "Rdc." mechanism (black, plain) mechanisms under the non-ideal conditions. The average (blue, plain) and maximum FB (blue, shadow) between MCM and the "Rdc." mechanism is also presented.

The average SOA concentrations per volatility class on the pre-testing dataset at two simulation times (8 h and 72 h) are listed in Table 7. At both 8 h and 72 h, the "Rdc." mechanism accurately reproduces the total SOA mass with a relative difference lower than 0.1 % compared to MCM. An accumulation of the SOA mass into the ELVOC class is observed (51 % of the total SOA mass at 8h and 66 % at 72h), with both the MCM and the "Rdc" mechanisms. The aging of SOA produces compounds of low and extremely low volatility. Regarding the volatility classes, the "Rdc." mechanism tends to slightly overestimate the SOA resulting from ELVOCs and underestimate the SOA resulting from LVOCs, especially at 72 h. This suggests that ageing leads to "Rdc." condensables of slightly lower volatility than the MCM ones. However, the differences are low, up to 0.4 μg m^{-3} difference (10 %) at 72 h.

The average SOA composition per functional group simulated on the pre-testing dataset at 72 h is displayed in Fig.8. No significant change in the functional group distributions is found between 8 h and 72 h of oxidation. The alkyl (RC) and carbonyl groups (RCO) contribute the most to the SOA mass, by more than 1 μg m^{-3}, whereas the other functional groups contribute by less than 1 μg m^{-3}. Overall, the "Rdc." mechanism satisfactorily reproduces the composition of the MCM-simulated SOA

https://doi.org/10.5194/egusphere-2022-245
Preprint. Discussion started: 10 June 2022

Table 7. Average SOA concentrations per volatility class simulated with MCM and the "Rdc." mechanisms on pre-testing dataset at 8 h and 72 h.[a]

Conditions	SVOC	LVOC	ELVOC	Total
MCM at 8 h	0.18	1.91	2.17	4.26
"Rdc." at 8 h	0.13	1.80	2.31	4.24
MCM at 72 h	0.02	1.90	3.69	5.61
"Rdc." at 72 h	0.02	1.51	4.12	5.64

[a] unit in μg m^{-3}.

415 composition for most functional groups, except for nitrogen-containing groups. In comparison to MCM, only two condensable species containing nitrogen are retained in the "Rdc." mechanism: NBCOOH and C131PAN, leading to an overestimation of the nitrate mass of the organic nitrate group (0.10 μg m^{-3} in MCM and 0.30 in "Rdc.") and an underestimation of the peroxyacetyl nitrate group (0.31 in MCM and 0.04 in "Rdc."). To obtain better results on the reproduction of nitrogen groups, GENOA may be further restricted to distinguish nitrogen compounds in the training. Additionally, the peroxyacetyl acid group 420 results in an extremely low SOA mass in MCM (less than 0.01 %), and therefore, is not kept in the "Rdc." mechanism.

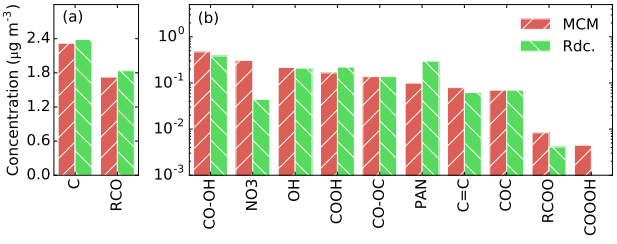

Figure 8. Average SOA mass per functional group simulated with the pre-testing dataset and MCM (red bar with slash), the "Rdc." mechanism (green bar with backslash) at 72 h. The figure is divided into two panels (a) and (b) due to the large gap in mass between the groups. The labels of the functional groups from left to right are RC carbon bond, RCO carbonyls (ketone and aldehyde), CO-OH hydroxy peroxide, NO3 organic nitrates, OH alcohol, COOH carbonyl acid, CO-OC peroxide, PAN peroxyacyl nitrates, C=C carbon double bond, COC ether, RCOO ester, COOOH peroxyacetyl acid.

Moreover, the temporal profiles of the OM/OC ratio, as well as the H/C, O/C, and N/C atomic ratios are presented in Fig. 9. Comparable patterns are observed in the OM/OC (1.65 in MCM and 1.63 in "Rdc." on average), the O/C (0.37 in MCM and 0.36 in "Rdc."), as well as the H/C ratios (1.62 in MCM and 1.60 in "Rdc."). During the first 8-hour simulation, "Rdc." tends to slightly overestimate the OM/OC and O/C ratios, while the H/C ratio remains fairly stable throughout the entire simulation with 425 a negligible difference (0.02) between MCM and "Rdc.". The N/C ratio, however, is underestimated by the "Rdc." mechanism by 37 % on average (ratio equal to 0.019 in MCM and to 0.012 in "Rdc."). The N/C ratio is consistent with the nitrogen-

https://doi.org/10.5194/egusphere-2022-245
Preprint. Discussion started: 10 June 2022

containing SOA mass distribution discussed previously, as a result of over reducing organic nitrites in the building process of "Rdc.".

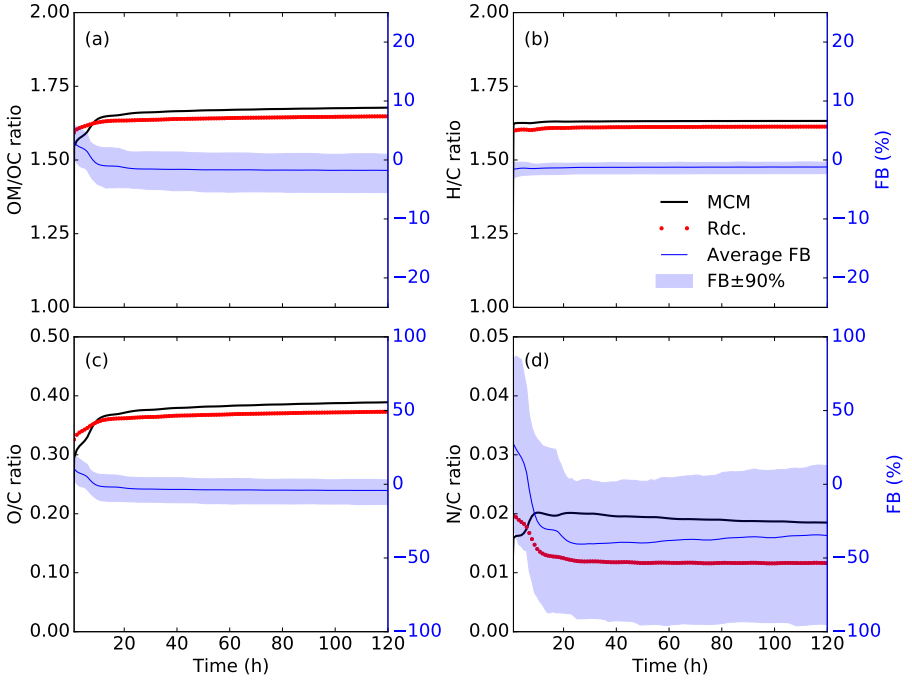

Figure 9. Time variations of (a) average organic mass to organic carbon mass (OM/OC) ratio, (b) hydrogen to carbon (H/C) atomic ratio, (C) oxygen to carbon (O/C) atomic ratio, and (d) nitrogen to carbon (N/C) atomic ratio, simulated with MCM (solid black curves) and the "Rdc." mechanism (dotted red curves) on the pre-testing dataset. The average FB (blue plain line) with the 90 % range of FB (blue shadow) is also presented.

3.2.3 Sensitivity on environmental parameters

The sensitivities of the "Rdc." mechanism to temperature, relative humidity (RH), and SOA mass conditions are investigated with the pre-testing dataset. The default value of BCARY concentration is 5 μg m^{-3}, and the default RH and T are set to constant 50 % and 298 K, respectively. As presented in Fig. 10, the SOA yields simulated by the "Rdc." mechanism with different environmental parameters show a remarkable resemblance with the SOA yields simulated by MCM.

Under 10 μg m^{-3}, the simulated SOA yields are not affected by the SOA mass loading. This result is consistent with the large contribution of ELVOC reported in Table 7. A discrepancy of 25 % in the average SOA yield at 1 h with an SOA mass loading of 10^3 μg m^{-3} at 1 h and a discrepancy of 8 % at 72 h with an SOA mass loading of 10^{-3} μg m^{-3} are observed. The result indicates that the "Rdc." mechanism may introduce relatively large uncertainty with extreme SOA loading (conditions that were outside the range of conditions used for the construction of the "Rdc." mechanism). Results indicate that the "Rdc." mechanism may delay the SOA formation under large mass loading (larger than 500 μg m^{-3}), because of differences of the

https://doi.org/10.5194/egusphere-2022-245
Preprint. Discussion started: 10 June 2022

volatility of the oxidation products of the "Rdc" mechanism and MCM. SOA formation is affected by RH, because of both the gas-phase chemistry (reaction with H_2O vapors) and the gas-particle transfer (condensation of hydrophilic SOA precursors on aqueous aerosols). The sensitivity tests show that the "Rdc." mechanism reproduces well (differences lower than 2 %) the SOA yields of MCM with relative humidity ranging from 5 % to 95 %. For temperature, the "Rdc." mechanism reproduces very well the SOA aging at 72 h, but larger discrepancies are observed at earlier time, when the the oxidation products are more volatile. However, the discrepancies on SOA yield stay low: differences up to 7 % (at 1 h and 72 h) and 10 % (at 8 h) are observed for temperatures equal to 263 K and 323 K, respectively. This finding is consistent with the testing results. To sum up, the discrepancies suggest that the reduced mechanism performs quite well, although larger discrepancies with MCM are observed under conditions that are outside the range of conditions used during training.

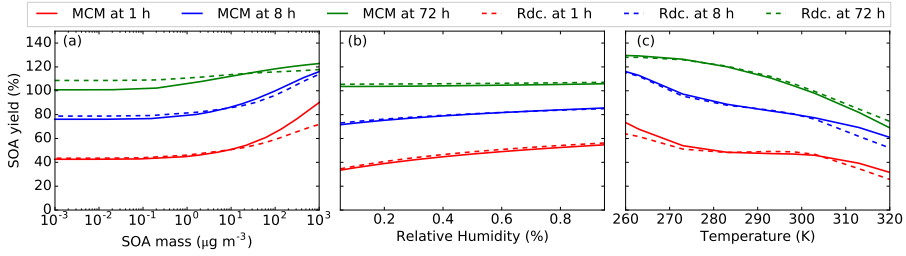

Figure 10. Dependence of the average SOA yield simulated by the pre-testing dataset with MCM (solid line) and the "Rdc." mechanism (dotted line) on (a) BCARY-SOA mass, (b) relative humidity, and (c) temperature at 1 h (red), 8 h (blue) and 72 h (green).

4 Conclusions

The development and application of the GENerator of Reduced Organic Aerosol Mechanisms (GENOA) have been presented in this study. GENOA generates semi-explicit SOA mechanisms designed for large-scale air quality modeling by reducing explicit VOC mechanisms with a series of automatic training and testing processes. During the training procedure of GENOA, four types of reduction strategies (lumping, replacing, jumping, and removing) are adopted to locate the potential reduction in the mechanism. Each reduction attempt is evaluated against the explicit mechanism under a sequence of near-realistic atmospheric conditions (the training dataset, and/or the pre-testing dataset at the late stage of reduction). Finally, the reduced mechanism is evaluated on various conditions of a testing dataset. Under each condition, two 5-day 0D simulations starting at midnight and noon are conducted with the aerosol model SSH-aerosol to evaluate the performance of the reduce SOA mechanism.

GENOA successfully generated semi-explicit SOA chemical mechanisms for the degradation of sesquiterpene, for which the explicit β-caryophellene mechanism of the Master Chemical Mechanism serves as the reference mechanism and the starting point. The final reduced SQT-SOA mechanism, contains 23 reactions (down from 1626 reactions in MCM), 15 gas-phase species (down from 579 gases), and 6 aerosol species (down from 356 aerosols). It reproduces the SOA formation and aging by introducing an average error of 2.7 % under conditions over Europe with only 2 % of the size of MCM. The SOA volatility is well reproduced with the reduced mechanism, as well as the decomposition into functional groups, and the OM/OC (1.55

https://doi.org/10.5194/egusphere-2022-245
Preprint. Discussion started: 10 June 2022

in the "Rdc." mechanism and 1.60 in MCM), H/C and O/C ratios. Nitrogen containing SOA, which contributes to only 7 % of the total mass, is not as well represented as other groups, and the ratio N/C is slightly underestimated in the "Rdc." mechanism (0.016 against 0.021 in MCM). The similarity of the representation of the functional group decomposition allows to reproduce the non-ideality of SOA similarly in the "Rdc." mechanism and in MCM. Additionally, the sensitivity tests on RH, Temperature, and organic mass loading show that the SOA simulated with the "Rdc." mechanism is in good agreement with MCM results under most of conditions (except for conditions with extremely high temperature or with massive organic aerosol loading where discrepancies in the SOA yields may reach 8 % (temperature) and 25 % (massive mass loading)). It indicates that the reduced mechanism performs well for conditions in the training range, but the performance may decrease for conditions outside of this range. To improve the performance of the semi-explicit SOA mechanism under conditions outside the training range, two methods can be employed: the first is to include the outlier conditions in the training procedure if they are considered influential to SOA formation, and the second is to adopt strict error tolerance to restrict the reduction.

Code and data availability. The source code for GENOA v1.0 is hosted on GitHub at https://github.com/tool-genoa/GENOA/tree/v1.0 (last access: 25 April 2022). The associated Zenodo DOI is https://doi.org/10.5281/zenodo.6482978. The dataset we used to run the BCARY MCM reduction is publicly available online on Zenodo: https://doi.org/10.5281/zenodo.6483088.

Appendix A: The computation of saturation vapor pressure of BCARY-SVOCs

The ozonolysis experimental data reported in Tasoglou and Pandis (2015) and Chen et al. (2012) are used to evaluate the performances of different computation methods for the saturation vapor pressure of BCARY oxidation products. In our simulations, the saturation vapor pressure is computed by UManSysProp with the SMILES structures of organic compounds. Eight methods are provided in UManSysProp, including SIMPOL.1 of Pankow and Asher (2008) ("sim"), EVAPORATION of Compernolle et al. (2011) ("evp"), and six methods out of the combination of two methods to compute the vapor pressure ("v0": Myrdal and Yalkowsky (1997) and "v1": Nannoolal et al. (2004)) and three methods to compute the boiling point ("b0": Nannoolal et al. (2008), "b1": Stein and Brown (1994), and "b2": Joback and Reid (1987)). As shown in Fig. A1, the SOA distribution simulated with "v1b2" agrees best with the experimental data. Therefore, this method with the vapor pressure computed by Myrdal and Yalkowsky (1997) and the boiling point computed by Joback and Reid (1987) is used in the BCARY reduction. The results simulated with the final reduced mechanism "Rdc." is also presented in Fig. A1, which has a great resemblance to the experimental data.

https://doi.org/10.5194/egusphere-2022-245
Preprint. Discussion started: 10 June 2022

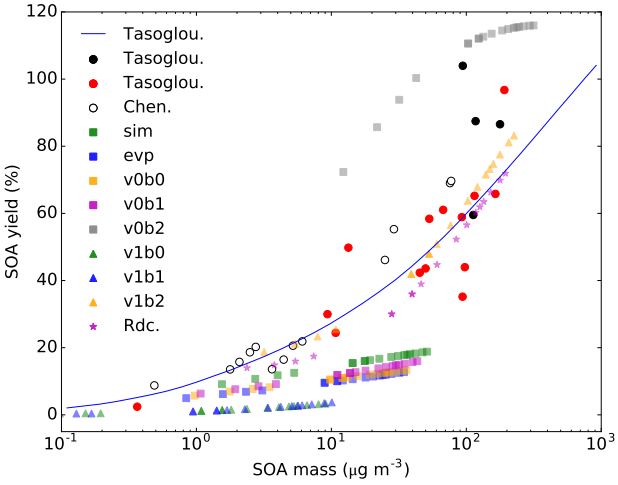

Figure A1. The SOA yields versus the total SOA mass from the experimental data reported by Chen et al. (2012); Tasoglou and Pandis (2015), simulated in SSH-aerosol with the MCM mechanism and different saturation vapor pressures methods, and simulated with the "Rdc." mechanism. The "Rdc." mechanism is trained from the MCM mechanism with the "v1b2" method.

Appendix B: Information related to the reduction of the beta-caryophyllene mechanism

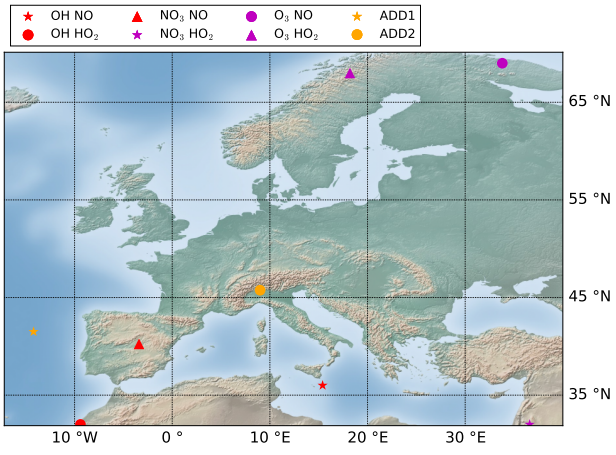

Figure C1. Simulation domain and locations of the training dataset for the reduction.

Author contributions. ZW developed the model code and performed the simulations. ZW, FC, KS designed the research, developed the methodology. ZW wrote the manuscript with contributions from FC and KS. FC and KS were responsible for funding acquisition.

https://doi.org/10.5194/egusphere-2022-245
Preprint. Discussion started: 10 June 2022

Table A1. The computation of estimating chemical activity ratios used to display training dataset in Fig. 2

Name[a]	Reactant[b]	Computation[c]	Kinetic rate coefficient[d]
r_{OH}	OH	k_{OH}[OH] / (k_{OH}[OH] + k_{O_3}[O_3] + k_{NO_3} [NO_3])	k_{OH} = 1.97×10^{-10}
r_{O_3}	O_3	k_{O_3}[O_3] / (k_{OH}[OH] + k_{O_3}[O_3] + k_{NO_3} [NO_3])	k_{O_3} = 1.20×10^{-14}
r_{NO_3}	NO_3	k_{NO_3} [NO_3] / (k_{OH}[OH] + k_{O_3}[O_3] + k_{NO_3} [NO_3])	k_{NO_3} = 1.90×10^{-11}
r_{NO}	NO	k_{NO} [NO]/(k_{NO} [NO] + k_{HO_2} [HO_2] + k_{RNO_3} [NO_3] + k_{RO_2} [RO_2])	k_{NO} = $2.70\times10^{-12}\times\exp(\frac{360}{T})$
r_{HO_2}	HO_2	k_{HO_2} [HO_2]/(k_{NO} [NO] + k_{HO_2} [HO_2] + k_{RNO_3} [NO_3] + k_{RO_2} [RO_2])	k_{HO_2} = $2.91\times10^{-13}\times\exp(\frac{1300}{T})$
r_{RO_2}	RO_2	k_{RO_2} [HO_2]/(k_{NO} [NO] + k_{HO_2} [HO_2] + k_{RNO_3} [NO_3] + k_{RO_2} [RO_2])	k_{RO_2} = 9.20×10^{-14}
r_{RNO_3}	NO_3 + RO_2	k_{NO_3} [NO_3]/(k_{NO} [NO] + k_{HO_2} [HO_2] + k_{RNO_3} [NO_3] + k_{RO_2} [RO_2])	k_{RNO_3} = 2.30×10^{-12}

[a] names of the reacting ratio of OH radical, O_3, and NO_3 radical reacted with BCARY. (r_{OH} + r_{O_3} + r_{NO_3} = 1); and of the reacting ratio of NO, HO_2 radical, RO_2 radical, and NO_3 radical (at the presence of RO_2) reacted with RO_2 species (r_{NO} + r_{HO_2} + r_{RO_2} + r_{RNO_3} = 1).

[b] Reactions with those compounds are preferred when the corresponding reacting ratios are high.

[c] [species_name] (e.g.,[OH]) is the monthly average concentration of oxidants concentration extracted from CHIMERE.

[d] kinetic rate coefficient are provided by MCM, where k_{OH}, k_{O_3}, and k_{NO_3} are the kinetic rate coefficient of first-generation BCARY reaction with OH, O_3, and NO_3, respectively; k_{NO}, k_{HO_2}, and k_{RNO_3} are the simple rate coefficients KRO2NO, KRO2HO2, and KRO2NO3, respectively; k_{RO_2} is self-reaction rate coefficients for the tertiary peroxy radicals (e.g., BCAO2, BCCO2). T: temperature (K).

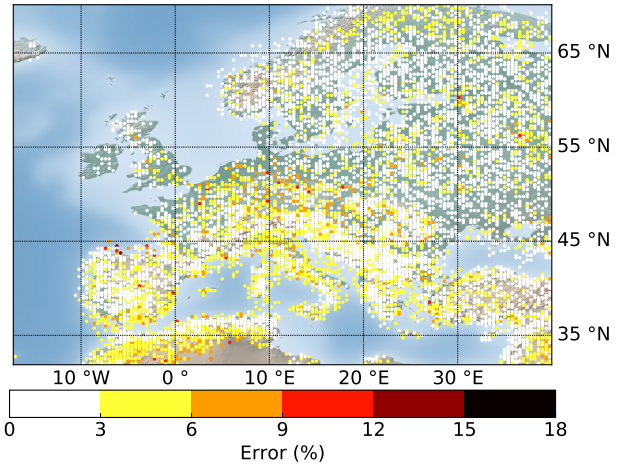

Figure D1. Error map of the testing results of the "Rdc." mechanism on all 12 159 testing conditions.

https://doi.org/10.5194/egusphere-2022-245
Preprint. Discussion started: 10 June 2022

Table B1. Species list of the "Rdc." mechanism.

Notice that the species in the reduced case may be different from the MCM species with identical names.

Surrogate[a]	Type[b]	Molecular formula	MW[c]	P^d_{sat}	ΔH^e_{vap}	H[f]	γ^g
BCARY	VOC	$C_{15}H_{24}$	204.4				
NBCO2	Radical	$C_{15}H_{24}NO_5$	298.4				
BCBOO	Radical	$C_{15}H_{24}O_3$	252.3				
mBCALO2	Radical	$C_{14.68}H_{24.08}O_{4.87}$	278.5				
mBCSOZ	VOC	$C_{15}H_{24}O_{2.74}$	248.2				
NBCOOH	LVOC	$C_{15}H_{25}NO_5$	299.4	4.04×10^{-12}	1.19×10^2	7.63×10^6	1.8×10^6
mC141CO2H	LVOC	$C_{14.94}H_{23.85}O_{3.06}$	252.4	4.43×10^{-11}	1.14×10^2	2.53×10^4	4.95×10^7
mBCKSOZ	SVOC	$C_{14}H_{22}O_{3.9}$	252.7	2.02×10^{-8}	9.12×10^1	1.89×10^5	1.45×10^4
mC131O2	Radical	$C_{13.14}H_{21.22}O_{4.14}$	245.5				
mC133CO	ELVOC	$C_{13.42}H_{20.83}O_{4.59}$	255.6	1.62×10^{-12}	1.25×10^2	2.45×10^2	1.4×10^{11}
mC132OOH	ELVOC	$C_{13.97}H_{23.92}O_{4.59}$	279.5	4.13×10^{-14}	1.36×10^2	5.70×10^3	2.36×10^{11}
C131PAN	LVOC	$C_{14}H_{21}NO_7$	315.3	4.39×10^{-11}	1.13×10^2	2.07×10^4	6.09×10^7
mC133O2	Radical	$C_{13}H_{21}O_{5.97}$	272.8				
C133O	Radical	$C_{13}H_{21}O_5$	257.3				

[a] species with "m" are the new surrogates that merged multiple MCM-BCARY species. [b] VOC: stable species without gas-particle partitioning; The volatility classes is defined in Sect. 3. [c] Molar weight ($g \cdot mol^{-1}$).
Only compute for SOA precursors: [d] saturation vapor pressure at 298 K (atm) ; [e] Enthalpy of vaporation ($KJ \cdot mol^{-1}$); [f] Henry's law constant ($m \cdot atm^{-1}$); [g] activity coefficient at infinite dilution in water.

https://doi.org/10.5194/egusphere-2022-245
Preprint. Discussion started: 10 June 2022

Table C1. Reaction list of the "Rdc." mechanism.

No.	Reactions	Kinetic Rate constant[a]
1	BCARY + NO3 -> NBCO2	1.9×10^{-11}
2	BCARY + O3 -> 0.874 BCBOO + 0.111 mBCALO2	1.19×10^{-14}
3	BCARY + OH -> mBCSOZ	1.97×10^{-10}
4	NBCO2 + HO2 -> NBCOOH	$2.837 \times 10^{-13} \times \exp(\frac{1300}{T})$
5	NBCO2 + NO -> mBCSOZ	$2.7 \times 10^{-12} \times \exp(\frac{360}{T})$
6	NBCO2 + NO3 -> mBCSOZ	2.3×10^{-12}
7	BCBOO -> 0.5 mC141CO2H + 0.5 mBCSOZ	$[H_2O] \times 4 \times 10^{-16}$
8	BCBOO -> mBCSOZ	2×10^{2}
9	mBCSOZ + O3 -> 0.915 mBCKSOZ + 0.085 mBCALO2	1.1×10^{-16}
10	mBCSOZ + OH -> 0.92 mBCALO2 + 0.08 mC131CO3	7.6×10^{-11}
11	mC141CO2H + OH -> mC131O2	6.494×10^{-11}
12	mBCALO2 + NO -> 0.505 mBCKSOZ + 0.353 mC131CO3 + 0.099 mC133CO	$6.56 \times 10^{-12} \times \exp(\frac{360}{T})$
13	mBCALO2 -> 0.6 mC131CO3	$[RO_2] \times 1.711 \times 10^{-12}$
14	mBCALO2 + HO2 -> mC132OOH	$1.939 \times 10^{-13} \times \exp(\frac{1300}{T})$
15	mBCKSOZ + OH -> mC131O2	3.28×10^{-11}
16	mC131CO3 + NO -> mC131O2	$6.377 \times 10^{-12} \times \exp(\frac{290}{T})$
17	mC131CO3 + NO2 -> C131PAN	0.8502 × KFPAN
18	mC131O2 + HO2 -> mC132OOH	$2.288 \times 10^{-13} \times \exp(\frac{1300}{T})$
19	mC131O2 + NO -> mC133O2	$2.213 \times 10^{-12} \times \exp(\frac{360}{T})$
20	mC133O2 + HO2 -> mC132OOH	$2.695 \times 10^{-13} \times \exp(\frac{1300}{T})$
21	mC133O2 + NO -> 0.757 C133O + 0.243 mC133CO	$2.61 \times 10^{-12} \times \exp(\frac{360}{T})$
22	C133O -> mC133CO	$[O_2] * 2.5 \times 10^{-14} \times \exp(\frac{-300}{T})$
23	C133O ->	$2.7 \times 10^{14} \times \exp(\frac{-6643}{T})$

[a] $[H_2O]$ concentration of H_2O, [RO2] the total concentration of RO2 species pool,$[O_2]$ concentration of O_2. KFPAN is one of the Complex rate coefficients from the MCM mechanism.

https://doi.org/10.5194/egusphere-2022-245
Preprint. Discussion started: 10 June 2022

Competing interests. The authors declare that no competing interests are present.

Acknowledgements. This work was financially supported by INERIS and DIM QI^2 (Air Quality Research Network on air quality in the Île-de-France region). The authors would like to thank Youngseob Kim from CEREA for his help in using the SSH-aerosol model.

https://doi.org/10.5194/egusphere-2022-245
Preprint. Discussion started: 10 June 2022

References

Aumont, B., Szopa, S., and Madronich, S.: Modelling the evolution of organic carbon during its gas-phase tropospheric oxidation: development of an explicit model based on a self generating approach, Atmos. Chem. Phys., 5, 2497–2517, https://doi.org/10.5194/acp-5-2497-2005, 2005.

Breysse, P. N., Delfino, R. J., Dominici, F., Elder, A. C., Frampton, M. W., Froines, J. R., Geyh, A. S., Godleski, J. J., Gold, D. R., Hopke, P. K., et al.: US EPA particulate matter research centers: summary of research results for 2005–2011, Air Qual. Atmos. Health, 6, 333–355, https://doi.org/10.1007/s11869-012-0181-8, 2013.

Carter, W. P.: Development of the SAPRC-07 chemical mechanism, Atmos. Environ., 44, 5324–5335, https://doi.org/10.1016/j.atmosenv.2010.01.026, 2010.

Chen, Q., Li, Y., McKinney, K., Kuwata, M., and Martin, S.: Particle mass yield from β-caryophyllene ozonolysis, Atmos. Chem. Phys., 12, 3165–3179, https://doi.org/10.5194/acp-12-3165-2012, 2012.

Compernolle, S., Ceulemans, K., and Müller, J.-F.: EVAPORATION: a new vapour pressure estimation methodfor organic molecules including non-additivity and intramolecular interactions, Atmos. Chem. Phys., 11, 9431–9450, https://doi.org/10.5194/acp-11-9431-2011, 2011.

Couvidat, F. and Sartelet, K.: The Secondary Organic Aerosol Processor (SOAP v1. 0) model: a unified model with different ranges of complexity based on the molecular surrogate approach, Geosci. Model Dev, 8, 1111–1138, https://doi.org/10.5194/gmd-8-1111-2015, 2015.

Couvidat, F., Debry, E., Sartelet, K., and Seigneur, C.: A hydrophilic/hydrophobic organic (H2O) aerosol model: Development, evaluation and sensitivity analysis, J. Geophys. Res.-Atmos., 117, https://doi.org/10.1029/2011JD017214, 2012.

Donahue, N. M., Robinson, A. L., Stanier, C. O., and Pandis, S. N.: Coupled Partitioning, Dilution, and Chemical Aging of Semivolatile Organics, Env. Sc. and Tech., 40, 2635–2643, https://doi.org/10.1021/es052297c, 2006.

Fredenslund, A., Jones, R. L., and Prausnitz, J. M.: Group-contribution estimation of activity coefficients in nonideal liquid mixtures, AIChE J., 21, 1086–1099, 1975.

Goliff, W., Stockwell, W., and Lawson, C.: The Regional Atmospheric Chemistry Mechanism, version 2, Atmos. Environ., 68, 174–185, https://doi.org/10.1016/j.atmosenv.2012.11.038, 2013.

Griffin, R. J., Nguyen, K., Dabdub, D., and Seinfeld, J. H.: A Coupled Hydrophobic-Hydrophilic Model for Predicting Secondary Organic Aerosol Formation, J. Atmos. Chem., 44, 171–190, https://doi.org/10.1023/A:1022436813699, 2003.

Hallquist, M., Wenger, J. C., Baltensperger, U., Rudich, Y., Simpson, D., Claeys, M., Dommen, J., Donahue, N. M., George, C., Goldstein, A. H., Hamilton, J. F., Herrmann, H., Hoffmann, T., Iinuma, Y., Jang, M., Jenkin, M. E., Jimenez, J. L., Kiendler-Scharr, A., Maenhaut, W., McFiggans, G., Mentel, T. F., Monod, A., Prévôt, A. S. H., Seinfeld, J. H., Surratt, J. D., Szmigielski, R., and Wildt, J.: The formation, properties and impact of secondary organic aerosol: current and emerging issues, Atmos. Chem. Phys., 9, 5155–5236, https://doi.org/10.5194/acp-9-5155-2009, 2009.

Huang, X., Ding, A., Gao, J., Zheng, B., Zhou, D., Qi, X., Tang, R., Wang, J., Ren, C., Nie, W., Chi, X., Xu, Z., Chen, L., Li, Y., Che, F., Pang, N., Wang, H., Tong, D., Qin, W., Cheng, W., Liu, W., Fu, Q., Liu, B., Chai, F., Davis, S. J., Zhang, Q., and He, K.: Enhanced secondary pollution offset reduction of primary emissions during COVID-19 lockdown in China, Natl. Sci. Rev., 8, https://doi.org/10.1093/nsr/nwaa137, 2020.

https://doi.org/10.5194/egusphere-2022-245
Preprint. Discussion started: 10 June 2022

Jenkin, M., Watson, L., Utembe, S., and Shallcross, D.: A Common Representative Intermediates (CRI) mechanism for VOC degradation. Part 1: Gas phase mechanism development, Atmos. Environ., 42, 7185–7195, https://doi.org/10.1016/j.atmosenv.2008.07.028, 2008.

Jenkin, M., Wyche, K., Evans, C., Carr, T., Monks, P., Alfarra, M., Barley, M., McFiggans, G., Young, J., and Rickard, A.: Development and chamber evaluation of the MCM v3. 2 degradation scheme for β-caryophyllene, Atmos. Chem. Phys., 12, 5275–5308, https://doi.org/10.5194/acp-12-5275-2012, 2012.

Jenkin, M. E., Saunders, S. M., and Pilling, M. J.: The tropospheric degradation of volatile organic compounds: a protocol for mechanism development, Atmos. Environ., 31, 81–104, https://doi.org/10.1016/S1352-2310(96)00105-7, 1997.

Joback, K. G. and Reid, R. C.: Estimation of pure-component properties from group-contributions, Chem. Eng. Comm., 57, 233–243, https://doi.org/10.1080/00986448708960487, 1987.

Kanakidou, M., Seinfeld, J., Pandis, S., Barnes, I., Dentener, F. J., Facchini, M. C., Dingenen, R. V., Ervens, B., Nenes, A., Nielsen, C., et al.: Organic aerosol and global climate modelling: a review, Atmos. Chem. Phys., 5, 1053–1123, https://doi.org/10.5194/acp-5-1053-2005, 2005.

Khan, M., Jenkin, M., Foulds, A., Derwent, R., Percival, C., and Shallcross, D.: A modeling study of secondary organic aerosol formation from sesquiterpenes using the STOCHEM global chemistry and transport model, J. Geophys. Res.-Atmos., 122, 4426–4439, https://doi.org/10.1002/2016JD026415, 2017.

Kim, Y., Sartelet, K., and Couvidat, F.: Modeling the effect of non-ideality, dynamic mass transfer and viscosity on SOA formation in a 3-D air quality model, Atmos. Chem. Phys., 19, 1241–1261, https://doi.org/10.5194/acp-19-1241-2019, 2019.

Lannuque, V., Camredon, M., Couvidat, F., Hodzic, A., Valorso, R., Madronich, S., Bessagnet, B., and Aumont, B.: Exploration of the influence of environmental conditions on secondary organic aerosol formation and organic species properties using explicit simulations: development of the VBS-GECKO parameterization, Atmos. Chem. Phys., 18, 13 411–13 428, https://doi.org/10.5194/acp-18-13411-2018, 2018.

Lanzafame, G. M., Bessagnet, B., Srivastava, D., Jaffrezo, J. L., Favez, O., Albinet, A., and Couvidat, F.: Modelling aerosol molecular markers in a 3D air quality model: Focus on anthropogenic organic markers, Sci. Total Environ., p. 155360, https://doi.org/https://doi.org/10.1016/j.scitotenv.2022.155360, 2022.

Li, J., Cleveland, M., Ziemba, L. D., Griffin, R. J., Barsanti, K. C., Pankow, J. F., and Ying, Q.: Modeling regional secondary organic aerosol using the Master Chemical Mechanism, Atmos. Environ., 102, 52–61, https://doi.org/10.1016/j.atmosenv.2014.11.054, 2015.

McNeill, V. F.: Atmospheric Aerosols: Clouds, Chemistry, and Climate, Annual Rev. Chem. Biomol. Eng., 8, 427–444, https://doi.org/10.1146/annurev-chembioeng-060816-101538, 2017.

Myrdal, P. B. and Yalkowsky, S. H.: Estimating Pure Component Vapor Pressures of Complex Organic Molecules, Ind. Eng. Chem. Res., 36, 2494–2499, https://doi.org/10.1021/ie950242l, 1997.

Nannoolal, Y., Rarey, J., Ramjugernath, D., and Cordes, W.: Estimation of pure component properties: Part 1. Estimation of the normal boiling point of non-electrolyte organic compounds via group contributions and group interactions, Fluid Phase Equilibr., 226, 45–63, https://doi.org/10.1016/j.fluid.2004.09.001, 2004.

Nannoolal, Y., Rarey, J., and Ramjugernath, D.: Estimation of pure component properties: Part 3. Estimation of the vapor pressure of non-electrolyte organic compounds via group contributions and group interactions, Fluid Phase Equilibr., 269, 117–133, https://doi.org/10.1016/j.fluid.2008.04.020, 2008.

Odum, J. R., Hoffmann, T., Bowman, F., Collins, D., Flagan, R. C., and Seinfeld, J. H.: Gas/Particle Partitioning and Secondary Organic Aerosol Yields, Env. Sc. and Tech., 30, 2580–2585, https://doi.org/10.1021/es950943+, 1996.

https://doi.org/10.5194/egusphere-2022-245
Preprint. Discussion started: 10 June 2022

Pankow, J. F. and Asher, W. E.: SIMPOL.1: a simple group contribution method for predicting vapor pressures and enthalpies of vaporization of multifunctional organic compounds, Atmos. Chem. Phys., 8, 2773–2796, https://doi.org/10.5194/acp-8-2773-2008, 2008.

Pun, B. K., Seigneur, C., and Lohman, K.: Modeling Secondary Organic Aerosol Formation via Multiphase Partitioning with Molecular Data, Env. Sc. and Tech., 40, 4722–4731, https://doi.org/10.1021/es0522736, 2006.

Sartelet, K., Couvidat, F., Wang, Z., Flageul, C., and Kim, Y.: SSH-Aerosol v1. 1: A Modular Box Model to Simulate the Evolution of Primary and Secondary Aerosols, Atmosphere, 11, 525, https://doi.org/10.3390/atmos11050525, 2020.

Sarwar, G., Luecken, D., Yarwood, G., Whitten, G. Z., and Carter, W. P. L.: Impact of an Updated Carbon Bond Mechanism on Predictions from the CMAQ Modeling System: Preliminary Assessment, J. Applied Meteor., 47, 3 – 14, https://doi.org/10.1175/2007JAMC1393.1, 2008.

Saunders, S. M., Jenkin, M. E., Derwent, R., and Pilling, M.: Protocol for the development of the Master Chemical Mechanism, MCM v3 (Part A): tropospheric degradation of non-aromatic volatile organic compounds, Atmos. Chem. Phys., 3, 161–180, https://doi.org/10.5194/acp-3-161-2003, 2003.

Seinfeld, J. H. and Pandis, S. N.: Atmospheric chemistry and physics: from air pollution to climate change, John Wiley & Sons, 2016.

Seinfeld, J. H., Bretherton, C., Carslaw, K. S., Coe, H., DeMott, P. J., Dunlea, E. J., Feingold, G., Ghan, S., Guenther, A. B., Kahn, R., et al.: Improving our fundamental understanding of the role of aerosol-cloud interactions in the climate system, Proc. Nat. Acad. Sci., 113, 5781–5790, https://doi.org/10.1073/pnas.1514043113, 2016.

Stein, S. E. and Brown, R. L.: Estimation of normal boiling points from group contributions, J. Chem. Inf. Comput. Sci., 34, 581–587, https://doi.org/10.1021/ci00019a016, 1994.

Szopa, S., Aumont, B., and Madronich, S.: Assessment of the reduction methods used to develop chemical schemes: building of a new chemical scheme for VOC oxidation suited to three-dimensional multiscale HO_x-NO_x-VOC chemistry simulations, Atmos. Chem. Phys., 5, 2519–2538, https://doi.org/10.5194/acp-5-2519-2005, 2005.

Tasoglou, A. and Pandis, S. N.: Formation and chemical aging of secondary organic aerosol during the β-caryophyllene oxidation, Atmos. Chem. Phys., 15, 6035–6046, https://doi.org/10.5194/acp-15-6035-2015, 2015.

Topping, D., Barley, M., Bane, M. K., Higham, N., Aumont, B., Dingle, N., and McFiggans, G.: UManSysProp v1.0: an online and open-source facility for molecular property prediction and atmospheric aerosol calculations, Geosci. Model Dev, 9, 899–914, https://doi.org/10.5194/gmd-9-899-2016, 2016.

Watson, L., Shallcross, D., Utembe, S., and Jenkin, M.: A Common Representative Intermediates (CRI) mechanism for VOC degradation. Part 2: Gas phase mechanism reduction, Atmos. Environ., 42, 7196–7204, https://doi.org/10.1016/j.atmosenv.2008.07.034, 2008.

Xavier, C., Rusanen, A., Zhou, P., Dean, C., Pichelstorfer, L., Roldin, P., and Boy, M.: Aerosol mass yields of selected biogenic volatile organic compounds–a theoretical study with nearly explicit gas-phase chemistry, Atmos. Chem. Phys., 19, 13 741–13 758, https://doi.org/10.5194/acp-19-13741-2019, 2019.

Ying, Q. and Li, J.: Implementation and initial application of the near-explicit Master Chemical Mechanism in the 3D Community Multiscale Air Quality (CMAQ) model, Atmos. Environ., 45, 3244–3256, https://doi.org/10.1016/j.atmosenv.2011.03.043, 2011.

Yuan, J., Ling, Z., Wang, Z., Lu, X., Fan, S., He, Z., Guo, H., Wang, X., and Wang, N.: PAN–Precursor Relationship and Process Analysis of PAN Variations in the Pearl River Delta Region, Atmosphere, 9, https://doi.org/10.3390/atmos9100372, 2018.

Zuend, A., Marcolli, C., Luo, B., and Peter, T.: A thermodynamic model of mixed organic-inorganic aerosols to predict activity coefficients, Atmos. Chem. Phys., 8, 4559–4593, https://doi.org/10.5194/acp-8-4559-2008, 2008.