Implementation of a parallel reduction algorithm in the GENerator of reduced Organic Aerosol mechanisms (GENOA v2.0): Application to multiple monoterpene aerosol precursors

Zhizhao Wang, Florian Couvidat, Karine Sartelet

Journal of Aerosol Science 174 (2023) 106248

Contents lists available at ScienceDirect

Journal of Aerosol Science

journal homepage: www.elsevier.com/locate/jaerosci

Implementation of a parallel reduction algorithm in the GENerator of reduced Organic Aerosol mechanisms (GENOA v2.0): Application to multiple monoterpene aerosol precursors

Zhizhao Wang [a,b,*], Florian Couvidat [b], Karine Sartelet [a]

[a] Centre d'Enseignement et de Recherche en Environnement Atmosphérique (CEREA), Ecole des Ponts ParisTech, EdF R&D, IPSL, 6-8 avenue Blaise Pascal, Cité Descartes, Champs-sur-Marne, 77455, Marne-la-Vallée, France

[b] Institut National de l'Environnement Industriel et des Risques (INERIS), Parc Technologique ALATA, BP 2, Verneuil-en-Halatte, 60550, Oise, France

ARTICLE INFO

Editor: Dr. Chris Hogan

Dataset link: https://doi.org/10.5281/zenodo.8187593

Keywords:
Aerosol mechanism
Mechanism reduction algorithm
Secondary organic aerosol
Air quality modeling

ABSTRACT

Explicit gas-phase chemical mechanisms represent the state of knowledge regarding the chemistry of volatile organic compounds (VOCs), which are crucial in the formation of secondary organic aerosols (SOAs). However, these chemical mechanisms are computationally expensive, which limits their practical use in large-scale air quality modeling. Mechanism reduction is therefore required for computational efficiency while preserving the accuracy of the detailed gas-phase chemical mechanisms.

This paper presents a new version of the Generator of Reduced Organic Aerosol Mechanisms (GENOA v2.0), which reduces mechanisms at a size suitable for three-dimensional (3-D) modeling while preserving the accuracy of detailed chemical mechanisms for simulating aerosol concentrations. GENOA v2.0 adopts a parallel reduction framework to identify the most optimal reductions from competitive candidates, and can reduce chemical mechanisms from multiple aerosol precursors. To demonstrate the reduction efficiency, GENOA v2.0 is applied to the reduction of monoterpene chemistry from the Master Chemical Mechanism (MCM) combined with the Peroxy Radical Autoxidation Mechanism (PRAM) mechanism. The original mechanism, consisting of 3 001 reactions and 1 227 species (including 738 condensable species), is reduced by 93% to 197 reactions and 110 species (including 23 condensable species), inducing an average error of only 3% in aerosol concentrations. Sensitivity tests showed that this reduced mechanism behaved similarly to the original mechanism in response to changes in environmental conditions such as temperature, relative humidity, and SOA mass loading. Moreover, if the error tolerance is increased to 20% — which can still be acceptable for 3-D air quality modeling — the mechanism can be further simplified to 40 reactions and 24 species (including 5 condensable species). Consequently, the GENOA-generated aerosol mechanism preserves the complexity of the detailed gas-phase chemical mechanisms on SOA formation while increasing computational efficiency, which makes it suitable for most environmental conditions encountered in the atmosphere.

* Corresponding author at: Centre d'Enseignement et de Recherche en Environnement Atmosphérique (CEREA), Ecole des Ponts ParisTech, EdF R&D, IPSL, 6-8 avenue Blaise Pascal, Cité Descartes, Champs-sur-Marne, 77455, Marne-la-Vallée, France.
E-mail address: zhizhao.wang@enpc.fr (Z. Wang).

https://doi.org/10.1016/j.jaerosci.2023.106248
Received 14 April 2023; Received in revised form 26 July 2023; Accepted 9 August 2023
Available online 25 August 2023

1. Introduction

Organic aerosols affect air quality, climate change, and human health (Hallquist et al., 2009; Kanakidou et al., 2005). Some organic aerosols are emitted in the atmosphere as primary aerosols, while the majority are secondary organic aerosols (SOA) formed in the atmosphere by the oxidation of volatile organic compounds (VOCs) (Atkinson & Arey, 2003; Gelencsér et al., 2007). VOC global emissions are about 1 300 tons of carbon dioxide per year (Tg C/yr) (Goldstein & Galbally, 2007). A large amount of VOCs (approximately 1 000 Tg C/yr) originates from biogenic emissions (about 50% from isoprene, 15% from monoterpenes, and 3% from sesquiterpenes estimated by Guenther et al. (2012)), while the remaining is derived from anthropogenic sources and biomass burning. When exposed to atmospheric oxidants (e.g., ozone, OH radical, and NO_3 radical), VOCs may undergo multi-generational oxidation affected by the chemical regime of the atmosphere (Fry et al., 2009; Han & Jang, 2023). The low-volatility oxidation products from VOC degradation may further condense on existing particles, forming SOA (Hallquist et al., 2009). Depending on the NO_x concentrations, the chemical regime can be divided into high- and low-NO_x regimes, which may favor different SOA formation pathways (Atkinson, 2000). Consequently, aerosol formation is highly influenced by VOCs, atmospheric oxidants, NO_x concentrations, as well as other environmental parameters (e.g., temperature) that affect gas-phase chemistry and gas-particle partitioning (Porter, Jimenez, & Barsanti, 2021).

Due to the complexity of the phenomena involved in SOA formation, chemical transport models (CTMs) typically adopt simplified representations of organic aerosols. This is achieved by utilizing implicit mechanisms, such as the two-product empirical approach of Odum et al. (1996), the molecular surrogate approach (e.g., Couvidat, Debry, Sartelet, & Seigneur, 2012; Griffin, Nguyen, Dabdub, & Seinfeld, 2003; Pun, Seigneur, & Lohman, 2006) or the volatility basis set (VBS) approach (Donahue, Robinson, Stanier, & Pandis, 2006; Stolzenburg, Wang, Schervish, & Donahue, 2022). In the molecular surrogate approach, SOA formation is represented by only a few model species for each major SOA precursor, whereas in the VBS approaches, organic compounds are grouped according to volatility or, in the case of the VBS two-dimensional approach, based on volatility and the oxygen content (e.g., Donahue, Epstein, Pandis, & Robinson, 2011). However, the actual gas-phase chemistry governing SOA formation is highly complex and involves numerous reactions and organics interacting with one another. Hence, CTM models struggle to capture the non-linear interactions between gas-phase chemistry and SOA formation with current SOA mechanisms (Shrivastava et al., 2017).

Many theoretical and experimental studies have been conducted to understand VOC degradation and its influence on atmospheric oxidants and other pollutants, including aerosols (e.g., Chen et al., 2022; Ehn et al., 2014). Accordingly, the development of detailed state-of-the-art gas-phase chemical mechanisms, also known as explicit VOC mechanisms, has progressed rapidly based on our current knowledge of VOC chemistry (Stockwell, Saunders, Goliff, & Fitzgerald, 2020). Explicit mechanisms are either written manually, such as the Master Chemical Mechanism (MCM, Jenkin, Saunders, & Pilling, 1997; Saunders, Jenkin, Derwent, & Pilling, 2003), or generated automatically based on predefined protocols, such as the Generator for Explicit Chemistry and Kinetics of Organics in the Atmosphere (GECKO-A, Aumont, Szopa, & Madronich, 2005; Camredon, Aumont, Lee-Taylor, & Madronich, 2007). As our understanding of VOC chemistry advances, the development of explicit mechanisms is still ongoing (e.g., Coggon et al., 2019; Jenkin, Valorso, Aumont, Newland, & Rickard, 2020; Jenkin et al., 2012; Newland et al., 2022). For example, as a follow-up to the latest research on Highly Oxygenated organic Molecules (HOMs) chemistry and its significant contribution to SOA formation (e.g., Bianchi et al., 2019; Ehn et al., 2014; McFiggans et al., 2019), Roldin et al. (2019) has developed the Peroxy Radical Autoxidation Mechanism (PRAM). Designed as a complement to MCM, PRAM contains reactions of autoxidation and dimerization, leading to the formation of extremely low-volatility organic compounds (ELVOCs) from monoterpenes that are missing in MCM.

Several studies have demonstrated the reliability of explicit chemical mechanisms to simulate major atmospheric oxidant and pollutant concentrations (e.g., Li et al., 2015; Li, Jiang, Afreh, Barsanti, & Cocker III, 2022; Mouchel-Vallon et al., 2020; Ying & Li, 2011). However, the direct use of explicit mechanisms in three-dimensional (3-D) air quality models such as regional CTMs is impractical from a numerical perspective. As explicit mechanisms contain thousands of gas-phase species with varying lifetimes, simulating the evolution of all those species is computationally expensive. Furthermore, to accurately simulate SOA formation, the gas-particle partitioning needs to be resolved for each condensable species (i.e., species that can condense into particles) and particle size bin. This process involves not only aerosol dynamics but also other atmospheric processes (e.g., transport and deposition), resulting in an overwhelming computational expense for aerosol modeling with explicit mechanisms.

The development of mechanisms reduction algorithms is therefore essential (Kaduwela, Luecken, Carter, & Derwent, 2015) to reduce explicit VOC mechanisms to sizes suitable for use in CTMs while still ensuring reliable simulations of pollutant concentrations. Several methods to reduce explicit VOC mechanisms have been investigated alongside the development of explicit mechanisms (e.g., Szopa, Aumont, & Madronich, 2005; Whitehouse, Tomlin, & Pilling, 2004a, 2004b; Xia, Michelangeli, & Makar, 2009). Regardless of the varying effectiveness of these reduction methods, reduced VOC mechanisms have been developed for simulating ozone concentrations, semi-volatile organic compounds, and other secondary gas-phase compounds (e.g., hydroperoxyl, formaldehyde). Notable examples include the Common Representative Intermediates (CRI) mechanisms (Jenkin, Watson, Utembe, & Shallcross, 2008; Watson, Shallcross, Utembe, & Jenkin, 2008; Weber et al., 2020) and recently AMORE mechanism (Wiser et al., 2023). Additionally, emerging data-driven approaches have been developed to enhance the computational efficiency of key processes involved in 3-D CTM modeling, including chemical integration (Kelp, Jacob, Lin, & Sulprizio, 2022; Shen et al., 2022) and transport (Sturm et al., 2023).

Particularly addressing aerosol formation, Lannuque et al. (2018) developed VBS-GECKO (a volatility basis set-type parameterization based on the GECKO-A mechanism) and integrated it into the 3-D CTM model CHIMERE (Lannuque et al., 2018). Recent investigations have also been underway to explore the application of machine learning approaches in training SOA parameterizations from GECKO-A organic aerosol chemistry (e.g., Mouchel-Vallon & Hodzic, 2022; Schreck et al., 2022). Furthermore, Wang, Couvidat,

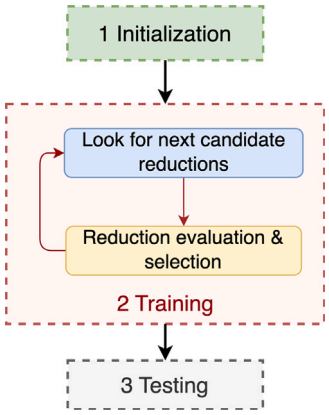

Fig. 1. Schematic diagram showing the main processes of the GENOA algorithm.

& Sartelet, 2022 developed the GENerator of Reduced Organic Aerosol Mechanisms (GENOA v1.0), which generates condensed SOA mechanisms from MCM following a predefined reduction protocol. Compared to VBS-GECKO and neural network training, the GENOA algorithm has the advantage of preserving the major chemical pathways related to SOA formation, as well as SOA properties. Although GENOA v1.0 was capable of reducing SOA mechanisms from the oxidation of sesquiterpenes, it could not efficiently reduce mechanisms from multiple SOA precursors simultaneously. To address this issue, a new version of GENOA, GENOA v2.0, has been developed, focusing on enhancing reduction efficiency.

This paper presents the second version of GENOA (GENOA v2.0), which enables the simultaneous reduction of mechanisms from several SOA precursors with a parallel reduction framework. The GENOA v2.0 algorithm and its configuration used to reduce monoterpene SOA mechanisms are described in Section 2. The resulting reduced monoterpene SOA mechanism is presented and discussed in Section 3, along with a sensitivity analysis. Finally, the conclusion is drawn in Section 4.

2. Methods

This section begins with an overview of GENOA v1.0 in Section 2.1, along with a discussion of its limitations in Section 2.2. The major updates of GENOA v2.0 are summarized in Section 2.3, followed by the configuration setups for its application to monoterpene mechanism reduction in Section 2.4. For simplicity, we refer to GENOA without mentioning its version number when describing the common features shared by v1.0 and v2.0.

2.1. Presentation of GENOA v1.0

GENOA is a reduction algorithm that generates concise SOA mechanisms based on detailed VOC mechanisms under representative atmospheric conditions. Since a detailed description of the first version of the GENOA algorithm (GENOA v1.0) is available in Wang et al. (2022), only the main concepts are reviewed here.

As illustrated in Fig. 1, three major processes are involved in GENOA:

- Initialization — the process of initializing reduction parameters and performing the prereduction. In practice, some minor reductions in the explicit mechanism may be necessary to ensure numerical stability (see Section 2.4.1). The mechanism after this prereduction is referred to as the reference mechanism and serves as a benchmark and a starting point for the main reduction.
- Training — the main process of reduction. During training, reductions are performed in loops (referred to hereafter as reduction cycles) over species and reactions. A reduction cycle consists of successive reduction steps, each step representing one evaluation of a set of candidate reductions. At each step, GENOA searches for candidate reductions based on targeted species or reactions using the predefined protocol and evaluates them under selected near-realistic atmospheric conditions. Candidate reductions that meet the evaluation criteria are referred to as "approved reductions". These approved reductions are accepted in the reduced mechanism and serve as the basis for the next reduction step. When no candidate reduction meets the criteria, the next step is to search for new candidates based on other species or reactions. The search-evaluation process continues until all species and reactions have been explored and no further reductions can be approved. Afterward, the training may be stopped or moved to a new cycle.

  In GENOA v1.0, each reduction cycle uses a single reduction strategy, which means that only one candidate reduction is evaluated for each reduction step. The evaluation is conducted by comparing an error indicating the differences in total SOA concentrations due to the candidate reduction against predefined user-chosen error tolerances.

- Testing — the process of evaluating the performance of the final reduced SOA mechanism under various conditions, encompassing all potential situations in which the mechanism may be employed. Meanwhile, as the entire training process is traceable, all mechanisms obtained during training can be evaluated in the testing process.

In GENOA, SOA concentration and composition are simulated using the box model SSH-aerosol v1.3 (Sartelet, Couvidat, Wang, Flageul, & Kim, 2020). After the training and testing processes, the reduced SOA mechanisms can provide an accurate simulation of SOA concentration and composition compared to explicit mechanisms, within a specified tolerance and under the tested environmental conditions. As the main purpose of the GENOA algorithm is to preserve reliable SOA concentrations from specific SOA precursors, the influences of VOC mechanisms on other compounds (i.e., O_3, OH, NO, NO_3, HO_2, SO_2, CO, inorganic aerosols) are currently not tracked. Their concentrations are based on the diurnal profiles derived from 3-D CTM simulations and remain unchanged during the reduction. Therefore, for 3-D applications, it is necessary to combine the reduced SOA mechanism with implicit gas-phase mechanisms (e.g., CB05 Sarwar, Luecken, Yarwood, Whitten, & Carter, 2008 and RACM2 Goliff, Stockwell, & Lawson, 2013) for accurate estimation of inorganic concentrations.

GENOA is also a user-parameterized algorithm, allowing users to design their own reductions by adjusting reduction parameters and options. One example of customization is the selection of evaluation datasets. An evaluation dataset is a set of atmospheric conditions for evaluating candidate reductions (see Section 2.4.2). By selecting the appropriate evaluation dataset, users can train mechanisms for specific purposes, such as focusing on SOA modeling in urban areas. Mechanisms trained under specific conditions are expected to be more compact than those trained under general atmospheric conditions with the same level of accuracy. To facilitate customization, the training process can be divided into several sequential stages, referred to as training stages. For each stage, the users can select a set of appropriate reduction parameters and options for all reduction cycles inside the stage.

GENOA v1.0 was applied to the degradation scheme of sesquiterpene of MCM v3.3.1 (Jenkin et al., 2012), resulting in a reduction from 1 625 reactions and 579 species to 23 reactions and 15 species (2% of the original size) with less than 3% error on average (Wang et al., 2022).

2.2. Limitations

Two main limitations of GENOA v1.0 are identified, which may prevent its application to more complex VOC mechanisms.

The primary limitation of GENOA v1.0 is its incapability to evaluate competitive reductions (i.e., to evaluate the best reduction among several possibilities). At each step of the reduction cycle, GENOA v1.0 examines one candidate reduction with the same reduction strategy (one strategy per cycle). Due to the lack of exploration of other options, this reduction may not be the optimal choice. As the reduction is performed in series (Fig. 2(a)), the competitive reductions acting on the same reaction/species with other reduction strategies cannot be investigated in GENOA v1.0. Since other possible reductions have not been assessed, the approved reduction may not be optimal, even though it satisfies the evaluation criteria. Due to this single find-select approach, the reduction choice in GENOA v1.0 is strongly influenced by the order in which the reduction strategies are employed. Each reduction strategy may also require a specific search order based on its features. Therefore, reduction outcomes may vary significantly depending on the order in which reduction strategy and searching are applied, requiring investigation for each application of GENOA v1.0.

The second limitation of GENOA v1.0 is its inability to reduce mechanisms with multiple SOA precursors simultaneously, whereas some precursors may share common or similar chemical pathways. As presented in Wang et al. (2022), only one initial SOA precursor (i.e., beta-caryophyllene) is considered in the sesquiterpene mechanism adopted in GENOA v1.0. The reduction was performed with a single initial concentration of sesquiterpene of 10 $\mu g/m^3$ for all conditions. Despite training with one precursor initial concentration condition, sensitivity tests showed that the reduced mechanism was efficient when initial SOA precursor mass loading varied from 10^{-3} $\mu g/m^3$ to 1 000 $\mu g/m^3$. However, when dealing with mechanisms involving multiple SOA precursors, a single set of initial concentrations is not sufficient. The algorithm needs, therefore, to account for different initial SOA precursor sets when reducing mechanisms of several SOA precursors simultaneously.

For the above reasons, the use of GENOA v1.0 may be problematic when applied to more complex and extensive VOC mechanisms involving multiple SOA precursors. For example, for the reduction of the SOA mechanism from a single monoterpene precursor, α-pinene, the reduced mechanism generated by GENOA v1.0 has less reduction and larger errors (size of the mechanism reduced by 86% with an error of 4%) compared to the reduced sesquiterpene mechanisms (size reduced by 98% with an error of 3%). With such a reduction, the final reduced SOA mechanism for all monoterpenes would probably contain more than 100 condensable species, which may not be acceptable for regional-scale CTM modeling.

Two possible reasons can explain the suboptimal performance of the monoterpene mechanism reduction with GENOA v1.0. One reason is related to the properties of products formed by the oxidation of monoterpenes. As monoterpene oxidation products are dominated by semi-volatile species that tend to exist in both gas and particle phases, its mechanism is more difficult to reduce than the mechanism of sesquiterpene, whose oxidation products are mainly low-volatile and preferably remain in one phase (gas or particle). Another reason may be due to bias introduced by the order of reduction strategies and searching, which are arbitrarily defined in GENOA v1.0, as previously detailed.

2.3. New features in GENOA v2.0

Based on GENOA v1.0, the second version of GENOA (GENOA v2.0) has been developed for the reduction of detailed gas-phase mechanisms of multiple SOA precursors. The training process in GENOA v2.0 has been improved in several ways. A novel parallel reduction structure is adopted in GENOA v2.0 to locate and evaluate multiple competitive reductions with all reduction strategies at each reduction step, as described in Section 2.3.1. Along with the parallel structure, the reduction search order (Section 2.3.2), the reduction strategy (Section 2.3.3), and the evaluation method (Section 2.3.4) are also updated in GENOA v2.0.

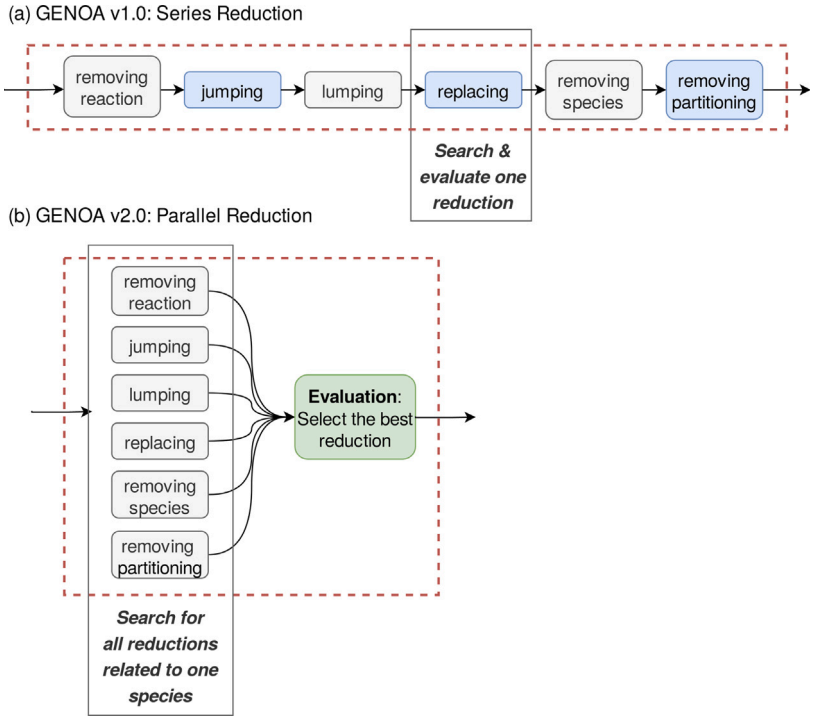

Fig. 2. Schematic diagram that shows the reduction frameworks in series (a) and in parallel (b) used for GENOA v1.0 and v2.0, respectively.

2.3.1. Parallel reduction

Fig. 2 illustrates the series reduction adopted in GENOA v1.0 and the parallel reduction in GENOA v2.0. In contrast to the series reduction, which tests one candidate reduction per step and accepts it if the accuracy criteria are met, GENOA v2.0 investigates multiple candidate reductions simultaneously and selects the one with the highest reduction score (see Section 2.3.4). With this parallel reduction framework, GENOA v2.0 searches all competing candidate reductions related to the targeted species or reactions using all reduction strategies at each reduction step. Afterward, the candidate reductions are examined under the applied evaluation dataset and different sets of initial SOA precursor concentrations. The best reduction is then selected as the starting point for the next reduction.

Due to the large number of candidate reductions that need to be examined, parallel reductions can incur considerable computational costs. To improve computational efficiency, parallelization is implemented in the code. This allows for the simultaneous investigations of candidate reductions by multiple processors per reduction step, thereby maximizing the utilization of the processing power available on a given machine.

2.3.2. Reduction search order

The search for candidate reductions is based on the reduction search order. As accepted reductions affect subsequent reductions, the reduction search order directly affects the mechanism reduction. The search orders of GENOA v1.0 are determined based on the reduction strategy. For example, a reduction cycle using the removal of reaction strategy searches for candidate reductions based on the reaction list, from those involved in higher-generation oxidation to those of lower-generation. Meanwhile, a reduction cycle via removing species is from the species with the smallest molar mass to the species with the largest molar mass.

To align with the parallel reduction approach, GENOA v2.0 features a revised search order that is based on the impact of species on SOA concentrations. The search prioritizes species that have minimal influence on SOA formation and proceeds to more influential species. Prior to training, all organic species are ranked according to their contribution to SOA formation, based on the difference in total SOA concentration caused by their removal. Upon accepting one reduction, this species list is adjusted accordingly. For example, if one reduction merges species, those species are replaced by the new surrogate, located in the list at the place of the merged species that had the greatest influence.

For each species, the candidate reductions include all potential reductions associated with that species, as well as reactions in which the target species is a product (or a reactant if there is no product in that reaction). An example of the searching related to a species with all reduction strategies is presented in Section 2.3.3. Once all potential reductions for the current species have been examined, the search moves on to the next species on the list. This new species-based search order suits the parallel reduction and requires initializing the species list only once. Furthermore, it is more efficient than the reaction-based search order in which all strategies related to one reaction are investigated at one reduction step.

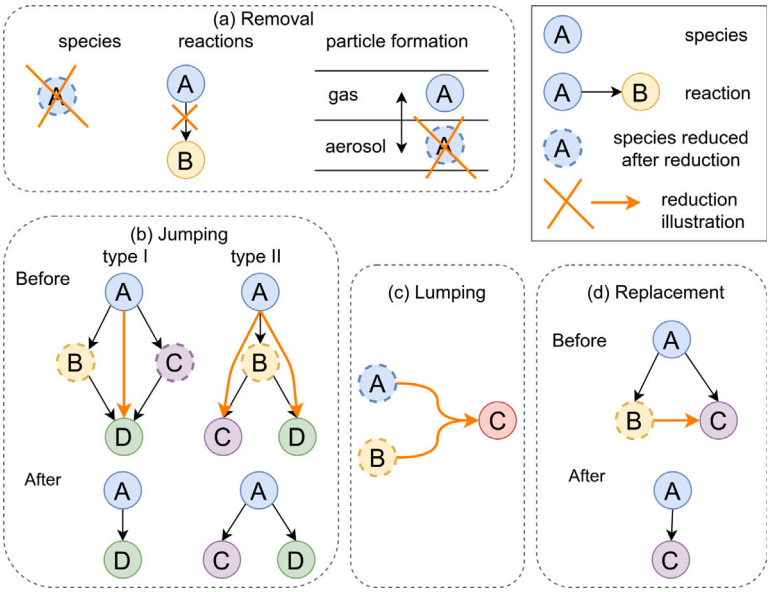

Fig. 3. Schematic diagram illustrating the different reduction strategies with examples of candidate reductions related to species "A", including (a) removal (of species, reactions, and gas-particle partitioning), (b) jumping (types I and II), (c) lumping, and (d) replacement.

2.3.3. Reduction strategy

Reduction strategies are protocols used to lead the reduction searching. Four basic reduction strategies are available in GENOA v2.0: removal, jumping, lumping, and replacement. As shown in Fig. 3, each strategy offers several candidates for reductions related to the species "A":

- Removal — an element of the mechanism (a species, a reaction, or the formation of particle species) is removed from the current mechanism. A reduction via removal (Fig. 3(a)) can either remove a species (e.g., species "A"), one or several reactions related to that species (e.g., the reaction that "A" forms "B"), or the gas-particle partitioning of the species (e.g., condensation of species "A").
- Jumping — a species is jumped over and replaced by its oxidation products. Jumping can happen when an intermediate is formed but can be replaced directly by its oxidation products without affecting the performance of the mechanism. As an example shown in Fig. 3(b), when a species "A" is destroyed and a species "D" is formed with intermediate species "B" and "C", jumping can combine several reactions involving the formation of "D" from "A" (i.e., reactions of "A" to "B", "A" to "C", "B" to "D", and "C" to "D") into one reaction (i.e., the reaction of "A" to "D"), thereby removing middle reactions. The jumping strategy can be separated into two types. The first-type jumping (type I in Fig. 3(a)) removes more intermediates (species "B" and "C") than products (species "D"), while the second type of jumping (type II in Fig. 3(b)) involves more products than intermediates.
- Lumping — a species is merged with other species to form a new surrogate. In the reduction via lumping illustrated in Fig. 3(c), species "A" and "B" merge into a new surrogate "C", where reactions with species "A" and "C" are combined, and rewritten to reactions with only species "C".
- Replacement — a species is replaced by another species with similar properties. A reduction via replacement (Fig. 3(d)) can replace a species "B" with a species "C" if both of them are formed from the same species "A". In this case, the species "B" is eliminated from the scheme, and all its reactions are rewritten as those of the species "C".

Overall, GENOA v2.0 adopts similar reduction strategies as GENOA v1.0, except for jumping strategies, which are revised based on the relationships between species involved in successive reactions. In GENOA v1.0, jumping can only skip over one species if it directly forms another species. Therefore, one species that forms several species cannot be reduced via jumping, even if its products are highly unstable and naturally appropriate for jumping. This is now achievable with the two-type jumping strategy in GENOA v2.0, where the number of jumped species is no longer limited.

In addition to the restrictions for the sesquiterpene reduction of GENOA v1.0 (detailed in Wang et al., 2022), more stringent restrictions are used for the monoterpene mechanism reduction of GENOA v2.0. Lumping is now restricted according to the saturation vapor pressure (P_{sat}) and oxygen atom numbers (indicating the oxidation state) of the species. Species that differ in P_{sat} by more than a factor of 100 or in oxygen atoms by more than three cannot be merged by lumping. As for jumping and replacement, the relative difference in molar masses is restricted and should not exceed 50%.

Table 1
Size of the different detailed monoterpene chemical mechanisms in terms of the numbers of reactions, species, and condensables[a].

Mechanism	Reaction	Species	Condensable
MCM - API	894	293	171
MCM - BPI	1 190	382	248
MCM - LIM	1 576	519	433
MCM - MT	3 454	1 143	738
Full PRAM	1 773	307	176
Reduced PRAM	192	107	69
MCM + PRAM	3 001	1 227	738
Ref. mechanism	2 608	975	643

[a]Mechanisms with elementary reactions from top to bottom are API, BPI, LIM, and MT degradation schemes in MCM v3.3.1, the full and reduced PRAM. Mechanisms with combined elementary reactions are the reference mechanism used for the reduction before (MCM + PRAM) and after (Ref.) prereduction.

2.3.4. Reduction score

Total SOA concentrations are simulated to evaluate the performance of the candidate reductions under the applied evaluation dataset. Simulations are conducted for five days, starting at midnight and noon, in order to account for both daytime and nighttime chemistry. To estimate the uncertainty induced in the current mechanism by the candidate reduction, a reduction error (ϵ) is computed based on the total SOA mass. This error is defined as the larger value of the fractional mean error (FME) between the day 1 and day 2–5 simulation results. The FME is computed according to Eq. (1), where $C_{val,i}$ and $C_{cmp,i}$ are the SOA mass concentrations at time step i simulated with the current mechanism and the mechanism used for comparison, respectively. Here, this comparison mechanism is either the original mechanism after prereduction (see Section 2.4.1) or the mechanism with the latest approved candidate reduction (noted as the previously reduced mechanism). Reduction errors are classified as ϵ_{ref} and ϵ_{pre}, where the comparison mechanisms considered are the mechanism after prereduction and the previously reduced mechanism, respectively. The reduction error (ϵ) is restricted by the predefined error tolerance (ε). During the training process, the average and maximum reduction errors (ϵ_{ref} and ϵ_{pre}), as well as the differences of ϵ_{ref} induced between current and previously reduced reductions ($\delta\epsilon$) are restricted for each candidate reduction. For clarity, errors (ϵ) with subscripts "ave" and "max" represents the average and maximum values under the evaluation dataset, while the superscripts "ref" and "pre" detail the mechanism used to compute the error (original reference after prereduction or previously reduced). A variety of error tolerances may be used to evaluate training at different stages. The error tolerance setting employed in this study for each stage can be found in Section 2.4.4 and Table 4.

$$FME = \frac{2\sum_{n}^{i=1} abs(C_{val,i} - C_{cmp,i})}{n\sum_{n}^{i=1}(C_{val,i} + C_{cmp,i})}. \quad (1)$$

With parallel reduction, several candidate reductions can respect the prescribed error tolerances. Therefore, to determine the most appropriate reduction, a reduction score (S_{rdc}) is calculated in GENOA v2.0 using Eq. (2). The terms N^{now} and N^{pre} denote the numbers in the current and previously reduced mechanisms, and the subscripts "rcn", "sps", and "aero" represent reaction, species, and condensables respectively. The reduction score provides a measure of the effectiveness of the candidate reduction in decreasing the number of reactions (N^{rcn}), species (N^{sps}), and condensables (N^{aero}). In order to emphasize the importance of reducing aerosols, which impose the greatest computational burden in 3-D modeling, the term N^{aero} in S_{rdc} is multiplied by ten (i.e., the typical number of aerosol size bins used in CHIMERE). The reduction candidate that satisfies the error tolerances and achieves the highest reduction score is deemed the best reduction and is accepted for further training.

$$S_{rdc} = (N_{rcn}^{now} - N_{rcn}^{pre}) + (N_{sps}^{now} - N_{sps}^{pre}) + (N_{aero}^{now} - N_{aero}^{pre}) \times 10 \quad (2)$$

2.4. Application to monoterpene SOA reduction

GENOA v2.0 is applied to the reduction of the monoterpene SOA mechanisms. The reference mechanism that serves as a starting point and benchmark for reduction is described in Section 2.4.1. Details on the evaluation dataset, which includes the atmospheric conditions used for reduction evaluation, can be found in Section 2.4.2, while Section 2.4.3 covers the initial SOA precursor conditions. Finally, the reduction parameters and options used in each training stage are outlined in Section 2.4.4.

2.4.1. Reference mechanism

The reference mechanism (MCM + PRAM) for the monoterpene mechanism reduction combined the Master Chemical Mechanism (MCM v3.3.1) for the oxidant-initiated chemistry and the Peroxy Radical Autoxidation Mechanism (PRAM) (Roldin et al., 2019) for HOMs formation, involving the degradation of three of the most abundant and representative monoterpene SOA precursors (i.e., α-pinene, β-pinene, limonene; hereafter referred to as "API", "BPI", and "LIM" respectively). The term "MT" is used when all three monoterpene species are involved.

Table 1 summarized the sizes of the detailed mechanisms associated with the reference mechanism. The MCM and PRAM reactions mentioned in the table are elementary reactions with a single product and an integer stoichiometric coefficient, and the

species contain both gas-phase organic radicals and stable species. Since all stable gas-phase species might partition on particles, they are initially assumed to be condensable. The MCM mechanism for LIM (MCM - LIM) exhibits higher complexity compared to the MCM mechanisms for API (MCM - API) and BPI (MCM - BPI), with a size that is comparable to the sum of these two MCM mechanisms. Despite these differences, the degradation of all three monoterpenes (MCM - MT) shares common pathways in MCM due to the similarity of their chemical structures and properties. However, upon comparing the total number of reactions for the three MCM mechanisms (i.e., MCM - API, MCM - BPI, and MCM - LIM) to MCM - MT, a small difference is observed in the number of reactions (3 660 versus 3 454), suggesting that the majority of MCM pathways may be specific to one monoterpene only. Hence, an efficient reduction that merges unique pathways of each monoterpene into common reaction pathways is necessary to reduce the size of the MCM mechanism.

PRAM mechanisms that contribute to the formation of HOMs are also presented in Table 1. The reduced PRAM contains the same reactions and species as the full PRAM, with the exception of HOM dimerization. The full PRAM undergoes thousands of autooxidation reactions between the PRAM organic peroxy radical (RO_2) and MCM RO_2 species, resulting in the formation of HOM dimers. However, in the reduced PRAM, the dimerization is simplified by several reactions between PRAM RO_2 and a so-call MCM "RO_2 pool", which is the sum of MCM RO_2 species. Notably, this pool is also used in the MCM mechanism to describe autoxidation reactions between MCM RO_2 species. By introducing the MCM RO_2 pool to PRAM, the size of the PRAM mechanism is significantly decreased from 1 773 reactions to 192, while the SOA production remains similar to that of the full PRAM (Roldin et al., 2019).

Consequently, the reference mechanism adopted for the reduction is constructed by combining the reduced PRAM and the MCM - MT mechanisms, followed by a prereduction process. For simplicity, the reduced PRAM is hereafter referred to as "PRAM". To build a numerically stable and faster reference mechanism, a prereduction process is carried out as follows:

- Combine elementary reactions from the original MCM and PRAM with the same reactants and kinetic constants into reactions with non-integer stoichiometric coefficients. After combining MCM and PRAM, the reference mechanism (MCM + PRAM in Table 1) contains 3 001 reactions, 1 227 species, and 738 condensables.
- Jump rapidly-degraded species with a chemical lifetime of less than 1 s under the training dataset (see Section 2.4.2) or a kinetic rate constant greater than 10^6 s^{-1}. As a result, 194 species are removed from the reference mechanism.
- Assume organic compounds with saturation vapor pressure larger than 10^{-3} atm are completely in the gas phase and do not condense. A total of 45 species with high volatility are then set as fully volatile species.
- Remove very high-generation reactions and species (generation order higher than 13) that have a negligible impact on SOA formation. This removal excludes 144 reactions and 58 species, amongst which 50 are condensables.

After the prereduction, the reference mechanism, hereafter referred to as the "Ref". mechanism, contains 2 608 reactions and 975 species (332 radicals and 643 stable species). The saturation vapor pressure (P_{sat}) of MCM condensables is estimated using UManSysProp (Topping et al., 2016), employing the method of Nannoolal, Rarey, and Ramjugernath (2008) for vapor pressure and the method of Nannoolal, Rarey, Ramjugernath, and Cordes (2004) for boiling point. For PRAM condensables, the estimation of P_{sat} is based on the SIMPOL method (Pankow & Asher, 2008), as originally used in the PRAM evaluation performed by Roldin et al. (2019). These methods were adopted by Xavier et al. (2019), where they replicated the biogenic SOA mass yields from an oxidative flow reactor (OFR) and an idealized smog chamber with 0-D simulations using near-explicit mechanisms (MCM + PRAM). Other aerosol properties, including the activity coefficient and Henry's law constant, are computed by the SSH-aerosol model based on the surrogate structures using the same method as in GENOA v1.0. As presented in Fig. S2, the SOA yields simulated with the Ref. mechanism closely match the reported values by Xavier et al. (2019), confirming the reliability of the selected methods for SOA simulations.

2.4.2. Evaluation dataset

The evaluation datasets used for assessing the mechanism reductions comprise different groups of representative near-realistic atmospheric conditions. These datasets consist of the training and pre-testing datasets for different stages in the training process, as well as the testing dataset for the testing process. Fig. 4 illustrates the locations of the training, pre-testing, and testing datasets adopted for the monoterpene mechanism reduction. All the selected atmospheric conditions were extracted from the same database used for the sesquiterpene reductions reported in Wang et al. (2022). This database includes CHIMERE simulation results performed over a one-year period (2015) over Europe. A monthly diurnal profile of hourly meteorological data was extracted from each condition (e.g., temperature, relative humidity), as well as hourly concentrations of oxidants, radicals, and other inorganic species. The evaluation datasets were selected according to the methodology described in Wang et al. (2022). Specifically, different locations with varying API concentrations were selected to cover a range of concentrations of oxidants (OH, O_3, and NO_3) and radicals (HO_2, NO).

The training dataset contains extreme atmospheric conditions representing different chemical regimes (e.g., high NOx and low NOx conditions, high amount of O_3, OH or NO_3) in order to guarantee that any changes in the mechanisms are tested over a wide range of environmental conditions. As shown in Table 2, the training conditions are referred to as "TC1" to "TC8" and cover a broad range of environmental parameters, including variations in temperature from 270 K to 297 K and relative humidity (RH) from 9% to 100%. The selection of training conditions is based on the reaction ratios of different oxidants (OH, O_3, NO_3 radicals), NO_x, HO_2, and RO_2, as SOA formation depends on NO_x/HO_2 ratios and ozone concentrations (Porter et al., 2021). These ratios are calculated using the equations described in Table S1, with the kinetic constants extracted from the API degradation scheme in MCM. The training conditions cover high and low NOx regimes with NO ratios from 9% to 100% and ozone ratios from 17% to 67%.

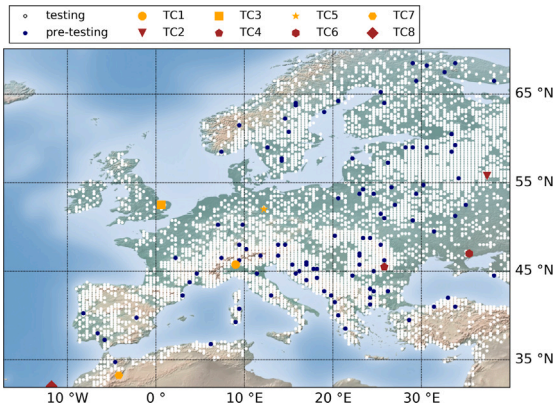

Fig. 4. Locations of conditions in the training (8 "TC" conditions, detailed in Table 2), pre-testing (100 conditions, navy dots), and testing datasets (9 818 conditions, white dots) for the monoterpene mechanism reduction. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2
Geographic and meteorological characteristics of the training dataset for the monoterpene mechanism reduction.

Condition[a]	Lat (°N)	Lon (°E)	Time (month)	TEMP (K)	RH (%)	R_{NO} [b] (%)	R_{O_3} (%)
TC1	45.75	9.0	Dec.	279	85	100	17
TC2	55.75	37.4	Feb.	270	81	100	67
TC3	52.5	0.6	May.	285	69	66	23
TC4	45.5	25.8	Nov.	277	76	68	40
TC5	52.0	12.2	Jun.	290	62	62	25
TC6	47.0	35.4	Aug.	297	46	51	21
TC7	33.25	−4.2	Aug.	297	40	16	65
TC8	32.0	−11.8	Jul.	295	83	9	26

[a]Columns from left to right list the condition identifier, latitude, longitude, time period, average temperature, average RH, average daily NO reaction ratio, and average daily O_3 reaction ratio.

[b]The calculation of reaction ratios is described in Table S1.

Once the mechanism has been reduced to a pre-defined accuracy, the training dataset is replaced by the pre-testing dataset, which is used to evaluate the general accuracy of the reduced mechanism under the most relevant conditions. The pre-testing dataset comprises 100 randomly selected conditions from the testing dataset (9 818 conditions). To ensure that the pre-testing dataset represents the general atmospheric conditions, the average reduction error using the pre-testing dataset is compared to the error using the testing dataset. The user can adjust the initially randomly selected pre-testing conditions so that the errors between the pre-testing and testing datasets are similar. This helps to ensure the accuracy and reliability of the reduced mechanism evaluation while minimizing any potential biases that may have arisen from the initial selection process.

The testing dataset initially is composed of 10 011 conditions, containing 4 011 conditions with high API concentrations ($\geq$0.2 μg/m^{-1}) and 3 000 randomly selected conditions with medium (between 0.1 μg/m^{-1} to 0.2 μg/m^{-1}) and low (between 0.01 μg/m^{-1} to 0.1 μg/m^{-1}) API concentrations, respectively. From the initially selected testing dataset, 193 conditions with low SOA concentrations (with an hourly maximum SOA concentration lower than 1 μg/m^{-1} during the testing process) are excluded to avoid significant numerical uncertainty during evaluation. The final testing dataset consists of 9 818 conditions that are relevant to monoterpene SOA formation over Europe.

2.4.3. Initial condition

As previously discussed in Section 2.2, the reduction process requires different initial conditions of precursor concentrations to accurately capture the distinct SOA formation pathways related to multiple precursors. For the monoterpene mechanism reduction, four initial conditions listed in Table 3 are employed to preserve both the specific and common reaction pathways of all three monoterpene precursors. In the initial condition labeled "iniMT", the total concentrations of the three representative monoterpene precursors are set to 5 ppb, ensuring high SOA production. The concentration of each monoterpene in "iniMT" is assigned proportionally to its emission reported in Guenther et al. (2012), i.e., 3 ppb for API, 1 ppb for BPI, and 1 ppb for LIM. For the other three initial conditions labeled "iniAPI", "iniBPI", and "iniLIM", the concentration of only one monoterpene precursor is set to 3 ppb, namely 3 ppb of API for iniAPI, 3 ppb of BPI for iniBPI, and 3 ppb of LIM for iniLIM. During the training process, candidate reductions are first evaluated under the initial condition iniMT. Subsequently, reductions that meet the criteria proceed to a second evaluation under the other initial conditions collectively. During this second evaluation, the average errors from these three conditions are calculated together to determine the acceptability of the reductions.

Table 3
Initial concentrations of SOA precursor used for the monoterpene mechanism reduction.

Set[a]	API	BPI	LIM
iniMT	3	1	1
iniAPI	3	0	0
iniBPI	0	3	0
iniLIM	0	0	3

[a]From left to right, the table lists the identifier of the initial conditions ("ini" + dominated precursors) and the corresponding concentrations of API, BPI, and LIM in parts per billion (ppb).

2.4.4. Training stages

The reduction process is divided into four training stages, namely stages I, II, III, and IV. A summary of the key reduction parameters and options adopted in each stage can be found in Table 4.

During each training stage, reduction cycles are performed, each consisting of multiple reduction steps that search for and evaluate all candidate reductions related to a targeted species. In each reduction step, only one candidate reduction is accepted if it meets all error tolerances and has the highest reduction score among all candidates. The predefined error tolerances adopted for reduction evaluation include thresholds for average and maximum reduction errors (ϵ) under the evaluation dataset, in comparison to the Ref. mechanism (i.e., tolerances ε_{ave}^{ref} and ε_{max}^{ref}), and to the previously reduced mechanism (i.e., tolerances ε_{ave}^{pre} and ε_{max}^{pre}). The increase of the error ϵ_{ave}^{ref} induced by one reduction is also limited by a tolerance change threshold $\delta\epsilon$.

The reduction performance is evaluated after each reduction cycle by the reduction errors ($\epsilon_{ave}^{pre-testing}$ and $\epsilon_{max}^{pre-testing}$) using the pre-testing dataset. The user-defined error tolerances for the final reduced mechanism, denoted as ε_{ave}^{usr} for the average error and ε_{max}^{usr} for the maximum error, are set to 3% and 30%, respectively. These two tolerances serve as criteria to constrain the errors $\epsilon_{ave}^{pre-testing}$ and $\epsilon_{max}^{pre-testing}$ of the final reduced mechanisms at each stage. If no further reduction can be achieved with the applied reduction parameters and options of one stage, the training progresses to the next stage or terminates after stage IV.

The settings for each training stage are explained below.

- Stage I: The reduction begins with the Ref. mechanism and undergoes evaluation against the training dataset with error tolerances ranging from 0.05% to 3%. Once the algorithm no longer identifies any further reduction candidates within all sets of error tolerances in stage I or when the current errors (ϵ_{ave}^{ref} or ϵ_{max}^{ref}) approach the final user-defined tolerances (ε_{ave}^{usr} and ε_{max}^{usr}), the reduction process advances from stage I to stage II.
- Stage II: Candidate reductions are tested against the pre-testing dataset. Given that the mechanism has already undergone a certain degree of reduction in stage I, the evaluation in stage II employs tolerances equal to ε_{ave}^{usr} and ε_{max}^{usr}, allowing reductions with errors up to 3% on average and 30% at maximum. To enhance the efficiency of the reduction process, the reduced mechanisms are solely compared to the reference mechanism against the pre-testing dataset in stage II and the subsequent stages. In other words, there are no further restrictions for errors ϵ_{ave}^{pre} and ϵ_{max}^{pre} during these stages.
- Stage III: Training in this stage adopts tolerances ε_{ave}^{ref} with values higher than the user-defined tolerance ε_{ave}^{usr}. By allowing a larger tolerance, stage III may temporarily accept mechanisms with errors that exceed ε_{ave}^{usr}. These high errors may be compensated by errors induced in subsequent reductions, resulting in a mechanism with errors falling under ε_{ave}^{usr} after several reductions. Such a mechanism, although trained with larger tolerances, is highly desirable as it enables more reduction than those directly trained with ε_{ave}^{usr}, while still meeting user-defined accuracy. For monoterpene reduction, the reduced mechanism trained with ε_{ave}^{ref} of 5% is selected as the final reduced mechanism for this stage.
- Stage IV: This final reduction stage is intended to finalize the reduction process. Therefore, the final user-defined tolerances are employed for reduction evaluation. The mechanism obtained after this stage is adopted as the final mechanism for the entire training process.

The training process also involves several specific treatments during different stages, which are now detailed.

- Efficient treatment is applied during stages I and II. At the end of one reduction cycle, the next cycle typically proceeds with the next set of reduction parameters (e.g., error tolerances) only if no reduction is accepted during the current cycle. With efficient treatment, the next cycle can move to the next set of parameters when the number of accepted reductions in the current cycle is less than five. This approach prevents the algorithm from spending excessive time reexamining candidate reductions with overly restrictive parameters in the early stages of reduction, which may impede reductions that might be easily accepted in subsequent reductions.
- Aerosol-oriented treatment is implemented during stages II and III. With this treatment, candidate reductions are only accepted if they lead to a reduction in the number of condensable species or a decrease in the average error. In other words, candidate reductions that meet all error criteria but do not reduce condensables or increase the current reduction errors are not accepted with the aerosol-oriented treatment. This approach prioritizes reductions that improve the accuracy of the reduced mechanism or reduce the number of condensables, thereby giving priority to the reduction of aerosol species.

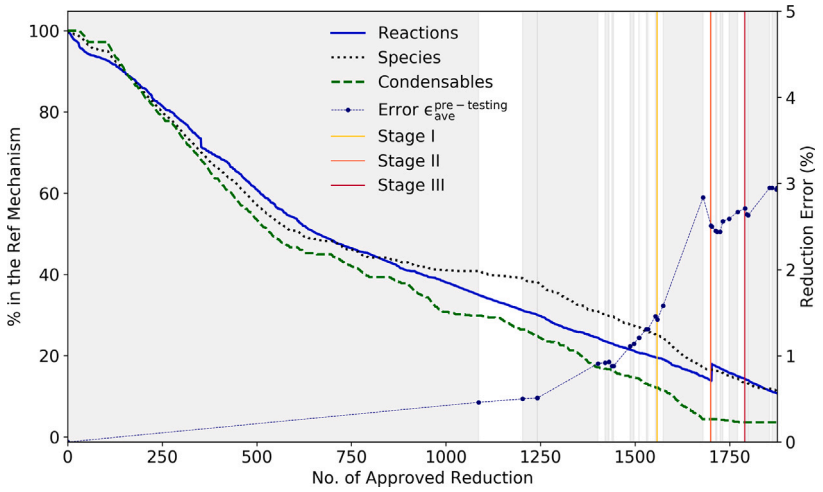

Fig. 5. Evolution of the size (measured as a percentage reduction in the number of reactions, species, and condensables) and accuracy (measured as $\epsilon_{ave}^{pre-testing}$) of the reduced mechanisms during training compared to the reference mechanism. Vertical gray and white intervals indicate reduction cycles. Vertical yellow, orange, and red bars indicate the end of training stages I, II, and III, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

- Elementary-like treatment is adopted in stages III and IV. This treatment involves reorganizing the reaction list, where reactions with multiple products are separated into elementary-like reactions with one product. This approach can increase the number of reactions, leading to a larger pool of candidate reductions via removing reactions to assess the possibility of removing each elementary-like reaction. With more reduction attempts, the mechanism may be further reduced. After stage IV, elementary-like reactions with the same reactants and kinetics are rewritten as combined reactions, resulting in a more concise and readable final mechanism.

3. Results and discussion

The monoterpene mechanism reduction was conducted using a 128-core computer and lasted for approximately four days. The final reduced monoterpene SOA mechanism, hereafter referred to as the "MT-rdc" mechanism, contains 197 reactions, 110 species, and 23 condensable species. As explained in Section 2.4.4, the training process for the monoterpene mechanism reduction can be divided into four consecutive training stages. The detailed training stages along with the resulting mechanism reductions are presented in Section 3.1. The final obtained MT-rdc mechanism is described in Section 3.2, and its complete lists of reactions and species are provided in the Supplementary Material. When evaluated against the testing dataset, MT-rdc introduces an average error of less than 3% over 9 818 conditions across Europe compared to the results simulated with Ref., as discussed in Section 3.3. MT-rdc can also well reproduce the API SOA yields of Xavier et al. (2019), as shown in Fig. S2. Finally, a discussion on the selection of error tolerance and a sensitivity test of MT-rdc on environmental parameters are presented in Section 3.4 and Section 3.5, respectively.

3.1. Mechanism evolution during reduction

Fig. 5 displays the evolution of the size of the mechanisms in terms of the number of reactions, species, and condensables during the training process, comprising 37 reduction cycles within 1 877 approved reductions. Table 4 summarizes the reduction setups and mechanism evolution for each training stage, highlighting the key reduction parameters and treatment, as well as the size and accuracy of the resulting mechanisms.

The majority of approved reductions occurred during training stage I, accounting for 83% of all approved reductions, consuming 65% of the total training time. More than half of the reductions (58%, 1 086 reductions) happened in the first reduction cycle, which utilized strict error tolerances with both ϵ_{max}^{ref} and ϵ_{max}^{pre} set to 0.5%. As the reduction process continued, error tolerances progressively increased. At the end of stage I, the reduced SOA mechanism consisted of 449 reactions and 239 species, with an average error (ϵ_{ave}^{ref}) of 1.6% under the pre-testing conditions. The reduction process in stage I resulted in significant reductions in the numbers of reactions, species, and condensables by 81%, 75%, and 88%, respectively, compared to the Ref. mechanism.

After 1 559 approved reductions, the reduction proceeded to stage II, where the pre-testing dataset was used for reduction evaluation and the aerosol-oriented treatment was activated. Since the mechanism had been strongly reduced in stage I, the reductions achieved in stage II accounted for less than 8% of the total reduction. Compared to the stage I mechanism, the number of reactions and species were reduced by 28% and 33%, respectively. Furthermore, the aerosol-oriented treatment led to a reduction of up to 62% in the number of condensable species, from 74 to 28, indicating the effectiveness of the treatment. At the end of this stage, the reduced mechanism consisted of 337 reactions and 144 species.

Table 4
Reduction configurations and results for different training stages of the monoterpene mechanism reduction.

Training stage	stage I	stage II	stage III	stage IV
Reduction parameter and option				
Evaluation dataset	training	pre-testing	pre-testing	pre-testing
Tolerance ϵ_{ave}^{ref} (%)	0.25–1.5[a]	3	3–5	3
Tolerance ϵ_{max}^{ref} (%)	0.5–3	30	30	30
Tolerance ϵ_{ave}^{pre} (%)	0.25–0.5	-[b]	–	–
Tolerance ϵ_{max}^{pre} (%)	0.5–1	–	–	–
Tolerance $\delta\epsilon$ (%)	0.05–0.3	0.3	0.3–1.5	0.3
Efficient treatment	Yes	Yes	No	No
Elementary-like treatment	No	No	Yes	Yes
Aerosol-oriented treatment	No	Yes	Yes	No
Training detail				
No. of approved reduction	1 559	142	90	86
No. of reduction cycle[c]	17	3	10	7
Training time (h)	57.7	8.7	16	5.6
Size and accuracy of mechanism[d]				
No. of reaction	499	358 (468)[e]	364	197 (279)
No. of species	239	160	130	110
No. of condensable	74	28	23	23
Error $\epsilon_{ave}^{pre-testing}$[f](%)	1.6	2.5	2.6	3.0
Error $\epsilon_{max}^{pre-testing}$ (%)	6.2	26.3	24.9	28.4

[a]For stage I, the values of ϵ_{max}^{ref}/ ϵ_{max}^{pre} are applied to the reduction in the following order: 0.5/ 0.5, 1/ 1, 1.5/ 1, 2/ 1, 2.5/ 1, 3/ 1. For all training under pre-testing dataset, other tolerances are assigned accordingly: $\epsilon_{ave}^{ref} = \epsilon_{max}^{ref}/2$, $\epsilon_{ave}^{pre} = \epsilon_{max}^{pre}/2$, $\delta\epsilon = \epsilon_{max}^{ref}/10$.
[b]Tolerance not used.
[c]Record only those with approved reductions.
[d]Mechanisms obtained after the respective training stages.
[e]The number of elementary-like reactions (the values within parentheses) is only noted at the start and end of elementary-like treatment (mechanisms of stages II and IV).
[f]Errors $\epsilon_{ave}^{pre-testing}$ and $\epsilon_{max}^{pre-testing}$ represent the average and maximum errors, respectively, obtained when evaluating the reduced mechanisms against the Ref. mechanism using the pre-testing dataset.

Table 5
Percentage and number of approved reductions per reduction strategy in different training stages.

Training stage[a]	stage I	stage II	stage III	stage IV	Overall
Removing reactions[b]	48	37	75	74	50
Jumping	6	16	9	10	7
Lumping	13	21	6	5	13
Replacement	4	5	2	5	4
Removing species	20	15	8	6	18
Removing partitioning	9	5	0	0	8
No. of reduction[c]	1 574	129	93	81	1 877

[a]Total percentage per training stage is 100%. The last column shows the percentage for the entire training process.
[b]includes reductions via removing elementary-like reactions.
[c]Number of approved reactions per training stage.

Afterward, in stage III, the reactions with multiple products were separated into multiple reactions with one product due to the elementary-like treatment. This resulted in a change in the number of reactions in the stage II mechanism, increasing from 358 to 468. The other settings remained identical to stage II, except for the use of higher error tolerances. After completing stage II, GENOA v2.0 selects the stage III mechanism trained with a tolerance ϵ_{ave}^{ref} of 5%, balancing accuracy and reduction extent. This mechanism exhibits an average error of 2.6% and a maximum error of 24.9% under pre-testing conditions, satisfying the final user-defined tolerances. Compared to the stage II mechanism, the use of the elementary-like treatment facilitated a reduction of 22% in the number of elementary-like reactions.

In the final training stage, the mechanism was trimmed with the final user-defined tolerances. While no condensables were removed in stage IV, 23% of reactions and 15% of species were reduced compared to the stage III mechanism. Finally, the elementary-like reactions were recombined, resulting in the MT-rdc mechanism with 197 reactions.

Table 5 presents the percentages and numbers of approved reductions achieved per reduction strategy at different training stages. Based on the results, removing reactions is the most effective reduction strategy, contributing significantly to the reduction at all stages, especially in stages III and IV, where the elementary-like treatment is activated, accounting for approximately three-quarters of the total reductions. Other removal strategies, including removing species and removing partitioning, together represent 26% of the total reductions. These two types of removal strategies are mainly effective for reducing pathways that have a negligible impact on SOA formation in the early stages of reduction. When reducing the pathways that are more sensitive to SOA formation,

Table 6
Percentage and number of organic species derived from different combinations of monoterpene precursors per reduced SOA mechanisms.[a].

Combination[b]	Ref.	stage I	stage II	stage III	MT-rdc
API	10	18	12	12	17
BPI	18	9	11	10	13
LIM	48	21	20	17	17
API+BPI	7	7	4	5	6
BPI+LIM	2	0	0	0	0
API+LIM	5	10	21	14	7
API+BPI+LIM	10	36	32	43	39
No. of species	975	239	160	130	110

[a]Total percentage per mechanism is 100 .
[b]Mechanisms from left to right are the reference mechanism (Ref.), mechanisms of training stages I, II, III, and the final reduced SOA mechanism (MT-rdc).

particularly in stage II, the reduction strategies of lumping and jumping stand out, together leading to a total reduction of 37%. These two strategies are favored for reducing the number of condensables due to aerosol-oriented treatments. Compared to other strategies, reduction via replacement contributes less, with an average contribution of 4%.

Table 6 displays the variation of the numbers of species derived from different SOA precursors during the reduction. The original mechanism considers the SOA formation and aging from three monoterpene precursors, i.e., API, BPI, and LIM, resulting in common and individual reaction pathways and formed species. Species are classified based on the combinations of precursors from which they can be formed, including one precursor (API, BPI, or LIM), two precursors (e.g., API + BPI, API + LIM, BPI + LIM), or all three precursors (e.g., API + BPI + LIM). The numbers of reactions and condensable species exhibiting similar variation patterns to those of the number of species are therefore not discussed further here.

As shown in Table 6, the number of species decreases significantly during stage I from 975 to 239, consistent with the largest number of reductions occurring during this stage. LIM leads to the formation of the largest number of monoterpene species (48%) in the Ref. mechanism. Its number decreases the most during reduction compared to the number of species formed from the other two precursors. As the mechanism is reduced, the percentage of species common to the three precursors (API + BPI + LIM) increases. This result is reasonable since the reduction via lumping and replacing tends to merge species from different precursors. While species common to the three precursors represent only 10% of species in Ref., they represent 39% of species in the final reduced mechanism. Meanwhile, the percentage of species derived from a single precursor decreases from 76% to 47%.

3.2. Description of the reduced mechanism

The MT-rdc mechanism, which comprises 197 reactions and 110 species, is illustrated in Fig. 6. As depicted in this figure, for each MT precursor, reactions with all three oxidants (i.e., OH, O_3, NO_3) are kept in MT-rdc. Key oxidation products involved in SOA formation are preserved but may be lumped with similar surrogates. Those lumped surrogates are designated with the prefix "m" followed by the dominant original species contributing to the lumped species. Reactions with radicals (NO, HO_2, RO_2) are also preserved, indicating the mechanism can account for SOA formation pathways under various atmospheric conditions, including both high- and low- NOx chemical regimes.

While high-generation oxidations from different precursors are significantly merged in MT-rdc, the disparities in the first two generations of oxidations are retained. For API, ten first-generation oxidant radicals are preserved, including two nitrate RO_2 (i.e., APINAO2 and NAPINBO2) from the API + NO_3 reaction, three RO_2 species from API + OH reactions (including one monomer C10H17O4O2 originating from the PRAM mechanism), and five from reactions with O_3 (including C10H15O2O2 from PRAM). Consequently, two first-generation condensables, namely mNAPINAOOH and C920PAN, specific to API oxidation, are retained with MT-rdc. With regard to BPI, the initial reactions with BPI yield seven radicals, five of which are specifically formed from BPI oxidation, such as NBPINAO2 and NBPINBO2 from NO_3 reactions; BPINAO2 and BPINBO2 from OH reactions; and NOPINDO2 from ozonolysis. BPI oxidation leads to the formation of two first-generation condensables (i.e., mBPINAOOH and mNBPINAOOH). As for LIM, six radicals result from oxidant-initiated reactions: NLIMO2 from the NO_3 reaction; mLIMAO2, mLIMCO2, and C19H17O4O2 from the OH reaction; and LIMALBO2 and C10H15O2O2 from ozonolysis. Two condensables (i.e., LIMALNO3 and LIMALOOH) are generated from the first two generations of LIM oxidation. As a consequence of the extensive merging of high-generation oxidation in the MT-rdc, 14 out of 23 condensables are formed from the common reaction pathways of all three precursors.

The MT-rdc mechanism also preserves the HOM formation from the PRAM mechanism. As illustrated in Fig. 7, three condensable species are retained from PRAM (including 69 condensables), derived from the initial reactions of monoterpenes with OH and O_3. The HOM condensables comprise two lumped monomer condensables (i.e., mC10H14O9 and mC10H14O11) formed from RO_2 reactions with autoxidation and one dimer condensable C20H30O13 formed from three dimerization reactions. Regarding HOM formation, MT-rdc records three RO_2 reactions with HO_2, nine RO_2 reactions with NO, and 12 RO_2-RO_2 reactions. It is worth noting that some of the reactions with radicals in the MT-rdc mechanism are competitive, which suggests that the mechanism is capable of maintaining the sensitivity of HOM formation to the chemical regime under both high- and low-NOx conditions.

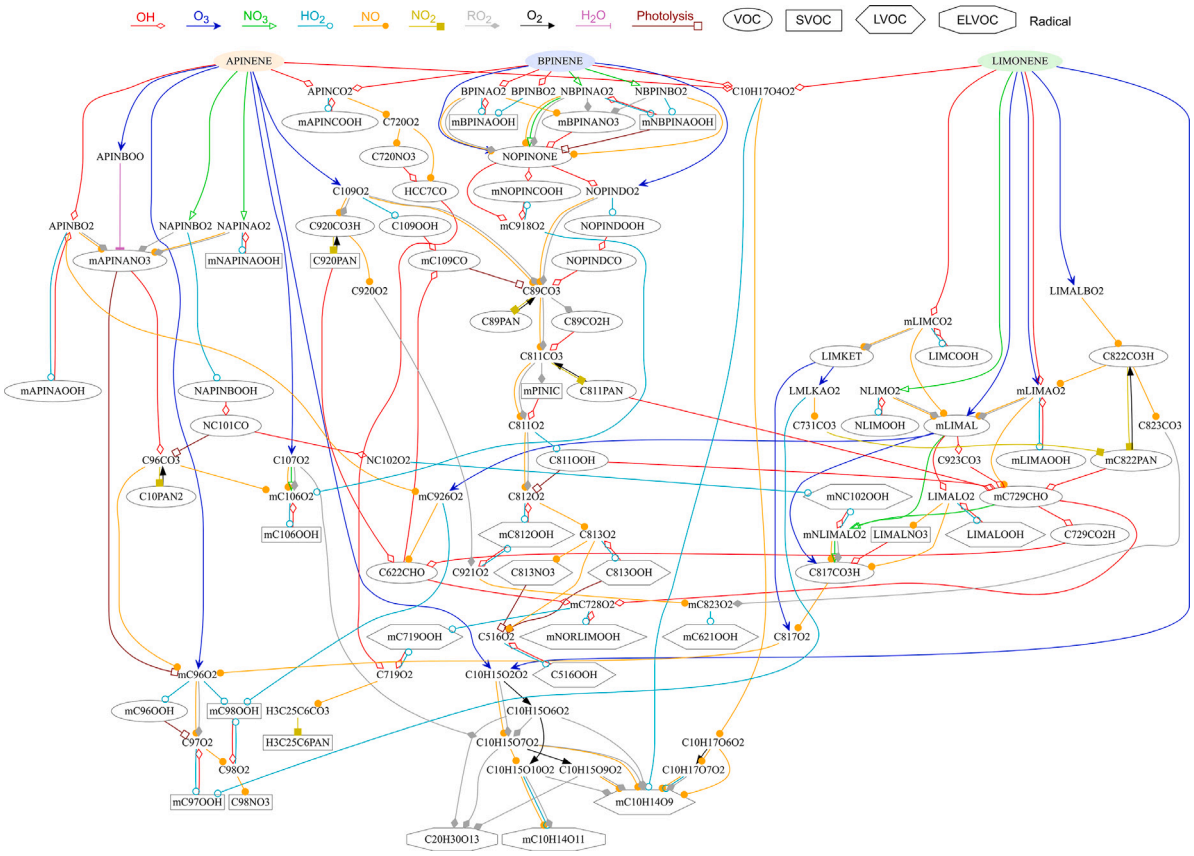

Fig. 6. Reaction pathways of the MT-rdc mechanism. The VOC oxidation reactions with OH radical, Ozone, and NO_3 radical are depicted by red, blue, and green lines, respectively, with different arrowhead symbols pointing to reaction products: empty diamond, filled "V"-type, and empty triangle. Reactions with HO_2, NO, RO_2, O_2, H_2O, and photolysis reactions are represented by cyan, orange, yellow, gray, black, and orchid lines, respectively, with different arrowhead symbols: empty dot, filled dot, filled square, filled diamond, filled triangle, tee, empty square. The shapes of species nodes indicate the species types: radicals with no outline, VOCs with ellipses, semi-volatile organic compounds with boxes (SVOCs: P_{sat} lower than 10^{-9} atm), low-volatile organic compounds with hexagons (LVOCs: P_{sat} between 10^{-9} atm and 10^{-13} atm), and ELVOCs with octagons (ELVOCs: P_{sat} lower than 10^{-13} atm). All P_{sat} values are at 298 K. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

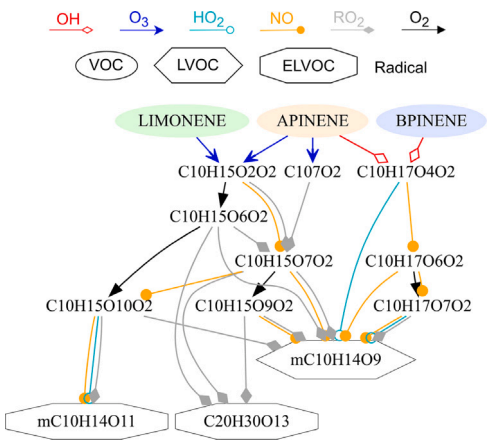

Fig. 7. Reaction pathway in the MT-rdc mechanism corresponding to the formation of HOMs trained from the PRAM mechanism. Refer to Fig. 6 for a more detailed description of the legend.

Table 7
Errors generated by the MT-rdc mechanism (compared to the reference mechanism) simulated over testing conditions with different initial SOA precursor conditions and two simulation starting times (i.e., 0 h and 12 h).

Error (%)	iniMT	iniAPI	iniBPI	iniLIM	Average
at 0 h	3.4	3.0	3.4	4.0	3.4
at 12 h	2.1	2.2	3.0	2.3	2.4
Average	2.7	2.6	3.2	3.1	2.9

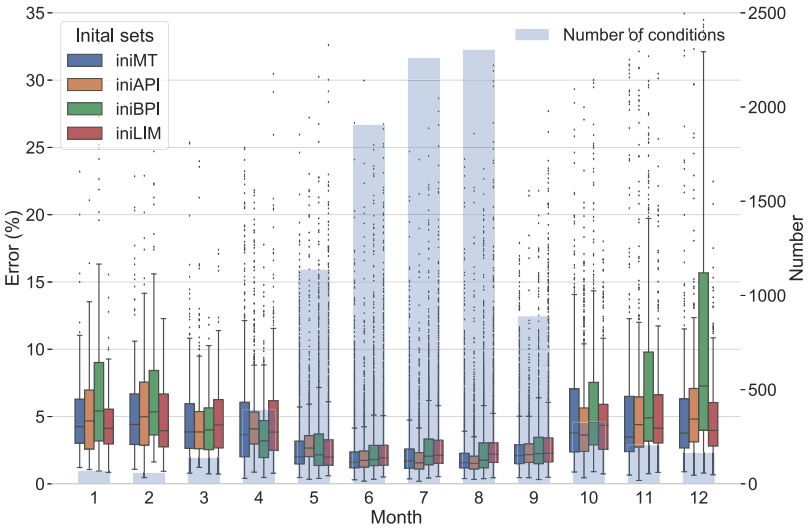

Fig. 8. Monthly distribution of errors over testing conditions (box plot) generated by the MT-rdc mechanism (compared to the reference mechanism) simulated with different initial SOA precursor conditions. The bars represent the number of testing conditions adopted for testing.

3.3. Mechanism performance during testing

The testing process is evaluated by comparing the differences in simulated total SOA concentrations between the reference mechanism (Ref.) and the final reduced SOA mechanism (MT-rdc) under the testing dataset (consisting 9 878 testing conditions, explained in Section 2.4.2). The reduction error (calculation detailed in Section 2.3.4) over testing conditions is used to evaluate the efficiency of the reduction.

In total, within 79 024 simulations, an average error of 2.9% is obtained, which is consistent with the predefined error tolerance set for training. The errors over testing conditions for different initial SOA precursor conditions and starting times are summarized in Table 7. Overall, the testing errors for all initial conditions are acceptable, ranging from 2.1% and 4.0%. MT-rdc performs slightly better for simulations beginning at noon (12 h) than those beginning at midnight (0 h), with an average error of 2.4% and 3.4%, respectively. This suggests that MT-rdc is slightly more efficient at predicting daytime SOA formation than nighttime.

The monthly distribution of the testing errors is presented in Fig. 8. 86% of testing conditions (8 509 conditions) are from May to September and the rest 14% (1 369 conditions) are from October to April. The majority of conditions correspond to warm weather that is likely to be associated with high monoterpene SOA concentrations. From May to September, a low error is obtained with MT-rdc with an average below 2.5% for testing with all initial conditions. The monthly errors are higher from October to April, averaging 5.4%. This result indicates that MT-rdc exhibits higher accuracy for predicting SOA formation under spring and summer conditions in Europe than for fall and winter conditions in Europe. This discrepancy may be attributed to the pre-testing dataset having a higher proportion of conditions from May to September (83 out of 100 conditions) than from October to April (17 out of 100 conditions).

Furthermore, significant errors are observed in the simulations conducted with the initial condition iniBPI compared to those performed with other initial conditions. This disparity is particularly prominent for the testing conditions during December, where the average error obtained with iniBPI is 11.4%. However, it is worth noting that the SOA yields from BPI oxidation are significantly low compared to those from other monoterpene precursors, especially in December. Under all testing conditions, an average SOA yield of 3.5% is obtained with iniBPI while simulations with other initial conditions lead to higher yields (15% for iniMT, 11% for iniAPI, and 35% for iniLIM). Consequently, despite the high testing errors, the absolute errors remain low when simulated with iniBPI, especially under the December testing conditions, when the average BPI SOA yield is only 2%.

The map distribution of errors induced by the reduction and the corresponding SOA yields simulated with the MT-rdc mechanism under testing conditions from May to September are presented in Fig. 9. As mentioned previously, LIM has the highest SOA productivity, leading to the highest SOA yields simulated with iniLIM compared to those simulated with other initial conditions.

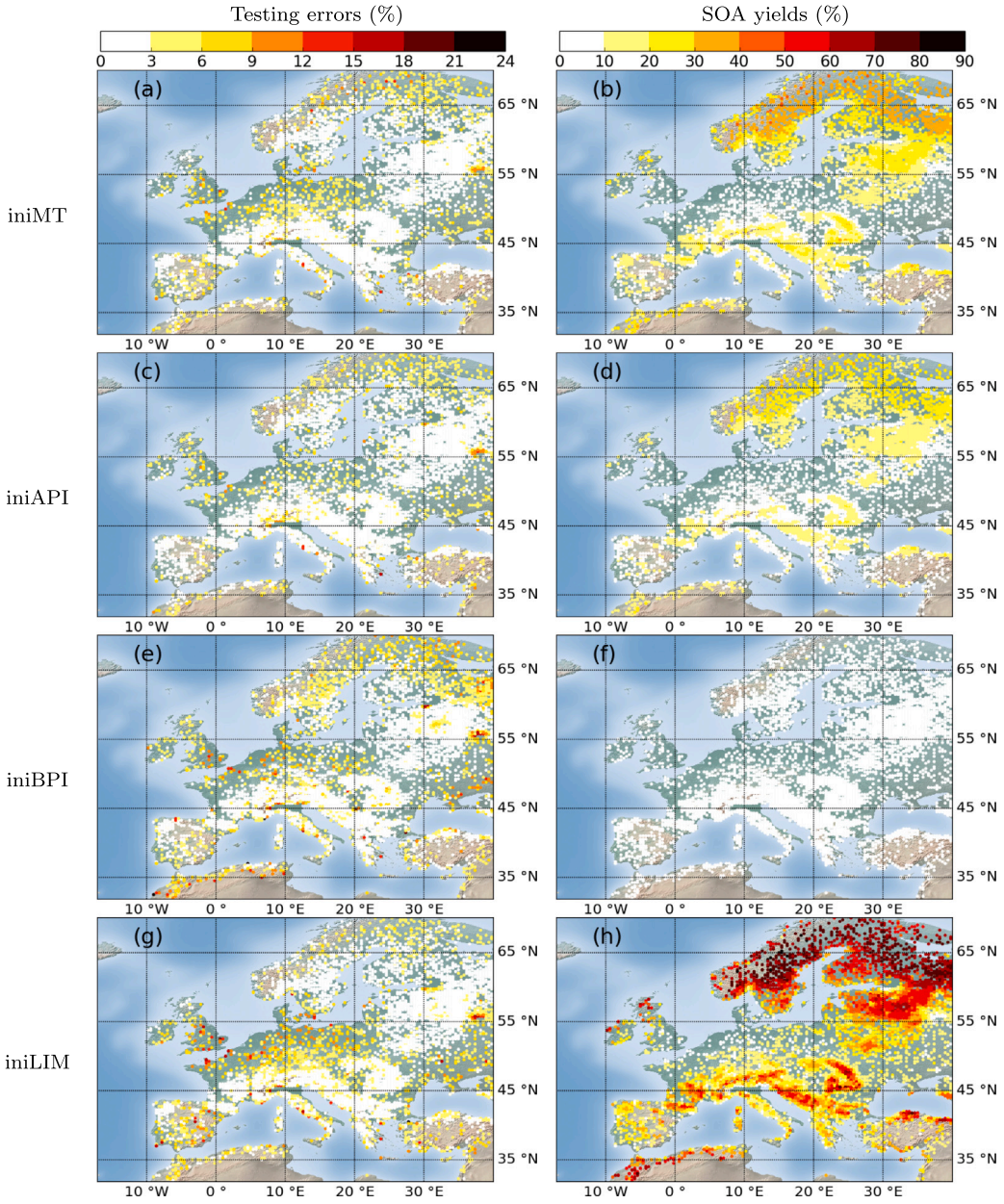

Fig. 9. Map distributions of MT-rdc testing errors (left panels) and SOA yields (right panels) simulated with the different initial precursor conditions under testing conditions from May to September (8 509 conditions, corresponding 86% of total testing conditions). The maps of all testing conditions are in Fig. S3.

The highest SOA yields are simulated in Northern and Eastern Europe around the Baltic Sea, followed by Central Europe around the Mediterranean Sea. From May to September, no obvious differences in errors are simulated between the different initial concentration conditions. The results are consistent with the monthly distribution of errors reported in Fig. 8 and are considerably lower than those across all testing conditions presented in Fig. S3. 95% of simulations are within an error lower than 6.3%, while 99% with an error lower than 12.6%. High errors (above 12%) are concentrated in a few particular areas with high-NOx conditions for all initial conditions. These areas, (e.g., near Moscow, Rome, and the English Channel) likely correspond to areas near large cities or shipping routes. However, these areas have very low monoterpene SOA production (average yield around 4% over these areas). Other conditions with high errors are distributed sporadically over Europe, with many also associated with low SOA production. If the performance of the reduced mechanism needs to be improved under these sporadic conditions, one or several conditions with high errors can be added to the pre-testing dataset to better constrain the mechanism. However, as a trade-off, the new reduced mechanism may contain more species and reactions than MT-rdc.

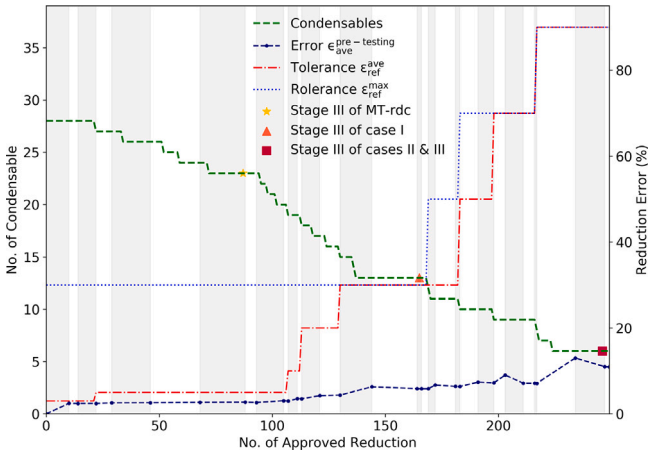

Fig. 10. Evolution of the size (measured as the number of condensables) and accuracy (measured as $\epsilon_{ave}^{pre-testing}$) of reduced mechanisms during training stage III with increasing error tolerances. The final reduced mechanisms after stage IV reduction, trained from three stage III mechanisms marked in the figure, can be found in Table 8.

Table 8
Size and accuracy of the reduced mechanisms trained with different error tolerances.

Mechanism[a]	MT-rdc	case I	case II	case III
Size and accuracy of mechanism[b]				
No. of reaction	197	153	81	40
No. of species	110	84	47	24
No. of condensable	23	13	6	5
Error $\epsilon_{ave}^{testing}$ (%)	2.9	6.0	11.7	19.1
Error $\epsilon_{99\%}^{testing}$ (%)	15.6	28.5	50.2	55.2
Error tolerance for stage IV				
Tolerance ϵ_{ave}^{usr} (%)	3	6	12	20
Tolerance ϵ_{max}^{usr} (%)	30	30	60	60

[a]Mechanisms are trained from stage III mechanisms noted in Fig. 10. The case IV mechanism is trained from the case III mechanism with a higher tolerance ϵ_{ave}^{usr} during stage IV.
[b]The errors $\epsilon_{ave}^{testing}$ and $\epsilon_{99\%}^{testing}$ represent the average error and the 99th percentile error, respectively, when compared to the Ref. mechanism under the testing dataset.

3.4. Reduction sensitivity to prescribed error tolerances

The reduction parameters and options adopted during training can significantly influence the resulting mechanisms. In particular, error tolerances directly affect reduction evaluation, thereby altering the size and accuracy of the reduced mechanisms. The effect of larger error tolerances on the extent of the reduction is investigated. As shown in Fig. 10, during the training stage III (detailed in Section 2.4.4), the error tolerances (i.e., ϵ_{ave}^{ref} and ϵ_{max}^{ref}) are increased up to 90% to explore further reduction. As the aerosol-oriented treatment is activated during stage III, a candidate reduction can only be accepted if it reduces condensables or current errors. Consequently, with increasing tolerance, the errors of the reduced mechanisms rise only when the number of condensables decreases. Due to the aerosol-oriented treatment and potential error compensation along with approved reductions, the obtained reduction errors are much lower than the tolerances, reaching only 11% at the end of stage III against 90% for the tolerance.

From all mechanisms obtained during stage III, three are selected (marked in Fig. 10) and undergo the final training stage IV with different user-defined error tolerances: ϵ_{ave}^{usr} is 3% for MT-rdc, 6% for case I, 12% for case II, and 20% for case III. Table 8 lists the size and accuracy of those reduced mechanisms. The reaction pathways of cases I to III can be found in the Supplementary Material. As expected, there is a trade-off between mechanism size and accuracy. Training with larger error tolerances leads to greater reductions in size but decreases accuracy. The more reduced mechanisms exhibit a lower number of condensable species (i.e., 13, 6, and 5 for cases I to III, respectively) compared to MT-rdc trained with a stricter error tolerance of 3%, which includes 23 condensables. Notably, a highly condensed SOA mechanism with only 40 reactions and 24 species is obtained with an average tolerance of 20%, which corresponds to the tolerance used in other reduction algorithms (e.g., in Lannuque et al., 2018). This case III mechanism undergoes a remarkable reduction in size compared to the reference mechanism (more than 98% reduction) and holds great potential for 3-D simulations. The selection of appropriate error tolerances is, therefore, crucial in order to achieve an optimal balance between the extent and accuracy of the reduction.

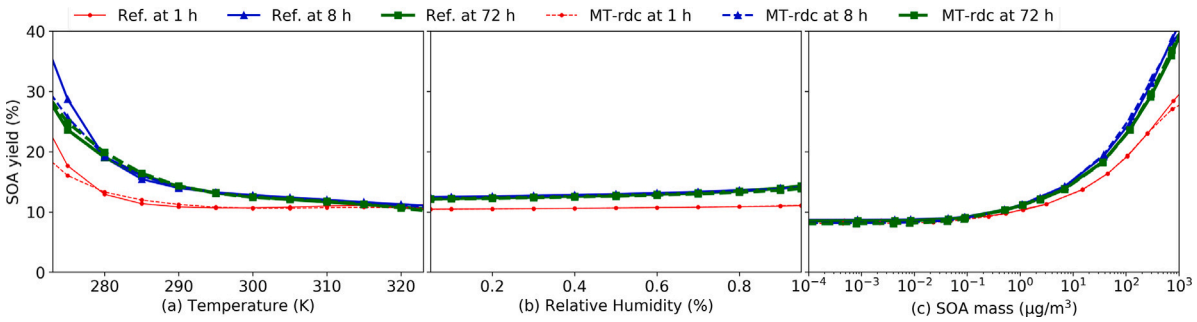

Fig. 11. Dependence of average SOA yields simulated under the pre-testing dataset with the reference mechanism (Ref., solid line) and the final reduced SOA mechanism (MT-rdc, dotted line) on (a) temperature, (b) relative humidity, and (c) SOA mass at 1 h (red point), 8 h (blue triangle) and 72 h (green square). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.5. Mechanism sensitivity to environmental parameters

The sensitivity of MT-rdc to several environmental parameters crucial for SOA formation is investigated. These parameters include temperature, relative humidity (RH), and SOA mass conditions. Average SOA yields under the pre-testing dataset are simulated for five days with two start times (0 h and 12 h), while varying environmental parameters. The default temperature and RH are held constant at 298 K and 50%, respectively, and the initial precursor condition follows the iniMT condition. The sensitivity to SOA mass is achieved by changing the concentration proportionally in the initial concentrations of precursors (the ratios between the different precursors are kept identical to iniMT). While changing one of the three parameters, the others are kept constant.

The results of the sensitivity test, presented in Fig. 11, are compared at three different simulation times (1 h, 8 h, and 72 h) considering the SOA formation evolution. Overall, the MT-rdc and Ref. mechanisms showed similar behaviors. No significant differences (lower than 2%) between the two mechanisms were found when varying RH from 5% to 95% or mass loading from 10^{-4} μg/m^3 to 10^3 μg/m^3. Regarding temperature variations, MT-rdc differs from the Ref. mechanism under a few extreme low-temperature conditions, with differences of up to 10% observed when the temperature was set constant to 270 K. The disparities between the two mechanisms were mainly noticeable for oxidation durations of 1 h and 8 h, whereas for an oxidation duration of 72 h, the variations were relatively minor (approximately 3%) even at 270 K.

4. Conclusion

This paper presents the development and application of GENOA v2.0, an algorithm designed to reduce the size and complexity of SOA mechanisms for multiple precursors. As part of the parallel reduction scheme of GENOA v2.0, multiple candidate reductions are evaluated simultaneously at each reduction step, with the optimal reduction chosen based on evaluation criteria such as reduction score and errors. To account for both the unique and common reaction pathways of multiple SOA precursors, several initial conditions with different compositions of SOA precursors are used in the training process. The training is divided into several stages, in which the user can adjust the reduction parameters and options in order to customize the reduction.

GENOA v2.0 has been applied to the reduction of monoterpene SOA mechanisms from MCM combined with the HOM formation mechanism from PRAM for three monoterpene precursors (α-pinene, β-pinene, limonene). The original MCM + PRAM mechanism contains 3 001 reactions and 1 143 gas-phase species, with 738 species that can be considered condensables. After reduction, the monoterpene SOA mechanism (MT-rdc) is reduced to 197 reactions and 110 species, with the number of condensable species decreasing from 738 to 23. When evaluated against the testing dataset, MT-rdc introduced a low average error of less than 3% over Europe compared to the reference mechanism. Sensitivity tests demonstrated that MT-rdc behaves similarly to the original mechanism in response to changes in temperature, relative humidity, and SOA mass loading. By allowing a larger error tolerance of up to 20%, the SOA mechanism could be further reduced to a mechanism consisting of 40 reactions and 24 species, including only 5 condensables. This work shows that GENOA v2.0 has the potential to generate reduced condensed SOA mechanisms from explicit VOC mechanisms while maintaining reasonable accuracy.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data used for this paper can be accessed publicly on Zenodo: https://doi.org/10.5281/zenodo.8187593.

Acknowledgments

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 Dr. Pontus Roldin for providing the decomposition into functional groups of the PRAM species.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jaerosci.2023.106248.

References

Atkinson, R. (2000). Atmospheric chemistry of VOCs and NOx. *Atmospheric Environment*, *34*(12–14), 2063–2101. http://dx.doi.org/10.1016/S1352-2310(99)00460-4.

Atkinson, R., & Arey, J. (2003). Atmospheric degradation of volatile organic compounds. *Chemical Reviews*, *103*(12), 4605–4638. http://dx.doi.org/10.1021/cr0206420, PMID: 14664626.

Aumont, B., Szopa, S., & Madronich, S. (2005). Modelling the evolution of organic carbon during its gas-phase tropospheric oxidation: Development of an explicit model based on a self generating approach. *Atmospheric Chemistry and Physics*, *5*(9), 2497–2517. http://dx.doi.org/10.5194/acp-5-2497-2005, URL https://acp.copernicus.org/articles/5/2497/2005/.

Bianchi, F., Kurtén, T., Riva, M., Mohr, C., Rissanen, M. P., Roldin, P., et al. (2019). Highly oxygenated organic molecules (HOM) from gas-phase autoxidation involving peroxy radicals: A key contributor to atmospheric aerosol. *Chemical Reviews*, *119*(6), 3472–3509. http://dx.doi.org/10.1021/acs.chemrev.8b00395.

Camredon, M., Aumont, B., Lee-Taylor, J., & Madronich, S. (2007). The SOA/VOC/NO x system: An explicit model of secondary organic aerosol formation. *Atmospheric Chemistry and Physics*, *7*(21), 5599–5610. http://dx.doi.org/10.5194/acp-7-5599-2007.

Chen, T., Zhang, P., Chu, B., Ma, Q., Ge, Y., Liu, J., et al. (2022). Secondary organic aerosol formation from mixed volatile organic compounds: Effect of RO2 chemistry and precursor concentration. *npj Climate and Atmospheric Science*, *5*(1), 1–8. http://dx.doi.org/10.1038/s41612-022-00321-y.

Coggon, M. M., Lim, C. Y., Koss, A. R., Sekimoto, K., Yuan, B., Gilman, J. B., et al. (2019). OH chemistry of non-methane organic gases (NMOGs) emitted from laboratory and ambient biomass burning smoke: evaluating the influence of furans and oxygenated aromatics on ozone and secondary NMOG formation. *Atmospheric Chemistry and Physics*, *19*(23), 14875–14899. http://dx.doi.org/10.5194/acp-19-14875-2019, URL https://acp.copernicus.org/articles/19/14875/2019/.

Couvidat, F., Debry, E., Sartelet, K., & Seigneur, C. (2012). A hydrophilic/hydrophobic organic (H2O) aerosol model: Development, evaluation and sensitivity analysis. *Journal of Geophysical Research: Atmospheres*, *117*(D10), http://dx.doi.org/10.1029/2011JD017214, URL https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2011JD017214.

Donahue, N. M., Epstein, S. A., Pandis, S. N., & Robinson, A. L. (2011). A two-dimensional volatility basis set: 1. Organic-aerosol mixing thermodynamics. *Atmospheric Chemistry and Physics*, *11*(7), 3303–3318. http://dx.doi.org/10.5194/acp-11-3303-2011, URL https://acp.copernicus.org/articles/11/3303/2011/.

Donahue, N. M., Robinson, A. L., Stanier, C. O., & Pandis, S. N. (2006). Coupled partitioning, dilution, and chemical aging of semivolatile organics. *Environmental Science & Technology*, *40*(8), 2635–2643. http://dx.doi.org/10.1021/es052297c.

Ehn, M., Thornton, J. A., Kleist, E., Sipilä, M., Junninen, H., Pullinen, I., et al. (2014). A large source of low-volatility secondary organic aerosol. *Nature*, *506*(7489), 476–479. http://dx.doi.org/10.1038/nature13032, URL https://www.nature.com/articles/nature13032.

Fry, J. L., Kiendler-Scharr, A., Rollins, A. W., Wooldridge, P. J., Brown, S. S., Fuchs, H., et al. (2009). Organic nitrate and secondary organic aerosol yield from NO_3 oxidation of β-pinene evaluated using a gas-phase kinetics/aerosol partitioning model. *Atmospheric Chemistry and Physics*, *9*(4), 1431–1449. http://dx.doi.org/10.5194/acp-9-1431-2009, URL https://acp.copernicus.org/articles/9/1431/2009/.

Gelencsér, A., May, B., Simpson, D., Sánchez-Ochoa, A., Kasper-Giebl, A., Puxbaum, H., et al. (2007). Source apportionment of PM2.5 organic aerosol over Europe: Primary/secondary, natural/anthropogenic, and fossil/biogenic origin. *Journal of Geophysical Research: Atmospheres*, *112*(D23), http://dx.doi.org/10.1029/2006JD008094.

Goldstein, A. H., & Galbally, I. E. (2007). Known and unexplored organic constituents in the Earth's atmosphere. *Environmental Science & Technology*, *41*(5), 1514–1521. http://dx.doi.org/10.1021/es072476p, URL https://pubs.acs.org/doi/pdf/10.1021/es072476p.

Goliff, W., Stockwell, W., & Lawson, C. (2013). The regional atmospheric chemistry mechanism, version 2. *Atmospheric Environment*, *68*, 174–185. http://dx.doi.org/10.1016/j.atmosenv.2012.11.038.

Griffin, R. J., Nguyen, K., Dabdub, D., & Seinfeld, J. H. (2003). A coupled hydrophobic-hydrophilic model for predicting secondary organic aerosol formation. *Journal of Atmospheric Chemistry*, *44*(2), 171–190. http://dx.doi.org/10.1023/A:1022436813699.

Guenther, A. B., Jiang, X., Heald, C. L., Sakulyanontvittaya, T., Duhl, T., Emmons, L. K., et al. (2012). The model of emissions of gases and aerosols from nature version 2.1 (MEGAN2.1): An extended and updated framework for modeling biogenic emissions. *Geoscientific Model Development*, *5*(6), 1471–1492. http://dx.doi.org/10.5194/gmd-5-1471-2012, URL https://gmd.copernicus.org/articles/5/1471/2012/.

Hallquist, M., Wenger, J. C., Baltensperger, U., Rudich, Y., Simpson, D., Claeys, M., et al. (2009). The formation, properties and impact of secondary organic aerosol: Current and emerging issues. *Atmospheric Chemistry and Physics*, *9*(14), 5155–5236. http://dx.doi.org/10.5194/acp-9-5155-2009, URL https://acp.copernicus.org/articles/9/5155/2009/.

Han, S., & Jang, M. (2023). Modeling daytime and nighttime secondary organic aerosol formation via multiphase reactions of biogenic hydrocarbons. *Atmospheric Chemistry and Physics*, *23*(2), 1209–1226. http://dx.doi.org/10.5194/acp-23-1209-2023, URL https://acp.copernicus.org/articles/23/1209/2023/.

Jenkin, M. E., Saunders, S. M., & Pilling, M. J. (1997). The tropospheric degradation of volatile organic compounds: A protocol for mechanism development. *Atmospheric Environment*, *31*(1), 81–104. http://dx.doi.org/10.1016/S1352-2310(96)00105-7, URL https://www.sciencedirect.com/science/article/abs/pii/S1352231096001057.

Jenkin, M. E., Valorso, R., Aumont, B., Newland, M. J., & Rickard, A. R. (2020). Estimation of rate coefficients for the reactions of O_3 with unsaturated organic compounds for use in automated mechanism construction. *Atmospheric Chemistry and Physics*, *20*(21), 12921–12937. http://dx.doi.org/10.5194/acp-20-12921-2020, URL https://acp.copernicus.org/articles/20/12921/2020/.

Jenkin, M., Watson, L., Utembe, S., & Shallcross, D. (2008). A common representative intermediates (CRI) mechanism for VOC degradation. Part 1: Gas phase mechanism development. *Atmospheric Environment*, *42*(31), 7185–7195. http://dx.doi.org/10.1016/j.atmosenv.2008.07.028.

Jenkin, M., Wyche, K., Evans, C., Carr, T., Monks, P., Alfarra, M., et al. (2012). Development and chamber evaluation of the MCM v3. 2 degradation scheme for β-caryophyllene. *Atmospheric Chemistry and Physics*, *12*(11), 5275–5308. http://dx.doi.org/10.5194/acp-12-5275-2012, URL https://acp.copernicus.org/articles/12/5275/2012/acp-12-5275-2012.pdf.

Kaduwela, A., Luecken, D., Carter, W., & Derwent, R. (2015). New directions: Atmospheric chemical mechanisms for the future. *Atmospheric Environment*, *122*, 609–610. http://dx.doi.org/10.1016/j.atmosenv.2015.10.031, URL https://www.sciencedirect.com/science/article/pii/S1352231015304507.

Kanakidou, M., Seinfeld, J., Pandis, S., Barnes, I., Dentener, F. J., Facchini, M. C., et al. (2005). Organic aerosol and global climate modelling: A review. *Atmospheric Chemistry and Physics*, *5*(4), 1053–1123. http://dx.doi.org/10.5194/acp-5-1053-2005, URL https://acp.copernicus.org/articles/5/1053/2005/acp-5-1053-2005.pdf.

Kelp, M. M., Jacob, D. J., Lin, H., & Sulprizio, M. P. (2022). An online-learned neural network chemical solver for stable long-term global simulations of atmospheric chemistry. *Journal of Advances in Modeling Earth Systems*, *14*(6), Article e2021MS002926. http://dx.doi.org/10.1029/2021MS002926.

Lannuque, V., Camredon, M., Couvidat, F., Hodzic, A., Valorso, R., Madronich, S., et al. (2018). 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. *Atmospheric Chemistry and Physics*, *18*(18), 13411–13428. http://dx.doi.org/10.5194/acp-18-13411-2018, URL https://acp.copernicus.org/articles/18/13411/2018/.

Li, J., Cleveland, M., Ziemba, L. D., Griffin, R. J., Barsanti, K. C., Pankow, J. F., et al. (2015). Modeling regional secondary organic aerosol using the master chemical mechanism. *Atmospheric Environment*, *102*, 52–61. http://dx.doi.org/10.1016/j.atmosenv.2014.11.054, URL https://www.sciencedirect.com/science/article/pii/S1352231014009194.

Li, Q., Jiang, J., Afreh, I. K., Barsanti, K. C., & Cocker III, D. R. (2022). Secondary organic aerosol formation from camphene oxidation: Measurements and modeling. *Atmospheric Chemistry and Physics*, *22*(5), 3131–3147. http://dx.doi.org/10.5194/acp-22-3131-2022.

McFiggans, G., Mentel, T. F., Wildt, J., Pullinen, I., Kang, S., Kleist, E., et al. (2019). Secondary organic aerosol reduced by mixture of atmospheric vapours. *Nature*, *565*(7741), 587–593. http://dx.doi.org/10.1038/s41586-018-0871-y.

Mouchel-Vallon, C., & Hodzic, A. (2022). Towards emulating an explicit organic chemistry mechanism with random forest models. *Journal of Geophysical Research*, Article e2022JD038227. http://dx.doi.org/10.1029/2022JD038227.

Mouchel-Vallon, C., Lee-Taylor, J., Hodzic, A., Artaxo, P., Aumont, B., Camredon, M., et al. (2020). Exploration of oxidative chemistry and secondary organic aerosol formation in the Amazon during the wet season: Explicit modeling of the manaus urban plume with GECKO-A. *Atmospheric Chemistry and Physics*, *20*(10), 5995–6014. http://dx.doi.org/10.5194/acp-20-5995-2020.

Nannoolal, Y., Rarey, J., & Ramjugernath, D. (2008). 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 Equilibria*, *269*(1), 117–133. http://dx.doi.org/10.1016/j.fluid.2008.04.020, URL https://www.sciencedirect.com/science/article/pii/S0378381208001611.

Nannoolal, Y., Rarey, J., Ramjugernath, D., & Cordes, W. (2004). 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 Equilibria*, *226*, 45–63. http://dx.doi.org/10.1016/j.fluid.2004.09.001.

Newland, M. J., Mouchel-Vallon, C., Valorso, R., Aumont, B., Vereecken, L., Jenkin, M. E., et al. (2022). Estimation of mechanistic parameters in the gas-phase reactions of ozone with alkenes for use in automated mechanism construction. *Atmospheric Chemistry and Physics*, *22*(9), 6167–6195. http://dx.doi.org/10.5194/acp-22-6167-2022, URL https://acp.copernicus.org/articles/22/6167/2022/.

Odum, J. R., Hoffmann, T., Bowman, F., Collins, D., Flagan, R. C., & Seinfeld, J. H. (1996). Gas/particle partitioning and secondary organic aerosol yields. *Environmental Science & Technology*, *30*(8), 2580–2585. http://dx.doi.org/10.1021/es950943+.

Pankow, J. F., & Asher, W. E. (2008). SIMPOL.1: A simple group contribution method for predicting vapor pressures and enthalpies of vaporization of multifunctional organic compounds. *Atmospheric Chemistry and Physics*, *8*(10), 2773–2796. http://dx.doi.org/10.5194/acp-8-2773-2008, URL https://acp.copernicus.org/articles/8/2773/2008/.

Porter, W. C., Jimenez, J. L., & Barsanti, K. C. (2021). Quantifying atmospheric parameter ranges for ambient secondary organic aerosol formation. *ACS Earth and Space Chemistry*, *5*(9), 2380–2397. http://dx.doi.org/10.1021/acsearthspacechem.1c00090.

Pun, B. K., Seigneur, C., & Lohman, K. (2006). Modeling secondary organic aerosol formation via multiphase partitioning with molecular data. *Environmental Science & Technology*, *40*(15), 4722–4731. http://dx.doi.org/10.1021/es0522736.

Roldin, P., Ehn, M., Kurtén, T., Olenius, T., Rissanen, M. P., Sarnela, N., et al. (2019). The role of highly oxygenated organic molecules in the Boreal aerosol-cloud-climate system. *Nature Communications*, *10*(1), 1–15. http://dx.doi.org/10.1038/s41467-019-12338-8.

Sartelet, K., Couvidat, F., Wang, Z., Flageul, C., & Kim, Y. (2020). SSH-aerosol v1. 1: A modular box model to simulate the evolution of primary and secondary aerosols. *Atmosphere*, *11*(5), 525. http://dx.doi.org/10.3390/atmos11050525, URL https://www.mdpi.com/2073-4433/11/5/525/htm.

Sarwar, G., Luecken, D., Yarwood, G., Whitten, G. Z., & Carter, W. P. L. (2008). Impact of an updated carbon bond mechanism on predictions from the CMAQ modeling system: Preliminary assessment. *Journal of Applied Meteorology*, *47*(1), 3–14. http://dx.doi.org/10.1175/2007JAMC1393.1, URL https://journals.ametsoc.org/view/journals/apme/47/1/2007jamc1393.1.xml.

Saunders, S. M., Jenkin, M. E., Derwent, R., & Pilling, M. (2003). Protocol for the development of the master chemical mechanism, MCM v3 (Part A): Tropospheric degradation of non-aromatic volatile organic compounds. *Atmospheric Chemistry and Physics*, *3*(1), 161–180. http://dx.doi.org/10.5194/acp-3-161-2003, URL https://acp.copernicus.org/articles/3/161/2003/acp-3-161-2003.pdf.

Schreck, J. S., Becker, C., Gagne, D. J., Lawrence, K., Wang, S., Mouchel-Vallon, C., et al. (2022). Neural network emulation of the formation of organic aerosols based on the explicit GECKO-A chemistry model. *Journal of Advances in Modeling Earth Systems*, *14*(10), Article e2021MS002974. http://dx.doi.org/10.1029/2021MS002974.

Shen, L., Jacob, D. J., Santillana, M., Bates, K., Zhuang, J., & Chen, W. (2022). A machine-learning-guided adaptive algorithm to reduce the computational cost of integrating kinetics in global atmospheric chemistry models: Application to GEOS-chem versions 12.0.0 and 12.9.1. *Geoscientific Model Development*, *15*(4), 1677–1687. http://dx.doi.org/10.5194/gmd-15-1677-2022, URL https://gmd.copernicus.org/articles/15/1677/2022/.

Shrivastava, M., Cappa, C. D., Fan, J., Goldstein, A. H., Guenther, A. B., Jimenez, J. L., et al. (2017). Recent advances in understanding secondary organic aerosol: Implications for global climate forcing. *Reviews of Geophysics*, *55*(2), 509–559. http://dx.doi.org/10.1002/2016RG000540, URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2016RG000540.

Stockwell, W. R., Saunders, E., Goliff, W. S., & Fitzgerald, R. M. (2020). A perspective on the development of gas-phase chemical mechanisms for Eulerian air quality models. *Journal of the Air & Waste Management Association*, *70*(1), 44–70. http://dx.doi.org/10.1080/10962247.2019.1694605.

Stolzenburg, D., Wang, M., Schervish, M., & Donahue, N. M. (2022). Tutorial: Dynamic organic growth modeling with a volatility basis set. *Journal of the Atmospheric Sciences*, *166*, Article 106063. http://dx.doi.org/10.1016/j.jaerosci.2022.106063.

Sturm, P. O., Manders, A., Janssen, R., Segers, A., Wexler, A. S., & Lin, H. X. (2023). Advecting superspecies: Efficiently modeling transport of organic aerosol with a mass-conserving dimensionality reduction method. *Journal of Advances in Modeling Earth Systems*, *15*(3), Article e2022MS003235.

Szopa, S., Aumont, B., & Madronich, S. (2005). 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. *Atmospheric Chemistry and Physics*, *5*(9), 2519–2538. http://dx.doi.org/10.5194/acp-5-2519-2005, URL https://acp.copernicus.org/articles/5/2519/2005/.

Topping, D., Barley, M., Bane, M. K., Higham, N., Aumont, B., Dingle, N., et al. (2016). UManSysProp v1.0: An online and open-source facility for molecular property prediction and atmospheric aerosol calculations. *Geoscientific Model Development*, *9*(2), 899–914. http://dx.doi.org/10.5194/gmd-9-899-2016, URL https://gmd.copernicus.org/articles/9/899/2016/.

Wang, Z., Couvidat, F., & Sartelet, K. (2022). GENerator of reduced organic aerosol mechanism (GENOA v1.0): An automatic generation tool of semi-explicit mechanisms. *Geoscientific Model Development*, *15*(24), 8957–8982. http://dx.doi.org/10.5194/gmd-15-8957-2022.

Watson, L., Shallcross, D., Utembe, S., & Jenkin, M. (2008). A common representative intermediates (CRI) mechanism for VOC degradation. Part 2: Gas phase mechanism reduction. *Atmospheric Environment*, *42*(31), 7196–7204. http://dx.doi.org/10.1016/j.atmosenv.2008.07.034, URL https://www.sciencedirect.com/science/article/pii/S1352231008006845.

Weber, J., Archer-Nicholls, S., Griffiths, P., Berndt, T., Jenkin, M., Gordon, H., et al. (2020). CRI-HOM: A novel chemical mechanism for simulating highly oxygenated organic molecules (HOMs) in global chemistry–aerosol–climate models. *Atmospheric Chemistry and Physics*, *20*(18), 10889–10910. http://dx.doi.org/10.5194/acp-20-10889-2020, URL https://acp.copernicus.org/articles/20/10889/2020/.

Whitehouse, L. E., Tomlin, A. S., & Pilling, M. J. (2004a). Systematic reduction of complex tropospheric chemical mechanisms, part I: Sensitivity and time-scale analyses. *Atmospheric Chemistry and Physics*, *4*(7), 2025–2056. http://dx.doi.org/10.5194/acp-4-2025-2004, URL https://acp.copernicus.org/articles/4/2025/2004/.

Whitehouse, L. E., Tomlin, A. S., & Pilling, M. J. (2004b). Systematic reduction of complex tropospheric chemical mechanisms, part II: Lumping using a time-scale based approach. *Atmospheric Chemistry and Physics*, *4*(7), 2057–2081. http://dx.doi.org/10.5194/acp-4-2057-2004, URL https://acp.copernicus.org/articles/4/2057/2004/.

Wiser, F., Place, B. K., Sen, S., Pye, H. O. T., Yang, B., Westervelt, D. M., et al. (2023). AMORE-isoprene v1.0: A new reduced mechanism for gas-phase isoprene oxidation. *Geoscientific Model Development*, *16*(6), 1801–1821. http://dx.doi.org/10.5194/gmd-16-1801-2023, URL https://gmd.copernicus.org/articles/16/1801/2023/.

Xavier, C., Rusanen, A., Zhou, P., Dean, C., Pichelstorfer, L., Roldin, P., et al. (2019). Aerosol mass yields of selected biogenic volatile organic compounds–A theoretical study with nearly explicit gas-phase chemistry. *Atmospheric Chemistry and Physics*, *19*(22), 13741–13758. http://dx.doi.org/10.5194/acp-19-13741-2019, URL https://acp.copernicus.org/articles/19/13741/2019/acp-19-13741-2019.pdf.

Xia, A. G., Michelangeli, D. V., & Makar, P. A. (2009). Mechanism reduction for the formation of secondary organic aerosol for integration into a 3-dimensional regional air quality model: α-pinene oxidation system. *Atmospheric Chemistry and Physics*, *9*(13), 4341–4362. http://dx.doi.org/10.5194/acp-9-4341-2009, URL https://acp.copernicus.org/articles/9/4341/2009/.

Ying, Q., & Li, J. (2011). Implementation and initial application of the near-explicit master chemical mechanism in the 3D community multiscale air quality (CMAQ) model. *Atmospheric Environment*, *45*(19), 3244–3256. http://dx.doi.org/10.1016/j.atmosenv.2011.03.043, URL https://www.sciencedirect.com/science/article/pii/S1352231011003037.