Influence of anthropogenic emissions on organic aerosol formation depending on the physico chemical characteristics of the environment

Zhizhao Wang

École des Ponts ParisTech

THÈSE DE DOCTORAT
de l'École des Ponts ParisTech

Influence of anthropogenic emissions on organic aerosol formation depending on the physicochemical characteristics of the environment

École doctorale N° 531 : SCIENCES, INGÉNIERIE ET ENVIRONNEMENT (SIE)

Sciences et technologies de l'environnement

Thèse préparée au CEREA et à l'INERIS

Thèse soutenue le 4 Mai 2023, par
Zhizhao WANG

Composition du jury:

Miikka, DAL MASO Professeur, Tampere University	*Rapporteur*
Eric, VILLENAVE Professeur, Université de Bordeaux	*Rapporteur*
Maud, LERICHE Professeure associée, LaMP	*Examinatrice*
Marie, CAMREDON Maîtresse de conférences, LISA	*Examinatrice*
Matthias, BEEKMANN Directeur de recherche, LISA	*Examinateur*
Antoine, WAKED Docteur, ELARD	*Examinateur*
Karine, SARTELET Directrice de recherche, CEREA	*Directeur de thèse*
Florian, COUVIDAT Ingénieur de recherche, INERIS	*Co-directeur de thèse*

THÈSE DE DOCTORAT DE l'ECOLE NATIONALE DES PONTS ET CHAUSSÉES

Spécialité

Science et Technologie de l'Environnement

École doctorale Sciences, Ingénierie et Environnement (Paris)

Présentée par

Zhizhao WANG

Pour obtenir le grade de

DOCTEUR DE l'ECOLE NATIONALE DES PONTS ET CHAUSSÉES

Sujet de thèse:

Influence of anthropogenic emissions on organic aerosol formation depending on the physicochemical characteristics of the environment

Soutenue le 4 Mai 2023, devant le jury composé de:

Miikka DAL MASO (Tampere University)	Rapporteur
Eric VILLENAVE (Université de Bordeaux)	Rapporteur
Maud LERICHE (LaMP)	Examinatrice
Marie CAMREDON (LISA)	Examinatrice
Matthias BEEKMANN (LISA)	Examinateur
Antoine WAKED (ELARD)	Examinateur
Karine SARTELET (CEREA)	Directrice de thèse
Florian COUVIDAT (INERIS)	Co-directeur de thèse

Remerciements

I would like to first express my sincere gratitude to the PhD jury for their interest and availability to participate in my PhD evaluation. Your comments and feedback are undoubtedly valuable to my work. Special thanks go to Dr. Miikka Dal Maso for traveling all the way from Finland to attend my PhD defense, and to Dr. Marie Camredon for her detailed suggestions and revisions on my thesis, which helped me a lot to improve it. I would also like to thank the jury members of my first and second PhD committee meetings for their time and guidance.

Next, I want to extend my deepest appreciation to my supervisors, Dr. Karine Sartelet and Dr. Florian Couvidat, who have been my guiding lights throughout this academic journey. I enjoyed the moments when we shared new and exciting ideas and brought them to life. Starting with my Master's internship in 2019, it was your support and guidance that allowed me to overcome the challenges I faced in my study. I feel incredibly fortunate to have had this opportunity to work with you.

I also want to express my gratitude to my colleagues and friends at CEREA. Special thanks to Youngseob, who was always around to help me with coding and technical issues, and to Victor, who helped me understand SOA mechanism reduction. My dear sweetie, Lya, has been like a sister to me, giving me a lot of love and support when I needed it most. I would also like to thank Thibaud, Quentin, Oscar, Aurelie, Pierre, Alice, Alexis, Soojin, Yijie, Aurélien, Maxence, Balvet, Lucas, Lydie, and Patrick, as all of you have made CEREA a second home to me.

A big thank you to the lovely people I met during my PhD, especially Yunyi, Sijia, Xin, Zhidian, and Jing. You have made my PhD life colorful and have motivated me to explore life outside the lab. I truly cannot imagine this journey without your friendship.

To my family, I am forever grateful for your unconditional support. Even though we have been far apart, you have always been the source of my energy, my rock, and my solid shield.

At one point during these four years, I felt like the journey was long, but in retrospect, the time seems to have flown by too quickly. I can't count how many times I worked until midnight, how many times I got frustrated with my work, and how many times I had to redo all of my simulations due to bugs found in the code. In the end, I want to thank myself for not giving up. From all the obstacles and mistakes, I have learned that not all my efforts will be rewarded, but I consider myself very lucky to have had the experience of being able to devote myself fully to this work, and to have had the encouragement of all of you throughout my doctoral journey.

Résumé

Les aérosols organiques secondaires (AOS) affectent la qualité de l'air, le climat et la santé humaine. Dans la troposphère, les composés organiques volatils (COV) peuvent subir une chimie multigénérationnelle, et leurs produits d'oxydation se condenser sur les particules existantes pour former des AOS. Ainsi, leur formation implique de nombreuses réactions et espèces, et dépend des conditions environnementales.

Notre compréhension actuelle de la formation des AOS peut être décrite par des mécanismes chimiques détaillés (par exemple, le Master Chemical Mechanism (MCM) et le Peroxy Radical Autoxidation Mechanism (PRAM)). Cependant, en raison de limitations en temps de calcul, les modèles de chimie-transport (MCT) ne peuvent pas les utiliser directement. Ils utilisent plutôt des mécanismes implicites avec seulement quelques espèces modèles et réactions. Ces mécanismes implicites sont généralement construits à partir de mesures en chambre et peuvent ne pas avoir la complexité nécessaire pour simuler avec précision les concentrations en aérosols organiques.

Généralement, on estime que les concentrations d'AOS diminueront en raison des réglementations sur les émissions, en particulier dans les zones rurales et périurbaines où les concentrations d'oxydants devraient diminuer. Cependant, certaines études suggèrent que la réduction des émissions anthropiques, en particulier des oxydes d'azote (NOx), peut ne pas conduire à une diminution efficace des concentrations d'AOS mais peut même les augmenter. Avec des mécanismes implicites d'AOS hautement simplifiés, cette interaction complexe entre la réduction des émissions et la formation d'AOS pourrait ne pas être simulée de manière fiable dans les MTC.

Pour améliorer la formation des AOS dans les MTC, le GENerator of Reduced Organic Aerosol Mechanisms (GENOA) a été développé. GENOA réduit les mécanismes chimiques détaillés en mécanismes d'AOS semi-explicites qui sont suffisamment petits pour être utilisés dans les MCT. Les mécanismes obtenus peuvent être personnalisés par les utilisateurs selon la précision souhaitée, et préservent les propriétés physico-chimiques des AOS. GENOA v1.0 a été appliqué au mécanisme de formation des AOS de sesquiterpène (SQT) à partir de MCM, résultant en un mécanisme dont la taille est réduite à moins de 2 % de celle du MCM avec une erreur moyenne inférieure à 3 %.

Pour améliorer l'efficacité de la réduction et traiter les mécanismes de plusieurs précurseurs d'AOS simultanément, une approche de réduction parallèle est utilisée dans GENOA v2.0. Pour les mécanismes (MCM + PRAM) de trois monoterpènes (MTs), le mécanisme global est ainsi réduit de 93 % avec une erreur inférieure à 3 %.

Le mécanisme d'AOS biogénique généré par GENOA v2.0 (GBM), comprenant les schémas d'AOS pour MT et SQT, a ensuite été implémenté dans le MCT CHIMERE. Les concentrations simulées avec GBM sur l'Europe pendant l'été (juin-août, 2018) sont plus élevées que celles simulées avec le mécanisme implicite Hydrophilic/Hydrophobic Organics (H^2O), et les AOS sont plus oxydés. Le mécanisme GBM améliore les comparaisons

modèle/mesures pour les concentrations en aérosol organique.

Avec une réduction de 50 % des émissions anthropiques de NOx, le mécanisme GBM prédit une augmentation des AOS (6,5 %) due à une augmentation des AOS provenant de MT (15 %). Les réductions de NOx favorisent la formation de molécules hautement oxygénées (HOM) par auto-oxydation, entraînant une augmentation de la concentration en AOS provenant de MT. La diminution des concentrations de NOx favorise également les voies chimiques entraînant une augmentation des concentrations d'AOS non-HOM provenant de MT.

Dans l'ensemble, ce travail montre que les mécanismes détaillés d'AOS sont nécessaires aux MCT pour simuler les variations des concentrations d'AOS selon l'environnement physico-chimique, et pour évaluer avec précision l'impact de scénarios de réduction des émissions.

Influence des émissions anthropiques sur la formation d'aérosol organique en fonction des caractéristiques physico-chimiques de l'environnement

Mots clés: Modélisation, Aérosols Organiques Secondaires, Qualité de l'Air

Abstract

Secondary organic aerosols (SOAs) affect air quality, climate, and human health. In the troposphere, volatile organic compounds (VOCs) can undergo multi-generation chemistry, and their oxidation products can condense onto existing particles to form SOAs. Consequently, SOA formation involves numerous reactions and species, depending on environmental conditions. Our up-to-date understanding of SOA formation can be described by detailed VOC mechanisms (e.g., the Master Chemical Mechanism (MCM) and the Peroxy Radical Autoxidation Mechanism (PRAM)). However, due to computational limitations, chemistry-transport models (CTMs) are unable to directly employ detailed SOA mechanisms but use rather implicit mechanisms with only a few model species and reactions. These implicit mechanisms are usually built from chamber measurements and may lack the necessary complexity to accurately represent the concentrations of organic particles.

Typically, SOA concentrations are predicted to decrease due to emission regulations, particularly in rural and peri-urban areas where oxidant concentrations are expected to decrease. However, some studies suggest that reducing anthropogenic emissions, especially nitrogen oxides (NOx), may not lead to an efficient decrease in SOA concentration but may even increase it.

With highly simplified implicit SOA mechanisms, this complex interaction between emission reduction and SOA formations may not be reliably simulated in CTMs. Therefore, there is a need to improve the representation of SOA formation in CTMs, especially for emission regulation evaluation.

To address this issue, the GENerator of Reduced Organic Aerosol Mechanisms (GENOA) has been developed. GENOA reduces detailed chemical mechanisms into semi-explicit SOA mechanisms that are small enough to be used for regional CTM simulations. The obtained SOA mechanisms can be customized by users to the desired accuracy, and preserve the physicochemical properties of SOA. GENOA v1.0 was applied to the sesquiterpene (SQT) SOA formation mechanism from MCM, resulting in a reduced SOA mechanism within 2 % of the MCM size and introducing an average error of less than 3 %. To improve the reduction efficiency and to process mechanisms of multiple SOA precursors simultaneously, a parallel reduction approach is employed in GENOA v2.0 GENOA v2.0 was applied to the mechanisms (MCM + PRAM) of three monoterpenes (MTs), where the mechanism is reduced up to 93 % with an error of less than 3 %.

The GENOA-generated biogenic SOA mechanism (GBM), including MT and SQT SOA schemes trained with GENOA v2.0, was then implemented in the CTM model CHIMERE. Simulations with GBM over Europe during summer (June-August, 2018) estimate more oxidized OAs with higher concentration than those simulated with the implicit Hydrophilic/Hydrophobic Organics (H^2O) mechanism. The GBM mechanism leads to an improvement of the model to measurement comparisons for organic aerosol concen-

trations. With a 50 % reduction in NOx anthropogenic emissions, the GBM mechanism predicts an increase in total SOA (6.5 %) due to an increase in MT SOA (15 %). When NOx is reduced, the formation of highly oxygenated molecules (HOMs) by auto-oxidation is enhanced, leading to an increase in MT SOA concentration. The decrease of NOx concentrations also favors chemical pathways resulting in an increase of MT non-HOM concentrations.

Overall, this work shows that detailed SOA mechanisms are necessary for CTMs to preserve the variations in the physical-chemical environment of the SOA concentrations, and to accurately evaluate the impact of emission reduction scenarios.

Influence of anthropogenic emissions on organic aerosol formation depending on the physicochemical characteristics of the environment

Keywords: Modeling, Secondary Organic Aerosols, Air Quality

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Air quality concerns everybody's life. Although air quality seems much improved compared to twenty years ago, it is still estimated that exposure to air pollution is responsible for millions of deaths and loss of healthy lifespan each year. Organic aerosols (OA) are crucial atmospheric pollutants and have various sources, including biogenic and anthropogenic emissions. They can be formed in the atmosphere through physical-chemical transformations, known as secondary organic aerosols (SOAs), which substantially affect air quality, climate, and human health.

This chapter provides a brief overview of air pollution, highlighting the gas-to-particle processes that lead to the formation of secondary organic aerosols (SOAs), as well as the current state of SOA modeling. Furthermore, the interaction between anthropogenic emissions and SOA formation is explored, underlining the highly non-linear effects caused by the complexity of gas-phase chemistry on SOA formation and aging. To better understand the formation of organic aerosols, it is important to investigate how anthropogenic emissions interact with the physicochemical characteristics of the environment.

Contents

1 Air pollution

Air pollution is driven by chemical, physical, and biological gaseous and particulate pollutants emitted or formed in the atmosphere, contaminating both the indoor and outdoor environment and the atmosphere's natural characteristics to a level that affects vegetation, climate, and human health. In this thesis, the discussion on air pollution focus on outdoor air pollution and its long-term impacts (independent of pollution source).

1.1 Sources

Air pollution mainly comes from anthropogenic emissions as side effects of human activities; biogenic emissions related to vegetation's natural growth and decay; and other sources such as biomass burning and sea salt activities.

1.1.1 Anthropogenic emission

A variety of human activities contribute to anthropogenic emissions, notably transportation, industry, power generation, agriculture, and construction. As an example, poor air quality is typically found in urban areas with high levels of traffic congestion, coastal areas near shipping routes, as well as the surrounding areas of industrial plants.

There are three major chemical and physical processes that drive anthropogenic emissions, including combustion, volatilization, and mechanical processes:

Combustion The combustion of fossil fuels (e.g., petroleum, coal, wood, and natural gas), such as those used to power transportation and generate electricity, contributes primarily to anthropogenic emissions. Presently, most combustion processes do not completely oxidize carbonaceous compounds. This results in a large amount of gaseous and particulate pollutants, including carbon monoxide (CO), anthropogenic volatile organic compounds (AnVOCs), and black carbon ([André, 2020]). Even when combustion is considered completed, carbonates still produce greenhouse gases (e.g., carbon dioxide - CO_2). Depending on the quality of the fuel, the combustion equipment, and the combustion technology used, various pollutants can be released from both incomplete and completed combustion, including nitrogen oxides (NOx) and sulfur dioxide (SO_2). There are also natural processes that cause air pollution through combustion, such as biomass burning.

Volatilization Various human activities (e.g., industrial processes, construction and demolition, shipping, and using household products) involve the application of solvents, chemicals, and other materials that can evaporate easily at room temperature. Those activities, therefore, release volatile organic compounds (VOCs) along with other pollutants (e.g., SO_2 from sulfur-containing products, Formaldehyde (HCHO) from certain building materials). Agricultural runoff is another typical source of air pollution caused by volatilization. As a result of the decomposition of animal manure, fertilizers, and other organic matter from livestock, pollutants such as ammonia (NH_3) and VOCs can be released into the atmosphere.

Mechanical Processes Mechanical processes are involved in many human activities (e.g., transportation, construction, and manufacturing), which cause material abrasion and particle fugitives, leading to particulate pollutants being released into the atmosphere. During transportation, abrasions, such as wear and tear on brakes, tires, and road surfaces, can release metals and particles into the atmosphere. Construction activities, unpaved roads, and open storage piles can aggravate dust pollution. Aside from anthropogenic pollution, natural mechanical activities, such as sea salt emissions, also contribute to air pollution ([Xu, 2021a]).

1.1.2 Biogenic emission

Biogenic emissions constitute one of the natural sources of air pollution. Plants and other living organisms can emit biogenic volatile organic compounds (BVOCs) through their metabolic processes. BVOCs, dominated by isoprenoids (isoprene and monoterpenes), are responsible for a significant fraction of atmospheric VOCs, especially in rural areas with high vegetation cover. Although biogenic sources of air pollution are natural and occur without human intervention, they still have a profound impact on air quality by interacting with anthropogenic pollutants and participating in tropospheric chemistry. Among all types of BVOCs, isoprene and terpenes (i.e., monoterpenes and sesquiterpenes) are particularly reactive under atmospheric conditions, contributing to ground-level ozone and secondary aerosol formation ([Kesselmeier, 1999; Sartelet, 2012]).

1.2 Gaseous pollutants

[Turner, 2020] summarized important gaseous air pollutants and their typical sources:

Gaseous pollutants	**Typical sources**
Nitrogen oxides (NOx)	Fossil fuel combustion (e.g., diesel vehicle)
Volatile organic compounds (VOCs)	Petrochemical solvents, evaporated fuels, biogenics, incomplete combustion, chemical processing
Ground-level Ozone (O_3)	Formed via photochemical reactions in the atmosphere from NOx and VOCs
Sulfur dioxide (SO_2)	Fuel combustion, smelters
Carbon monoxide (CO)	Fuel combustion, biomass burning wildfires
Ammonia (NH_3)	Livestock yards, agricultural activities

Table 1.1: Gaseous air pollutants and their typical origins. Source:[Turner, 2020]

Among them, NOx, VOCs, and O_3 are the key pollutants that have a substantial impact on severe air pollution such as photochemical smog and particle pollution. A concise introduction to these pollutants and their roles in tropospheric chemistry is, therefore, presented in this section.

1.2.1 Nitrogen oxides

Nitrogen oxides (NOx), commonly representing nitric oxide (NO) and nitrogen dioxide (NO_2), are typically released as exhaust from combustion processes, making significant contributions to air pollution. NOx are highly reactive gases that play a crucial role in the formation and destruction of many pollutants and oxidants in the atmosphere. They are, therefore, the precursors to many secondary pollutants, including ground-level ozone, nitric acid, and organic nitrates. Nitric acid and organic nitrates can further condense and form secondary aerosols. Additionally, the oxidation of NOx in the atmosphere results in the formation of free radicals such as hydroxyl (OH) and nitrate (NO_3) radicals, which are two of the most common atmospheric oxidants alongside ozone.

1.2.2 Volatile organic compounds

Anthropogenic and biogenic organic gases emitted into the atmosphere are known as volatile organic compounds (VOCs). They can undergo complex gas-phase oxidation in the troposphere, with oxidation products being less volatile and/or more soluble. These products can condense on existing particles, leading to the formation or reshaping of secondary organic aerosols (SOAs). Meanwhile, the degradation of VOCs consumes atmospheric oxidants and influences ground-level ozone and NOx concentrations. As a result, the ratio of NOx to VOC is frequently used to define the chemical regimes, where favorable ozone formation and VOC degradation pathways differ. ([Atkinson, 2000]).

1.2.3 Ground-level ozone

Ground-level ozone is formed through a series of chemical reactions involving NOx and VOCs in the presence of sunlight ([Atkinson, 2003]). Unlike the protective ozone in the upper atmosphere, which shields the Earth from harmful ultraviolet radiation, ground-level ozone is a major component of photochemical smog. For the purpose of this discussion on air pollution, ground-level ozone will be referred to as ozone or O_3, without further specification.

During sunny afternoons with high traffic volumes and industrial activity, ozone levels can often accumulate to unhealthy levels in urban areas. Even in colder weather or rural areas, ozone can still reach high concentrations due to tropospheric chemistry and long-range transportation ([Seigneur, 2019]). Additionally, it serves as one of the most abundant atmospheric oxidants, alongside NO_3 and OH radicals, actively participating in the degradation of various pollutants, such as VOCs.

1.3 Particulate Pollutants

Atmospheric aerosols, also known as particulate matter (PMs), are complex mixtures of tiny particles (solid or liquid) suspended in the air. The size distribution and chemical composition of these particles vary considerably depending on the emission source and the atmospheric conditions in which they are suspended.

1.3.1 Size distribution

Aerosols range widely in size, with aerodynamic diameters from a few nanometers (nm) to tens of micrometers (μm). Depending on their size, aerosols are typically classified as coarse, fine, and ultrafine particles.

Coarse particle Coarse particles have a large aerodynamic diameter between 2.5 μm and 10 μm, including particles such as dust and pollen, which are emitted directly into the air as primary pollutants. The term PM_{10} refers to aerosols with a diameter smaller than 10 μm, thus including coarse particles.

Fine particles or $PM_{2.5}$ Fine particles have an aerodynamic diameter of 2.5 μm or less. They are mainly formed in the atmosphere via chemical-physical processes from gas-phase species rather than being emitted directly. Fine particles encompass PM_1 (particles with a diameter of 1 μm or less), which is one of the commonly used terms in aerosol pollution monitoring and research, alongside PM_{10} and $PM_{2.5}$.

Ultrafine particles Ultrafine particles have an aerodynamic diameter of 0.1 μm or less. Due to their tiny sizes, ultrafine particles have unique properties and atmospheric behaviors. For instance, their large surface areas relative to their mass enable them to carry a substantial load of chemicals.

1.3.2 Chemical composition

Aside from liquid water, aerosols can contain a variety of organic, inorganic, and insert compounds, including primary fractions directly emitted from sources and secondary fractions formed in the atmosphere. As displayed in Fig. 1.1, the chemical composition and concentration of aerosols vary from location to location.

[McNeill, 2017] summarized the chemical composition of atmospheric aerosols at all sizes and the estimated annual production rates as presented in Table 1.2. It suggests that sea salts (e.g., chloride and sodium) and mineral dust from natural sources have much larger emissions than other compounds. They are the main natural sources of primary aerosols. The secondary compounds also have a noticeable production, which can be further classified into secondary organic and inorganic aerosols.

Composition	Production (Tg/yr)
Secondary inorganics	69.9
Mineral dust (natural)	10^3
Sea salt	1 - 3 $\times$ 10^4
Black carbon	7.5 (2 - 29)
Primary organics	33.9 (17 - 77)
Secondary organics	140 (50 - 380)

Table 1.2: The chemical composition of aerosols and annual production estimation (with range if available). Source: [McNeill, 2017]

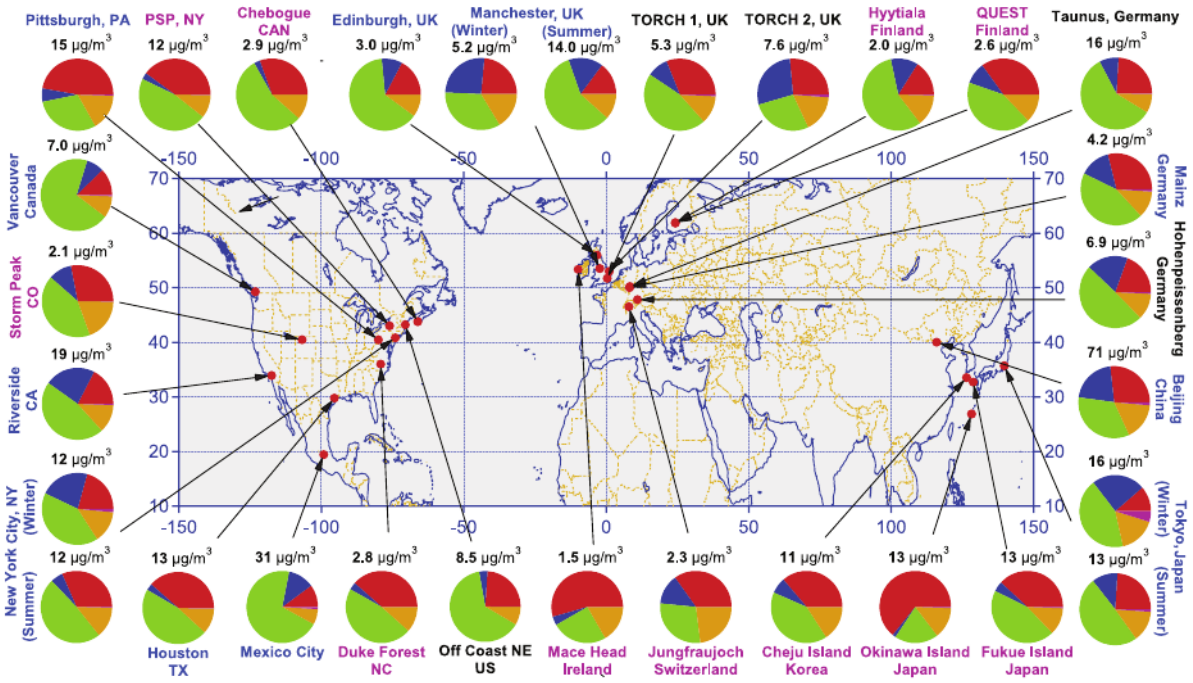


Figure 1.1: The chemical composition and concentrations of PM_1 in the Northern Hemisphere. Source: [Zhang, 2007]

Secondary inorganic aerosols Secondary inorganic aerosols (SIAs) includes nitrate, ammonium, sulfate, and chloride salts, which are formed through atmospheric chemical physical activities. Sulfate is formed through the oxidation of SO_2 into sulfuric acid (H_2SO_4), while nitrate is formed through the oxidation of NO_2 into nitric acid (HNO_3). Meanwhile, ammonium mainly originates from NH_3, and chlorides are formed due to marine activities. The formation of ammonium is associated with sulfate and nitrate concentrations ([Seigneur, 2019]), while in sea salt aerosol, a negative correlation is pronounced between chloride and nitrate ([Wu, 1994]).

Secondary organic aerosols Secondary organic aerosols (SOAs) result from the gas-particle mass transfer of numerous gas-phase organic compounds with low volatility that are either formed or emitted into the air. Due to the large amount and the difficulty in measurements, a large fraction of organic compounds is still unspeciated [Donahue, 2006].

As previously mentioned, the chemical compositions of aerosols vary significantly with size. Figure 1.2 presents an example of the concentration and chemical composition of urban PM_{10} at different sizes. It is observed that dust is the dominant component in coarse particles, while inorganics are mostly found in fine particles, and organics concentrate in ultrafine particles.

In the Northern Hemisphere, organic fractions contribute an average of 45 % total PM_1 (ranging from 18 % to 70 %), while inorganic fractions contribute the rest (sulfate with an average of 32 %, nitrate with 10 %, ammonium with 13 %, and chloride with 0.6 %) ([Zhang, 2007]).

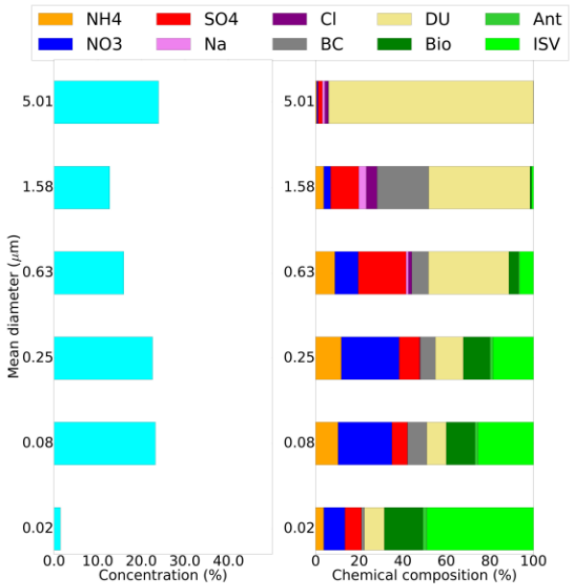


Figure 1.2: PM_{10} aerosol size distribution and chemical composition simulated in Paris streets. The analyzed aerosols include ammonium (NH_4), nitrate (NO_3), sulfate (SO_4), sodium (Na), chlorine (Cl), black carbon (BC), dust (DU), and organic aerosols. Organic aerosols contain biogenic emissions (Bio), anthropogenic emissions (Ant), and oxidant products of both biogenic and anthropogenic sources (ISV). Source: [Lugon, 2021].

1.4 Health and Environmental and health impacts

1.4.1 Environmental effects

Air pollution poses negative effects on all components of the environment, including groundwater, soil, and air. Severe environmental issues triggered by air pollution include haze, acid rain, and global warming.

Haze Haze is a type of air pollution that reduces the transparency of the atmosphere by dispersing fine particles in the air. Figure 1.3 shows one type of haze called photochemical "smog", which is often visible as a fog-like haze near urban areas during warm weather. Under favorable atmospheric conditions, sunlight triggers chemical reactions between gases and gas-to-particle transformation, leading to the formation of many pollutants that cause haze with reduced visibility. Fine particles are the main cause of haze formation in metropolitan areas, with ozone a byproduct involved in gas-phase chemistry. Haze is a harmful outdoor condition, while both ozone and fine particles are causing more serious health effects.

Acid rain Acid rain, also known as acid deposition, is precipitation that contains acidic compounds from the atmosphere. In the atmosphere, gases such as SO_2 and NO_2 can react and form acidic compounds such as sulfuric and nitric acids that eventually deposit as acid rain. The acidity of acid rain can alter nutrient balances in aquatic ecosystems, damage vegetation, and contaminate waterways. Constructions, outdoor sculptures, and buildings may also be damaged by acid rain. As a result of contaminated drinking water

Figure 1.3: Photos taken in Beijing, China, compare the visibility on a "haze" day (left, June 19^{th}, 2009) and on a clear day (right, June 22^{nd}, 2009). Courtesy of China Air Daily.

sources and the inhalation of acid aerosols, acid rain can also have a detrimental effect on human health.

Global warming Global warming is the result of the increase in the Earth's surface temperature caused by human activities. As greenhouse gases (e.g., CO_2) accumulate in the atmosphere, they absorb infrared radiation from the Earth's surface that has been heated by the sun, warming the atmosphere. This increase in temperature is a reflection of the disturbance of the balance between incoming and outgoing radiation energy. As a result of global warming, the melting ice caps and rising sea levels are contributing to a heightened risk of flooding and the spread of vector-borne diseases. It is also evident that in recent years, there has been a noticeable increase in the frequency of extreme weather conditions that may be due to global warming, including continuous heat waves and rainstorms. Without decisive action on global warming, there is no doubt that humans will continue to suffer its consequences.

Notably, aerosols have a significant impact on the climate, as illustrated in Fig. 1.4. Their chemical lifetime, optical and hygroscopic properties impact the cloud–aerosol–radiation interactions ([Seinfeld, 2016]), further affecting climate systems and contributing to climate change. One key effect is the influence of aerosols on cloud formation. Particles can serve as nuclei around which water droplets or ice crystals can form, altering the size, brightness, and lifespan of clouds, as well as affecting precipitation patterns. Additionally, by scattering and absorbing sunlight, aerosols can influence cloud brightness while also impacting the coalescence and growth of cloud droplets. Due to their small size, aerosols have an extended half-life in the atmosphere, which can result in prolonged suspension and possibly spread to distant areas where people and the environment might be exposed to the same level of pollution ([Manisalidis, 2020]).

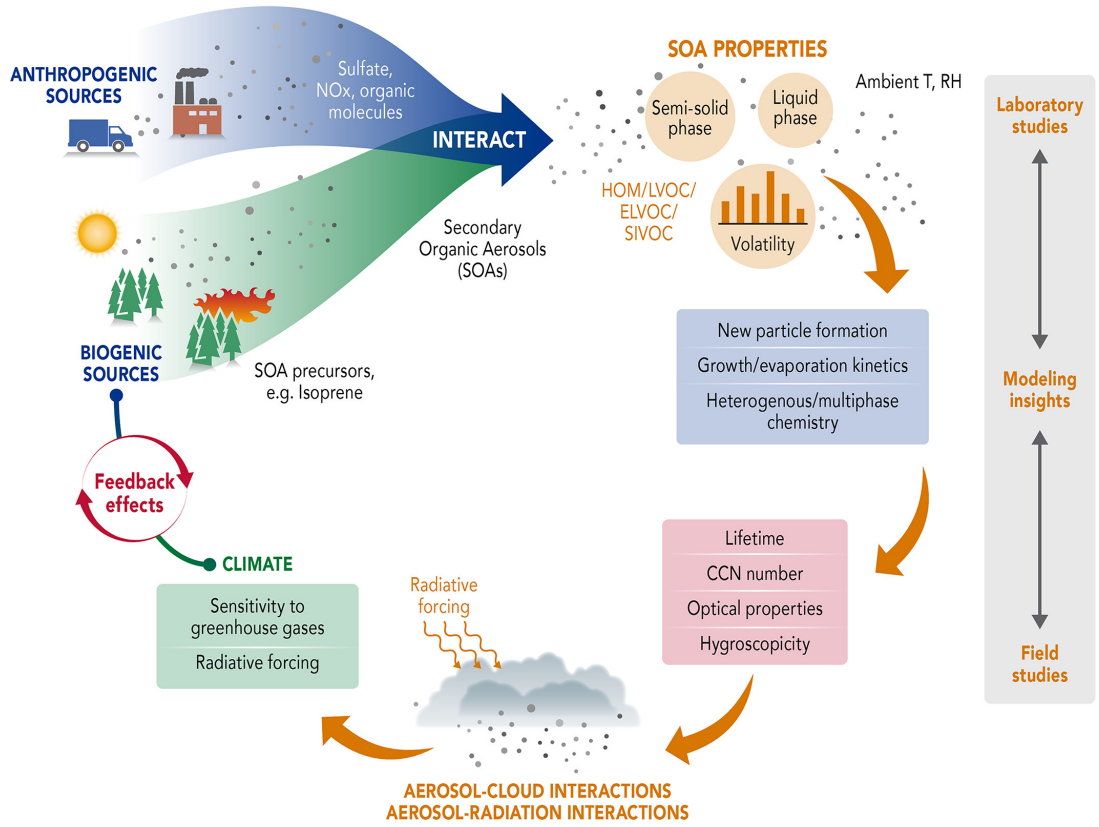


Figure 1.4: Processes related to the climatic impacts of SOAs, illustrating the complex interactions between aerosol formation and climate change. Source: [Shrivastava, 2017].

1.4.2 Health impacts

In addition to the environmental consequences, air pollution poses a severe threat to living organisms. Numerous studies, such as those conducted by [Kampa, 2008; Breysse, 2013; Schraufnagel, 2019], have demonstrated that air pollution poses numerous negative effects on human health. According to the World Health Organization (WHO), almost all people worldwide (99 %) breathe air containing high levels of pollutants exceeding guideline limits, with countries of lower and moderate economic conditions being disproportionately affected ([Goshua, 2022]). As a result of its adverse impacts, air pollution is a leading environmental health risk worldwide, causing numerous diseases.

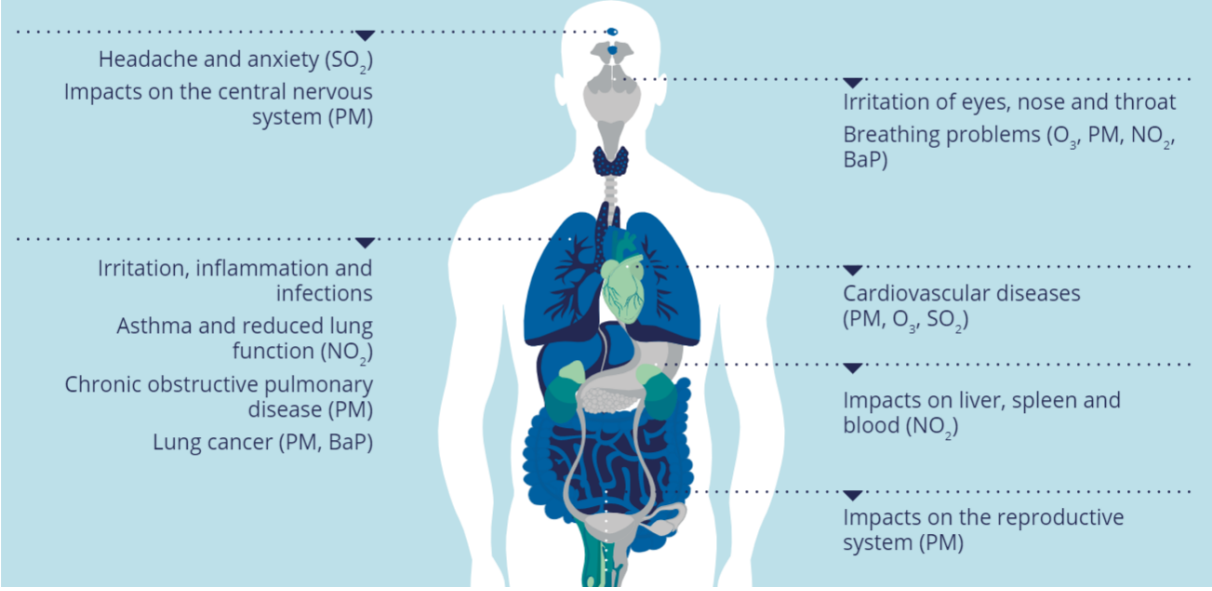


Figure 1.5: A visualization of health impacts of different air pollutants. Source: European Environment Agency (EEA).

Human bodies are vulnerable to the damaging effects of air pollution when inhaled or simply exposed to it. As illustrated in Fig. 1.5, numerous pollutants can trigger adverse health problems and diseases affecting multiple organs in the body, such as the head, lungs, and cardiovascular system. Exposure to unhealthy pollutants for a short period can result in headaches, dizziness, and respiratory problems, while long-term exposure even to low levels of pollutants can lead to irreversible damage to the nervous and liver systems ([Schraufnagel, 2019]).

A particular concern is the inhalation of atmospheric aerosols, as aerosols can carry various pollutants to the human body, such as heavy metals that can be extremely harmful. Upon inhalation, coarse particles are mostly deposited in the upper respiratory tract. Fine particles can penetrate profoundly into the lungs, while ultrafine particles may reach even deeper, entering the bloodstream and spreading into other organs. Consequently, pollutants transported through aerosols may penetrate the respiratory system and cause respiratory, cardiovascular, and central nervous system disorders, reproductive problems, and possibly cancer ([Manisalidis, 2020]).

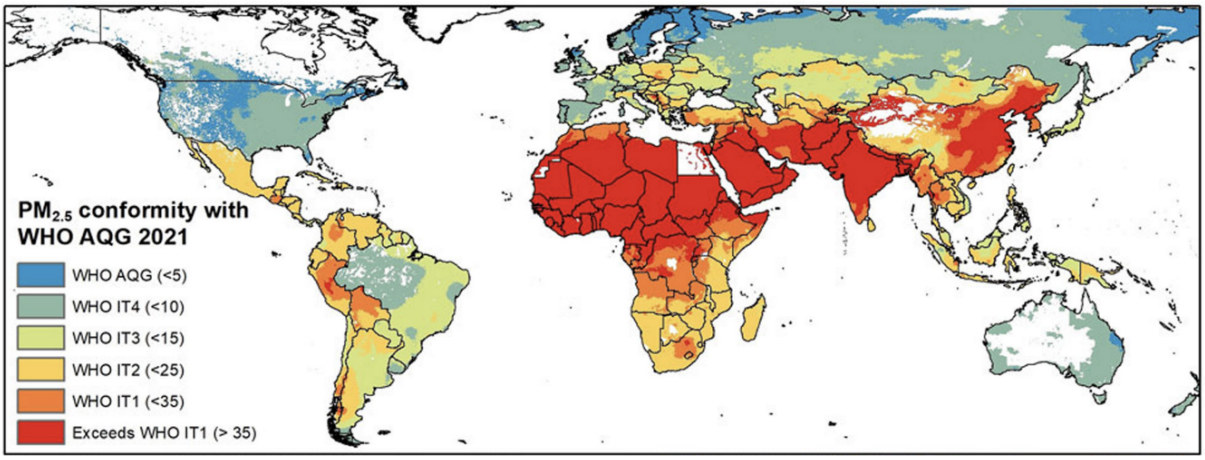


Figure 1.6: Comparison of $PM_{2.5}$ with WHO 2021 AQG, suggesting that current aerosol pollution is still severe to human life. Source: [Goshua, 2022].

1.5 Guideline and regulation

In response to the growing awareness of the harmful effects of air pollution, guidelines and regulations have been developed to help control and address the problem. This section includes an example of the World Health Organization (WHO) guidelines and European Union (EU) regulations on air quality.

1.5.1 WHO guideline

In 2021, WHO has updated its quantitative air quality guidelines (AQGs) and the corresponding interim target levels for six key pollutants: $PM_{2.5}$, PM_{10}, SO_2, NO_2 (indicator for NOx), O_3, and CO. These guidelines are summarized in Table 1.3, providing the recommended or lowest level of exposure known to cause health problems. Among all six pollutants, fine particle levels are the ones that currently exceed the AQGs the most. As evident from Fig. 1.6, a significant number of people reside in regions with high $PM_{2.5}$ exposure that greatly exceeds the WHO-recommended levels. This highlights the urgent need for further actions to regulate aerosol concentration.

1.5.2 EU regulation

EU has a comprehensive regulatory framework for air pollution, including publishing air quality stands (e.g., 2008/50/EC Directive shown in Table 1.4), developing real-time air quality monitoring tools (e.g., CAMs monitor [Granier, 2019]), and setting regulation targets for air pollutants and greenhouse gases (e.g., European Green Deal [Wolf, 2021]). Similar regulations and standards are also established in other countries and regions (e.g., US's NAAQS, China's GB 3095—2012 for ambient air quality standards).

Pollutant	Averaging time	Interim targets				AQG	Evidence certainty
		1	2	3	4		
$PM_{2.5}$, μg/m^3	Annual	35	25	15	10	5	High
	24-h[a]	75	50	37.5	25	15	High
PM_{10}, μg/m^3	Annual	70	50	30	20	15	High
	24-h[a]	150	100	75	50	45	High
O_3, μg/m^3	Peak season[b]	100	70	–	–	60	Low-moderate
	8-h[a]	160	120	–	–	100	High
NO_2, μg/m^3	Annual	40	30	20	–	10	Moderate-high
	24-h[a]	120	50	–	–	25	High
SO_2, μg/m^3	24-h[a]	125	50	–	–	40	Low-high
CO, mg/m^3	24-h[a]	7	–	–	–	4	Moderate

[a]99th percentile of the distribution of daily values (3–4 exceedance days per year).
[b]Average of daily maximum 8-h mean O_3 concentration in a period of six consecutive months with the highest 6-month running-average O_3 concentration.

Table 1.3: AQG guidelines for six critical air pollutants. Source: [Goshua, 2022].

Pollutant	**Concentration**	**Sampling duration**	**Statistical form (number of authorized exceedances)**
Pb	0.5 μg m^{-3}	1 year	(0)
CO	10 mg m^{-3}, 9 ppm	8 h	(0)
SO_2	350 μg m^{-3}, 133 ppb	1 h	99.7th percentile (24)
	125 μg m^{-3}, 47 ppb	24 h	99.2th percentile (3)
NO_2	200 μg m^{-3}, 106 ppb	1 h	99.8th percentile (18)
	40 μg m^{-3}, 21 ppb	1 year	(0)
O_3^h	120 μg m^{-3}, 61 ppb	8 h	93th percentile (25)
PM_{10}	50 μg m^{-3}	24 h	90.4th percentile (35)
	40 μg m^{-3}	1 year	(0)
$PM_{2.5}$	25 μg m^{-3}	1 year	(0)
C_6H_6	5 μg m^{-3}, 1.5 ppb	1 year	(0)

Table 1.4: EU air quality standards. Directive 2008/50/EC. Source: [Seigneur, 2019]

2 Formation of organic aerosols

As mentioned in the previous section, organic aerosols (OAs) are significant air particulate pollutants. While some are emitted directly (e.g., biogenic particles like pollen and spores or anthropogenic particles from combustion), the majority of atmospheric OAs are SOAs, formed through the oxidation of VOCs and subsequent aerosol dynamic processes (e.g., condensation).

The SOA formation from VOC oxidation in the troposphere is a complex process. After being emitted, VOCs undergo multi-generation oxidation in the troposphere, forming oxidation products with low volatility. Those low-volatile gas-phase organic may condense on existing particles via gas-particle partitioning and form SOA. This section summarizes three processes involved in SOA formation: the gas-phase chemistry that leads to the formation of gas-phase low-volatile compounds (Sect. 2.1), the gas-particle thermodynamics that drives the distribution of organics on both gas and particle phases depending on their volatility (Sect. 2.2), and the aerosol dynamics that form aerosols physical activities (Sect. 2.3).

2.1 Gas-phase chemistry: Low-volatile organic formation

2.1.1 VOC oxidation

The degradation of the emitted VOCs is initiated by atmospheric oxidants (i.e., Ozone, OH radical, and NO_3 radical) or triggered by photolysis. Reactions by photolysis and oxidation with OH are favorable during the daytime, with NO_3 during the nighttime and O_3 throughout the entire day.

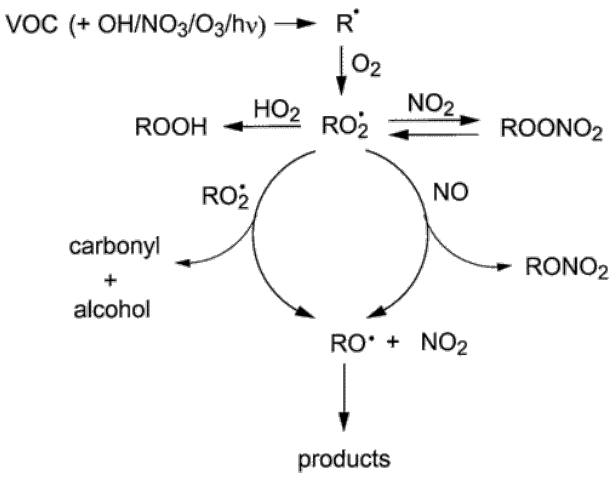


Figure 1.7: General gas-phase chemical pathways involved in VOC degradation. Source: [Atkinson, 2003].

As illustrated in Fig. 1.7, the degradation of VOCs results in the formation of alkyl or substituted alkyl radicals ($R^{\cdot}$), which then rapidly become key degradation intermediates: organic peroxy ($RO_2^{\cdot}$ or RO_2) and alkoxy ($RO^{\cdot}$ or RO) radicals. Subsequently, the radicals resulting from the initial VOC degradation undergo reactions with NOx, HO_2, or other RO_2s, leading to the formation of stable oxidation products such as carbonyl compounds, nitrates (RONO2), and peroxides (ROOH) ([Atkinson, 2003]). Typically, VOC

degradation follows either fragmentation or functionalization pathways. Fragmentation breaks carbon chains, leading to the formation of more volatile products and ultimately CO_2. Conversely, functionalization adds functional groups to the molecules, resulting in compounds with lower volatility and/or higher solubility that tend condense and form SOAs. Depending on the structure of the parent VOCs, competitive oxidation processes occurs between different degradation pathways. Meanwhile, there is also competition between different pathways for RO_2 reactions. The fate of RO_2 radicals is dependent on NOx and HO_2 concentration. Under high NOx conditions (RO_2 + NO is the dominant reaction), RO_2 further undergoes transformation to RONO2 and other compounds, while RO_2 leads to ROOH with low NOx conditions (RO_2 + HO_2 is the dominant reaction).

The oxidation products formed from the VOC initial degradation can further undergo oxidation with atmospheric oxidants. These successive reactions are also referred to as multi-generation reactions, where one generation stands for an ensemble of reactions leading to the formation of stable products. Figure 1.8 summarizes the multi-generation reactions into a cycle of three types of reactions ([Aumont, 2005]). The initial oxidation of a VOC with atmospheric oxidants or photolysis by sunlight promotes RO_2 formation. Afterward, reactions with RO_2 either lead to stable species participating in SOA formation or RO radicals. RO radicals react rapidly with O_2, monomolecular decomposition, or isomerization, resulting in the formation of the next generation of stable species and/or new RO_2 radicals that continue to oxidize. The fate of the multi-generation is terminated by fragmentation to CO_2.

The evolution of SOA concentrations due to the second or higher-generation reactions is referred to as SOA aging ([Donahue, 2006]).

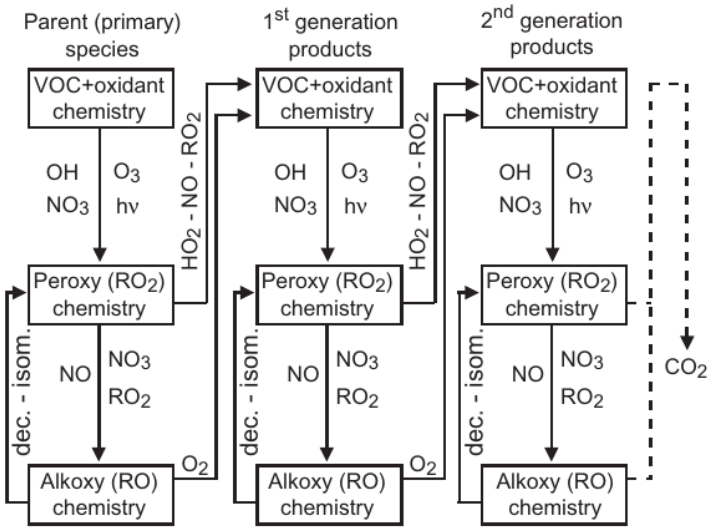


Figure 1.8: Processes of multi-generation degradation by [Aumont, 2005].

2.1.2 HOM formation

Due to SOA aging, some VOCs, such as monoterpenes, can rapidly form large amounts of oxidant products, including highly oxygenated organic molecules (HOMs) that contributes considerably to SOA production (up to 50 % of the total mass of SOA reported by [Roldin,

2019]). HOMs are commonly defined as molecules containing at least six oxygen atoms ([Bianchi, 2019]). Recent research on the mechanism of HOM formation has revealed that HOMs can be formed from autoxidation involving RO_2 reactions.

Autoxidation Autoxidation is a process that rapidly forms oxidized RO_2 with a large number of attached molecular oxygen. In this process, RO_2 can have one or several intramolecular hydrogen-atom shifts, each of which shifts the position of an unpaired electron and adds oxygen atoms.

An example is illustrated in Fig. 1.9 for a ketone molecule. After being initiated by the OH radical (C1 to C2), the ketone molecule undergoes a hydrogen-atom shift, leading to the formation of a molecule with ROOH and RO radical functional groups (C3). The C3 molecule then rapidly undergoes a rapid addition of O_2, resulting in the formation of a new and more oxidized RO_2, ultimately producing a stable dicarbonyl ROOH species.

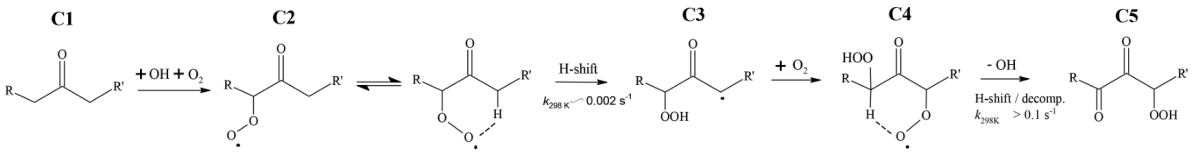


Figure 1.9: Autoxidation involved in the OH-initiated oxidation of a ketone. Copyright 2013 American Chemical Society. See more details in [Bianchi, 2019].

Reactions between organic peroxy radicals Reactions between RO_2, referred to as RO_2 - RO_2 reactions, also contribute to HOM formation. While RO_2-RO_2 reactions generally form two separate molecules (formation of two RO compounds in Eq. 1.1 or a carbonyl and an alcohol in Eq. 1.2), RO_2-RO_2 can in some cases lead to the formation of a ROOR dimer (Eq. 1.3).

$$RO_2 + RO_2 \rightarrow RO + RO + O_2 \tag{1.1}$$

$$RO_2 + RO_2 \rightarrow R_{-H}O + ROH + O_2 \tag{1.2}$$

$$RO_2 + RO_2 \rightarrow ROOR + O_2 \tag{1.3}$$

2.2 Gas-particle thermodynamic

Whether organics with low volatility can condense on existing particles depends on the gas-particle equilibrium, which is the equilibrium of concentrations between the gas and particle phases. Aqueous or organic liquids can absorb low-volatility organic compounds, leading to the formation of organic aerosols. The volatility of a compound is related to its saturation vapor pressure (P_{sat}), which signifies the pressure at which a gas becomes saturated and attempts to condense at a given condition. The gas-particle partitioning of a condensable organic compound is described by Raoult's law, which can be further extended under different assumptions to Pankow's law and Henry's law.

Raoult's law The amount of a compound in an aerosol can be calculated using Raoult's law, which relates the concentrations in the liquid phase to the concentrations in the gas phase at equilibrium. In general, Raoult's law can be written as Eq. 1.4, with the key parameter P_{sat}. This intrinsic parameter varies with temperature, with the compound becoming less volatile as the temperature decreases.

$$\gamma_{g,i} P_i = \gamma_{l,i} x_i P_{sat,i} \tag{1.4}$$

where P_i represents the partial pressure of compound i in the gas phase, $P_{sat,i}$ represents the P_{sat} of compound i, and x_i is the mole fraction of compound i in the liquid phase.

$\gamma_{g,i}$ and $\gamma_{l,i}$ are the activity coefficients of the compound i in the gas and liquid phases, respectively. Those activity coefficients characterize the non-ideality between compounds. For a compound i at the ideal state, it interacts with another compound in the same way as it interacts with itself. The activity coefficients are equal to 1 for an ideal mixture ($\gamma_{g,i}$ = 1 for ideal gas and $\gamma_{l,i}$ = 1 for ideal liquid).

In practice, in the atmosphere, the aqueous or organic solution of aerosols is rarely ideal because it results from the partitioning of many compounds of different origins and varied molecular structures. However, under atmospheric conditions, the gas mixture is ideal because the interactions between different gas molecules are very weak, unlike in a condensed liquid phase. The concept of non-ideality (solution deviating from ideal behavior) only applies to the particulate phase in atmospheric chemistry. Assuming ideal gas and non-ideal liquid phases, Raoult's law (Eq. 1.4) can be written as Eq. 1.5.

$$P_i = \gamma_i xi P_{sat,i} \tag{1.5}$$

A lower P_{sat} value indicates that the organic compound is less volatile and has the potential to form a higher amount of SOAs. As for the non-ideality, the lower the γ_i, the more stable the compound is in the liquid phase and the more affinity it has with the other compounds in the liquid phase. Conversely, the higher the γ_i value, the less stable the compound is and the less affinity it has with the other compounds.

Pankow's law [Pankow, 1994] developed the first aerosol model by rewriting Raoult's law (Eq. 1.5) to calculate the partitioning of an atmospheric organic compound between the gas phase and an organic aerosol phase. This model assumes that particulate organic matter is made up of a single phase, although it is possible that particulate organic matter is made up of multiple phases in a complex mixture of molecules with little affinity between them. [Pankow, 1994] defines a gas-particle partition coefficient $K_{p,i}$ (in unit $m^3/\mu g$) as Eq. 1.6

$$K_{p,i} M_o = \frac{A_{p,i}}{A_{g,i}} \tag{1.6}$$

where M_o is the total organic aerosol concentration ($\mu g/m^3$), $A_{g,i}$ and $A_{p,i}$ are the concentrations of compound i in the gas and organic aerosol phase, respectively. Using Raoult's law and the ideal gas law, the partitioning constant of compound i can be calculated by Eq. 1.7:

$$K_{p,i} = \frac{760 \times 8.202 \times 10^{-5} \times T}{M_{om}\gamma_i P_{sat,i} \times 10^6} \tag{1.7}$$

where M_{om} is the average molar mass of the organic phase (in g/mol), T is the temperature (in Kelvin), and $P_{sat,i}$ is in torr.

Henry's law Henry's law is used to describe the partitioning of soluble organic compounds between the gas and aqueous phases (Eq. 1.8). It applies to compounds that are infinitely dilute in water, assuming that compound i is surrounded only by water molecules and interacts only with water.

$$C_i = H_i \times P_i \tag{1.8}$$

where C_i is the concentration (in M: mol/L) of compound i in the aqueous phase and H_i is Henry's law constant (M/atm).

Similar to Pankow's law, a partition constant $K_{aq,i}$ of compound i in the aqueous phase can then be defined as Eq. 1.9.

$$K_{aq,i}AQ = \frac{A_{aq,i}}{A_{g,i}} \tag{1.9}$$

where AQ is the total mass in the aqueous phase (in $\mu g/m^3$) and $A_{aq,i}$ is the concentration of i in the aqueous phase. The partition constant $K_{aq,i}$ can be written as Eq. 1.10.

$$K_{aq,i} = \frac{18 \times H_i RT}{M_{aq}\rho\gamma_{aq,i} \times 1.013 \times 10^{11}} \tag{1.10}$$

where M_{aq} and ρ are the molar mass (g/mol) and density (Kg/L) of the aqueous phase. These values may differ from those for water due to absorbed compounds. $\gamma_{aq,i}$ is the activity coefficient of compound i at infinite dilution and R is the gas constant (8.314 J/(mol·K)). For a further explanation of SOA formation at equilibrium, see [Couvidat, 2015].

2.3 Aerosol dynamics

Along with meteorological process changes, aerosols grow or decay in size, mass, and composition. Aerosol dynamics refers to the study of how aerosols, which are tiny particles suspended in the air, are formed, evolve, and eventually removed within the general circulation of the atmosphere.

As shown in Fig. 1.10, aerosols can be formed through the nucleation and condensation processes of gas molecules. The formed or emitted particles can grow in size by coagulation or undergo reshaping due to the condensation/evaporation processes. Additionally, particles can be removed from the atmosphere by dry and wet deposition processes. Among all major aerosol dynamic processes, condensation is the most important aerosol dynamic contributing to SOA formation.

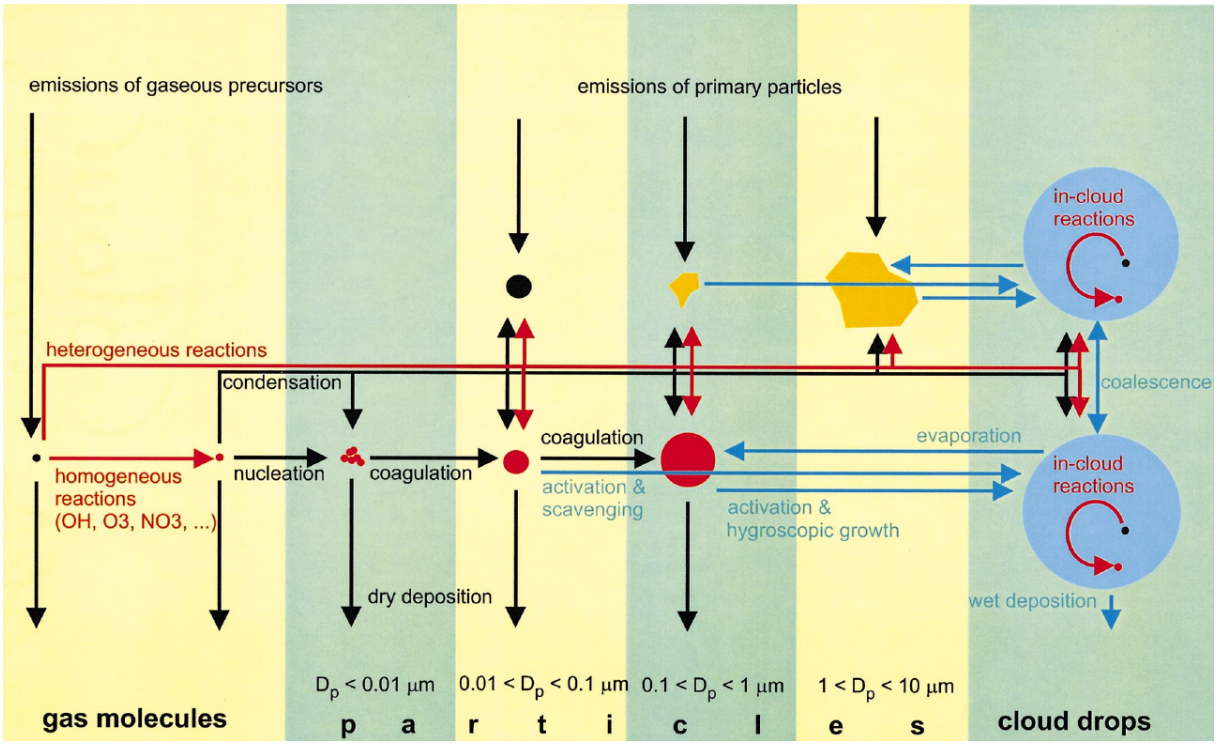


Figure 1.10: Processes involved in aerosol dynamic. Source: [Raes, 2000].

2.3.1 Condensation/evaporation

Condensation/evaporation occurs when a compound undergoes a phase change between gas and particles. As a result, these processes do not alter the number of particles in the air, but they can significantly affect particle size and chemical composition. The size of existing particles can either increase due to condensation, or decrease due to evaporation.

The driving force behind condensation/evaporation is the thermodynamic equilibrium shifts. When the gas's vapor pressure exceeds the surface vapor pressure, gas molecules tend to transfer to the particle phase, resulting in particle growth via condensation. Conversely, when the gas's vapor pressure falls below the surface vapor pressure, molecules from the particle phase transfer to the gas phase, causing particle shrinkage through evaporation. The calculation of the gas-to-particle thermodynamic is explained in Sect. 2.2.

2.3.2 Nucleation

Nucleation is the process that forms new particles. It occurs when gas molecules of aerosol precursors form clusters that are eventually large enough to become stable aerosols. The new particles produced from nucleation are typically very small in size (a few nanometers in diameter). After being formed, they may grow to sizes greater than about 0.1 μm in a few days to a few weeks by coagulation and condensation ([Raes, 2000]).

2.3.3 Coagulation

Coagulation occurs when two particles collide with each other due to Brownian motion and turbulence, resulting in aerosol growth into larger particles. During this process,

the total aerosol mass is conserved while the particle number is reduced. Meanwhile, coagulation leads to the formation of larger particles that can more easily settle out of the air (e.g., by wet deposition).

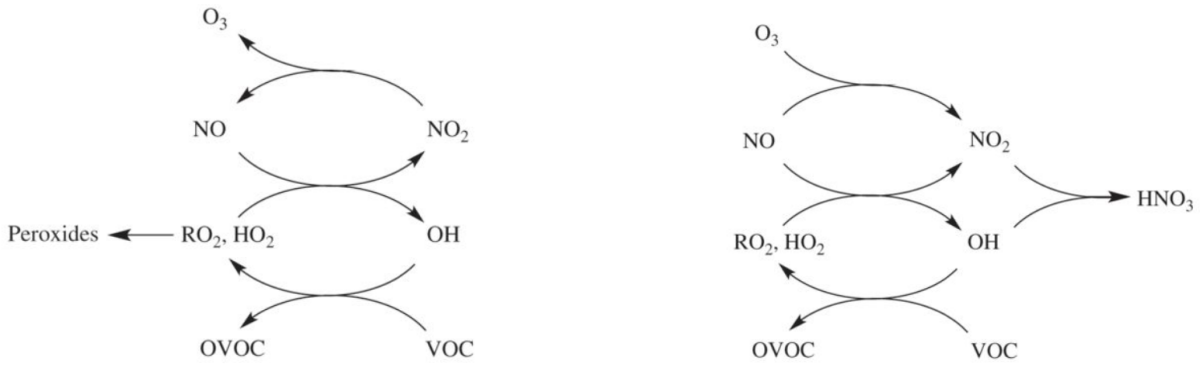
3 Interaction between anthropogenic emission and aerosol formation

The interaction between anthropogenic emissions and SOA formation is complex. Anthropogenic emissions contain primary aerosols that can contribute to SOA formation (e.g., by acting as a condensation base), and emitted AnVOCs are also precursors to SOAs. Other anthropogenic emissions, such as NOx, can indirectly affect SOA formation by influencing gas-phase chemistry and gas-particle partitioning, as described below.

3.1 Gas-phase chemistry

Anthropogenic emissions, such as NOx, can impact SOA formation by altering the oxidation pathways of VOCs. This may result in a complex SOA composition with different hydrocarbon precursors having different sensitivities to NOx levels. Furthermore, the fact that NOx and VOCs are both producers and destructors of ozone and oxidants, thereby creating a dilemma in developing effective reduction strategies to control air pollution.

High/low-NOx regime There is competition between the different oxidation pathways and two main chemical regimes can be distinguished: a regime where the peroxy radicals (i.e., HO_2 and RO_2) concentrations are not limited compared to NO (or more generally NOx) concentrations; and another regime where the peroxy radicals concentrations are limited compared to NO (or more generally, NOx) concentrations.



(a) Low NOx regime

(b) High NOx regime

Figure 1.11: Schematic representations of ozone chemistry in a low-NOx regime (a) and a high-NOx regime (b) by [Seigneur, 2019]. VOC: volatile organic compound, OVOC: oxidized volatile organic compound. NO, NO_2, and O_3 are involved in the Leighton photostationary-state reactions, and the directions of the arrows indicate the main effect of the change in NOx emissions on the O_3 concentration for each regime.

Under the different regimes, SOA formation is different. In a low-NOx regime, the formation of peroxides governs the termination reactions, therefore, O_3 formation is proportional to the concentration of NO and to the square root of the production of HOx radicals. Organic peroxide formation is favored. In a high-NOx regime, nitric acid formation governs the termination reactions, implying that the rate of propagation of the radicals is much greater than that of their termination. Organic nitrate formation is favored.

3.2 Gas-particle partitioning

The gas-to-particle process is very complex because accurate parameterization may need to take into account many of the phenomena involved in physicochemical transformations. For example, the viscosity of the aerosol, the interaction between gaseous chemistry and aqueous chemistry, the hydrophilicity and hydrophobicity of the aerosol, and oligomerization all affect the formation of SOAs from VOC oxidation products.

Anthropogenic aerosols can also play a role in SOA formation by acting as intermediates for gas-particle partitioning and particle-phase activities. They can alter various properties of the aerosol, including acidity, phase stage, and liquid water content, which can further affect the formation and aging of biogenic SOAs.

3.3 Effects of anthropogenic emission mitigation

Efforts to regulate air pollution are expected to lead to a reduction in anthropogenic emissions. Consequently, the impact of reducing anthropogenic emissions on biological SOA formation is of concern.

Anthropogenic-biogenic interaction arises when air masses enriched with anthropogenic emissions mix with regions having significant BVOC emissions. This mixing perturbs BVOC oxidation and the following processes linked to SOA formation. Implementing regulations and strategies to reduce emissions of key anthropogenic pollutants like NOx, SO_2, and primary parties may help mitigate negative impacts on biogenic SOA formation ([Sartelet, 2012]).

According to current estimates, biogenic organic aerosol concentrations may decrease in the future, particularly in rural or peri-urban areas where oxidant concentrations are expected to decrease. This estimate may be inaccurate due to simplifications made in the simulation of the SOA formation process. As a result, the reduction of anthropogenic emissions could potentially result in more or less significant reductions in biogenic organic aerosol concentrations or even an increase. For this reason, AQM and therefore aerosol models should improve the representation of the non-linear evolution of VOC oxidation leading to aerosol growth ([Huang, 2020]).

4 Modeling of organic aerosol formation

To achieve accurate SOA modeling, it is essential to consider both the gas-phase mechanism describing VOC degradation and formation of condensable species, as well as the SOA mechanism involved in transferring these condensables from the gas phase to the particulate phase.

4.1 Gas-phase chemical mechanism

As detailed in Sect. 2, the gas-phase chemical mechanism plays a crucial role in the formation of SOA by producing low-volatility organics that participate in SOA formation, and hence, it is an essential component of SOA modeling. The simplified and explicit chemical mechanisms are presented here.

4.1.1 Simplified chemical mechanism

Simplified chemical mechanisms contain inorganic reactions related and a limited number of organic reactions describing VOC degradation. Those mechanisms are generally built by two approaches, namely lumped-species and lumped-structure approaches.

Lumped-species approach The lumped-species approach uses surrogate molecules that are representative of groups of similar organic compounds with similar chemical properties. The approach is applied in simplified chemical mechanisms such as SAPRC-07 Carter [Carter, 2010] and RACM2 Goliff et al. [Goliff, 2013].

Lumped-structure approach The lumped-structure or the carbon-bond approach, breaks a molecule down into its functional groups and assumes that its chemical behavior is equivalent to the chemical behaviors of its constituent functional groups. This approach is used in "carbon-bond" mechanisms like CB05 ([Sarwar, 2008]).

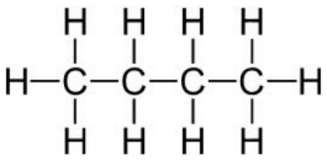


Figure 1.12: Structure of n-butane.

Here is an example illustrating the differences between the two approaches: In the RACM2 mechanism with the "lumped-species" approach), n-butane (structure see Fig. 1.12) is represented by a molecule "HC3", which is also the surrogate for alkanes, alcohols, esters, and alkynes with rate constants with OH less than 3.4×10^{-12} cm^3 s^{-1}. Alternatively, in the CB05 mechanism with the "lumped-structure" approach), n-butane is denoted by "4 PAR", indicating the reactivity of n-butane is represented by four groups containing one carbon atom with a single bond.

It is noteworthy that the lumped-structure approach may not be able to effectively track the oxidation of SVOC because the carbon number in the initial molecule can be lost during the decomposition into functional groups. Therefore, for SOA modeling, the lumped-species approach is considered more straightforward. To accurately represent the chemical reactions specific to SOA formation, additional reactions must be incorporated into a carbon-bond mechanism.

4.1.2 Explicit chemical mechanism

Explicit gas-phase chemical mechanisms reflect our understanding of VOC chemistry, describing detailed tropospheric reactions and species involved in the degradation of VOCs.

Master Chemical Mechanism The Master Chemical Mechanism (MCM) developed by [Jenkin, 1997] is a near-explicit chemical mechanism, containing the gas-phase degradation of a list of VOC species (142 non-methane VOCs) that are found in tropospheric chemistry. For each VOC, MCM contains hundreds to thousands of reactions and species.

Generator for Explicit Chemistry and Kinetics of Organics in the Atmosphere The Generator for Explicit Chemistry and Kinetics of Organics in the Atmosphere (GECKO-A) is a fully explicit SOA mechanism generator developed by [Aumont, 2005], which follows protocols to identify possible reactive pathways and organic properties based on the molecular structure of the species. It can generate millions to tens of millions of reactions from a VOC.

SAPRC Mechanism Generation system Similar to GECKO-A, the SAPRC Mechanism Generation System (MechGen) is also a generator that can produce fully explicit tropospheric degradation pathways for emitted VOCs and their oxidation products ([Carter, 2020]). While GECKO-A can generate all the reactions of a selected compound and its oxidation products (until terminated by the formation of CO or CO_2), MechGen focuses on generating reactions related to stable oxidation products and subsequent radical intermediates.

When comparing the near-explicit mechanism MCM to the fully-explicit mechanisms GECKO-A and MechGen, similarities emerge as they all provide detailed representations of VOC chemistry, including reaction pathways, kinetics data, and molecular structures. However, a major difference stands out: the fully explicit mechanism is generated through automated methods and can capture all reactions in several generations, while the near-explicit MCM mechanism is manually written, resulting in a much smaller size. Mechanisms generated by GECKO-A and MechGen are with more order of magnitude complexity than those in MCM, where reactions and species in MCM are highly lumped after the first two generations ([Mouchel-Vallon, 2020]).

4.2 SOA mechanism

Due to computational limitations and an incomplete understanding of physicochemical processes, simplified parameterizations derived from laboratory chamber data are employed first in air quality models (AQMs) for SOA formation. The two-product approach, volatility basis set, and surrogate approach are three major approaches to developing SOA mechanisms.

Two-product approach The first SOA model was proposed by [Odum, 1996], which determines the SOA yield (Y) formed from one VOC by a total of n semi-volatile products, as described in Eq. 1.11.

$$Y = \sum_{i}^{n} \frac{\alpha_i K_{p,i} M}{1 + M K_{p,i}} \tag{1.11}$$

where α and $K_{p,i}$ are the stoichiometric coefficient and the partitioning coefficient of each product i (out of n), respectively. M is the total organic aerosol mass. Y is a ratio defined as the mass of SOA produced divided by the mass of reacted VOC. Since both values of α and $K_{p,i}$ are fitted from chamber experiments, the so-called products are not necessarily to be related to actual oxidation compounds. As [Odum, 1996] sets the number of products n to 2, this SOA mechanism is also known as the "2-product" approach.

Volatility basis set The volatility basis set (VBS) method developed by [Donahue, 2006] is also a widely used method for SOA modeling. Condensable organic compounds are grouped into a series of predefined log-spaced "bins" based on their volatility. These volatility bins are spaced at exponential intervals with a constant ratio of 10 (e.g., 10^{-2}, 10^{-1}, 1, 10, ...) in terms of effective saturation concentration (C^* in unit $\mu g/m^3$), which is the reciprocal of K_p.

As the compounds have been grouped into bins, they are assumed to share the same properties and aging processes, evolving into bins of compounds with less volatility. This assumption, however, indicates that the VBS approach only models the functionalization process, leading to an overestimation of organic aerosol concentrations and the inability to capture the formation of compounds with higher volatility. To solve this issue, [Donahue, 2011] upgraded the VBS approach to a two-dimensional (2-D) VBS, whose parameters include, in addition to volatility, the atomic ratio of oxygen to carbon (O:C) or the average oxidation state. By incorporating the degree of oxygenation as a second VBS indicator, the 2-D VBS approach can be applied to explore SOA aging by both functionalization and fragmentation implicitly.

Surrogate approach The surrogate approach selects a few model species to represent the total SOA products from the oxidation of a specific VOC (e.g., [Pun, 2006; Pun, 2007]). The surrogates are chosen based on the species that are observed experimentally and that share the properties (volatility) determined using the 2-product approach. An example is the hydrophilic/hydrophobic organic (H^2O) mechanism developed by [Couvidat, 2012], which uses hydrophilic and hydrophobic surrogates to take into account different partitioning of SOAs on aqueous and organic aerosol phases.

One of the advantages of the surrogate approach is that the selected surrogates can be associated with molecular structures. By attaching a molecular structure to the surrogate, it can be used to study more complex aerosol phenomena (e.g., non-ideality and absorption into the aqueous phase). Since the choice of molecular structure can introduce large uncertainties in SOA modeling, the selection of a suitable molecular structure is critical and requires an accurate estimation of the surrogate.

It is important to emphasize the importance of molecular structure in SOA modeling. In gas-phase chemistry, the molecular structure of VOC determines their oxidation pathway, which is required in order to track the SOA aging and the interaction with other compounds (e.g., oxidants, inorganics, and other compounds) explicitly. Meanwhile, to calculate the thermodynamic properties of aerosols in SOA mechanisms (e.g.,

hydrophilicity, hydrophobicity, viscosity), it is necessary to know their molecular structure in order to consider important phenomena such as hygroscopicity and non-ideality. A number of atmospherically relevant physicochemical parameters can be estimated using organic molecular structure, including saturation vapor pressure, Henry's law constants, and activity coefficients ([Isaacman-VanWertz, 2021]).

4.3 VOC mechanism reduction for SOA modeling

Explicit mechanisms are powerful tools for modeling the complex chemical processes involved in SOA formation. However, because they are computationally intensive, they are typically used in box models or small regions to simulate SOA formation and aging. Their use for routine 3D air quality forecasting or modeling is currently not achievable. To use explicit mechanisms in 3D large-scale modeling, the complexity of explicit mechanisms must be reduced. Reducing explicit VOC mechanisms is one way to address the problems of long computational times required for explicit mechanisms and the inability of simplified mechanisms to accurately represent complex chemical phenomena.

Below are three condensed mechanisms reduced from explicit VOC mechanisms, all of which can be applied to SOA modeling.

Common Representative Intermediates mechanism for MCM The Common Representative Intermediates (CRI) mechanism is a series of condensed mechanisms reduced from the MCM mechanism with varying degrees of reduction ([Jenkin, 2008; Watson, 2008]). They were first developed focusing on accurate ozone formation by merging MCM species to new surrogates based on the number of carbon-carbon (C-C) and carbon-hydrogen (C-H) bonds active in the NO-to-NO_2 conversions concerning ozone formation.

Although the CRI mechanism is primarily designed to preserve ozone formation capacity, it still has good overall performance in characterizing VOC degradation. Therefore, CRI is also used in SOA modeling (e.g., [Utembe, 2009; Utembe, 2011; Khan, 2017; Weber, 2020]).

Volatility basis set for GECKO-A The volatility basis set – Generator for Explicit Chemistry and Kinetics of Organics in the Atmosphere (VBS-GECKO) developed by [Lannuque, 2018] is a VBS-type parameterization developed on the basis of GECKO-A, with a focus on SOA formation and aging. It has been optimized to simulate SOA variations in a box model using GECKO-A and has also been integrated and evaluated in a 3-D air quality model ([Lannuque, 2020]).

Neural network emulators for GECKO-A Recently, machine learning techniques have been explored in the context of mechanism reduction for SOA modeling. [Schreck, 2022] delved into the application of neural network methods in GECKO-A reduction and developed models that simulate SOA formation with an error range between 2 % – 8 %. This suggests that neural network models hold potential as promising tools for implementing the complexity of explicit VOC mechanisms in 3-D modeling. However, it is important to note that these models may have limitations and are constrained by the range of environmental conditions considered in training.

5 Numerical models employed in this work

This section describes the numerical model used in this study, including the chemical transport model CHIMERE and the aerosol box model SSH-aerosol.

5.1 Chemical transport model

The three-dimensional (3-D) chemical transport model (CTM) refers to Eulerian grid models that utilize a to simulate and forecast the evolution of concentrations and composition of air pollutants over a predefined 3-D grid at the street, regional, or global scale. With 3-D modeling, a detailed representation of the spatial and temporal distribution of pollutants in the atmosphere can be simulated by considering the influences due to emissions, meteorology, transportation, and other physical-chemical processes. As a result, complex variations in air quality can be captured, such as the dispersion of pollutants and the formation of secondary pollutants. When modeling large areas or extended time periods, 3-D modeling demands significant computational resources and is computationally intensive. Meanwhile, its accuracy can be affected by uncertainties in input data, such as emission inventories and meteorological data. Nonetheless, it is a powerful tool that has been widely used in air quality modeling.

CHIMERE The CHIMERE chemistry-transport model has been continuously developed and distributed since 1999, with the latest version v2020r1 ([Menut, 2021]). It follows the principle of using validated and up-to-date parameters to calculate concentrations of major pollutants with optimal accuracy. As illustrated in Fig. 1.13, The CHIMERE model requires meteorological data, emission information, as well as boundary and initial conditions to accurately simulate atmospheric concentrations of numerous gas-phase and aerosol species from local to continental scales. CHIMERE incorporates key processes (e.g., transport, emissions, chemistry, and deposition) and can be run in offline or online mode using various meteorological models. The CHIMERE model has been used in a wide variety of atmospheric research and operational forecasting for regional and national air quality networks.

5.2 Aerosol box model

Aerosol box models simulate variations in SOA concentration and composition due to physical and/or chemical processes in a fixed volume of an atmospheric multiphase mixture. They are zero-dimensional (0-D) models that can be integrated into 3-D CTMs to simulate aerosol evolution within a 3-D grid cell.

SSH-aerosol SSH-aerosol is a modular, multi-phase, plug-and-play aerosol box model ([Sartelet, 2020]). It combines start-of-the-art models and mechanisms including SCRAM (Size-Composition Resolved Aerosol Model by [Zhu, 2015]), SOAP (Secondary Organic Aerosol Processor model by [Couvidat, 2015], as well as H^2O (Hydrophobic/Hydrophilic Organic by [Couvidat, 2012]).

As illustrated in Fig. 1.14, SSH-aerosol can simulate the formation and evolution of aerosols, taking into account the effects of emissions, gas-phase chemistry, as well as

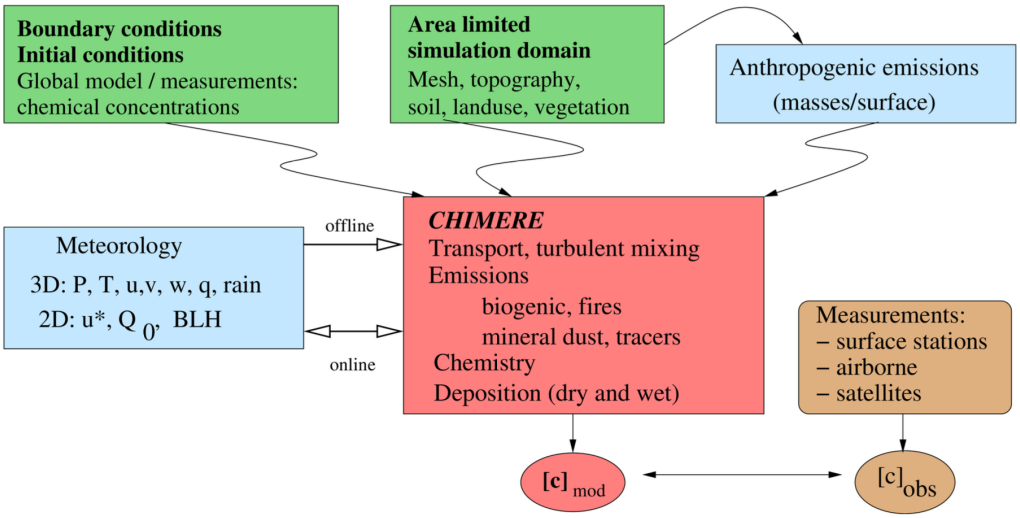


Figure 1.13: General principle of the CHIMERE model. Source: Chimere documentation v2020r3

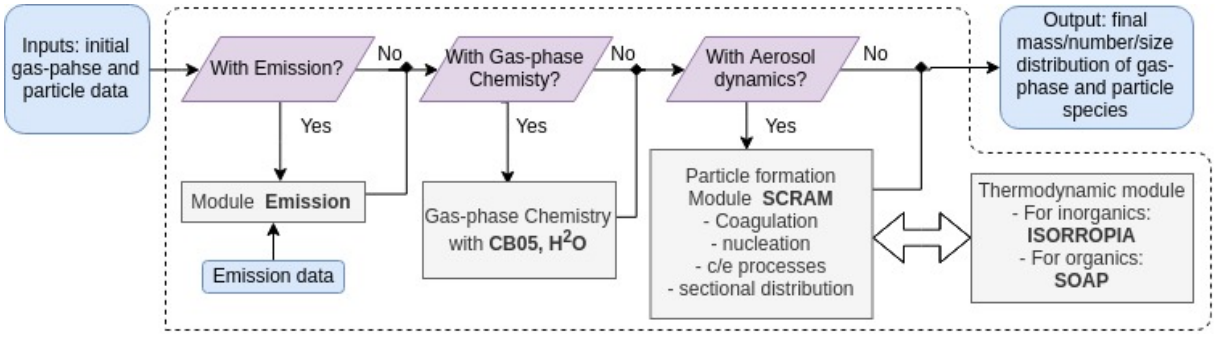


Figure 1.14: Framework of the SSH-aerosol model.

aerosol dynamics. Thanks to the embedded SCRAM model, SSH-aerosol is able to simulate aerosol dynamics processes and aerosol mixing states, which classifies aerosols by both chemical compositions and the mass fraction of the different species. The partitioning between the gas and particle phases is simulated with the thermodynamic module SOAP [Couvidat, 2015] for organic species. This thermodynamic module accounts for hygroscopicity, absorption into the aqueous phase, aerosol non-ideality (effect of molecular interactions on species partitioning), and phase separation. As for the gas-phase chemistry, it is simulated in SSH-aerosol by employing the SOA mechanism H^2O for the formation of condensable organic compounds with other gas-phase chemical mechanisms (e.g., CB05 by [Sarwar, 2008] and RACM2 by [Goliff, 2013]) to simulate the formation of ozone, radicals and the degradation of VOCs.

6 Objectives and the plan of the work

The formation of organic aerosols involves complex physical and chemical processes. It is influenced by anthropogenic emissions through emissions of organic compounds of various volatilities that may be SOA precursors, but also through emissions of NOx. There are highly non-linear interactions between anthropogenic emissions and aerosol formation, and these interactions may not be fully represented in simplified SOA mechanisms used in regional air quality modeling. Simplified SOA mechanisms are built on chamber measurements, and they do not represent the whole complexity of the chemical pathways involved in SOA formation. Therefore, the objectives of this study are to improve the representations of SOA mechanisms in regional air-quality models and to investigate the aerosol formation variation in response to anthropogenic emissions regulations with the revised SOA mechanisms.

In Chapter 2, the preliminary development of a mechanism reduction algorithm, the GENerator of Reduced Organic Aerosol Mechanisms (GENOA v1.0), is described. The algorithm is designed to generate condensed SOA mechanisms from reducing detailed VOC mechanisms (e.g., MCM) using a series of predefined strategies and evaluation criteria. The obtained SOA mechanisms are expected to preserve the complexity of gas-phase chemistry in SOA formation, with a number of reactions and species that are manageable and practical for regional modeling. The application of GENOA v1.0 to the degradation scheme of sesquiterpenes from MCM is investigated with the 0-D aerosol model SSH-aerosol.

Although GENOA v1.0 is tested to be effective for sesquiterpene SOA mechanism reduction, it may not have enough accuracy and generality required to reduce more complex mechanisms. In Chapter 3, the further development of the GENOA algorithm, i.e., the second version of GENOA (GENOA v2.0) is presented. Compared to GENOA v1.0, the new version implements a parallel reduction structure that allows more comprehensive and systematic reductions from more complex chemical mechanisms of multiple SOA precursors. The ability of GENOA v2.0 on SOA mechanism reduction is evaluated with reductions of two monoterpene chemical mechanisms (MCM+PRAM) from three monoterpene SOA precursors.

Subsequently, in Chapter 4, the condensed biogenic SOA mechanism generated by GENOA v2.0 is incorporated into the 3-D regional CTM model CHIMERE. The organic aerosol concentrations and compositions over Europe in 2018 summer (June-August) are

investigated using this detailed GENOA-generated Biogenic SOA Mechanism (GBM) and the simplified Hydrophilic/Hydrophobic Organics (H^2O) mechanism. Furthermore, the regional biogenic SOA formation in response to NOx emission regulations is explored using the GBM and H^2O mechanisms in CHIMERE.

Finally, in Chapter 5, I summarize and conclude my Ph.D. work, with a discussion of the perspectives on the application and development of GENOA algorithms.

Chapter 2

Development of GENOA v1.0 and its application to sesquiterpene SOAs

When coupling aerosol formation with other atmospheric processes (e.g., transport), air quality models, particularly three-dimensional (3-D) Chemical Transport Models (CTMs), are faced with the dilemma of spending an overwhelming computational capacity on the entire complexity of gas-phase chemistry to the formation of secondary organic aerosols (SOAs) or utilizing highly simplified parameterizations. To improve the efficiency and accuracy of the modeling of SOA formation and aging, the GENerator of reduced Organic Aerosol mechanisms (GENOA v1.0) algorithm has been developed.

A detailed presentation of GENOA v1.0 is described in this chapter. GENOA v1.0 is a reduction algorithm that produces semi-explicit mechanisms for simulating the formation and evolution of secondary organic aerosol (SOA) in regional-scale CTMs. GENOA uses a series of predefined reduction strategies and evaluation criteria to train and reduce SOA mechanisms from near-explicit chemical mechanisms (e.g., the Master Chemical Mechanism (MCM)) under representative atmospheric conditions. The reduction strategies, including removal, jumping, lumping, and replacement, are adopted to locate the potential reduction in the mechanism. Once found, the reduction attempt is evaluated against predefined criteria under selected evaluation conditions. Upon meeting the criteria, this reduction can be accepted and the mechanism with this change is used for the new search of reductions with the strategies. In this way, the semi-explicit mechanisms derived by GENOA retain the accuracy of explicit mechanisms for SOA formation with significantly fewer reactions and species, making them more computationally efficient than explicit mechanisms and appropriate for large-scale air-quality modeling.

The application of GENOA v1.0 to the degradation mechanism of the sesquiterpene (SQT) in MCM is also presented in this chapter. The obtained SQT SOA mechanism consists of 23 reactions and 15 species, with 6 of them being condensable. This mechanism is only 2 % of the size of the original MCM mechanism, which

includes 1626 reactions and 579 species with 356 of them being condensable. Compared with MCM, the reduced SOA mechanism presents a high level of accuracy and efficiency by inducing an average error of 2.66 % in a testing process under 12 159 conditions over Europe, and taking approximately 2 % of CPU time compared to the original MCM).

This paper has been published: **Wang, Z., Couvidat, F., and Sartelet, K.: GENerator of reduced Organic Aerosol mechanism (GENOA v1.0): an automatic generation tool of semi-explicit mechanisms, Geosci. Model Dev., 15, 8957–8982,** https://doi.org/10.5194/gmd-15-8957-2022,2022.

Contents

Geosci. Model Dev., 15, 8957–8982, 2022
https://doi.org/10.5194/gmd-15-8957-2022

Geoscientific Model Development
Open Access
EGU

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

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

[1]Centre d'Enseignement et de Recherche en Environnement Atmosphérique (CEREA), École des Ponts ParisTech, EDF R&D, IPSL, Marne-la-Vallée, France
[2]Institut National de l'Environnement Industriel et des Risques (INERIS), Verneuil-en-Halatte, France

Correspondence: Zhizhao Wang (zhizhao.wang@enpc.fr)

Received: 25 April 2022 – Discussion started: 10 June 2022
Revised: 7 November 2022 – Accepted: 14 November 2022 – Published: 14 December 2022

Abstract. This paper describes the GENerator of reduced Organic Aerosol mechanism (GENOA) that produces semi-explicit mechanisms for simulating the formation and evolution of secondary organic aerosol (SOA) in air quality models. Using a series of predefined reduction strategies and evaluation criteria, GENOA trains and reduces SOA mechanisms from near-explicit chemical mechanisms (e.g., the Master Chemical Mechanism – MCM) under representative atmospheric conditions. As a consequence, these trained SOA mechanisms can preserve the accuracy of detailed gas-phase chemical mechanisms on SOA formation (e.g., molecular structures of crucial organic compounds, the effect of "non-ideality", and the hydrophilic/hydrophobic partitioning of aerosols), with a size (in terms of reaction and species numbers) that is manageable for three-dimensional (3-D) aerosol modeling (e.g., regional chemical transport models). Applied to the degradation of sesquiterpenes (as β-caryophyllene) from MCM, GENOA builds a concise SOA mechanism (2 % of the MCM size) that consists of 23 reactions and 15 species, with 6 of them being condensable. The generated SOA mechanism has been evaluated regarding its ability to reproduce SOA concentrations under the varying atmospheric conditions encountered over Europe, with an average error lower than 3 %.

1 Introduction

Atmospheric aerosols have attracted attention due to their effects on climate and human health: they change the Earth's radiation balance and cloud formation processes (Ramanathan et al., 2001; McNeill, 2017), and they trigger a wide variety of acute and chronic diseases (Breysse et al., 2013). Because the effects of aerosols on health depend on their size and composition (Schwarze et al., 2006), adequate representations of aerosol composition, mass, and number concentrations are required in air quality models (AQMs).

Besides being directly emitted, aerosols can be secondary, i.e., formed in the atmosphere through chemical reactions and gas–particle mass transfer. Based on their chemical composition, they can be further divided into secondary inorganic aerosol (SIA) and secondary organic aerosol (SOA). SOA, which represents a significant fraction of aerosols (e.g., Gelencsér et al., 2007), is largely formed by the condensation of the oxidation products from the degradation of volatile organic compounds (VOCs). As SOA formation involves multiple processes such as the emission of SOA precursor gases, VOC gas-phase chemistry, and gas-to-particle partitioning (Kanakidou et al., 2005; Hallquist et al., 2009), great complexity and uncertainty are involved in accurately predicting SOA formation with the simplified representations currently used in air quality models (Porter et al., 2021).

The state of knowledge on VOC chemistry can be reflected by explicit gas-phase chemical mechanisms that contain all known essential reaction pathways of VOC degradation. For instance, Jenkin et al. (1997) and Saunders et al. (2003) developed the near-explicit Master Chemical Mechanism (MCM), which describes detailed gas-phase chemical processes related to VOC oxidation. Another example is the Generator for Explicit Chemistry and Kinetics of Organics in the Atmosphere (GECKO-A) (Aumont et al., 2005), which uses a prescribed protocol to assign complete reac-

tion pathways and kinetic data to the degradation of VOCs. Explicit mechanisms represent the current understanding of atmospheric chemistry, including information about reaction pathways, kinetics data, and chemical structures (which may be used to deduce thermodynamic properties based on structure–activity relationships).

The MCM mechanism has been used by two-dimensional (2-D) Lagrangian models to simulate the chemical evolution of major air pollutants and some SOAs in plumes (e.g., Evtyugina et al., 2007; Sommariva et al., 2008; Zhang et al., 2021). Moreover, it has been used for simulating the formation of more complex SOAs at a regional level in three-dimensional (3-D) models over a few weeks (e.g., modified MCM with 4642 species and 13 566 reactions in the simulations of Ying and Li, 2011, and with 5727 species and 16 930 reactions in the simulations of Li et al., 2015). Even so, explicit mechanisms of that size are too computationally intensive to be widely employed in 3-D AQMs for SOA formation.

For computational efficiency, AQMs generally use implicit gas-phase chemical mechanisms. Two major approaches are frequently adopted to build implicit chemical mechanisms:

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

Implicit gas-phase mechanisms have been developed and validated to simulate the concentrations of oxidants and other conventional air pollutants such as ozone and NO_2. In these mechanisms, VOCs have been grouped into a limited number of model species because of computational considerations, and the SOA formation is usually not considered.

To complete implicit gas-phase mechanisms, implicit SOA mechanisms have been developed (Kim et al., 2011) that model the SOA formation specifically without modifying ozone and radical concentrations. In 3-D modeling, implicit SOA mechanisms or parameterizations are usually added to implicit gas-phase mechanisms, conserving the oxidant chemistry of the implicit gas-phase mechanism.

Implicit SOA mechanisms are often established based on experimental data from smog chamber experiments to represent the formation and evolution of SOA, such as the two-product empirical SOA model (Odum et al., 1996) and the volatility basis set (VBS) that splits VOC oxidation products into a uniform set of volatility "bins" (Donahue et al., 2006). In the VBS approach, the successive evolution of oxidation products by aging is determined regardless of the chemical composition and structure of the species. Another approach is based on the molecular surrogate approach (e.g., Griffin et al., 2003; Pun et al., 2006; Couvidat et al., 2012). Similarly to the gas-phase chemistry lumped-species approach, the VOC oxidation products are represented via the formation of a few SOA surrogates that are attached to a molecular structure (assumed to be representative of a myriad of semi-volatile compounds). By attaching a molecular structure to the surrogate, several processes otherwise not accounted for (like "non-ideality", hygroscopicity, and condensation on the aqueous phase of particles) can be represented in this approach. However, the choice of adequate molecular structures, which could be highly uncertain, is crucial and requires a precise estimation.

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

As the current SOA representations in AQMs are implicit and may not accurately reflect the true SOA formation process, there is a need for improvement. This has led to the development of semi-explicit mechanisms of condensed sizes. The development of semi-explicit mechanisms is a compromise between the high computational time of explicit mechanisms and the lack of accuracy in the representation of chemical phenomena in the implicit SOA mechanisms. They are generated by reducing explicit mechanisms to a level of complexity suitable for the computational constraints of AQMs. Recent developments of reduced mechanisms include the Common Representative Intermediates (CRI) mechanism (Jenkin et al., 2008; Watson et al., 2008; Khan et al., 2017) from the MCM reduction (Szopa et al., 2005) and the volatility basis set – Generator for Explicit Chemistry and Kinetics of Organics in the Atmosphere (VBS-GECKO) (Lannuque et al., 2018) from a GECKO-A reduction. However, the reduced mechanisms mentioned above do not track the detailed molecular structure of surrogates, rather only considering some of their specific properties:

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

This study presents the development of the first version of the GENerator of reduced Organic Aerosol mechanism

(GENOA) that generates customized semi-explicit chemical mechanisms appropriate for AQMs from explicit mechanisms, using surrogates assigned to molecular structures. As described in Sect. 2, the new reduced mechanisms can effectively and efficiently reproduce the complexity of gas-phase oxidation, by training under various atmospheric conditions, and the non-ideality of gas–particle partitioning, using a molecular-structure-preserving approach. GENOA also provides practical user-defined options, enabling users to specify the required reduction scale or accuracy. For gas–particle partitioning, a 0-D box model "SSH-aerosol" (Sartelet et al., 2020) is modified and coupled with GENOA to simulate aerosol concentrations. With SSH-aerosol, the effects of mass transfer between the gas-phase and the organic/aqueous phases, hygroscopicity, and non-ideality are taken into account in the reduction.

The application of GENOA to the MCM degradation scheme of β-caryophyllene (BCARY) (Jenkin et al., 2012) is described in Sect. 3. β-Caryophyllene is selected to demonstrate the GENOA algorithm because it is one of the most abundant and representative sesquiterpenes (SQTs). Sesquiterpenes are a well-known source of SOAs (Hellén et al., 2020; Tasoglou and Pandis, 2015), and their degradation mechanism (as BCARY) is well documented in the near-explicit MCM mechanism (Jenkin et al., 2012). Studies have also compared SOA yields simulated using the MCM mechanism to chamber data for sesquiterpenes (e.g., Xavier et al., 2019). BCARY is, therefore, an ideal candidate for model development and demonstration of the reduction methodology. In this paper, the near-explicit MCM BCARY degradation scheme serves as a reliable benchmark for GENOA. The experiment data from Tasoglou and Pandis (2015) and Chen et al. (2012) are also compared to the newly developed reduced mechanism in Appendix A. Finally, conclusions are drawn in Sect. 4.

2 Model development

The GENerator of reduced Organic Aerosol mechanism (GENOA) is an algorithm that generates semi-explicit chemical mechanisms focusing on SOA formation. The generated semi-explicit mechanisms are designed to preserve the accuracy of explicit mechanisms for SOA formation while also keeping the number of reactions/species low enough to be suitable for large-scale modeling, particularly in 3-D AQMs. The focus of the semi-explicit mechanism is solely on the accurate modeling of SOA. Because ozone, major radicals, and other inorganics are also affected by inorganic and other VOC chemistry, their concentrations are not tracked with the semi-explicit mechanism. Instead, they are simulated using existing implicit gas-phase chemical mechanisms.

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

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

In the parameter selection, GENOA first assigns the error tolerance, defined as the largest acceptable error induced by each change in the mechanism (see Sect. 2.5), and then employs one of the reduction strategies along with its required parameters (see Sect. 2.2).

Afterward, in the reduction cycle, GENOA searches for potential reductions according to the selected reduction strategy. The new mechanism with the first found reduction is then simulated over the conditions from the training dataset (a limited set of conditions used through all of the reduction processes; see Sect. 2.3.1) or from the pre-testing dataset (a more extensive set of conditions used only at the end of the reduction process; see Sect. 2.3.2). The simulated total SOA concentrations are then compared with those simulated with the reference mechanism, where the differences are used to evaluate the potential reduction (see Sect. 2.5). If the SOA differences are under the predefined error tolerances, the mechanism with the current reduction is accepted and serves as the basis for the next search for reduction. If the reduction is refused, the following reduction attempt starts with the previously validated mechanism. Once no further reduction is found, the current reduction cycle ends. The next step is either selecting the subsequent error tolerance and/or reduction strategy in the next parameter selection or terminating the GENOA training section. Finally, the performance of the final reduced mechanism is evaluated under a variety of environmental conditions, denoted as the testing dataset (see Sect. 2.3.3). The 0-D aerosol model SSH-aerosol is used to simulate the SOA concentration and composition, which is required in all of the GENOA sections (e.g., the initialization of reduction parameters and the evaluation of the reduced mechanism).

2.1 Prereduction

A prereduction process is conducted on the original MCM mechanism before it is used as the reference mechanism for the reduction. This process skips extremely fast unimolecular reactions (i.e., the reaction rate constant of $10^6\ s^{-1}$ corresponding to a lifetime of 1 µs) to avoid numerical problems. For computational efficiency, the process also combines elementary reactions with the same reactants into combined reactions with non-integer stoichiometric coefficients.

An example is shown in Table 1, where the original MCM reaction nos. 1 to 7 have first been merged into the combined reaction nos. 8 to 10. The prereduction compacts the reaction

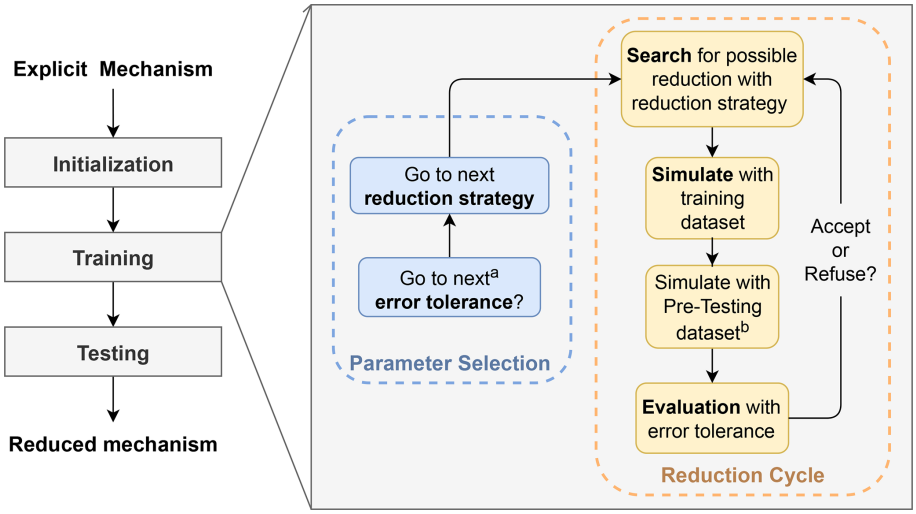


Figure 1. Flow chart indicating the three major procedures in GENOA and illustrating the main execution of the training section. [a] GENOA uses the first value of the targeted variables for initialization and then passes to the next values for subsequent parameter updates. [b] Simulation with the pre-testing dataset is only activated under certain circumstances.

list (from 1 626 to 1 242 reactions), improving the reduction efficiency. The prereduction also skips two biradicals (i.e., BCALOOA and CH2OOF) that are extremely reactive and disintegrate instantaneously with a kinetic rate coefficient of $10^6\,s^{-1}$. As a result, reaction nos. 8 to 10 can then be repented by one reaction, reaction no. 11, whose kinetic rate coefficient corresponds to that of the reaction producing the skipped species (in this case, the ozonolysis of BCAL, reaction no. 9).

2.2 Reduction strategies

GENOA supports four types of reduction strategies:

- removal – reactions, species, or gas–particle partitioning with negligible effects on SOA formation are removed from the mechanism;
- jumping – one compound is substituted by its oxidation product, as if the compound had been "jumped over" in the reaction pathway;
- lumping – compounds with similar properties are combined to form a new compound;
- replacement – one compound is replaced by another existing compound with similar properties.

The reduction strategies are illustrated with examples from the BCARY reduction in Sect. 2.2.1 to 2.2.4. A detailed list of all of the options and parameters controlling the BCARY reduction is summarized in the Supplement.

For the BCARY reduction, the reduction strategies are employed in the following order: removing reactions, jumping, lumping, replacement, removing species, and finally removing gas–particle partitioning. The reduction strategies are ordered based on their potential influences on the mechanism. The first applied strategies, removing reactions and jumping, trim trivial reactions and species without altering the properties of the species. They are followed by lumping and replacement (as an extension to lumping), which refine the mechanisms considerably by merging the species and reactions involved. Afterward, the removing species strategy attempts to delete all merged and unmerged species. Finally, the strategy of removing gas–particle partitioning is applied in order to remove the partitioning of condensable species, which cannot be removed by removing species. This current order has been tested and found to be efficient for the BCARY mechanism, but it can be changed by the user along with other user-defined parameters.

2.2.1 Removal strategy

The removal strategy assumes that chemical reactions and/or species with a low probability of contributing to the formation and evolution of SOA can be eliminated from the mechanism. In general, three types of removal are applied depending on the removed subject:

- removing reactions;
- removing compounds in both the gaseous and particle phases (completely removing a species from the scheme);
- removing the gas–particle partitioning of semi-volatile compounds (consider the semi-volatile compounds as VOCs that do not condense to the particle phase but retain their gas-phase chemistry).

There is no particular restriction to exclude species from the reduction attempt via the strategy of removing species or re-

Table 1. Reactions before and after prereduction, where the MCM (v3.3.1) species BCALOOA and CH2OOF are skipped over by their degradation products. The molecular structures of all mentioned MCM species can be found in Fig. C1.

No.	Reaction	Kinetic rate coefficient[a]
1	BCAL + O_3 → BCALOOA + HCHO	$1.1 \times 10^{-16} \times 0.670$
2	BCAL + O_3 → BCLKET + CH2OOF	$1.1 \times 10^{-16} \times 0.330$
3	BCALOOA → BCALOO	$1.0 \times 10^{6} \times 0.500$
4	BCALOOA → C146O2 + OH	$1.0 \times 10^{6} \times 0.500$
5	CH2OOF → CH2OO	$1.0 \times 10^{6} \times 0.370$
6	CH2OOF → CO	$1.0 \times 10^{6} \times 0.500$
7	CH2OOF → HO_2 + CO + OH	$1.0 \times 10^{6} \times 0.130$
8	BCAL + O_3 → 0.67 BCALOOA + 0.33 BCLKET + 0.33 CH2OOF + 0.67 HCHO	1.1×10^{-16}
9	BCALOOA → 0.5 BCALOO + 0.5 C146O2 + 0.5 OH	1.0×10^{6}
10	CH2OOF → 0.37 CH2OO + 0.63 CO + 0.13 HO_2 + 0.13 OH	1.0×10^{6}
11	BCAL + O_3 → 0.5 BCALOO + 0.5 C146O2 + 0.37 CH2OO + BCLKET + HCHO + 0.13 HO_2 + 0.63 CO + 1.13 OH	1.1×10^{-16}

[a] The kinetic rate coefficients are given in units of per second (s^{-1}) for unimolecular reactions and in units of cubic centimeters per molecule per second (cm^3 $molec.^{-1}$ s^{-1}) for bimolecular reactions.

moving gas–particle partitioning. However, for removing reactions, a threshold on the branching ratio of the reaction is applied to the reduction. The branching ratio is defined as the ratio of the destruction rate of one reaction to the sum of the destruction rates of all reactions of the targeted species. In the BCARY reduction, a maximum branching ratio (B_{rm}) is defined as a restriction criterion. All reactions with an hourly branching ratio (averaged over the training conditions) under this value (reactions that are likely to have a minimal effect on SOA formation) are considered candidates for removal.

To avoid over-reduction, a small B_{rm} is applied at the beginning of reduction. After going through the reductions for all reduction strategies, the value of B_{rm} is then incremented. In the reduction of BCARY, an ascending list of B_{rm} values equal to 5 %, 10 %, and 50 % is employed, which is changed to 10 %, 50 %, and 100 % at the late stage (explained in Sect. 2.5). When B_{rm} equals 100 %, GENOA evaluates the removal of each reaction.

2.2.2 Jumping strategy

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

As shown in Table 2, the Criegee intermediate BCALOO, formed during the ozonolysis of BCAL (reaction no. 11 in Table 1), is jumped over to its only destruction product BCLKET. Consequently, reaction nos. 12 to 16 are removed, and reaction no. 11 is updated to Reaction (R1) ("R" for reaction after reduction strategy).

There are similarities between reduction by jumping and prereduction in the sense that both can jump reactions without affecting organic compounds. However, the two processes serve different purposes, as prereduction is intended to provide a reliable reference mechanism for training, whereas jumping is used in training to search for possible reductions. On the one hand, the current prereduction only reduces very fast degraded radicals that undergo a single unimolecular reaction with a constant kinetic rate coefficient (e.g., no temperature effect). In this case, one species may lead to several degradation products. As these reactions are extremely fast and independent of atmospheric conditions, they only cause numerical issues in simulation and should be removed from the reference mechanism. On the other hand, jumping may be relatively slow or affected by environmental conditions; therefore, an evaluation is necessary. Jumping is currently limited from one species to another at a time. The difference in carbon numbers between reduced species can not exceed three in order to prevent significant differences in organic mass before and after jumping. As shown in Table 2, the degradation of BCALOO into BCLKET involves five bimolecular reactions, which may affect SOA formation under different atmospheric conditions (e.g., with different inorganic concentrations and relative humidity, RH).

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

No.	Reaction	Kinetic rate coefficient[a]
12	$BCALOO + CO \rightarrow BCLKET$	1.2×10^{-15}
13	$BCALOO + NO \rightarrow BCLKET + NO_2$	1.0×10^{-14}
14	$BCALOO + NO_2 \rightarrow BCLKET + NO_3$	1.0×10^{-15}
15	$BCALOO + SO_2 \rightarrow BCLKET + SO_3$	7.0×10^{-14}
16	$BCALOO \rightarrow BCLKET + H_2O_2$	$1.4 \times 10^{-17} \times [H_2O]$
R1[b]	$BCAL + O_3 \rightarrow 1.5\ BCLKET + 0.5\ C146O2 + 0.37\ CH2OO + HCHO + 0.13\ HO_2 + 0.63\ CO + 1.13\ OH$	1.1×10^{-16}

[a] $[H_2O]$ is the concentration of H_2O. [b] Reaction (R1) is updated from reaction no. 11 in Table 1.

Table 3. Explicit reactions of the MCM species BCAO2, BCBO2, and BCCO2 in the degradation scheme of β-caryophyllene (BCARY).

No.	Reaction [a]	Kinetic rate coefficient[b]
17[c]	$BCARY + OH \rightarrow 0.408\ BCAO2 + 0.222\ BCBO2 + 0.37\ BCCO2$	1.97×10^{-10}
18	$BCAO2 + HO_2 \rightarrow BCAOOH$	$KAHO2 = KRO2HO2 \times 0.975$
19	$BCAO2 + NO \rightarrow 0.753\ BCAO + 0.753\ NO_2 + 0.247\ BCANO3$	$KANO = KRO2NO$
20	$BCAO2 + NO_3 \rightarrow BCAO + NO_2$	$KANO3 = KRO2NO3$
21	$BCAO2 + RO_2 \rightarrow 0.7\ BCAO + 0.3\ BCAOH$	$KARO2 = 9.2 \times 10^{-14}$
22	$BCBO2 + HO_2 \rightarrow BCBOOH$	$KBHO2 = KRO2HO2 \times 0.975$
23	$BCBO2 + NO \rightarrow 0.753\ BCBO + 0.753\ NO_2 + 0.247\ BCBNO3$	$KBNO = KRO2NO$
24	$BCBO2 + NO_3 \rightarrow BCBO + NO_2$	$KBNO3 = KRO2NO3$
25	$BCBO2 + RO_2 \rightarrow 0.6\ BCBO + 0.2\ BCAOH + 0.2\ BCBCO$	$KBRO2 = 8.8 \times 10^{-13}$
26	$BCCO2 + HO_2 \rightarrow BCCOOH$	$KCHO2 = KRO2HO2 \times 0.975$
27	$BCCO2 + NO \rightarrow 0.753\ BCCO + 0.753\ NO_2 + 0.247\ BCCNO3$	$KCNO = KRO2NO$
28	$BCCO2 + NO_3 \rightarrow BCCO + NO_2$	$KCNO3 = KRO2NO3$
29	$BCCO2 + RO_2 \rightarrow 0.7\ BCCO + 0.3\ BCCOH$	$KCRO2 = 9.2 \times 10^{-14}$

[a] Species RO_2 represents the sum of all peroxy radicals. [b] The same symbols are used to demonstrate the reduction strategies shown in Tables 4 and 5. The precise values of kinetic rate coefficients (i.e., KRO2HO2, KRO2NO, and KRO2NO3) can be found on the MCM website (v3.3.1, http://mcm.york.ac.uk/home.htt, last access: 25 April 2022) (in cm^3 $molec.^{-1}$ s^{-1}). [c] Reaction no. 17 shows the production of BCAO2, BCBO2, and BCCO2, whereas the other reactions (nos. 18 to 29) depict their destruction processes.

2.2.3 Lumping strategy

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

The lumping strategy is illustrated by the comparison of Table 3 (reactions before lumping) and Table 4 (reactions after lumping). In this example, a total of 13 chemical reactions (nos. 17 to 29) involving three organic compounds are reduced to five reactions (a production reaction, Reaction R2, and four destruction reactions , Reactions (R3) to (R6), of the new surrogate).

As demonstrated in the tables, the organic compounds BCAO2, BCBO2, and BCCO2 from the original MCM scheme are the peroxy radicals formed from the OH-initiated oxidation of β-caryophyllene (Table 3). It is evident from their structures (shown in Fig. C1) that they are isomers and may share similar chemical properties. When applying the lumping strategy, BCAO2, BCBO2, and BCCO2 are merged into a new surrogate named "mBCAO2" (Table 4). Additional lumping examples are provided in Appendix C1, describing the lumping of compounds with differing structural groups derived from different oxidation reactions.

The key parameter that drives the reduction accuracy is the "weighting ratio" of lumping (f_w), corresponding to the weight of the original species in the new surrogate compound. As detailed in Table 4, f_w is computed as a function of the chemical lifetime τ following the computation of Seinfeld and Pandis (2016), and the reference concentrations C_r that are the arithmetic mean concentrations calculated from 0-D simulations using the reference mechanism. Both τ and

Table 4. Reduced reactions of Table 3 via the lumping strategy, in the case of lumping BCAO2, BCBO2, and BCCO2 into a new surrogate mBCAO2. The name of the new surrogate contains the letter "m" for "merged" and the name of the relatively dominant lumped species. This notation of lumping is used hereafter.

No.[a]	Lumped[b]	Reaction	Kinetic rate coefficient
R2	17	BCARY + OH → mBCAO2	1.97×10^{-10}
R3	18, 22, 26	mBCAO2 + HO_2 → $f_{w,a}$ BCAOOH + $f_{w,b}$ BCBOOH + $f_{w,c}$ BCCOOH	$f_{w,a}$ KAHO2 + $f_{w,b}$ KBHO2 + $f_{w,c}$ KCHO2
R4	19, 23, 27	mBCAO2 + NO → 0.753 × ($f_{w,a}$ BCAO + $f_{w,b}$ BCBO + $f_{w,c}$ BCCO) + 0.247 × ($f_{w,a}$ BCANO3 + $f_{w,b}$ BCBNO3 + $f_{w,c}$ BCCNO3) + 0.753 × ($f_{w,a}$ + $f_{w,b}$ + $f_{w,c}$) NO_2	$f_{w,a}$ KANO + $f_{w,b}$ KBNO + $f_{w,c}$ KCNO
R5	20, 24, 28	mBCAO2 + NO_3 → $f_{w,a}$ BCAO + $f_{w,b}$ BCBO + $f_{w,c}$ BCCO + ($f_{w,a}$ + $f_{w,b}$ + $f_{w,c}$) NO_2	$f_{w,a}$ KANO3 + $f_{w,b}$ KBNO3 + $f_{w,c}$KCNO3
R6	21, 25, 29	mBCAO2 + RO_2 → 0.7 × $f_{w,a}$ BCAO + 0.8 × $f_{w,b}$ BCBO + 0.7× $f_{w,c}$ BCCO + 0.2 × $f_{w,b}$ BCBCO + (0.3 × $f_{w,a}$ + 0.2 × $f_{w,b}$) BCAOH + 0.3× $f_{w,c}$ BCCOH	$f_{w,a}$ KARO2 + $f_{w,b}$ KBRO2 + $f_{w,c}$KCRO2

Symbol[c]	Meaning	Computation
$C_{r,a}$	Reference concentration of BCAO2[d]	Average BCAO2 concentrations from 5 d 0-D simulations under the training dataset
$C_{r,b}$	Reference concentration of BCBO2	Average BCBO2 concentrations from 5 d 0-D simulations under the training dataset
$C_{r,c}$	Reference concentration of BCCO2	Average BCCO2 concentrations from 5 d 0-D simulations under the training dataset
τ_a	Chemical lifetime of BCAO2[e]	$1/(KAHO2\,[HO_2] + KANO\,[NO] + KANO3\,[NO_3] + KARO2\,[RO_2])$
τ_b	Chemical lifetime of BCBO2	$1/(KBHO2\,[HO_2] + KBNO\,[NO] + KBNO3\,[NO_3] + KBRO2\,[RO_2])$
τ_c	Chemical lifetime of BCCO2	$1/(KCHO2\,[HO_2] + KCNO\,[NO] + KCNO3\,[NO_3] + KCRO2\,[RO_2])$
$f_{w,a}$	Weighting ratio of BCAO2	$\tau_a C_{r,a}/(\tau_a C_{r,a} + \tau_b C_{r,b} + \tau_c C_{r,c})$
$f_{w,b}$	Weighting ratio of BCBO2	$\tau_b C_{r,b}/(\tau_a C_{r,a} + \tau_b C_{r,b} + \tau_c C_{r,c})$
$f_{w,c}$	Weighting ratio of BCCO2	$\tau_c C_{r,c}/(\tau_a C_{r,a} + \tau_b C_{r,b} + \tau_c C_{r,c})$

[a] The reaction number after lumping, where Reactions (R3) to (R6) preserve the destruction of BCAO2, BCBO2, and BCBO2, and Reaction (R2) presents the production. [b] The reaction numbers before lumping as presented in Table 3. [c] The subscript letters a, b, and c stand for BCAO2, BCBO2, and BCCO2, respectively. [d] The calculation method also applies to other BCARY-derived organics. [e] $[X]$ in the calculations is the reference concentration of radical and other inorganic species, where X is HO_2, NO, NO_3, or RO_2 in this case. For radicals derived from the SOA precursor, the reference concentration is the produced concentration without considering their rapid destruction.

C_r are based on the averages of simulations across all training conditions. The properties of the new surrogate compound (e.g., molecular structure, saturation vapor pressure, molar mass, and degradation kinetics) are estimated by weighing the properties of the initial compounds, whereas the stoichiometric coefficients and the kinetic rate coefficient of the new reaction are obtained by weighing those of the initial reactions.

Chemical lifetimes and reference concentrations may be close for species that share similar structures and undergo analogous reactions. In cases where these species originate from the same reaction, they can be lumped directly, with the branching ratios of the formation reaction serving as weighting ratios. As an example, BCAO2, BCBO2, and BCCO2 undergo equivalent reactions, with the exception of the RO_2 reaction of BCBO2. As the BCARY degradation is not very sensitive to RO_2, BCAO2, BCBO2, and BCCO2 can be lumped together with $f_{w,a}$, $f_{w,b}$, and $f_{w,c}$ equal to the branching ratios of reaction no. 17, i.e., 0.408, 0.222, and 0.37, respectively.

Most lumping involves species that are not isomers and undergo different reactions, which makes lumping multiple species at the same time highly uncertain. Therefore, in practice, GENOA attempts to lump only two species in a single reduction in order to ensure the effectiveness of computation. A lumping of multiple species can be achieved by combining several reductions (e.g., first lumping BCAO2 with BCCO2 to form mBCAO2 and then lumping BCBO2 into mBCAO2).

In BCARY reduction, lumping is subject to certain restrictions:

- There should be no lumping between a compound and its oxidation products.
- Compounds with specific structural groups sharing common chemical behavior may be more appropriately merged together. Thus, compounds containing peroxyacetyl nitrate (PAN), organic nitrate (RONO2), organic radical (R), oxy radical (RO), peroxy radical (RO_2), carboxylic acid (RC(O)OH), and peroxycarboxylic acid (RC(O)OOH) functional groups can only be lumped with compounds containing the same groups.
- The difference in the molecular weight should be negligible (i.e., smaller than 100 g mol^{-1}).
- The difference in the carbon number should be no more than two.
- The difference in the chemical lifetime should be less than 10-fold.
- Lumping is not considered for biradicals (ROO) that degrade rapidly into closed-shell molecules, as jumping is considered to be more appropriate for these compounds.

The difference in saturation vapor pressure between "lumpable" condensables is not explicitly restricted in BCARY reduction. However, it is implicitly considered, as GENOA searches and attempts to lump species with similar saturation vapor pressures first. Nonetheless, the user can activate the option to limit the range of saturated vapor pressure differentials between lumpable condensables, along with other user-defined reduction options listed in the Supplement.

2.2.4 Replacement strategy

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

Table 5 illustrates a reduction occurring via the replacement strategy (to be compared to the original mechanism in Table 3), assuming that BCAO2 is predominant in SOA formation. By substituting both BCBO2 and BCCO2 with BCAO2, the OH reaction of BCARY only leads to the production of BCAO2. The MCM reaction nos. 17 to 29 can then be reduced to Reactions (R2′) to (R6′) via replacement.

The replacement strategy (Table 5) is expected to reduce the computational time more than the lumping strategy (Table 4), as all reactions originating from the replaced species are removed from the mechanism. Hence, it does not require the computation of weighting ratios and new surrogates. However, as a compromise, replacement could be less accurate than lumping, as replacement may discard some compounds and part of the mechanism, thereby leading to more error.

Thus, in an effort to prioritize the accuracy of reduction, GENOA currently employs replacement only after lumping and exclusively on species from the same reaction. In this way, species that were not lumped (because lumping was rejected or because they do not respect the lumping restriction) can be reduced by replacement. During the training of BCARY reduction, a restriction is applied on small organic compounds with a molar mass of less than $100\,\mathrm{g\,mol^{-1}}$, which are excluded from replacement. The difference in carbon number is no more than three.

Overall, the searches for viable reductions are conducted in reverse order of the reaction/species list. For removal, GENOA attempts to remove reactions from the bottom of the list and moves to the previous reactions. The same reverse sequence is followed for other strategies. When applied to the jumping strategy, for instance, GENOA tries to jump the species that has the highest generation and then move down to the species that has the lowest generation. Among all reduction strategies, only lumping alters the saturation vapor pressure of condensable species. Therefore, a rank of saturation vapor pressure is used exclusively in lumping to determine the most appropriate lumpable species. At each reduction, GENOA attempts to reduce only one species/reaction via removal or one pair of compounds via lumping/replacing/jumping. This restriction allows exhaustive tracking of every detailed modification and its effect on SOA concentrations.

2.3 Datasets of atmospheric conditions applied to reduction

All of the atmospheric conditions applied to the reduction are extracted from a 3-D simulation spanning the latitudes from 32 to 79° N and the longitudes from 17° W to 39.8° E over continental Europe in a 1-year period (2015) using the CHIMERE chemistry transport model. The CHIMERE model and the configuration used for the simulation are described in Lanzafame et al. (2022). The 3-D CHIMERE simulation was conducted with the implicit gas-phase MELCHIOR2 mechanism (Derognat et al., 2003), which contains 120 reactions and less than 80 lumped species. The MELCHIOR2 mechanism describes the degradation of sesquiterpenes by three oxidant-initiated reactions (HUMULE reacts with OH, O_3, and NO_3, respectively), where the species HUMULE represents the lumped class of all sesquiterpenes.

The monthly diurnal profiles of hourly meteorological data (e.g., temperature and RH) as well as the hourly concentrations of oxidant, radical, and other inorganic species were extracted from each location. This information is required in the 0-D simulations with SSH-aerosol (see Sect. 2.4) to reproduce SOA concentrations and compositions under near-realistic conditions. As the reduced SOA mechanism focuses only on SOA formation, the meteorological data and the concentrations of oxidants, radicals, and inorganics are assumed to remain intact during the 0-D SOA simulation. The coordinates and time of each condition are also provided to calculate the solar zenith angle. The concentration of HUMULE (denoted C_{SQT} as the CHIMERE surrogate for sesquiterpene) is used to estimate the SQT concentration. For the purpose of calculating reduction parameters (e.g., the weighting ratio f_{w} and the branching ratio B) and evaluating the reduced mechanisms, a dataset of representative physiochemical conditions extracted from CHIMERE simulation results is employed in GENOA. Depending on their usage, three groups of conditions are defined: the training dataset, the pretesting dataset, and the testing dataset.

2.3.1 Training dataset

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

Table 5. Reduced reactions of Table 3 via the replacement strategy, in the case of replacing BCBO2 and BCCO2 with one existing species – BCAO2.

No.[a]	Replaced	Reaction	Kinetic rate coefficient
R2′	17	BCARY + OH → BCAO2	1.97×10^{-10}
R3′	18, 22, 26	BCAO2 + HO_2 → BCAOOH	KAHO2
R4′	19, 23, 27	BCAO2 + NO → 0.753 BCAO + 0.753 NO_2 + 0.247 BCANO3	KANO
R5′	20, 24, 28	BCAO2 + NO_3 → BCAO + NO_2	KANO3
R6′	21, 25, 29	BCAO2 + RO_2 → 0.7 BCAO + 0.3 BCAOH	KARO2

[a] The symbol ′ is used to distinguish the reactions in this table from the corresponding number of lumping reactions in Table 4.

number of conditions can be evaluated at each reduction attempt.

For the reduction of BCARY degradation, a training dataset of eight conditions is selected, which contains six chemistry-relevant conditions and two additional meteorological conditions. The geographic and meteorological information of each condition is described in Table 6, where the conditions cover a broad range in time (summer and winter conditions), temperatures ranging from 260 to 302 K, and RH values from 39 % to 89 %.

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

The chemical regimes of the different conditions can be illustrated by seven competitive reaction ratios (equations are listed in Appendix Table C1):

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

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

Figure 2 describes the reaction ratios at midnight (00:00 h) and noon (12:00 h) for the training conditions. Under the majority of atmospheric conditions, O_3 is the dominant oxidant of BCARY due to the carbon–carbon double bonds that are subject to ozonolysis. The high-O_3 training conditions have a R_{O_3} ratio exceeding 98 % at both noon and midnight. The bimolecular reactions with NO and HO_2 dominate RO_2 reactions in the MCM mechanism. Due to the low kinetic rate constants and low concentrations, the ratios of OH and NO_3 reacting with BCARY are relatively low (under 40 %). The high-OH conditions are determined by the OH ratio at noon, whereas the high-NO_3 conditions are determined by $R_{RO_2-NO_3}$ at midnight. One specific exception is the additional condition ADD2, which is located in the northern part of Italy, within the Alpine arch, close to the metropolitan city of Milan. This condition is in the extremely high NO_x regime, as high concentrations of NO are transported from polluted areas. These high NO concentrations consume O_3 and NO_3, causing low concentrations of O_3 and NO_3. At night, ADD2 has a high R_{OH} ratio of 95 % at midnight that is not due to an abundance of OH but rather to extremely low concentrations of O_3 (2.9×10^{-4} ppb) and NO_3 (1.1×10^{-9} ppb) which lead to an absence of nighttime reactivity.

2.3.2 Pre-testing dataset

The pre-testing dataset contains a greater number of conditions than the training dataset, covering the major atmospheric conditions encountered across the domain. After the mechanism has been significantly reduced, the pre-testing dataset is included along with the training dataset in order to evaluate the reduction attempts at the late-stage reduction. At this point of reduction, a slight change in the mechanism significantly impacts the SOA concentrations; therefore, merely evaluating reduction based on the training dataset may not be adequate. Meanwhile, the size of the mechanism has already been significantly reduced, which makes the evaluation of each reduction attempt on the pre-testing dataset less computationally expensive.

In principle, the pre-testing dataset should be able to provide a fairly accurate representation of the testing dataset. However, this may not always be the case, as the pre-testing dataset is selected almost randomly from the testing dataset. Therefore, an adjustment may be required to increase the representativeness of the pre-testing dataset by adding or removing a few conditions. For the application to BCARY, a pre-testing dataset with 150 atmospheric conditions is selected

Table 6. Geographic and meteorological conditions of the training dataset. The table headings, from left to right, indicate the name, latitude, longitude, time period, average temperature, average RH, average daily NO reaction ratio, and simulated total SOA concentration of the training conditions.

Condition name	Lat (° N)	Long (° E)	Time (month)	Temp. (K)	RH (%)	R_{RO_2-NO} [a] (%)	SOA[b] ($\mu g\,m^{-3}$)
OH NO	36.0	15.4	Jul	299	79	60	4.1
OH HO_2	32.0	−9.4	Jul	296	77	20	6.1
NO_3 NO	40.25	−3.4	Jul	302	28	69	4.4
NO_3 HO_2	32.0	36.6	Aug	302	38	29	5.7
O_3 NO	69.0	33.8	Jan	261	84	99	5.2
O_3 HO_2	68.0	18.2	Dec	266	89	25	4.6
ADD1	41.5	−14.2	Dec	289	76	20	5.5
ADD2	45.75	9.0	Dec	279	85	100.	4.4

[a] The average daily NO reaction ratio is calculated using the RO_2 reactivity of NO, HO_2, NO_3, and RO_2. Conditions with a high R_{NO} ratio are considered in the high-NO_x regime. [b] SOA is simulated with an initial BCARY concentration of $5\,\mu g\,m^{-3}$.

Figure 2. A bar plot showing the occupancy of seven reaction ratios in the BCARY initiation reactions and RO_2 reactions, under the training conditions at midnight (00:00 GMT+1, top bar) and noon (12:00 GMT+1, bottom bar with hatching). From left to right, six ratios are presented on each bar in the following order: R_{O_3}, R_{OH}, R_{NO_3}, R_{RO_2-NO}, $R_{RO_2-HO_2}$, $R_{RO_2-NO_3}$, and $R_{RO_2-RO_2}$ (no display if ratio is zero). Table C1 provides the equations for calculating the reaction ratios.

from the testing dataset. The pre-testing dataset consists of 50 conditions for each level (low, medium, and high) of SQT emissions (see Sect. 2.3.3). The locations of the training and pre-testing conditions are presented in Fig. 3.

2.3.3 Testing dataset

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

In the BCARY reduction, the testing dataset is selected based on the concentrations of the CHIMERE sesquiterpene surrogate. Its maximum hourly concentration C_{SQT} in parts per billion (ppb) is used to exclude conditions with a negligible SQT concentration. A testing dataset within a total of 12 159 conditions is applied (see Sect. 3.2), including all conditions (2159 conditions) with a high SQT concentration ($C_{SQT} \geq 0.1$ ppb), 5000 random select conditions with a medium SQT concentration ($C_{SQT}\epsilon$ between 0.01 and 0.1 ppb), and 5000 random select conditions with a low SQT concentration ($C_{SQT}\epsilon$ (0.001, 0.01]). The conditions with an extremely low SQT concentration ($C_{SQT} < 0.001$ ppb) are not included in the testing dataset. Figure B1 indicates the locations of the testing dataset as well as the testing results for BCARY reduction.

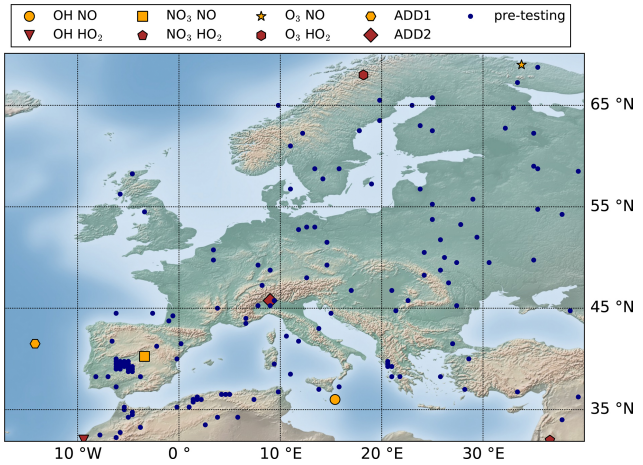


Figure 3. Simulation domain and locations of training (see the figure legend) and pre-testing (blue scattered dots) datasets applied to the reduction.

2.4 Settings for SOA simulations

The chemical composition and temporal variation in SOA due to gas-phase chemistry and condensation/evaporation are simulated using the 0-D aerosol module SSH-aerosol (Sartelet et al., 2020). As detailed in Couvidat and Sartelet (2015), the gas–particle partitioning is estimated with Raoult's law (for the partitioning between the gas phase and the organic phase) and Henry's law (for the partitioning between the gas phase and the aqueous phase). Therefore, some properties of condensable compounds, such as the saturation vapor pressure P_{sat} and the decomposition into functional groups, are crucial for modeling. For BCARY-derived organics, P_{sat} is calculated using UManSysProp (Topping et al., 2016). The vapor pressure is computed using the method of Nannoolal et al. (2008) and the boiling point estimation from Joback and Reid (1987). These methods were selected because they provide the best performance when compared with the chamber experiment data of Chen et al. (2012) and Tasoglou and Pandis (2015), as discussed in Appendix A. Furthermore, the activity coefficient γ is calculated with the UNIQUAC Functional-group Activity Coefficients (UNIFAC) thermodynamic model (Fredenslund et al., 1975) for short-range interactions and the Aerosol Inorganic–Organic Mixtures Functional groups Activity Coefficients (AIOMFAC) model for medium-range and long-range interactions (Zuend et al., 2008).

Unless stated otherwise, two simulations are performed for each condition starting at midnight (00:00 h) and noon (12:00 h), considering both the daytime and nighttime chemistry. All 0-D simulations are run for 5 d in order to adequately consider SOA formation and aging processes. The initial BCARY concentration is set to 5 $\mu g\,m^{-3}$ in order to ensure high SOA production (the SOA concentration is always greater than 1 $\mu g\,m^{-3}$ under all evaluated conditions). For optimal computational efficiency, the gas–particle partitioning is assumed to be at thermodynamic equilibrium.

2.5 Settings for evaluation

For the different datasets, the performance of the reduced mechanism on SOA concentrations is evaluated using the fractional mean error (FME) computed with Eq. (1), where $C_{\text{val},i}$ and $C_{\text{ref},i}$ denote the SOA mass concentration at time step i simulated with the reduced and the reference mechanisms, respectively.

The error of one simulation is defined as the larger of the FME on day 1 and the FME on days 2 to 5, in order to address the difference in the performance of the reduced mechanisms in the early stage of the simulations (SOA formation dominates) and in the later stage (SOA aging dominates). This error is used to evaluate reduction by comparing it to the error tolerance specified in training. For the evaluation on the training dataset, two errors are estimated compared to the previously verified reduced mechanism with a tolerance denoted ϵ_{pre} and to the reference mechanism with a tolerance denoted ϵ_{ref}. The error tolerances are used to restrict both the maximum and the average (half of the tolerance) errors of the training conditions. As for the evaluation on the pre-testing dataset, the error compared to the reference mechanism is calculated. The error tolerances $\epsilon^{\text{ave}}_{\text{pre-testing}}$ and $\epsilon^{\text{max}}_{\text{pre-testing}}$ are set to the average and maximum errors, respectively.

$$\text{FME} = \frac{2\sum_n^{i=1}\text{abs}(C_{\text{val},i} - C_{\text{ref},i})}{n\sum_n^{i=1}(C_{\text{val},i} + C_{\text{ref},i})} \quad (1)$$

In order to begin with a conservative BCARY reduction, the initial values of ϵ_{pre} and ϵ_{ref} are both set to 1 %. The values of these error tolerances are then increased to larger values, reflecting the looser criteria used throughout the reduction. ϵ_{ref} is used to track the performance of the reduction, while ϵ_{pre} is used to avoid large errors introduced by one reduction attempt. Therefore, ϵ_{pre} is lower than or equal to ϵ_{ref}. For every 1 % increase in ϵ_{ref}, ϵ_{pre} is stepped up by 1 % from 1 % to the value of ϵ_{ref}. By doing this, GENOA first accepts reductions that introduce small errors compared with the previously validated mechanism and then accepts reductions that introduce larger errors up to ϵ_{ref}.

The maximum values for both ϵ_{ref} and ϵ_{pre} are set to 10 %. When ϵ_{ref} reaches 3 %, the mechanism is expected to be largely reduced. From then, the evaluation under the pre-testing dataset is considered to be added to the reduction. This means that all subsequent reductions are evaluated using both the training and pre-testing datasets. The average and maximum errors ($\epsilon^{\text{ave}}_{\text{pre-testing}}$ and $\epsilon^{\text{max}}_{\text{pre-testing}}$) are restricted to be lower than 3 % and 20 %, respectively. As a result of the above error tolerances, a reduced SQT SOA mechanism with an average inaccuracy on SOA formation lower than 3 % (maximum 20 %) is expected.

Additionally, another error factor noted as the fractional bias (FB, computed as detailed in Eq. 2) is used to visualize

the temporal performance of the reduced mechanism at each simulation time step. As examples, Figs. 8 and 10 show the average FB at each time step for the pre-testing conditions.

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

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

2.6 Settings for aerosol-oriented treatments

In late-stage training, an intense competition between different potential reductions is observed, and a minor modification may induce significant uncertainty in the mechanism and prevent further reduction. Moreover, because the formation of aerosols costs more CPU time than gas-phase chemistry, specific treatments are employed in the late stage of training to reduce the number of condensable species preferentially. These treatments, which reduce species rather than reactions, are done when the size of the mechanism is below a certain threshold. For BCARY reduction, the treatments are activated once the number of condensable species has decreased to 20. Consequently, late-stage treatments encourage reduction via the removal of condensable species and are referred to as aerosol-oriented treatments. The treatments consist of the following:

- Restriction of the reduction of the number of reactions is applied; thus, strategies that reduce the number of aerosols are favored to result in fewer condensable species.
- The evaluation of aerosol-oriented reductions on the training dataset is bypassed when applied to jumping, lumping, and replacement. As a result, the aerosol-oriented reduction is evaluated only on the pre-testing dataset to avoid being rejected under some of the extreme conditions in the training dataset (which are less representative of average atmospheric conditions than the conditions of the pre-testing dataset).
- An additional type of removal is applied – removing elementary-like reactions.

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

3 Application to the β-caryophyllene mechanism

GENOA is applied to the SQT degradation mechanism of v3.3.1 of the Master Chemical Mechanism (Jenkin et al., 2012). Here, β-caryophyllene (BCARY) is considered a surrogate for SQT primary VOCs. The degradation of β-caryophyllene in the original MCM mechanism consists of 1626 reactions and 579 species (223 radicals and 356 stable species). After prereduction, the mechanism contains 1241 reactions and 493 species (137 radicals and 356 stable species); this is employed as the starting point and the reference for the reduction (hereafter referred to as MCM).

Moreover, at the beginning of the GENOA training, all of the stable species are assumed to be condensable (referred to as condensables), and their saturation vapor pressures and activity coefficients are calculated based on their molecular structures (as detailed in Sect. 2). Applying the effective partitioning coefficients (K_{p} at 298 K) described by Seinfeld and Pandis (2016), condensables can be classified into semi-volatile organic compounds (SVOCs; K_{p} between 10^{-2} and $10\,\mathrm{m^3\,\mu g^{-1}}$), low-volatility organic compounds (LVOCs; K_{p} between 10 and $10^4\,\mathrm{m^3\,\mu g^{-1}}$), and extremely low volatility organic compounds (ELVOCs; K_{p} larger than $10^4\,\mathrm{m^3\,\mu g^{-1}}$).

The semi-explicit SQT SOA mechanism "Rdc." presented in this section is trained from MCM with GENOA. Detailed descriptions of the building process and its chemical scheme are provided in Sect. 3.1. By the end of the training, Rdc. is reduced from MCM to only 23 reactions and 15 species (see Appendix B for the reaction and species lists). The size of the Rdc. mechanism is of the same order of magnitude as the BCARY degradation scheme of Khan et al. (2017) (28 reactions and 15 species) used for global modeling. As presented in Sect. 3.2, the Rdc. mechanism accurately reproduces the SOA concentration and composition simulated by MCM with only six condensables. Table B3 summarizes the new surrogates and the lumped MCM species that are included in the final Rdc. mechanism.

3.1 Building of the reduced SOA mechanism

As shown in Fig. 4, the Rdc. mechanism is built from 113 validated reduction steps. In GENOA, a reduction step refers to all reduction attempts based on the performed reduction strategy and reduction parameters, while a validated reduction step indicates at least one reduction attempt has been accepted at this step. The entire building process can be divided into three stages:

- *Early stage* refers to the period from the 1st to the 74th reduction step. By the end of the 74th reduction step, the mechanism is reduced to 68 reactions and 41 species (including 20 condensables). The early-stage reduction is trained only on the training dataset with the seven predescribed reduction strategies. After ϵ_{ref} reaches 3 %, the list of B_{rm} is changed from [5 %, 10 %, 50 %] to [10 %, 50 %, 100 %].

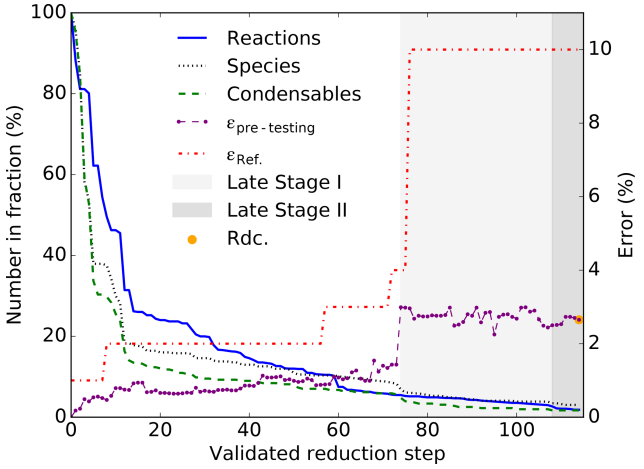


Figure 4. Reduction process of the Rdc. mechanism showing the decrease in the number of reactions, species, and condensables; the evolution of the average error on the pre-testing dataset ($\epsilon_{\text{pre-testing}}$, with an error tolerance $\epsilon^{\text{ave}}_{\text{pre-testing}}$ of 3 %); and the error tolerance compared with MCM (ϵ_{ref}).

Table 7. Reduction accomplished per each reduction strategy during the building process of the Rdc. mechanism.

Strategy	No. (fraction[a], %)		
	Reaction[b]	Species	Condensables
Removing reactions	594 (48)	38 (8)	26 (7)
Removing elementary-like reactions[c]	8 (1)	0 (0)	0 (0)
Jumping	138 (11)	79 (16)	43 (12)
Lumping	0 (0)	171 (35)	110 (31)
Replacement	25 (2)	39 (8)	31 (9)
Removing species	453 (37)	151 (31)	108 (30)
Removing partitioning	0 (0)	0 (0)	32 (9)
Removed in total	1 218 (98)	478 (97)	350 (98)

[a] The fraction of the original number (of reactions or species) that is reduced by the strategy. [b] The columns, from left to right, are the number (and fraction) of reduced chemical reactions, reduced total gas-phase species, and reduced gas-phase species that can condense on the particle phase, compared with MCM with 1241 reactions and 493 species (356 condensables). [c] This step is only applied in the reduction at late stage II.

– *Late stage I* spans from the 75th to the 107th reduction step. By the end of the 107th reduction step, the reduced mechanism consists of 38 reactions and 19 species (including seven condensables), and no further reduction can be found within $\epsilon_{\text{ref}} \leq 10\,\%$ and $\epsilon_{\text{pre}} \leq 10\,\%$. In this stage, the reduction is trained on the pre-testing dataset if condensables are removed with jumping, lumping, or replacement. For reduction with other types of reduction strategies, it is first trained on the training dataset and then on the pre-testing datasets. From all of the reduced mechanisms with seven condensables, GENOA selects the one with the minimum average errors on the pre-testing dataset (2.44 %) to start the next stage.

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

The extent of the reduction due to each strategy is summarized in Table 7. Compared with MCM, up to 98 % of reactions and 97 % of species are reduced in Rdc. As expected, the reduction strategy of removing reactions contributes the most to the decrease in the number of reactions (48 %), followed by the strategy of removing species (with a contribution of 37 %). Meanwhile, both lumping and removing species are significant in the reduction of species (by 35 % and 31 %, respectively). The number of condensables decreases in proportion to the number of species, except for the strategy of removing partitioning. In that case, the gas–particle partitioning is removed and the species remains in the gas phase with no changes in the chemical mechanism.

As shown in Fig. 5, which describes the chemical scheme of the Rdc. mechanism, the three oxidants (i.e., O_3, OH, and NO_3) initiated reactions, leading to common oxidation products (e.g., mBCSOZ and mBCALO2) that dominate the successive oxidations. The different reaction pathways under high- or low-NO_x regimes are presented in Rdc. with reactions with NO or HO_2, respectively, which results in the formation of different types of SOAs: mBCKSOZ, mC133O, and C131PAN (in the presence of NO_2) under high-NO_x conditions and mC132OOH under low-NO_x conditions. Other pathways, such as the bimolecular reactions of the Criegee intermediate BCBOO with water vapor and the RO_2 reaction of mBCALO2, are also preserved in the Rdc. mechanism. The six condensables in Rdc. can be categorized into one SVOC, four LVOCs, and one ELVOC, according to the effective partitioning coefficient calculated on the pre-testing dataset. The SOA concentration per volatility class is discussed in Sect. 3.2.

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

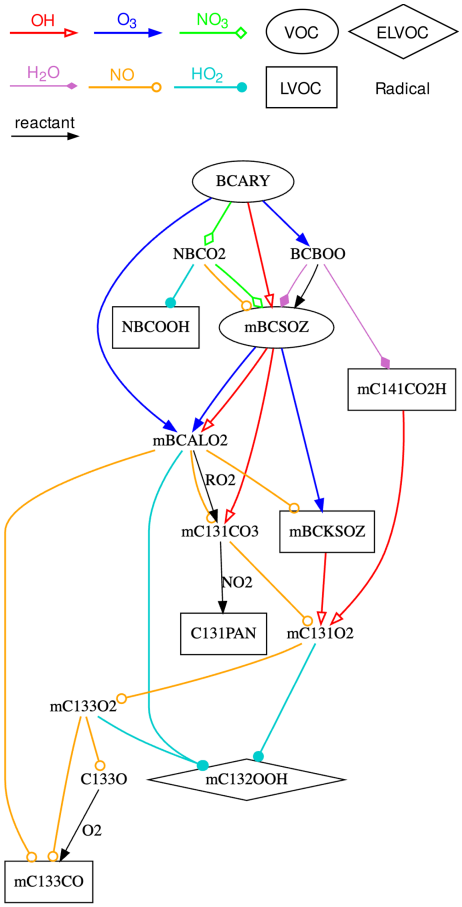


Figure 5. Representation of the chemical scheme of the Rdc. mechanism. VOCs, LVOCs, and ELVOCs are presented in ellipse, square, and diamond boxes, respectively. Radicals are written in plain text, without boxes. Reactions with OH, O_3, NO_3, NO, HO_2, and H_2O are shown using arrows with different colors and heads (see the figure legend). Other reactants (if any) are labeled near the edges. The complete species and reaction lists of the Rdc. mechanism are given in Appendix Tables B1 and B2, respectively.

uct formed from both OH- and O_3-initiated chemistry), and BCKET (from OH-initiated reactions).

3.2 Evaluation of the reduced SOA mechanism

3.2.1 Reproduction of the SOA concentrations

During the testing procedure, the Rdc. mechanism is evaluated at 12 159 locations, with two different starting times (00:00 and 12:00 h). The testing for Rdc. took approximately 2 % of the CPU time consumed for MCM.

Compared with MCM, Rdc. presents a high level of accuracy with an average error of 2.66 % and a maximum error of 17.29 %. The monthly distribution of the number of the testing conditions as well as the testing errors are described in Fig. 6. The error is lower than 10 % for more than 99 % of the simulations. The summer conditions, between June and

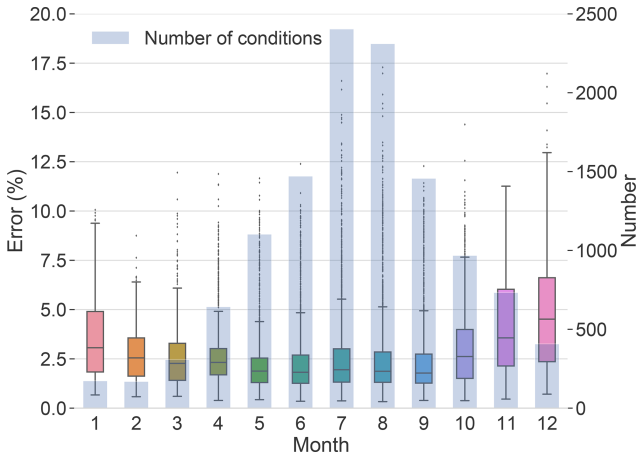


Figure 6. Monthly distribution of the testing results (errors compared with MCM) of the Rdc. mechanism in the box plot as well as the number of testing conditions in the histogram.

September, covering more than half of the testing conditions (63 %, 7647 conditions), result in an average error of 2.37 % and a 3rd quartile error of 2.85 %. Compared with the summer conditions, testing results under winter conditions, from October to January (19 % of the testing dataset, 2285 conditions), display slightly higher uncertainty, with an average error of 3.79 % and a 3rd quartile error of 5.36 %.

An error map of testing conditions in July and August is displayed in Fig. 7. It indicates the locations of testing conditions and the errors of each condition, especially highlighting outliers during this period. Detailed error maps of all testing conditions can be found in Appendix B. It shows that the Rdc. mechanism induces low errors (lower than 6 %) for most of the testing conditions. The conditions with errors over 6 % are mainly concentrated in northern Africa near the Atlas Mountains and in the eastern Mediterranean, where the conditions most likely correspond to a dry Mediterranean climate with low RH and high temperature. Other conditions with errors above 6 % are dispersed in the Po Valley of northern Italy and along the coasts of southern Spain. More accurate results could be obtained with stricter parameters for reduction (e.g., lower error tolerance) or by updating the conditions (e.g., training and pre-testing datasets) covering more extreme conditions in the training process.

3.2.2 Reproduction of the SOA composition

The SOA concentrations and chemical composition simulated with the Rdc. mechanism and with MCM are compared in this section. The temporal profiles of the total SOA concentrations on an average of the pre-testing dataset and non-ideal conditions are displayed in Fig. 8. Throughout the entire 5 d simulation period, there is excellent agreement between hourly SOA concentrations simulated with MCM and those obtained from the Rdc. mechanism. The SOA concen-

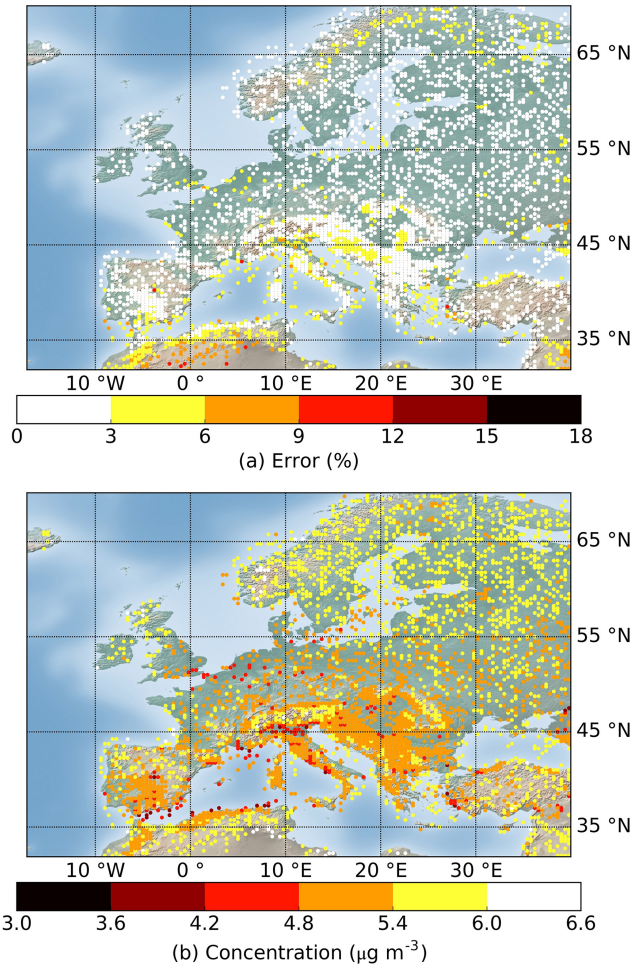


Figure 7. Geographic distributions of the **(a)** error and **(b)** average SOA concentration of the testing results in July and August simulated using the Rdc. mechanism. The total number of conditions displayed is 4717 out of the 12 159 that were tested. The results of all testing conditions are shown in Appendix B for reference.

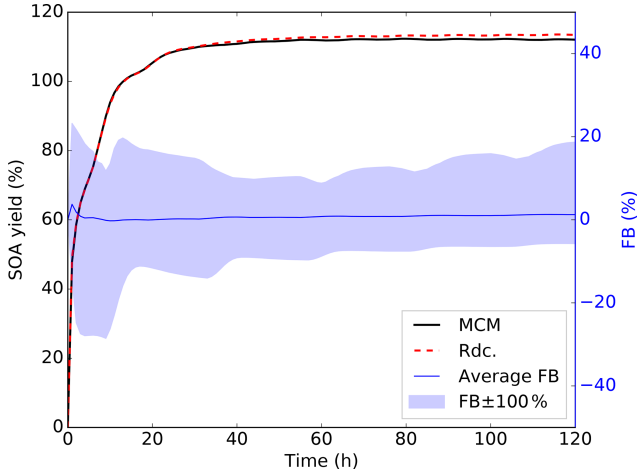


Figure 8. Temporal variation in the total SOA concentration simulated with the pre-testing dataset using the MCM (red dashed line) and Rdc. (solid black line) mechanisms under nonideal conditions. The average (solid blue line) and maximum (blue shading) FB values between the MCM and the Rdc. mechanisms are also presented.

Table 8. Average SOA concentrations per volatility class simulated with the MCM and the Rdc. mechanisms on the pre-testing dataset at 8 and 72 h (in $\mu g\,m^{-3}$).

Conditions	SVOCs	LVOCs	ELVOCs	Total
MCM at 8 h	0.18	1.91	2.17	4.26
Rdc. at 8 h	0.13	1.80	2.31	4.24
MCM at 72 h	0.02	1.90	3.69	5.61
Rdc. at 72 h	0.02	1.51	4.12	5.65

tration builds up rapidly in the first few hours, where the results of the Rdc. mechanism present relatively larger fluctuations (the maximum FB of 3.74 % is observed at 1 h on the average pre-testing results).

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

The average SOA composition per functional group simulated on the pre-testing dataset at 72 h is displayed in Fig. 9. No significant change in the functional group distributions is found between 8 and 72 h of oxidation. The alkyl (C) and carbonyl groups (RCO) contribute the most to the SOA mass, by more than $1\,\mu g\,m^{-3}$, whereas the other functional groups contribute by less than $1\,\mu g\,m^{-3}$. Overall, the Rdc. mechanism satisfactorily reproduces the composition of the MCM-simulated SOA composition for most functional groups, except for nitrogen-containing groups. In comparison to MCM, only two condensables containing nitrogen are retained in the Rdc. mechanism – NBCOOH and C131PAN – leading to an underestimation of the organic nitrate group ($0.31\,\mu g\,m^{-3}$ in MCM and $0.04\,\mu g\,m^{-3}$ in Rdc.) and an overestimation of the nitrate mass of the peroxyacetyl nitrate group ($0.10\,\mu g\,m^{-3}$ in MCM and $0.30\,\mu g\,m^{-3}$ in Rdc.). To obtain better results on the reproduction of nitrogen groups, GENOA may be further restricted to distinguish nitrogen compounds in training.

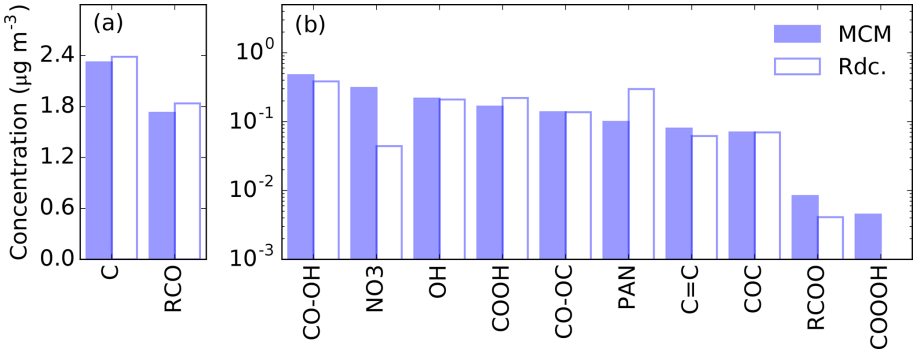


Figure 9. Average SOA mass per functional group simulated with the pre-testing dataset using the MCM (blue bar) and Rdc. mechanisms (white bar) at 72 h. The figure is divided into two panels, **(a)** and **(b)**, due to the large gap in mass between the groups. The labels of the functional groups, from left to right, are as follows: C – carbon bond, RCO – carbonyls (ketone and aldehyde), CO–OH – hydroxy peroxide, NO_3 – organic nitrates, OH – alcohol, COOH – carbonyl acid, CO–OC – peroxide, PAN – peroxyacyl nitrates, C=C – carbon double bond, COC – ether, RCOO – ester, and COOOH – peroxyacetyl acid.

Additionally, the peroxyacetyl acid group results in an extremely low SOA mass in MCM (less than 0.01 %); therefore, it is not retained in the Rdc. mechanism.

Moreover, the temporal profiles of the organic mass to organic carbon mass (OM/OC) ratio as well as the H/C, O/C, and N/C atomic ratios are presented in Fig. 10. Comparable patterns are observed in the OM/OC (1.65 in MCM and 1.63 in Rdc. on average), the O/C (0.37 in MCM and 0.36 in Rdc.), and the H/C (1.62 in MCM and 1.60 in Rdc.) ratios. During the first 8 h of simulation, Rdc. tends to slightly overestimate the OM/OC and O/C ratios, while the H/C ratio remains fairly stable throughout the entire simulation with a negligible difference (0.02) between MCM and Rdc. The N/C ratio, however, is underestimated by the Rdc. mechanism by 37 % on average (ratio equal to 0.019 in MCM and to 0.012 in Rdc.), indicating the over-reduction of organic nitrites in Rdc. A total of three nitrogen-containing organics (NBCO2, NBCOOH, and C131PAN) are preserved in Rdc., two of which (NBCO2 and NBCOOH) are first-generation products. Therefore, during the first 10 h, the N/C ratio curve simulated by Rdc. drops, whereas it increases in MCM as higher-generation nitrates are produced.

3.2.3 Sensitivity on environmental parameters

The sensitivities of the Rdc. mechanism to temperature, RH, and SOA mass conditions are investigated with the pre-testing dataset. The default value of the BCARY concentration is 5 $\mu g\,m^{-3}$, and the default RH and temperature are set to constant values of 50 % and 298 K, respectively. As presented in Fig. 11, the SOA yields simulated by the Rdc. mechanism with different environmental parameters show a remarkable resemblance to the SOA yields simulated by MCM.

Under 10 $\mu g\,m^{-3}$, the simulated SOA yields are not affected by the SOA mass loading. This result is consistent with the large contribution of ELVOCs reported in Table 8. A discrepancy of 25 % in the average SOA yield at 1 h with an SOA mass loading of 10^3 $\mu g\,m^{-3}$ at 1 h and a discrepancy of 8 % at 72 h with an SOA mass loading of 10^{-3} $\mu g\,m^{-3}$ are observed. The result indicates that the Rdc. mechanism may introduce relatively large uncertainty with extreme SOA loading (larger than 500 $\mu g\,m^{-3}$), which was outside the range of conditions used for the construction of the Rdc. mechanism. SOA formation is affected by RH, due to both the gas-phase chemistry (reaction with H_2O vapors) and the gas–particle transfer (condensation of hydrophilic SOA precursors on aqueous aerosols). The sensitivity tests show that the Rdc. mechanism reproduces (differences lower than 2 %) the SOA yields of MCM well with RH values ranging from 5 % to 95 %. For temperature, the Rdc. mechanism reproduces the SOA aging at 72 h very well, but larger discrepancies are observed in the earlier period, when the oxidation products are more volatile. However, the discrepancies in the SOA yield stay low: differences up to 7 % (at 1 and 72 h) and 10 % (at 8 h) are observed for temperatures of 263 and 323 K, respectively. This finding is consistent with the testing results. In summary, the discrepancies suggest that the reduced mechanism performs quite well, although larger discrepancies with MCM are observed under conditions that are outside the range of conditions used during training.

4 Conclusions

The development and application of the GENerator of reduced Organic Aerosol mechanism (GENOA v1.0) have been presented in this study. GENOA generates semi-explicit SOA mechanisms designed for large-scale air quality modeling by reducing explicit VOC mechanisms with a series of automatic training and testing processes. During the training procedure of GENOA, four types of reduction strategies (removal, jumping, lumping, and replacement) are adopted to locate the potential reduction in the mechanism. Each reduc-

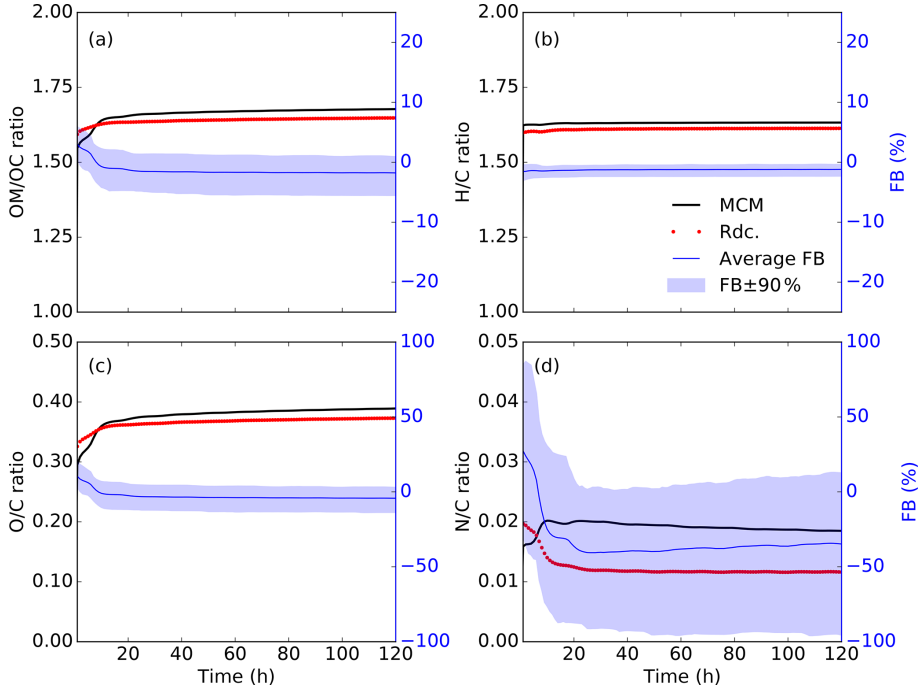


Figure 10. Temporal variations in the **(a)** average organic mass to organic carbon mass (OM/OC) ratio, **(b)** hydrogen to carbon (H/C) atomic ratio, **(c)** oxygen to carbon (O/C) atomic ratio, and **(d)** nitrogen to carbon (N/C) atomic ratio, simulated with MCM (solid black curves) and the Rdc. mechanism (dotted red curves) on the pre-testing dataset. The average FB (solid blue line) and the 90 % range of the FB (blue shading) are also presented.

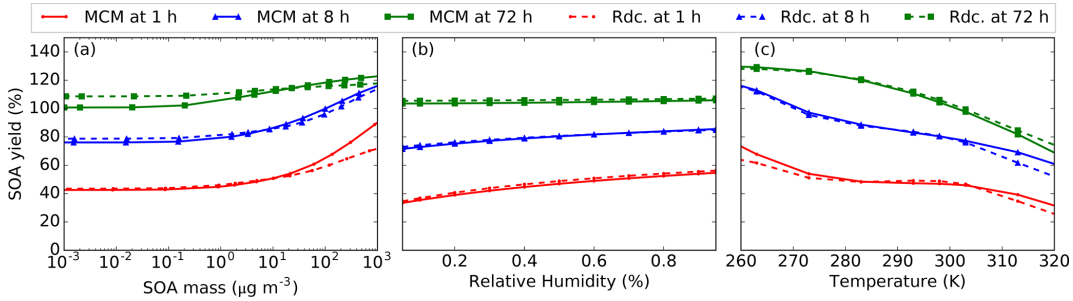


Figure 11. Dependence of the average SOA yield simulated by the pre-testing dataset with MCM (solid line) and the Rdc. mechanism (dashed line) on **(a)** BCARY SOA mass; **(b)** relative humidity (RH); and **(c)** temperature at 1 h (red points), 8 h (blue triangles), and 72 h (green squares).

tion attempt is evaluated against the explicit mechanism under a sequence of near-realistic atmospheric conditions (the training dataset, and/or the pre-testing dataset at the late stage of reduction). Finally, the reduced mechanism is evaluated under various conditions of a testing dataset. Under each condition, two 5 d 0-D simulations starting at midnight and noon are conducted with the SSH-aerosol model to simulate SOA concentrations and compositions for reduction evaluation.

GENOA successfully generated semi-explicit SOA chemical mechanisms for the degradation of sesquiterpene, for which the explicit β-caryophyllene mechanism of the Master Chemical Mechanism serves as the reference mechanism and the starting point. The final reduced SQT SOA mechanism contains 23 reactions (down from 1626 reactions in MCM), 15 gas-phase species (down from 579 gases), and 6 aerosol species (down from 356 aerosols). It reproduces the SOA formation and aging by introducing an average error of 2.7 % under conditions over Europe with only 2 % of the size of MCM. The SOA volatility is well reproduced with the reduced mechanism, as well as the decomposition into functional groups, and the OM/OC (1.55 in the Rdc. mechanism and 1.60 in MCM), H/C, and O/C ratios. Nitrogen-containing SOA, which contributes to only 7 % of the total mass, is not as well represented as other groups, and the N/C ratio is slightly underestimated in the Rdc. mechanism (0.016 compared with 0.021 in MCM). The similarity of the

representation of the functional group decomposition allows for the similar reproduction of the non-ideality of SOA in the Rdc. mechanism and in MCM. Additionally, the sensitivity tests on RH, temperature, and organic mass loading show that the SOA simulated with the Rdc. mechanism is in good agreement with MCM results under most conditions (except for conditions with extremely high temperature or with massive organic aerosol loading where discrepancies in the SOA yields may reach 8 % (temperature) and 25 % (massive mass loading)). This indicates that the reduced mechanism performs well for conditions in the training range, but its performance may deteriorate for conditions outside of this range. To improve the performance of the semi-explicit SOA mechanism under conditions outside of the training range, two methods can be employed: the first is to include the outlier conditions in the training procedure if they are considered influential to SOA formation, and the second is to adopt strict error tolerance to restrict the reduction.

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

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

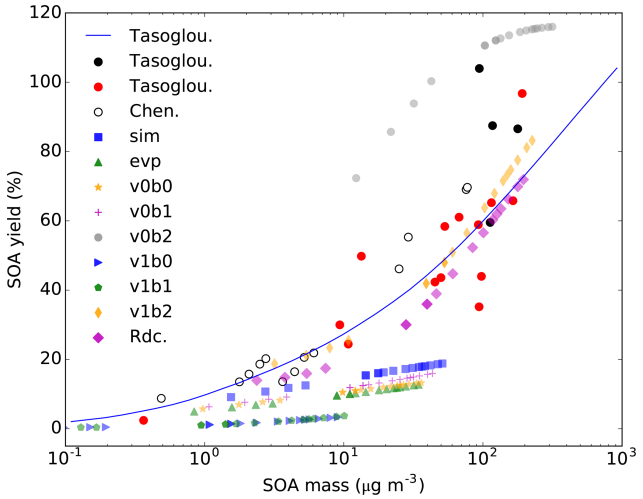


Figure A1. The SOA yields versus the total SOA mass from the experimental data reported by Chen et al. (2012) and Tasoglou and Pandis (2015), simulated in SSH-aerosol with the MCM mechanism and different saturation vapor pressures methods (see the figure legend) and simulated with the Rdc. mechanism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

Appendix B: An overview of the Rdc. mechanism

Table B1. Species list of the Rdc. mechanism. Notice that the species in the reduced case may be different from the MCM species with identical names.

Surrogate[a]	Type[b]	Molecular formula	MW[c]	P^{d}_{sat}	ΔH^{e}_{vap}	H^{f}	γ^{g}
BCARY	VOC	$C_{15}H_{24}$	204.4				
NBCO2	Radical	$C_{15}H_{24}NO_5$	298.4				
BCBOO	Radical	$C_{15}H_{24}O_3$	252.3				
mBCALO2	Radical	$C_{14.68}H_{24.08}O_{4.87}$	278.5				
mBCSOZ	VOC	$C_{15}H_{24}O_{2.74}$	248.2				
NBCOOH	LVOC	$C_{15}H_{25}NO_5$	299.4	4.04×10^{-12}	119	7.63×10^{6}	1.8×10^{6}
mC141CO2H	LVOC	$C_{14.94}H_{23.85}O_{3.06}$	252.4	4.43×10^{-11}	114	2.53×10^{4}	4.95×10^{7}
mBCKSOZ	SVOC	$C_{14}H_{22}O_{3.9}$	252.7	2.02×10^{-8}	91.2	1.89×10^{5}	1.45×10^{4}
mC131CO3	Radical	$C_{14.09}H_{21.28}O_{4.91}$	269.1				
mC131O2	Radical	$C_{13.14}H_{21.22}O_{4.14}$	245.5				
mC133CO	ELVOC	$C_{13.42}H_{20.83}O_{4.59}$	255.6	1.62×10^{-12}	125	245	1.4×10^{11}
mC132OOH	ELVOC	$C_{13.97}H_{23.92}O_{4.59}$	279.5	4.13×10^{-14}	136	5.70×10^{3}	2.36×10^{11}
C131PAN	LVOC	$C_{14}H_{21}NO_7$	315.3	4.39×10^{-11}	113	2.07×10^{4}	6.09×10^{7}
mC133O2	Radical	$C_{13}H_{21}O_{5.97}$	272.8				
C133O	Radical	$C_{13}H_{21}O_5$	257.3				

[a] Species with "m" are the new surrogates that merged multiple MCM BCARY species. [b] VOCs (stable gas-phase species) and radicals (unstable gas-phase species) are assumed not to undergo gas–particle partitioning. The volatility classes of condensable species are defined in Sect. 3. [c] Molar weight ($g\,mol^{-1}$). The properties calculated for condensable substances only are as follows: [d] saturation vapor pressure at 298 K (atm), [e] enthalpy of vaporization ($kJ\,mol^{-1}$), [f] Henry's law constant ($mol\,L^{-1}\,atm^{-1}$), and [g] activity coefficient at infinite dilution in water.

Table B2. Reaction list of the Rdc. mechanism.

No.	Reactions	Kinetic rate constant[a]
1	$BCARY + NO_3 \rightarrow NBCO2 + NO_3$	1.9×10^{-11}
2	$BCARY + O_3 \rightarrow 0.874\ BCBOO + 0.111\ mBCALO2 + O_3$	1.19×10^{-14}
3	$BCARY + OH \rightarrow mBCSOZ + OH$	1.97×10^{-10}
4	$NBCO2 + HO_2 \rightarrow NBCOOH + HO_2$	$2.837 \times 10^{-13} \times \exp(\frac{1300}{T})$
5	$NBCO2 + NO \rightarrow mBCSOZ + NO$	$2.7 \times 10^{-12} \times \exp(\frac{360}{T})$
6	$NBCO2 + NO_3 \rightarrow mBCSOZ + NO_3$	2.3×10^{-12}
7	$BCBOO \rightarrow 0.5\ mC141CO2H + 0.5\ mBCSOZ$	$[H_2O] \times 4 \times 10^{-16}$
8	$BCBOO \rightarrow mBCSOZ$	2×10^{2}
9	$mBCSOZ + O_3 \rightarrow 0.915\ mBCKSOZ + 0.085\ mBCALO2 + O_3$	1.1×10^{-16}
10	$mBCSOZ + OH \rightarrow 0.92\ mBCALO2 + 0.08\ mC131CO3 + OH$	7.6×10^{-11}
11	$mC141CO2H + OH \rightarrow mC131O2 + OH$	6.494×10^{-11}
12	$mBCALO2 + NO \rightarrow 0.505\ mBCKSOZ + 0.353\ mC131CO3 + 0.099\ mC133CO + NO$	$6.56 \times 10^{-12} \times \exp(\frac{360}{T})$
13	$mBCALO2 \rightarrow 0.6\ mC131CO3$	$[RO_2] \times 1.711 \times 10^{-12}$
14	$mBCALO2 + HO_2 \rightarrow mC132OOH + HO_2$	$1.939 \times 10^{-13} \times \exp(\frac{1300}{T})$
15	$mBCKSOZ + OH \rightarrow mC131O2 + OH$	3.28×10^{-11}
16	$mC131CO3 + NO \rightarrow mC131O2 + NO$	$6.377 \times 10^{-12} \times \exp(\frac{290}{T})$
17	$mC131CO3 + NO_2 \rightarrow C131PAN + NO_2$	$0.8502 \times$ KFPAN
18	$mC131O2 + HO_2 \rightarrow mC132OOH + HO_2$	$2.288 \times 10^{-13} \times \exp(\frac{1300}{T})$
19	$mC131O2 + NO \rightarrow mC133O2 + NO$	$2.213 \times 10^{-12} \times \exp(\frac{360}{T})$
20	$mC133O2 + HO_2 \rightarrow mC132OOH + HO_2$	$2.695 \times 10^{-13} \times \exp(\frac{1300}{T})$
21	$mC133O2 + NO \rightarrow 0.757\ C133O + 0.243\ mC133CO + NO$	$2.61 \times 10^{-12} \times \exp(\frac{360}{T})$
22	$C133O \rightarrow mC133CO$	$[O_2] \times 2.5 \times 10^{-14} \times \exp(\frac{-300}{T})$
23	$C133O \rightarrow$	$2.7 \times 10^{14} \times \exp(\frac{-6643}{T})$

[a] $[H_2O]$ is the concentration of H_2O, $[RO_2]$ is the total concentration of the RO_2 species pool, $[O_2]$ is the concentration of O_2, and KFPAN is one of the complex rate coefficients from the MCM mechanism.

Table B3. The new surrogates in the Rdc. mechanism and the corresponding lumped species in the original MCM mechanism. Notice that the Rdc. surrogates may also go through other reductions (i.e., jumping, replacement, and removal) that do not affect their molecular structures.

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

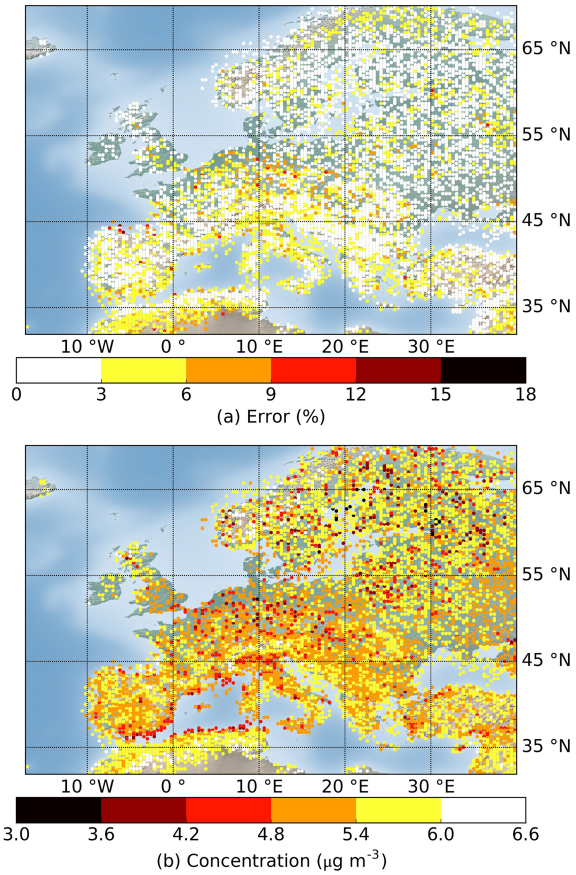


Figure B1. Maps of the **(a)** error and **(b)** average SOA concentration of the testing results simulated using the Rdc. mechanism on all (i.e., 12 159) testing conditions.

Appendix C: Information related to the reduction

C1 Additional examples of lumping

Besides the example shown in Sect. 2.2.3, two additional examples have been added from the BCARY reduction: one illustrates the lumping of two similar compounds formed by different reactions, and the other illustrates the lumping of two more distinct compounds. The first example is the MCM species C1313NO3 and C152NO3 (see Table C2). These two species come from different reactions. The molecular structures of both compounds are similar (they contain organic nitrates, aldehydes, and alcohols), but C152NO3 contains an additional carboxylic acid where C1313NO3 contains an aldehyde. The corresponding reactions before and after lumping are summarized in Table C2, where the new surrogate "mC1313NO3" is built from C1313NO3 with a weighting ratio of 83 % and C152NO3 with a weighting ratio of 17 %. As a result of this lumping, the average error increase under training conditions is 0.001 % (the tolerance is 0.01 %).

Another example of lumping is the MCM species BCALBOC and C1310OH (see Table C3). Unlike the previous example, these two species are more distinct. According to MCM, BCALBOC is generated through O_3-initiated reactions, whereas C1310OH is generated through high-generation oxidations. There is less similarity in the structures or chemical reactions of the two molecules. MCM contains the OH reaction of BCALBOC as well as the O_3 and OH reactions of C1310OH. However, this reduction was accepted because lumping them only increased the average error by 0.01 % under training conditions (the tolerance was 1 %). The new surrogate "mBCALBOC" is constructed from BCALBOC with a weighting ratio of 98 % and C1310OH with a weighting ratio of 2 %.

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

Table C1. The computation of estimating chemical activity ratios used to display training dataset in Fig. 2.

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

[a] Names of the reacting ratio of OH radical, O_3, and NO_3 radical reacted with BCARY ($R_{OH} + R_{O_3} + R_{NO_3} = 1$) and of the reacting ratio of NO, HO_2 radical, RO_2 radical, and NO_3 radical (in the presence of RO_2) reacted with RO_2 species ($R_{RO_2-NO} + R_{RO_2-HO_2} + R_{RO_2-RO_2} + R_{RO_2-NO_3} = 1$). [b] Reactions with those compounds are preferred when the corresponding reaction ratios are high. [c] [species_name] (e.g.,[OH]) is the monthly average concentration of oxidant concentrations extracted from CHIMERE. [d] Kinetic rate coefficients are provided by MCM, where k_{OH}, k_{O_3}, and k_{NO_3} are the kinetic rate coefficient of first-generation BCARY reaction with OH, O_3, and NO_3, respectively; k_{NO}, k_{HO_2}, and k_{RNO_3} are the simple rate coefficients KRO2NO, KRO2HO2, and KRO2NO3, respectively; and k_{RO_2} represents the self-reaction rate coefficients for the tertiary peroxy radicals (e.g., BCAO2 and BCCO2). T is temperature (K).

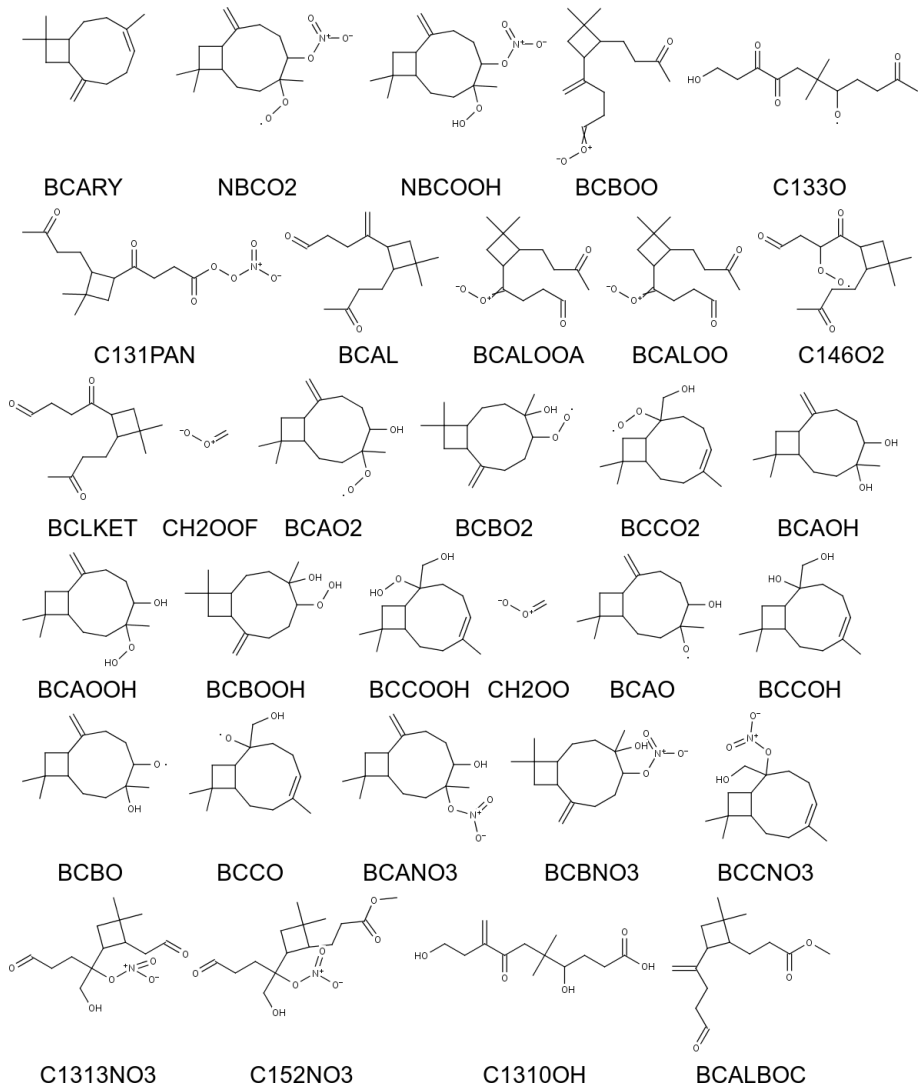


Figure C1. Molecular structures of the MCM species that are mentioned in the paper. For more information, please visit the MCM website.

Table C2. Reactions related to the reduction of the MCM species C1313NO3 and C152NO3 via lumping. The exact weighting ratio of C1313NO3 is 0.82945, and the exact weighting ratio of C152NO3 is 0.17055.

Reactions before lumping	Kinetic coefficient
Production	
$C1313O2 + NO \rightarrow C1313NO3$	$KRO2NO \times 0.134$
$C152O2 + NO \rightarrow C152NO3$	$KRO2NO \times 0.136$
Destruction	
$C1313NO3 + OH \rightarrow C116CHO + HCHO + NO_2$	5.59×10^{-11}
$C152NO3 + OH \rightarrow BCLKBOC + HCHO + NO_2$	1.58×10^{-11}
Reactions after lumping	
Production	
$C1313O2 + NO \rightarrow 0.134\ mC1313NO3 + 0.866\ C1313O + 0.866\ NO_2$	KRO2NO
$C152O2 + NO \rightarrow 0.136\ mC1313NO3 + 0.864\ C152O + 0.864\ NO_2$	KRO2NO
Destruction	
$mC1313NO3 + OH \rightarrow C116CHO + HCHO + NO_2$	$5.59 \times 10^{-11} \times 0.82945$
$mC1313NO3 + OH \rightarrow BCLKBOC + HCHO + NO_2$	$5.59 \times 10^{-11} \times 0.17055$

Table C3. Reactions related to the reduction of the MCM species BCALBOC and C1310OH via lumping. The exact weighting ratio of BCALBOC is 0.97675, and the exact weighting ratio of C1310OH is 0.023251.

Reactions before lumping	Kinetic coefficient
Production	
$BCOOA \rightarrow BCALBOC$	$1.0 \times 10^6 \times 0.15$
$C1310O2 \rightarrow C1310OH$	$2.5 \times 10^{-13} \times [RO_2] \times 0.2$
Destruction	
$BCALBOC + O_3 \rightarrow BCBOOA + HCHO$	$1.1 \times 10^{-16} \times 0.670$
$BCALBOC + O_3 \rightarrow BCLKBOC + CH2OOF$	$1.1 \times 10^{-16} \times 0.330$
$BCALBOC + OH \rightarrow C152O2$	6.98×10^{-11}
$C1310OH + OH \rightarrow C1310CO + HO_2$	6.2×10^{-11}
Reactions after lumping	
Production	
$BCOOA \rightarrow mBCALBOC$	$1.0 \times 10^6 \times 0.15$
$C1310O2 \rightarrow mBCALBOC$	$2.5 \times 10^{-13} \times [RO_2] \times 0.2$
Destruction	
$mBCALBOC + O_3 \rightarrow BCBOOA + HCHO$	$1.10 \times 10^{-16} \times 0.670 \times 0.97675$
$mBCALBOC + O_3 \rightarrow BCLKBOC + CH2OOF$	$1.10 \times 10^{-16} \times 0.330 \times 0.97675$
$mBCALBOC + OH \rightarrow C1310CO + HO_2$	$6.20 \times 10^{-11} \times 0.97675$
$mBCALBOC + OH \rightarrow C152O2$	$6.98 \times 10^{-11} \times 0.023251$

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

Supplement. The supplement related to this article is available online at: https://doi.org/10.5194/gmd-15-8957-2022-supplement.

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

Competing interests. The contact author has declared that none of the authors has any competing interests.

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Acknowledgements. The authors would like to thank Youngseob Kim from CEREA for his help with using the SSH-aerosol model.

Financial support. This research has been supported by INERIS and DIM QI2 (Air Quality Research Network on air quality in the Île-de-France region).

Review statement. This paper was edited by Andrea Stenke and reviewed by William Carter and two anonymous referees.

References

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

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

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

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

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

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

Couvidat, F., Debry, E., Sartelet, K., and Seigneur, C.: A hydrophilic/hydrophobic organic (H2O) aerosol model: Development, evaluation and sensitivity analysis, J. Geophys. Res.-Atmos., 117, D10, https://doi.org/10.1029/2011JD017214, 2012.

Derognat, C., Beekmann, M., Baeumle, M., Martin, D., and Schmidt, H.: Effect of biogenic volatile organic compound emissions on tropospheric chemistry during the Atmospheric Pollution Over the Paris Area (ESQUIF) campaign in the Ile-de-France region, J. Geophys. Res.-Atmos., 108, D17, https://doi.org/10.1029/2001JD001421, 2003.

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

Evtyugina, M., Pio, C., Nunes, T., Pinho, P., and Costa, C.: Photochemical ozone formation at Portugal West Coast under sea breeze conditions as assessed by master chemical mechanism model, Atmos. Environ., 41, 2171–2182, https://doi.org/10.1016/j.atmosenv.2006.10.059, 2007.

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

Gelencsér, A., May, B., Simpson, D., Sánchez-Ochoa, A., Kasper-Giebl, A., Puxbaum, H., Caseiro, A., Pio, C., and Legrand, M.: Source apportionment of PM2.5 organic aerosol over Europe: Primary/secondary, natural/anthropogenic, and fossil/biogenic origin, J. Geophys. Res.-Atmos., 112, D23, https://doi.org/10.1029/2006JD008094, 2007.

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

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

Hallquist, M., Wenger, J. C., Baltensperger, U., Rudich, Y., Simpson, D., Claeys, M., Dommen, J., Donahue, N. M., George, C., Goldstein, A. H., Hamilton, J. F., Herrmann, H., Hoffmann, T., Iinuma, Y., Jang, M., Jenkin, M. E., Jimenez, J. L., Kiendler-Scharr, A., Maenhaut, W., McFiggans, G., Mentel, Th. F., Monod, A., Prévôt, A. S. H., Seinfeld, J. H., Surratt, J. D., Szmigielski, R., and Wildt, J.: The formation, properties and im-

pact of secondary organic aerosol: current and emerging issues, Atmos. Chem. Phys., 9, 5155–5236, https://doi.org/10.5194/acp-9-5155-2009, 2009.

Hellén, H., Schallhart, S., Praplan, A. P., Tykkä, T., Aurela, M., Lohila, A., and Hakola, H.: Sesquiterpenes dominate monoterpenes in northern wetland emissions, Atmos. Chem. Phys., 20, 7021–7034, https://doi.org/10.5194/acp-20-7021-2020, 2020.

Huang, X., Ding, A., Gao, J., Zheng, B., Zhou, D., Qi, X., Tang, R., Wang, J., Ren, C., Nie, W., Chi, X., Xu, Z., Chen, L., Li, Y., Che, F., Pang, N., Wang, H., Tong, D., Qin, W., Cheng, W., Liu, W., Fu, Q., Liu, B., Chai, F., Davis, S. J., Zhang, Q., and He, K.: Enhanced secondary pollution offset reduction of primary emissions during COVID-19 lockdown in China, Natl. Sci. Rev., 8, 2, https://doi.org/10.1093/nsr/nwaa137, 2020.

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

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

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

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

Kanakidou, M., Seinfeld, J. H., Pandis, S. N., Barnes, I., Dentener, F. J., Facchini, M. C., Van Dingenen, R., Ervens, B., Nenes, A., Nielsen, C. J., Swietlicki, E., Putaud, J. P., Balkanski, Y., Fuzzi, S., Horth, J., Moortgat, G. K., Winterhalter, R., Myhre, C. E. L., Tsigaridis, K., Vignati, E., Stephanou, E. G., and Wilson, J.: Organic aerosol and global climate modelling: a review, Atmos. Chem. Phys., 5, 1053–1123, https://doi.org/10.5194/acp-5-1053-2005, 2005.

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

Kim, Y., Couvidat, F., Sartelet, K., and Seigneur, C.: Comparison of Different Gas-Phase Mechanisms and Aerosol Modules for Simulating Particulate Matter Formation, J. Air Waste Manag. Assoc., 61, 1218–1226, https://doi.org/10.1080/10473289.2011.603999, 2011.

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

Lannuque, V., Camredon, M., Couvidat, F., Hodzic, A., Valorso, R., Madronich, S., Bessagnet, B., and Aumont, B.: Exploration of the influence of environmental conditions on secondary organic aerosol formation and organic species properties using explicit simulations: development of the VBS-GECKO parameterization, Atmos. Chem. Phys., 18, 13411–13428, https://doi.org/10.5194/acp-18-13411-2018, 2018.

Lanzafame, G. M., Bessagnet, B., Srivastava, D., Jaffrezo, J. L., Favez, O., Albinet, A., and Couvidat, F.: Modelling aerosol molecular markers in a 3D air quality model: Focus on anthropogenic organic markers, Sci. Total Environ., 835, 155360, https://doi.org/10.1016/j.scitotenv.2022.155360, 2022.

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

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

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

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

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

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

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

Porter, W. C., Jimenez, J. L., and Barsanti, K. C.: Quantifying atmospheric parameter ranges for ambient secondary organic aerosol formation, ACS Earth Space Chem., 5, 2380–2397, https://doi.org/10.1021/acsearthspacechem.1c00090, 2021.

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

Ramanathan, V., Crutzen, P. J., Kiehl, J., and Rosenfeld, D.: Aerosols, climate, and the hydrological cycle, Science, 294, 2119–2124, https://doi.org/10.1126/science.1064034, 2001.

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

Sarwar, G., Luecken, D., Yarwood, G., Whitten, G. Z., and Carter, W. P. L.: Impact of an Updated Carbon Bond Mechanism on Predictions from the CMAQ Modeling Sys-

tem: Preliminary Assessment, J. Appl. Meteorol., 47, 3–14, https://doi.org/10.1175/2007JAMC1393.1, 2008.

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

Schwarze, P. E., Øvrevik, J., Låg, M., Refsnes, M., Nafstad, P., Hetland, R. B., and Dybing, E.: Particulate matter properties and health effects: consistency of epidemiological and toxicological studies, Hum. Exp. Toxicol., 25, 559–579, https://doi.org/10.1177/096032706072520, 2006.

Seinfeld, J. H. and Pandis, S. N.: Atmospheric chemistry and physics: from air pollution to climate change, John Wiley & Sons, ISBN 978-1-119-22117-3, 2016.

Sommariva, R., Trainer, M., de Gouw, J. A., Roberts, J. M., Warneke, C., Atlas, E., Flocke, F., Goldan, P. D., Kuster, W. C., Swanson, A. L., and Fehsenfeld, F. C.: A study of organic nitrates formation in an urban plume using a Master Chemical Mechanism, Atmos. Environ., 42, 5771–5786, https://doi.org/10.1016/j.atmosenv.2007.12.031, 2008.

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

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

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

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

Wang, Z.: tool-genoa/GENOA: GENOA v1.0 (v1.0), Zenodo [code], https://doi.org/10.5281/zenodo.6482978, 2022.

Wang, Z., Couvidat, F., and Sartelet, K.: The electronic supplement of the article "GENerator of reduced Organic Aerosol mechanism (GENOA v1.0): An automatic generation tool of semi-explicit mechanisms" (v1.1), Zenodo [data set], https://doi.org/10.5281/zenodo.6483088, 2022.

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

Xavier, C., Rusanen, A., Zhou, P., Dean, C., Pichelstorfer, L., Roldin, P., and Boy, M.: Aerosol mass yields of selected biogenic volatile organic compounds – a theoretical study with nearly explicit gas-phase chemistry, Atmos. Chem. Phys., 19, 13741–13758, https://doi.org/10.5194/acp-19-13741-2019, 2019.

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

Zhang, Y., Xue, L., Li, H., Chen, T., Mu, J., Dong, C., Sun, L., Liu, H., Zhao, Y., Wu, D., Wang, X., and Wang, W.: Source Apportionment of Regional Ozone Pollution Observed at Mount Tai, North China: Application of Lagrangian Photochemical Trajectory Model and Implications for Control Policy, J. Geophys. Res.-Atmos., 126, e2020JD033519, https://doi.org/10.1029/2020JD033519, 2021.

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

Chapter 3

Parallel reduction algorithm in GENOA v2.0: application to monoterpene SOAs

While the GENerator of reduced Organic Aerosol mechanisms (GENOA v1.0) has been applied to reduce mechanisms describing SOA formation from SQT oxidation, it may not be able to reduce more complex chemical mechanisms from multiple precursors of secondary organic aerosols (SOAs) with the necessary accuracy and efficiency. To address this issue, the second version of the GENOA algorithm (GENOA v2.0) has been developed, with a focus on enhancing reduction efficiency and allowing the simultaneous reduction of mechanisms from several SOA precursors.

The new features in GENOA v2.0 are presented in this chapter. Compared to GENOA v1.0, GENOA v2.0 adopts a novel parallel reduction structure, which locates and evaluates multiple competitive reductions simultaneously. Also included are updates on the reduction search order, the SOA precursor concentration, the reduction strategy, and the evaluation method.

GENOA v2.0 is applied to reducing monoterpene (MT) SOA mechanisms from the Master Chemical Mechanism (MCM) and the Peroxy Radical Autoxidation Mechanism (PRAM), containing 3 001 reactions and 1 227 species, with 738 of them being condensable. The original MCM + PRAM mechanism consists of the individual and common degradation pathways of three MT precursors (i.e., α-pinene, β-pinene, limonene), taking into account the Highly Oxygenated organic Molecules (HOMs) formation which may contribute to up to 50 % of the total MT SOA mass. After reduction, a reduced MT SOA mechanism with 197 reactions and 110 species (23 of them being condensables) is obtained with GENOA v2.0. The obtained mechanism induces an average error of less than 3 % under 9 818 testing conditions over Europe, preserving higher accuracy (2.5 %) under warm conditions (May-to-September) that are likely with higher MT emissions. Furthermore, by increasing the average error tolerance to 20 %, the mechanism can be further reduced to 40 reactions and 24 species (including 5 condensables).

This paper is under review by the Journal of Aerosol Science: **Wang, Z., Couvidat, F., & Sartelet, K. Implementation of a Parallel Reduction Algorithm in the Generator of Reduced Organic Aerosol Mechanisms (Genoa V2. 0): Application to Multiple Monoterpene Aerosol Precursors. Available at http://dx.doi.org/10.2139/ssrn.4431972.**

Contents

Implementation of a parallel reduction algorithm in the Generator of reduced Organic Aerosol Mechanisms (GENOA v2.0): application to SOA from multiple monoterpene precursors

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.

1 Introduction

Organic aerosols affect air quality, climate change, and human health ([Kanakidou, 2005; Hallquist, 2009]). 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, 2003; Gelencsér, 2007]). VOC global emissions are about 1 300 tons of carbon dioxide per year (Tg C/yr) ([Goldstein, 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, 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, 2009; Han, 2023]). The low-volatility oxidation products from VOC degradation may further condense on existing particles, forming SOA ([Hallquist, 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, 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, 1996], the molecular surrogate approach (e.g., [Griffin, 2003; Pun, 2006; Couvidat, 2012]) or the volatility basis set (VBS) approach ([Donahue, 2006; Stolzenburg, 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, 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, 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., [Ehn, 2014; Chen, 2022]). 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, 2020]). Explicit mechanisms are either written manually, such as the Master Chemical Mechanism (MCM, [Jenkin, 1997; Saunders, 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, 2005; Camredon, 2007]). As our understanding of VOC chemistry advances, the development of explicit mechanisms is still ongoing (e.g., [Jenkin, 2012; Coggon, 2019; Jenkin, 2020; Newland, 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., [Ehn, 2014; McFiggans, 2019; Bianchi, 2019]), [Roldin, 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., [Ying, 2011; Li, 2015; Mouchel-Vallon, 2020; Li, 2022]). 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, 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., [Whitehouse, 2004a; Whitehouse, 2004b; Szopa, 2005; Xia, 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, 2008; Watson, 2008; Weber, 2020]) and recently AMORE mechanism ([Wiser, 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 ([Shen, 2022; Kelp, 2022]) and transport ([Sturm, 2023]).

Particularly addressing aerosol formation, [Lannuque, 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, 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., [Schreck, 2022; Mouchel-Vallon, 2022]). Furthermore, [Wang, 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 Sect. 2. The resulting reduced monoterpene SOA mechanism is presented and discussed in Sect. 3, along with a sensitivity analysis. Finally, the conclusion is drawn in Sect. 4.

2 Methods

This section begins with an overview of GENOA v1.0 in Sect. 2.1, along with a discussion of its limitations in Sect. 2.2. The major updates of GENOA v2.0 are summarized in Sect. 2.3, followed by the configuration setups for its application to monoterpene mechanism reduction in Sect. 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, 2022], only the main concepts are reviewed here.

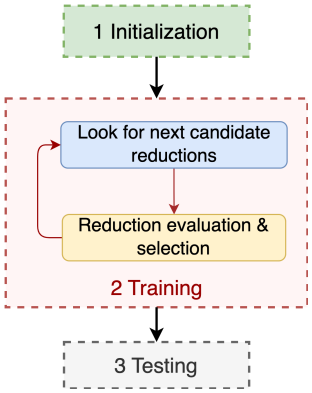


Figure 3.1: Schematic diagram showing the main processes of the GENOA algorithm.

As illustrated in Fig. 3.1, 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 Sect. 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, 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, 2008] and RACM2 [Goliff, 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 Sect. 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 sesquiterpene degradation scheme (as beta-caryophyllene) in MCM v3.3.1 ([Jenkin, 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 on SOA mass ([Wang, 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. 3.2 (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, 2022], only one initial SOA precursor (i.e., β-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 Sect. 2.3.1. Along with the parallel structure, the reduction search order (Sect. 2.3.2), the reduction strategy (Sect. 2.3.3), and the evaluation method (Sect. 2.3.4) are also updated in GENOA v2.0.

2.3.1 Parallel reduction

Figure 3.2 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

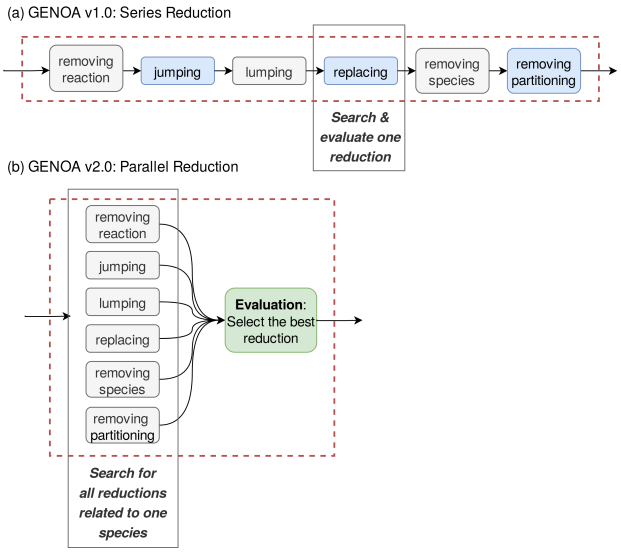


Figure 3.2: Schematic diagram that shows the reduction frameworks in series (a) and in parallel (b) used for GENOA v1.0 and v2.0, respectively.

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 Sect. 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 Sect. 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.

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. 3.3, each strategy offers several candidates for reductions related to the species “A":

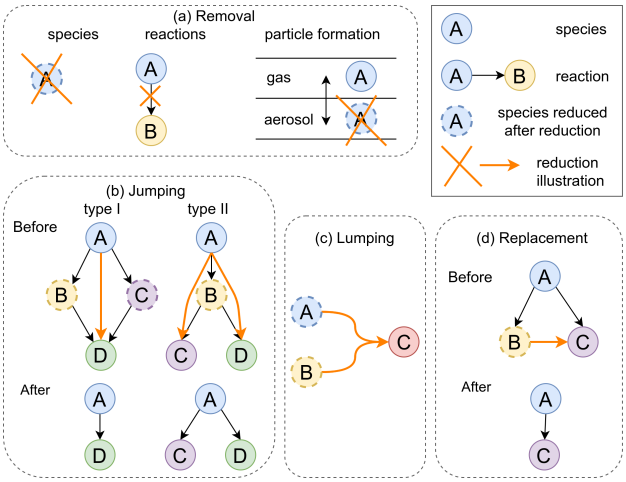


Figure 3.3: Schematic diagram illustrating the different reduction strategies with examples of candidate reductions related to species “A", including (a) removal (of species, reactions, and gas-particle partitioning), (b) jumping (types I and II), (c) lumping, and (d) replacement.

- 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. 3.3 (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. 3.3 (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. 3.3 (a)) removes more intermediates (species “B" and “C") than products (species “D"), while the second type of jumping (type II in Fig. 3.3 (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. 3.3 (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. 3.3 (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, 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 %.

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. 3.1, 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 Sect. 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 Sect. 2.4.4 and Table 3.4.

$$FME = \frac{2\sum_{n}^{i=1} abs(C_{val,i} - C_{cmp,i})}{n\sum_{n}^{i=1}(C_{val,i} + C_{cmp,i})}. \tag{3.1}$$

With the 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. 3.2. 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.

$$S_{rdc} = (N^{now}_{rcn} - N^{pre}_{rcn}) + (N^{now}_{sps} - N^{pre}_{sps}) + (N^{now}_{aero} - N^{pre}_{aero}) \times 10 \tag{3.2}$$

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 Sect. 2.4.1. Details on the evaluation dataset, which includes the atmospheric conditions used for reduction evaluation, can be found in Sect. 2.4.2, while Sect. 2.4.3 covers the initial SOA precursor conditions. Finally, the reduction parameters and options used in each training stage are outlined in Sect. 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, 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 3.1: 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.

Table 3.1 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 3.1. 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, 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 3.1) contains 3 001 reactions, 1 227 species, and 738 condensables.
- Jump rapidly-degraded species with a chemical lifetime of less than 1 second under the training dataset (see Sect. 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, 2016]), employing the method of [Nannoolal, 2008] for vapor pressure and the method of [Nannoolal, 2004] for boiling point. For PRAM condensables, the estimation of P_{sat} is based on the SIMPOL method ([Pankow, 2008]), as originally used in the PRAM evaluation performed by [Roldin, 2019]. These methods were adopted by [Xavier, 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, 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. Figure 3.4 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, 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, 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).

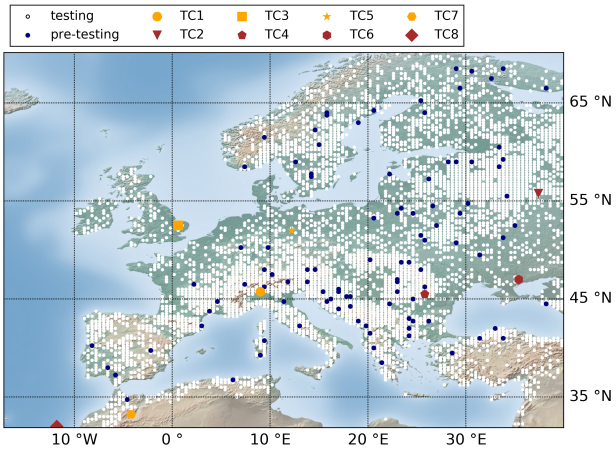


Figure 3.4: Locations of conditions in the training (8 “TC" conditions, detailed in Table 3.2), pre-testing (100 conditions, navy dots), and testing datasets (9 818 conditions, white dots) for the monoterpene mechanism reduction.

Table 3.2: 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.

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 3.2, 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, 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 %.

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 (≥ 0.2 $\mu g/m^{-1}$) and 3 000 randomly selected conditions with medium (between 0.1 $\mu g/m^{-1}$ to 0.2 $\mu g/m^{-1}$) and low (between 0.01 $\mu g/m^{-1}$ to 0.1 $\mu g/m^{-1}$) API concentrations, respectively. From the initially selected testing dataset, 193 conditions with low SOA concentrations (with an hourly maximum SOA concentration lower than 1 $\mu g/m^{-1}$ 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 Sect. 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 3.3 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, 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 3.3: 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 3.4.

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 $\delta\varepsilon$.

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.

- 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 Sect. 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 Sect. 3.1. The final obtained MT-rdc mechanism is described in Sect. 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 Sect. 3.3. MT-rdc can also well reproduce the API SOA yields of [Xavier, 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 Sect. 3.4 and Sect. 3.5, respectively.

3.1 Mechanism evolution during reduction

Figure 3.5 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 3.4 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.

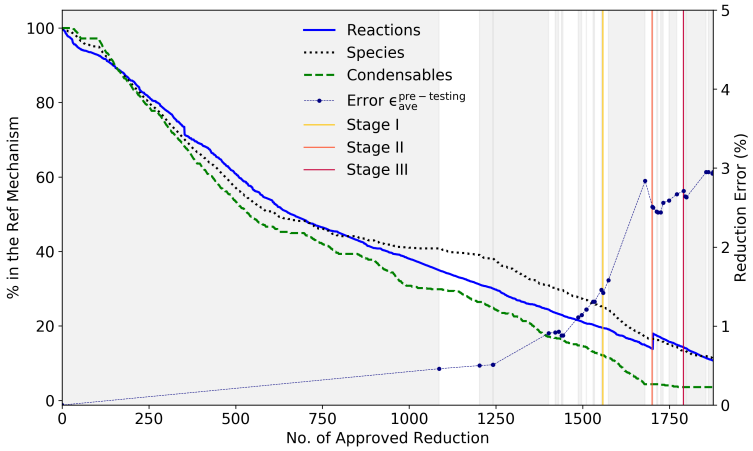


Figure 3.5: Evolution of the size (measured as a percentage reduction in the number of reactions, species, and condensables) and accuracy (measured as $\epsilon_{ave}^{pre-testing}$) of the reduced mechanisms during training compared to the reference mechanism. Vertical gray and white intervals indicate reduction cycles. Vertical yellow, orange, and red bars indicate the end of training stages I, II, and III, respectively.

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

Table 3.4: 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 $\delta\varepsilon$ (%)	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: $\varepsilon_{ave}^{ref} = \varepsilon_{max}^{ref}$ / 2, $\varepsilon_{ave}^{pre} = \varepsilon_{max}^{pre}$ / 2, $\delta\varepsilon = \varepsilon_{max}^{ref}$ / 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.

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.

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 3.5: 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.

Table 3.5 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, 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 3.6: 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).

Table 3.6 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 3.6, 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. 3.6. 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. 3.7, 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.

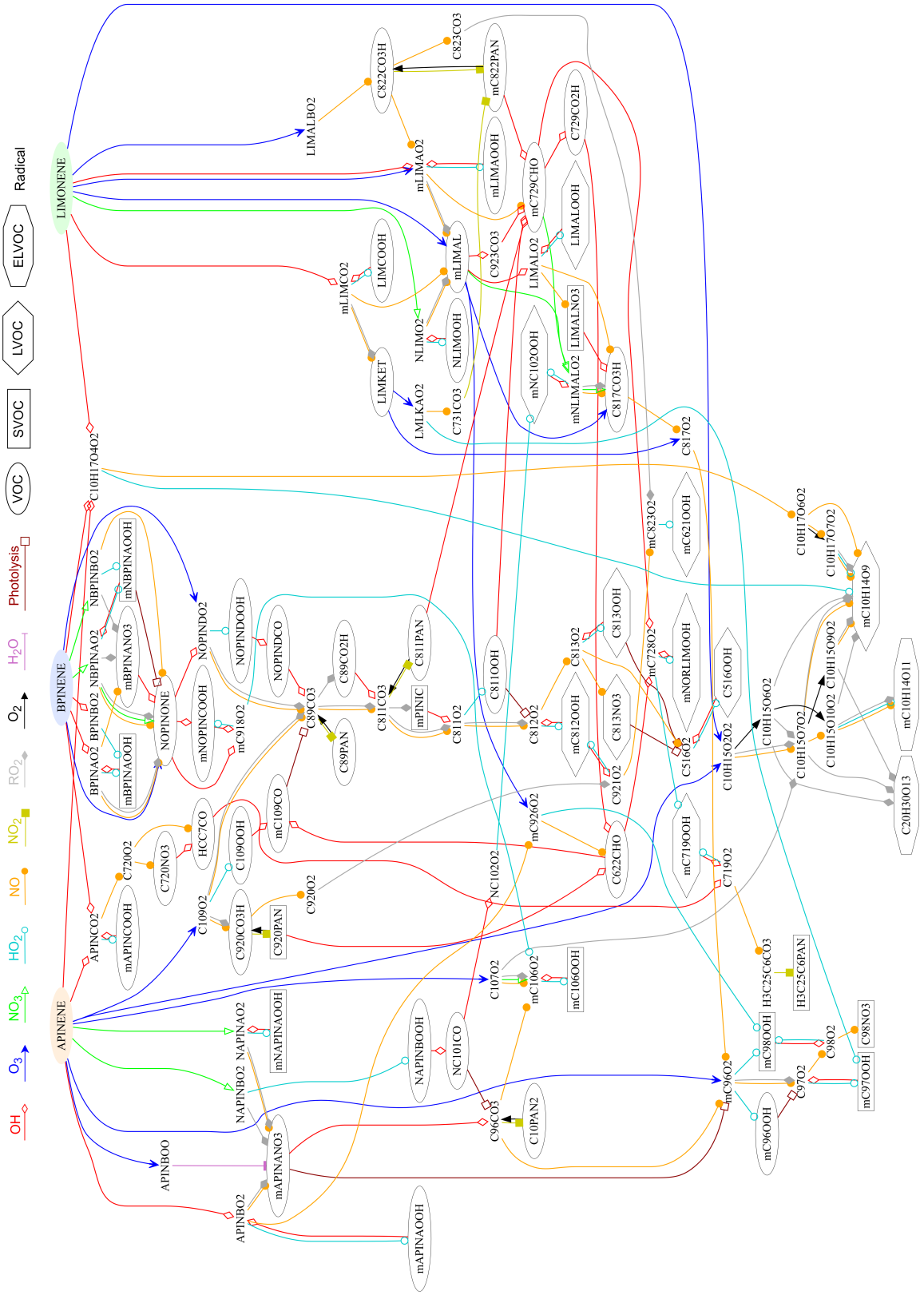


Figure 3.6: Reaction pathways of the MT-rdc mechanism. The VOC oxidation reactions with OH radical, Ozone, and NO_3 radical are depicted by red, blue, and green lines, respectively, with different arrowhead symbols pointing to reaction products: empty diamond, filled "V"-type, and empty triangle. Reactions 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.

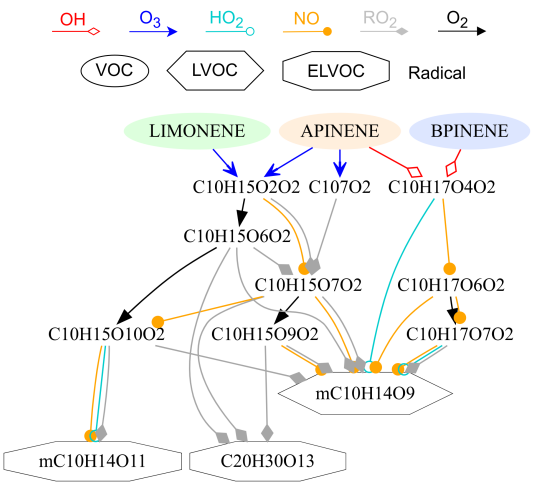


Figure 3.7: Reaction pathway in the MT-rdc mechanism corresponding to the formation of HOMs trained from the PRAM mechanism. Refer to Fig. 3.6 for a more detailed description of the legend.

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 Sect. 2.4.2). The reduction error (calculation detailed in Sect. 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 3.7. 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.

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

Table 3.7: 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).

The monthly distribution of the testing errors is presented in Fig. 3.8. 86 % of testing conditions (8 509 conditions) are between 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. between 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 between 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 %.

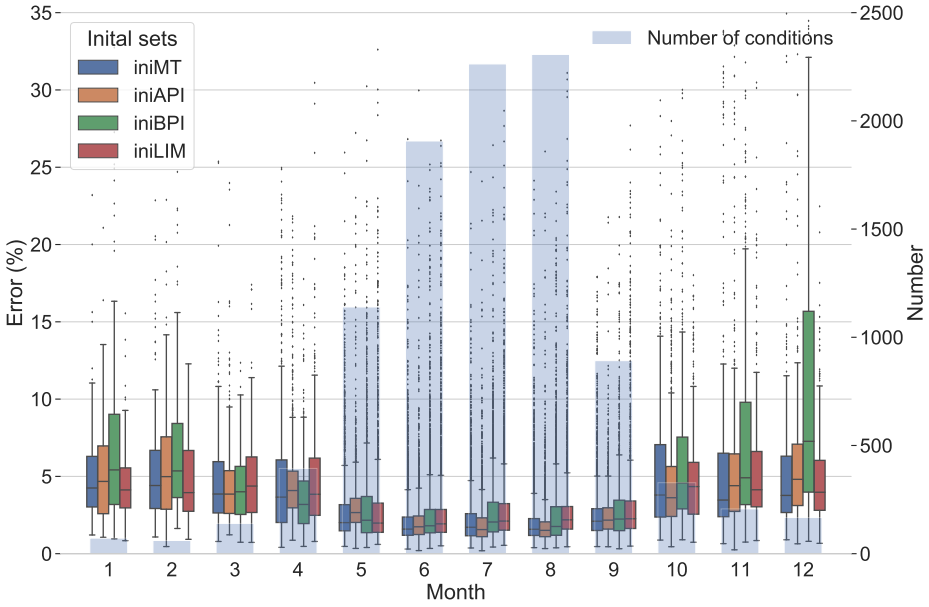


Figure 3.8: Monthly distribution of errors over testing conditions (box plot) generated by the MT-rdc mechanism (compared to the reference mechanism) simulated with different initial SOA precursor conditions. The bars represent the number of testing conditions adopted for testing.

The map distribution of errors induced by the reduction and the corresponding SOA yields simulated with the MT-rdc mechanism under testing conditions between May to September are presented in Fig. 3.9. 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. The highest SOA yields are simulated in Northern and Eastern Europe around the Baltic Sea, followed by Central Europe around the Mediterranean Sea. between 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. 3.8 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.

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. 3.10, during the training stage III (detailed in Sect. 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. 3.10) 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 3.8 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, 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.

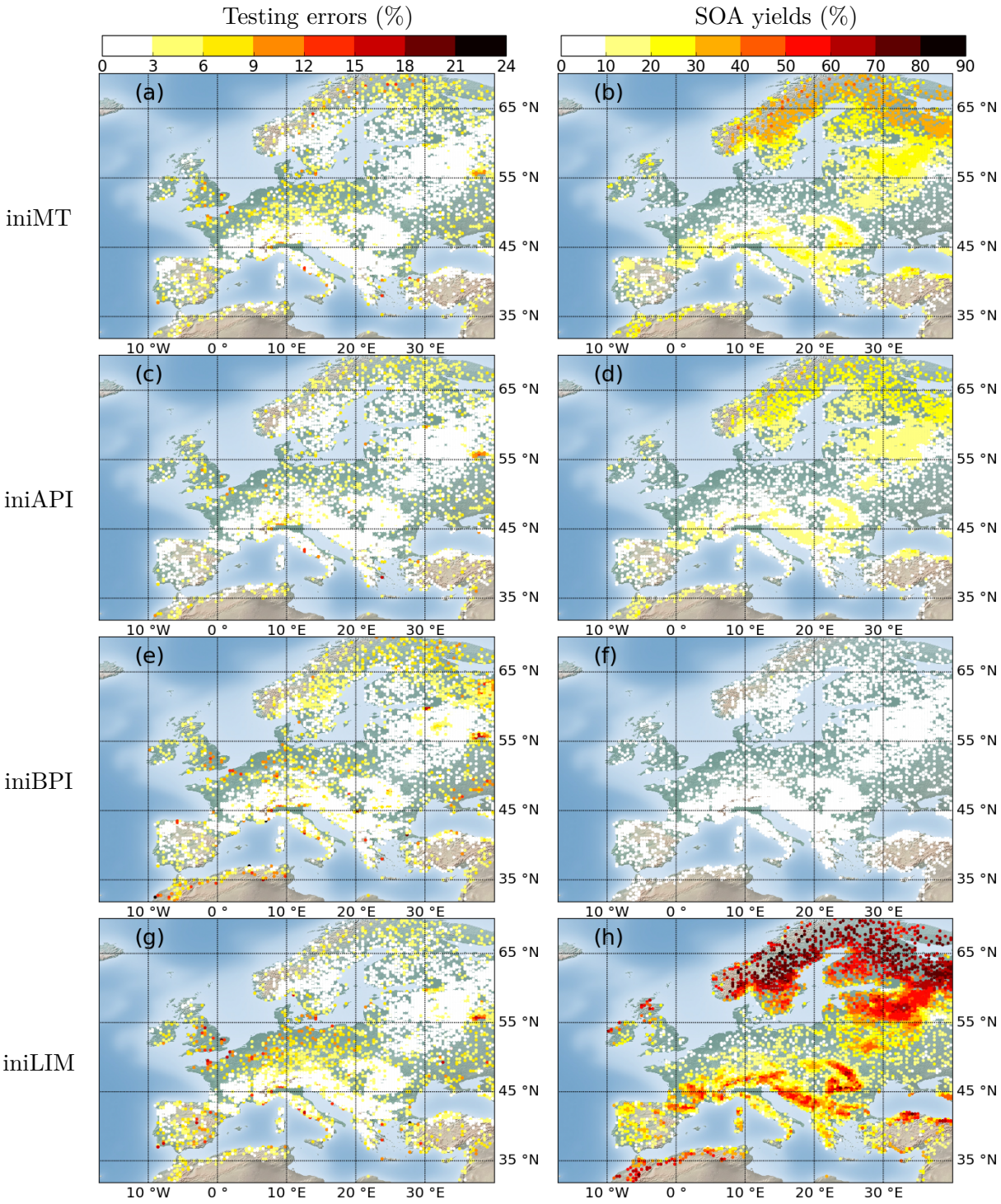


Figure 3.9: Map distributions of MT-rdc testing errors (left panels) and SOA yields (right panels) simulated with the different initial precursor conditions under testing conditions between May to September (8 509 conditions, corresponding 86 % of total testing conditions). The maps of all testing conditions are in Fig. S3.

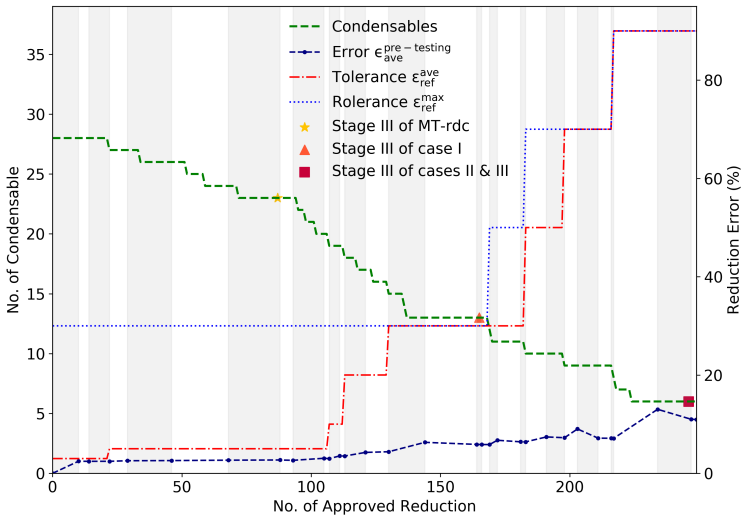


Figure 3.10: Evolution of the size (measured as the number of condensables) and accuracy (measured as $\epsilon_{ave}^{pre-testing}$) of reduced mechanisms during training stage III with increasing error tolerances. The final reduced mechanisms after stage IV reduction, trained from three stage III mechanisms marked in the figure, can be found in Table 3.8.

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

Table 3.8: Size and accuracy of the reduced mechanisms trained with different error tolerances.

[a] Mechanisms are trained from stage III mechanisms noted in Fig. 3.10. 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 99^{th} percentile error, respectively, when compared to the Ref. mechanism under the testing dataset.

3.5 Mechanism sensitivity to environmental parameters

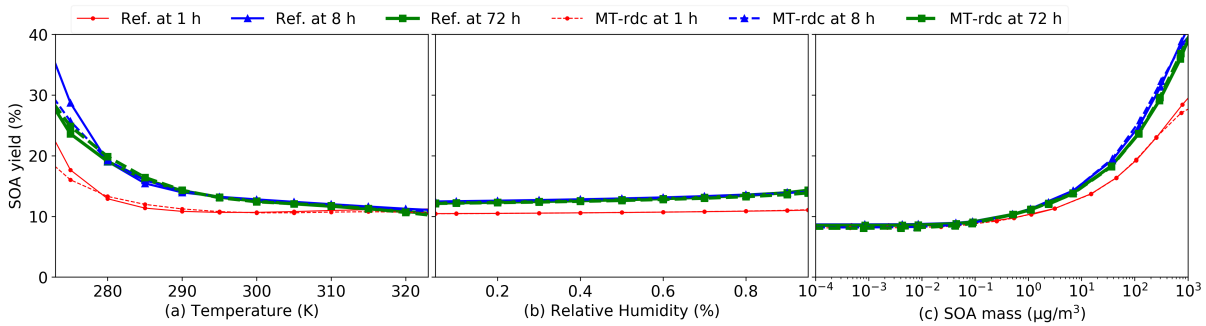


Figure 3.11: Dependence of average SOA yields simulated under the pre-testing dataset with the reference mechanism (Ref., solid line) and the final reduced SOA mechanism (MT-rdc, dotted line) on (a) temperature, (b) relative humidity, and (c) SOA mass at 1 h (red point), 8 h (blue triangle) and 72 h (green square).

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. 3.11, 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.

Code and data availability

The source code for GENOA v2.0 is hosted on GitHub at https://github.com/tool-genoa/GENOA-SSH-aerosol/tree/v2.0 (last access: 26 July 2023). The associated Zenodo DOI is https://doi.org/10.5281/zenodo.8187339. The data for the monoterpene MCM reductions mentioned in this paper can be accessed publicly on Zenodo:
https://10.5281/zenodo.8187593.

Supplemental materials for

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

S1 Information related to the training dataset

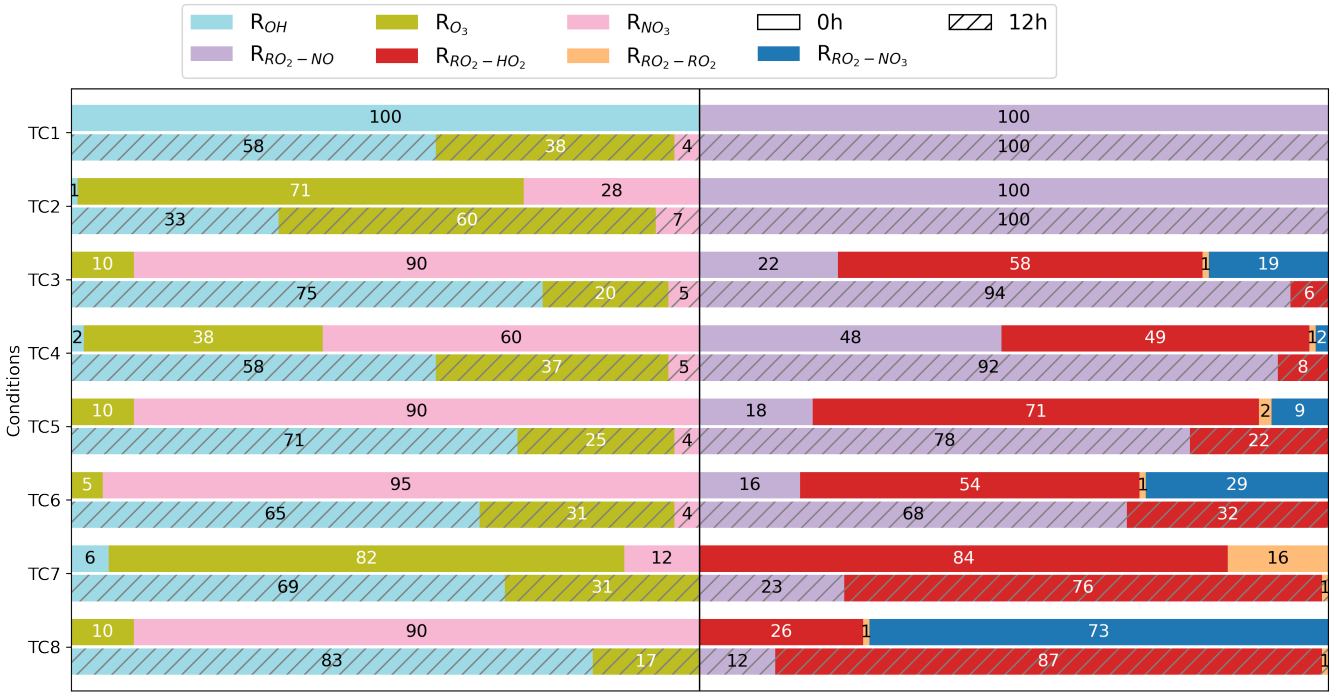


Figure S1: Bar plot showing the seven reaction ratios under the eight training conditions used for monoterpene mechanism reduction at midnight (0 h, top bar) and noon (12 h, bottom bar with slash). From left to right, six ratios are presented on each bar in the following order: R_{O_3}, R_{OH}, R_{NO_3}, R_{RO_2-NO}, $R_{RO_2-HO_2}$, $R_{RO_2-NO_3}$, and $R_{RO_2-RO_2}$ (No display if the ratio is zero). The computation of ratios is explained in Table 1.

Table S1: Computation of reaction ratios used in selecting training dataset.

Ratio [a]	Equation [b]	Kinetic rate coefficient [c]
R_{OH}	k_{OH}[OH] / f1	$k_{OH} = 1.97 \times 10^{-10}$
R_{O_3}	k_{O_3}[O_3] / f1	$k_{O_3} = 1.20 \times 10^{-14}$
R_{NO_3}	k_{NO_3} [NO_3] / f1	$k_{NO_3} = 1.90 \times 10^{-11}$
	f1 = k_{OH}[OH] + k_{O_3}[O_3] + k_{NO_3} [NO_3]	
R_{RO_2-NO}	k_{NO} [NO] / f2	$k_{NO} = 2.70 \times 10^{-12} \times \exp(\frac{360}{T})$
$R_{RO_2-HO_2}$	k_{HO_2} [HO_2] / f2	$k_{HO_2} = 2.91 \times 10^{-13} \times \exp(\frac{1300}{T})$
$R_{RO_2-RO_2}$	k_{RO_2} [RO_2] / f2	$k_{RO_2} = 9.20 \times 10^{-14}$
$R_{RO_2-NO_3}$	k_{NO_3} [NO_3] / f2	$k_{RNO_3} = 2.30 \times 10^{-12}$
	f2 = k_{NO} [NO] + k_{HO_2} [HO_2] + k_{RNO_3} [NO_3] + k_{RO_2} [RO_2]	

[a] names of the reaction ratio of OH radical, O_3, and NO_3 radical reacted withPI. (R_{OH} + R_{O_3} + R_{NO_3} = 1); and of the reaction ratio of NO, HO_2 radical, RO_2 radical, and NO_3 radical (at the presence of RO_2) reacted with RO_2 species (R_{RO_2-NO} + $R_{RO_2-HO_2}$ + $R_{RO_2-RO_2}$ + $R_{RO_2-NO_3}$ = 1).

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

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

S2 Information related to the MT-rdc mechanism: MT-rdc

S2.1 An overview of the MT-rdc mechanism

Table S2: Reaction list of the MT-rdc mechanism.

1	APINENE + NO_3 → 0.65 NAPINAO2 + 0.35 NAPINBO2	$1.2\times10^{-12} \times \exp(\frac{490}{T})$
2	APINENE + O_3 → 0.297 C107O2 + 0.243 C109O2 + 0.2 mC96O2 + 0.2 APINBOO + 0.06 C10H15O2O2	$8.1\times10^{-16} \times \exp(\frac{-640}{T})$
3	APINENE + OH → 0.902 APINBO2 + 0.073 APINCO2 + 0.025 C10H17O4O2	$1.2\times10^{-11} \times \exp(\frac{440}{T})$
4	BPINENE + NO_3 → 0.8 NBPINAO2 + 0.2 NBPINBO2	2.5×10^{-12}
5	BPINENE + O_3 → 0.502 NOPINONE + 0.3 NOPINDO2	$1.4\times10^{-15} \times \exp(\frac{-1270}{T})$
6	BPINENE + OH → 0.841 BPINAO2 + 0.075 BPINBO2 + 0.074 APINCO2 + 0.01 C10H17O4O2	$2.4\times10^{-11} \times \exp(\frac{357}{T})$
7	LIMONENE + NO_3 → NLIMO2	1.2×10^{-11}
8	LIMONENE + O_3 → 0.256 mLIMAO2 + 0.256 LIMALBO2 + 0.219 C10H15O2O2 + 0.135 mLIMAL	$2.8\times10^{-15} \times \exp(\frac{-770}{T})$
9	LIMONENE + OH → 0.586 mLIMCO2 + 0.404 mLIMAO2 + 0.01 C10H17O4O2	$4.3\times10^{-11} \times \exp(\frac{401}{T})$
10	NAPINAO2 + HO_2 → mNAPINAOOH	$2.7\times10^{-13} \times \exp(\frac{1300}{T})$
11	NAPINAO2 + NO → mAPINANO3	$2.7\times10^{-12} \times \exp(\frac{360}{T})$
12	NAPINAO2 → mAPINANO3	[RO_2] × 6.7×10^{-15}
13	NAPINBO2 + HO_2 → NAPINBOOH	$2.7\times10^{-13} \times \exp(\frac{1300}{T})$
14	NAPINBO2 + NO →	$2.7\times10^{-12} \times \exp(\frac{360}{T})$
15	NAPINBO2 → 0.111 mAPINANO3	[RO_2] × 2.2×10^{-13}
16	APINBO2 + HO_2 → mAPINAOOH	$2.3\times10^{-13} \times \exp(\frac{1300}{T})$
17	APINBO2 + NO → 0.877 mAPINANO3 + 0.104 mC926O2	$3.1\times10^{-12} \times \exp(\frac{360}{T})$
18	APINBO2 → mAPINANO3	[RO_2] × 5.3×10^{-13}
19	APINCO2 + HO_2 → mAPINCOOH	$5.3\times10^{-13} \times \exp(\frac{1300}{T})$

20	APINCO2 + NO → C720O2	$4.7\times10^{-12} \times \exp(\frac{360}{T})$
21	NBPINAO2 + HO_2 → mNBPINAOOH	$2.7\times10^{-13} \times \exp(\frac{1300}{T})$
22	NBPINAO2 + NO → NOPINONE	$2.7\times10^{-12} \times \exp(\frac{360}{T})$
23	NBPINAO2 + NO_3 → NOPINONE	2.3×10^{-12}
24	NBPINAO2 → 0.7 NOPINONE + 0.3 mBPINANO3	$[RO_2] \times 9.2\times10^{-14}$
25	NBPINBO2 + HO_2 → mNBPINAOOH	$2.7\times10^{-13} \times \exp(\frac{1300}{T})$
26	NBPINBO2 + NO → NOPINONE	$2.7\times10^{-12} \times \exp(\frac{360}{T})$
27	NBPINBO2 → mBPINANO3	$[RO_2] \times 8.0\times10^{-13}$
28	NOPINONE + OH → 0.367 mNOPINCOOH2 + 0.136 NOPINDO + 0.13 mC918O2	1.6×10^{-11}
29	BPINAO2 + HO_2 → mBPINAOOH	$2.7\times10^{-13} \times \exp(\frac{1300}{T})$
30	BPINAO2 + NO → 0.76 NOPINONE + 0.24 mBPINANO3	$2.7\times10^{-12} \times \exp(\frac{360}{T})$
31	BPINAO2 → NOPINONE	$[RO_2] \times 6.4\times10^{-14}$
32	BPINBO2 + HO_2 → mBPINAOOH	$2.7\times10^{-13} \times \exp(\frac{1300}{T})$
33	NLIMO2 + HO_2 → NLIMOOH	$2.7\times10^{-13} \times \exp(\frac{1300}{T})$
34	NLIMO2 + NO → mLIMAL	$2.7\times10^{-12} \times \exp(\frac{360}{T})$
35	NLIMO2 → mLIMAL	$[RO_2] \times 6.4\times10^{-14}$
36	mLIMAO2 + HO_2 → mLIMAOOH	$2.1\times10^{-13} \times \exp(\frac{1300}{T})$
37	mLIMAO2 + NO → 0.771 mLIMAL + 0.229 mC729CHO	$2.2\times10^{-12} \times \exp(\frac{360}{T})$
38	mLIMAO2 → mLIMAL	$[RO_2] \times 7.3\times10^{-14}$
39	mLIMCO2 + NO → 0.766 LIMKET + 0.234 mLIMAL	$2.1\times10^{-12} \times \exp(\frac{360}{T})$
40	mLIMCO2 + HO_2 → LIMCOOH	$2.0\times10^{-13} \times \exp(\frac{1300}{T})$
41	mLIMCO2 → LIMKET	$[RO_2] \times 4.9\times10^{-14}$
42	mNAPINAOOH + OH → NAPINAO2	6.3×10^{-12}
43	NAPINBOOH + OH → NC101CO	1.2×10^{-11}
44	NC101CO + OH → NC102O2	5.5×10^{-12}
45	NC101CO → C96CO3	J(1.1×10^{-5}, 0.974, 0.309)
46	C107O2 + HO_2 →	$2.7\times10^{-13} \times \exp(\frac{1300}{T})$
47	C107O2 + NO → mC106O2	$2.7\times10^{-12} \times \exp(\frac{360}{T})$
48	C107O2 + NO_3 → mC106O2	2.3×10^{-12}
49	C107O2 → 0.863 mC106O2 + 0.137 C10H15O7O2	$[RO_2] \times 6.7\times10^{-14}$
50	C109O2 + HO_2 → C109OOH	$2.7\times10^{-13} \times \exp(\frac{1300}{T})$
51	C109O2 + NO → 0.8 C89CO3 + 0.2 C920CO3H	$2.7\times10^{-12} \times \exp(\frac{360}{T})$
52	C109O2 → 0.8 C89CO3 + 0.2 C920CO3H	$[RO_2] \times 1.8\times10^{-12}$
53	APINBOO → 0.875 mAPINANO3	$[H_2O] \times 1.6\times10^{-17}$
54	mC96O2 + HO_2 → 0.903 mC96OOH + 0.097 mC98OOH	$2.6\times10^{-13} \times \exp(\frac{1300}{T})$
55	mC96O2 + NO → 0.843 C97O2	$2.4\times10^{-12} \times \exp(\frac{360}{T})$
56	mC96O2 → C97O2	$[RO_2] \times 7.0\times10^{-13}$
57	mAPINAOOH + OH → APINBO2	1.7×10^{-11}
58	mAPINCOOH + OH → APINCO2	6.6×10^{-11}
59	mNBPINAOOH + OH → NBPINAO2	9.0×10^{-12}
60	mNBPINAOOH → NOPINONE	J(7.2×10^{-6}, 0.682, 0.279)
61	mBPINANO3 + OH → NOPINONE	4.1×10^{-12}
62	mC918O2 + HO_2 → 0.7 mNOPINCOOH + 0.3 mC106O2	$1.8\times10^{-13} \times \exp(\frac{1300}{T})$
63	NOPINDO2 + HO_2 → NOPINDOOH	$2.6\times10^{-13} \times \exp(\frac{1300}{T})$
64	NOPINDO2 + NO → C89CO3	$2.7\times10^{-12} \times \exp(\frac{360}{T})$
65	NOPINDO2 → C89CO3	$[RO_2] \times 1.8\times10^{-12}$

66	mBPINAOOH + OH → BPINAO2	1.3×10^{-11}
67	NLIMOOH + OH → NLIMO2	4.3×10^{-11}
68	LIMALBO2 + NO → C822CO3H	$2.7\times10^{-12} \times \exp(\frac{360}{T})$
69	mLIMAOOH + OH → mLIMAO2	6.8×10^{-11}
70	LIMCOOH + OH → mLIMCO2	1.0×10^{-10}
71	mAPINANO3 + OH → C96CO3	$5.2\times10^{-12} \times \exp(\frac{600}{T})$
72	mAPINANO3 → mC96O2	J(2.8×10^{-5}, 0.805, 0.338)
73	C96CO3 + NO → mC96O2	$7.5\times10^{-12} \times \exp(\frac{290}{T})$
74	C96CO3 + NO_2 → C10PAN2	KFPAN
75	C109OOH + OH → mC109CO	5.5×10^{-11}
76	mC109CO → C89CO3	J(1.5×10^{-4}, 0.170, 0.208)
77	C622CHO + OH → 0.407 mC728O2 + 0.198 mC109CO	1.6×10^{-10}
78	mC96OOH → C97O2	J(5.5×10^{-6}, 0.682, 0.279)
79	HCC7CO + OH → C719O2	1.2×10^{-10}
80	C720O2 + NO → 0.722 HCC7CO + 0.278 C720NO3	$2.7\times10^{-12} \times \exp(\frac{360}{T})$
81	mNOPINCOOH + OH → mC918O2	7.5×10^{-12}
82	NOPINDOOH + OH → NOPINDCO	2.6×10^{-11}
83	NOPINDCO + OH → C89CO3	3.1×10^{-12}
84	mLIMAL + NO_3 → mNLIMALO2	2.5×10^{-13}
85	mLIMAL + O_3 → 0.67 mC926O2 + 0.33 C817CO3H	8.2×10^{-18}
86	mLIMAL + OH → 0.712 LIMALO2 + 0.288 C923CO3	1.1×10^{-10}
87	mC729CHO + OH → 0.798 mC728O2 + 0.202 C729CO2H	5.0×10^{-11}
88	LIMKET + O_3 → 0.844 LMLKAO2 + 0.156 C817O2	1.3×10^{-16}
89	C96CO3 + NO → 0.95 mC106O2	$2.7\times10^{-12} \times \exp(\frac{360}{T})$
90	C10PAN2 → C96CO3	KBPAN
91	mC106O2 + HO_2 → mC106OOH	$2.7\times10^{-13} \times \exp(\frac{1300}{T})$
92	mC106O2 + NO →	$4.5\times10^{-12} \times \exp(\frac{360}{T})$
93	C89CO3 + NO → 0.8 C811CO3	$7.5\times10^{-12} \times \exp(\frac{290}{T})$
94	C89CO3 + NO_2 → C89PAN	KFPAN
95	C89CO3 → 0.651 C811CO3 + 0.349 C89CO2H	$[RO_2] \times 8.6\times10^{-12}$
96	C920CO3H + NO → C920O2	$7.5\times10^{-12} \times \exp(\frac{290}{T})$
97	C920CO3H + NO_2 → C920PAN	KFPAN
98	C920O2 → C921O2	$[RO_2] \times 1.3\times10^{-12}$
99	C97O2 + HO_2 → mC97OOH	$2.6\times10^{-13} \times \exp(\frac{1300}{T})$
100	C97O2 + NO → C98O2	$2.7\times10^{-12} \times \exp(\frac{360}{T})$
101	C719O2 + HO_2 → mC719OOH	$2.4\times10^{-13} \times \exp(\frac{1300}{T})$
102	C719O2 + NO → 0.958 H3C25C6CO3	$2.7\times10^{-12} \times \exp(\frac{360}{T})$
103	C720NO3 + OH → HCC7CO	9.6×10^{-11}
104	mNLIMALO2 + HO_2 → mNC102OOH	$2.5\times10^{-13} \times \exp(\frac{1300}{T})$
105	mNLIMALO2 + NO → C817CO3H	$2.6\times10^{-12} \times \exp(\frac{360}{T})$
106	mNLIMALO2 + NO_3 → C817CO3H	2.2×10^{-12}
107	mNLIMALO2 → 0.7 C817CO3H	$[RO_2] \times 8.8\times10^{-14}$
108	LIMALO2 + HO_2 → LIMALOOH	$2.7\times10^{-13} \times \exp(\frac{1300}{T})$
109	LIMALO2 + NO → 0.941 C817CO3H + 0.059 LIMALNO3	$2.7\times10^{-12} \times \exp(\frac{360}{T})$
110	mC729CHO + NO_3 → mNLIMALO2	7.1×10^{-14}
111	C822CO3H + NO → 0.8 C823CO3 + 0.2 mLIMAO2	$7.5\times10^{-12} \times \exp(\frac{290}{T})$
112	C822CO3H + NO_2 → mC822PAN	KFPAN

113	mC926O2 + NO → 0.57 C622CHO	$2.1\times10^{-12} \times \exp(\frac{360}{T})$
114	mC728O2 + HO_2 → 0.517 mNORLIMOOH + 0.483 mC719OOH	$1.4\times10^{-12} \times \exp(\frac{1300}{T})$
115	mC728O2 + NO →	$1.2\times10^{-11} \times \exp(\frac{360}{T})$
116	C923CO3 + OH → mC729CHO	6.9×10^{-11}
117	NC102O2 + HO_2 → mNC102OOH	$2.7\times10^{-13} \times \exp(\frac{1300}{T})$
118	NC102O2 + NO →	$2.7\times10^{-12} \times \exp(\frac{360}{T})$
119	C89CO2H + OH → C811CO3	2.2×10^{-11}
120	C89PAN + OH →	2.5×10^{-11}
121	C89PAN → C89CO3	KBPAN
122	C920PAN + OH → C622CHO	5.6×10^{-12}
123	C920PAN → C920CO3H	KBPAN
124	mC97OOH + OH → C97O2	9.8×10^{-12}
125	mC719OOH + OH → C719O2	4.8×10^{-11}
126	mNC102OOH + OH → mNLIMALO2	2.0×10^{-11}
127	mC926O2 + HO_2 → mC98OOH	$1.4\times10^{-13} \times \exp(\frac{1300}{T})$
128	C817CO3H + NO → C817O2	$7.5\times10^{-12} \times \exp(\frac{290}{T})$
129	C817CO3H + NO_2 →	KFPAN
130	C817O2 + NO → 0.862 mC96O2	$2.7\times10^{-12} \times \exp(\frac{360}{T})$
131	LIMALOOH + OH → LIMALO2	4.7×10^{-11}
132	LIMALNO3 + OH → C817CO3H	3.0×10^{-11}
133	C729CO2H + OH → C622CHO	8.3×10^{-11}
134	mC822PAN + OH → mC729CHO	6.4×10^{-11}
135	mC822PAN → C822CO3H	7.6× KBPAN$\times10^{-1}$
136	mNORLIMOOH + OH → mC728O2	2.2×10^{-11}
137	LMLKAO2 + HO_2 → mC97OOH	$2.7\times10^{-13} \times \exp(\frac{1300}{T})$
138	LMLKAO2 + NO → 0.5 C731CO3	$5.4\times10^{-12} \times \exp(\frac{360}{T})$
139	C811CO3 + HO_2 →	$2.1\times10^{-13} \times \exp(\frac{980}{T})$
140	C811CO3 + NO → C811O2	$7.5\times10^{-12} \times \exp(\frac{290}{T})$
141	C811CO3 + NO_2 → C811PAN	KFPAN
142	C811CO3 → 0.7 C811O2 + 0.3 mPINIC	$[RO_2] \times 1.0\times10^{-11}$
143	C921O2 + HO_2 → mC812OOH	$2.6\times10^{-13} \times \exp(\frac{1300}{T})$
144	C98O2 + HO_2 → mC98OOH	$2.6\times10^{-13} \times \exp(\frac{1300}{T})$
145	C98O2 + NO → 0.118 C98NO3	$2.7\times10^{-12} \times \exp(\frac{360}{T})$
146	C823CO3 → mC823O2	$[RO_2] \times 7.0\times10^{-12}$
147	mC106OOH + OH → mC106O2	5.6×10^{-11}
148	C811O2 + HO_2 → C811OOH	$2.5\times10^{-13} \times \exp(\frac{1300}{T})$
149	C811O2 + NO → 0.862 C812O2	$2.7\times10^{-12} \times \exp(\frac{360}{T})$
150	C811O2 → C812O2	$[RO_2] \times 7.8\times10^{-13}$
151	mPINIC + OH → C811O2	5.6×10^{-12}
152	C811PAN + OH → mC729CHO	6.8×10^{-12}
153	C811PAN → C811CO3	KBPAN
154	mC812OOH + OH → 0.637 C812O2 + 0.363 C921O2	1.1×10^{-11}
155	mC98OOH + OH → C98O2	1.8×10^{-11}
156	mC98OOH →	J(2.9×10^{-4}, 0.148, 0.215)
157	C98NO3 →	J(3.3×10^{-4}, 0.148, 0.215)
158	H3C25C6CO3 + NO →	$7.5\times10^{-12} \times \exp(\frac{290}{T})$
159	H3C25C6CO3 + NO_2 → H3C25C6PAN	KFPAN

160	C731CO3 + NO_2 → mC822PAN	KFPAN
161	C811OOH + OH → mC729CHO	1.7×10^{-11}
162	C811OOH → C812O2	J(7.6×10^{-6}, 0.682, 0.279)
163	C921O2 + NO → mC823O2	$1.9\times10^{-12} \times \exp(\frac{360}{T})$
164	H3C25C6PAN + OH →	2.3×10^{-11}
165	C812O2 + HO_2 → mC812OOH	$2.5\times10^{-13} \times \exp(\frac{1300}{T})$
166	C812O2 + NO → C813O2	$2.7\times10^{-12} \times \exp(\frac{360}{T})$
167	mC823O2 + HO_2 → mC621OOH	$9.0\times10^{-14} \times \exp(\frac{1300}{T})$
168	mC823O2 + NO →	$1.1\times10^{-12} \times \exp(\frac{360}{T})$
169	mNORLIMOOH →	J(4.5×10^{-5}, 0.148, 0.215)
170	C813O2 + HO_2 → C813OOH	$2.5\times10^{-13} \times \exp(\frac{1300}{T})$
171	C813O2 + NO → 0.896 C516O2 + 0.104 C813NO3	$2.7\times10^{-12} \times \exp(\frac{360}{T})$
172	C813OOH + OH → C813O2	1.9×10^{-11}
173	C813OOH → C516O2	J(1.5×10^{-4}, 0.170, 0.208)
174	C813NO3 → C516O2	J(1.5×10^{-4}, 0.170, 0.208)
175	C516O2 + HO_2 → C516OOH	$2.1\times10^{-13} \times \exp(\frac{1300}{T})$
176	C516O2 + NO →	$2.7\times10^{-12} \times \exp(\frac{360}{T})$
177	C516OOH + OH → C516O2	3.4×10^{-11}
178	C10H15O2O2 → C10H15O6O2	$1.2\times10^{17} \times \exp(\frac{-12077}{T})$
179	C10H15O6O2 → C10H15O10O2	$2.0\times10^{17} \times \exp(\frac{-12077}{T})$
180	C10H15O7O2 → C10H15O9O2	$1.2\times10^{17} \times \exp(\frac{-12077}{T})$
181	C10H15O2O2 + NO → C10H15O7O2	$2.7\times10^{-12} \times \exp(\frac{360}{T})$
182	C10H15O7O2 + NO → 0.87 C10H15O10O2 + 0.13 mC10H14O9	$2.5\times10^{-12} \times \exp(\frac{360}{T})$
183	C10H15O9O2 + NO → 0.6 mC10H14O9	$1.4\times10^{-12} \times \exp(\frac{360}{T})$
184	C10H15O10O2 + NO → 0.6 mC10H14O11	$1.9\times10^{-12} \times \exp(\frac{360}{T})$
185	C10H15O10O2 + HO_2 → mC10H14O11	$2.9\times10^{-13} \times \exp(\frac{1300}{T})$
186	C10H15O2O2 → 0.6 C10H15O7O2	[RO_2] × 1.0×10^{-12}
187	C10H15O6O2 → 0.467 mC10H14O9 + 0.311 C10H15O7O2 + 0.222 C20H30O13	[RO_2] × 9.0×10^{-12}
188	C10H15O7O2 → 0.556 C20H30O13 + 0.444 mC10H14O9	[RO_2] × 1.1×10^{-11}
189	C10H15O9O2 → 0.714 C20H30O13 + 0.286 mC10H14O9	[RO_2] × 2.8×10^{-11}
190	C10H15O10O2 → 0.553 mC10H14O11 + 0.447 mC10H14O9	[RO_2] × 1.4×10^{-11}
191	C10H17O6O2 → C10H17O7O2	$6.0\times10^{16} \times \exp(\frac{-12077}{T})$
192	C10H17O4O2 + NO → 0.957 C10H17O6O2	$2.5\times10^{-12} \times \exp(\frac{360}{T})$
193	C10H17O6O2 + NO → 0.646 C10H17O7O2 + 0.166 mC10H14O9	$2.9\times10^{-12} \times \exp(\frac{360}{T})$
194	C10H17O7O2 + NO → 0.6 mC10H14O9	$1.4\times10^{-12} \times \exp(\frac{360}{T})$
195	C10H17O4O2 + HO_2 → mC10H14O9	$2.9\times10^{-13} \times \exp(\frac{1300}{T})$
196	C10H17O7O2 + HO_2 → mC10H14O9	$2.9\times10^{-13} \times \exp(\frac{1300}{T})$
197	C10H17O7O2 → mC10H14O9	[RO_2] × 8.0×10^{-12}

[a] [H_2O] is the concentration of H_2O, [RO_2] is the concentration of RO_2 pool, and [O_2] is the concentration of O_2. Kinetic rate constants, i.e., KFPAN and KBPAN, are the complex rate coefficients from the MCM mechanism v3.3.1. Photolysis rates are in the format J(l,m,n) = l×cosX^m×exp(-n×secX) with solar zenith angle X in radians.

Table S3: Species list of the MT-rdc mechanism.

Surrogate[a]	Type[b]	Molecular formula	MW[c]	P^{d}_{sat}	ΔH^{e}_{vap}	H[f]	γ^{g}	C*
HCC7CO	VOC	$C_7H_{10}O_2$	126.2					
C622CHO	VOC	$C_7H_{12}O_2$	128.2					
APINENE	VOC	$C_{10}H_{16}$	136.2					
BPINENE	VOC	$C_{10}H_{16}$	136.2					
LIMONENE	VOC	$C_{10}H_{16}$	136.2					
LIMKET	VOC	$C_9H_{14}O$	138.2					
NOPINONE	VOC	$C_9H_{14}O$	138.2					
NOPINDCO	VOC	$C_9H_{12}O_2$	152.2					
mC729CHO	VOC	$C_{8.24}H_{12.52}O_{2.62}$	153.5					
C729CO2H	VOC	$C_8H_{12}O_3$	156.2					
mLIMAL	VOC	$C_{10}H_{16.00}O_2$	168.3					
C89CO2H	VOC	$C_9H_{14}O_3$	170.2					
mNOPINCOOH	VOC	$C_9H_{14}O_3$	170.2					
NOPINDOOH	VOC	$C_9H_{14}O_3$	170.2					
C720NO3	VOC	$C_7H_{11}NO_4$	173.2					
C811OOH	VOC	$C_8H_{14}O_4$	174.2					
mC96OOH	VOC	$C_{9.28}H_{16}O_{3.28}$	180.2					
mC109CO	VOC	$C_{9.95}H_{13.95}O_3$	181.6					
C822CO3H	VOC	$C_9H_{14}O_4$	186.2					
mAPINAOOH	VOC	$C_{10.00}H_{17.98}O_3$	186.2					
LIMCOOH	VOC	$C_{10}H_{18}O_3$	186.2					
mAPINCOOH	VOC	$C_{10}H_{18}O_3$	186.2					
mLIMAOOH	VOC	$C_{10.00}H_{17.89}O_{3.06}$	187					
C109OOH	VOC	$C_{10}H_{16}O_4$	200.2					
C817CO3H	VOC	$C_9H_{14}O_5$	202.2					
NC101CO	VOC	$C_{10}H_{15}NO_4$	213.2					
mAPINANO3	VOC	$C_{9.98}H_{16.91}NO_4$	214.9					
mBPINANO3	VOC	$C_{10}H_{16.96}NO_4$	215.2					
C920CO3H	VOC	$C_{10}H_{16}O_5$	216.2					
C89PAN	VOC	$C_9H_{13}NO_6$	231.2					
NAPINBOOH	VOC	$C_{10}H_{17}NO_5$	231.2					
NLIMOOH	VOC	$C_{10}H_{17}NO_5$	231.2					
mC822PAN	VOC	$C_{8.76}H_{12.51}NO_{6.24}$	231.7					
C10PAN2	VOC	$C_{10}H_{15}NO_6$	245.2					
C811PAN	VOC	$C_9H_{13}NO_7$	247.2					
C720O2	Radical	$C_7H_{11}O_3$	143.2					
C516O2	Radical	$C_5H_7O_6$	163.1					
NOPINDO2	Radical	$C_9H_{13}O_3$	169.2					
mC96O2	Radical	$C_{8.90}H_{14.80}O_{3.20}$	173					
C811O2	Radical	$C_8H_{13}O_4$	173.2					
C817O2	Radical	$C_8H_{13}O_4$	173.2					
mC823O2	Radical	$C_{7.20}H_{11.39}O_{4.80}$	174.8					
C719O2	Radical	$C_7H_{11}O_5$	175.2					
mC728O2	Radical	$C_{6.88}H_{12.77}O_5$	175.5					
mC918O2	Radical	$C_9H_{13}O_{3.50}$	177.2					
APINBOO	Radical	$C_{10}H_{16}O_3$	184.2					
C89CO3	Radical	$C_9H_{13}O_4$	185.2					
APINBO2	Radical	$C_{10}H_{17}O_3$	185.2					
BPINAO2	Radical	$C_{10}H_{17}O_3$	185.2					

BPINBO2	Radical	$C_{10}H_{17}O_3$	185.2					
APINCO2	Radical	$C_{10}H_{17}O_3$	185.2					
mLIMCO2	Radical	$C_{10}H_{17}O_3$	185.2					
C731CO3	Radical	$C_8H_{11}O_5$	187.2					
C920O2	Radical	$C_9H_{15}O_4$	187.2					
C97O2	Radical	$C_9H_{15}O_4$	187.2					
mLIMAO2	Radical	$C_{9.96}H_{16.56}O_{3.18}$	187.3					
H3C25C6CO3	Radical	$C_7H_9O_6$	189.1					
C812O2	Radical	$C_8H_{13}O_5$	189.2					
mC926O2	Radical	$C_9H_{13.89}O_{4.56}$	195					
C10H15O2O2	Radical	$C_{10}H_{15}O_4$	199					
C107O2	Radical	$C_{10}H_{15}O_4$	199.2					
C96CO3	Radical	$C_{10}H_{15}O_4$	199.2					
C109O2	Radical	$C_{10}H_{15}O_4$	199.2					
C923CO3	Radical	$C_{10}H_{15}O_4$	199.2					
LIMALBO2	Radical	$C_{10}H_{15}O_4$	199.2					
C811CO3	Radical	$C_9H_{13}O_5$	201.2					
C823CO3	Radical	$C_9H_{13}O_5$	201.2					
LMLKAO2	Radical	$C_9H_{13}O_5$	201.2					
C921O2	Radical	$C_9H_{15}O_5$	203.2					
C98O2	Radical	$C_9H_{15}O_5$	203.2					
C813O2	Radical	$C_8H_{13}O_6$	205.2					
mC106O2	Radical	$C_{9.62}H_{13.94}O_{4.80}$	206.4					
LIMALO2	Radical	$C_{10}H_{17}O_5$	217.2					
NAPINAO2	Radical	$C_{10}H_{16}NO_5$	230.2					
NAPINBO2	Radical	$C_{10}H_{16}NO_5$	230.2					
NBPINAO2	Radical	$C_{10}H_{16}NO_5$	230.2					
NBPINBO2	Radical	$C_{10}H_{16}NO_5$	230.2					
NLIMO2	Radical	$C_{10}H_{16}NO_5$	230.2					
C10H17O4O2	Radical	$C_{10}H_{17}O_6$	233					
NC102O2	Radical	$C_{10}H_{14}NO_7$	260.2					
mNLIMALO2	Radical	$C_{9.92}H_{15.83}NO_7$	261.1					
C10H15O6O2	Radical	$C_{10}H_{15}O_8$	263					
C10H17O6O2	Radical	$C_{10}H_{17}O_8$	265					
C10H15O7O2	Radical	$C_{10}H_{15}O_9$	279					
C10H17O7O2	Radical	$C_{10}H_{17}O_9$	281					
C10H15O9O2	Radical	$C_{10}H_{15}O_{11}$	311					
C10H15O10O2	Radical	$C_{10}H_{15}O_{12}$	327					
mNAPINAOOH	SVOC	$C_{10}H_{16.83}NO_{5.08}$	232.4	6.06×10^{-8}	96.1	4.96×10^{4}	1.85×10^{4}	2.46×10^{7}
mBPINAOOH	SVOC	$C_{10}H_{18}O_3$	186.2	4.61×10^{-8}	97.4	4.68×10^{3}	2.57×10^{5}	1.77×10^{6}
mNBPINAOOH	SVOC	$C_{10}H_{17}NO_5$	231.2	3.61×10^{-8}	98.3	3.86×10^{4}	4.00×10^{4}	1.14×10^{7}
C920PAN	SVOC	$C_{10}H_{15}NO_7$	261.2	2.56×10^{-8}	99.3	1.08×10^{3}	2.00×10^{6}	2.27×10^{5}
mC97OOH	SVOC	$C_9H_{15.99}O_4$	188.3	2.28×10^{-8}	101	166	1.46×10^{7}	3.11×10^{4}
H3C25C6PAN	SVOC	$C_7H_9NO_8$	235.1	2.27×10^{-8}	101	11.6	2.10×10^{8}	2.16×10^{3}
C98NO3	SVOC	$C_9H_{15}NO_6$	233.2	7.98×10^{-9}	105	204	3.42×10^{7}	1.33×10^{4}
mC106OOH	SVOC	$C_{10}H_{16}O_5$	216.2	7.08×10^{-9}	105	212	3.70×10^{7}	1.23×10^{4}
mPINIC	SVOC	$C_{9.22}H_{14.44}O_4$	189.3	4.77×10^{-9}	108	50.8	2.30×10^{8}	1.98×10^{3}
LIMALNO3	SVOC	$C_{10}H_{17}NO_6$	247.2	1.48×10^{-9}	113	650	5.75×10^{7}	7.88×10^{3}
mC98OOH	SVOC	$C_{8.94}H_{15.75}O_5$	203.3	1.13×10^{-9}	114	37.5	1.30×10^{9}	349
mC719OOH	LVOC	$C_{7.15}H_{12.30}O_5$	178.2	7.14×10^{-10}	118	7.52	1.03×10^{10}	43.9
mNORLIMOOH	LVOC	$C_{8.37}H_{14.96}O_{5.14}$	197.8	2.56×10^{-10}	121	21.5	1.01×10^{10}	45

mNC102OOH	LVOC	$C_{10}H_{15.95}NO_7$	262.2	2.12×10^{-10}	121	4.11×10^{3}	6.38×10^{7}	7.14×10^{3}
LIMALOOH	LVOC	$C_{10}H_{18}O_5$	218.2	1.50×10^{-10}	124	137	2.70×10^{9}	168
mC812OOH	LVOC	$C_{8.32}H_{14.65}O_5$	194.8	3.23×10^{-11}	132	40.9	4.20×10^{10}	10.8
C813NO3	LVOC	$C_8H_{13}NO_7$	235.2	1.65×10^{-11}	135	30.8	1.10×10^{11}	4.16
mC621OOH	LVOC	$C_{6.34}H_{11.34}O_{5.66}$	178.2	1.56×10^{-11}	135	1.93	1.84×10^{12}	0.246
C516OOH	LVOC	$C_5H_8O_6$	164.1	7.12×10^{-12}	139	0.716	1.09×10^{13}	0.0417
mC10H14O9	LVOC	$C_{10}H_{15.47}O_{8.40}$	269.8	3.41×10^{-12}	75.9	63.2	2.58×10^{11}	1.76
C813OOH	LVOC	$C_8H_{14}O_6$	206.2	8.76×10^{-13}	148	9.03	7.02×10^{12}	0.0648
mC10H14O11	ELVOC	$C_{10}H_{14.77}O_{11.25}$	314.8	1.80×10^{-15}	81.3	12.6	2.45×10^{15}	1.85×10^{-4}
C20H30O13	ELVOC	$C_{20}H_{30}O_{13}$	478	1.97×10^{-20}	105	4.86×10^{3}	5.79×10^{17}	7.83×10^{-7}

[a] species with "m" are the new surrogates that merged multiple MCM species. [b] VOCs (stable gas-phase species) and radicals (unstable gas-phase species) are assumed not to undergo gas-particle partitioning; SVOCs: semi-volatile organic compounds with saturation vapor pressure (P_{sat} at 298 K) lower than 10^{-9} atm; LVOCs: low-volatile organic compounds with P_{sat} between 10^{-9} atm and 10^{-13} atm; and ELVOCs: low-volatile organic compounds with P_{sat} lower than 10^{-13} atm. [c] Molar weight (g mol^{-1}). Properties calculated for condensable substances only: [d] saturation vapor pressure at 298 K (atm) ; [e] Enthalpy of vaporation (kJ mol^{-1}); [f] Henry's law constant (mol L^{-1} atm^{-1}); [g] activity coefficient at infinite dilution in water.

S2.2 Companions to other studies

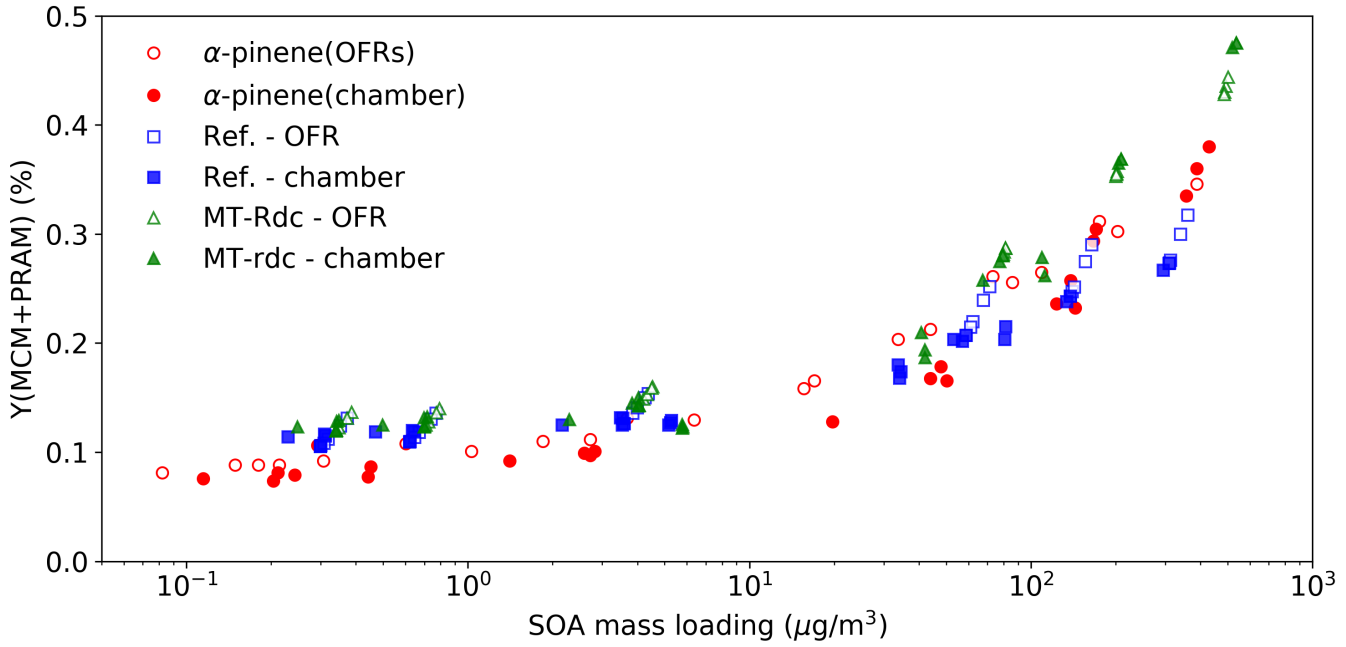


Figure S2: Reproduction of the SOA yields from the ozonolysis of α-pinene reported by [Xavier et al., 2019] (red circles) with the reference mechanism (MCM+PRAM, blue square), and the reduced monoterpene SOA mechanism MT-rdc (green triangle). The gas-particle partitioning is calculated using UManSysProp ([Topping et al., 2016]), with the method of [Myrdal and Yalkowsky, 1997] for vapor pressure and the method of [Nannoolal et al., 2004] for the boiling point. The legend "chamber" and "OFR" stands for SOA yields simulated using configurations of an idealized smog chamber (open symbols) and an oxidative flow reactor (filled symbols), respectively, as detailed in [Xavier et al., 2019].

S2.3 Testing results and SOA yields under all testing conditions

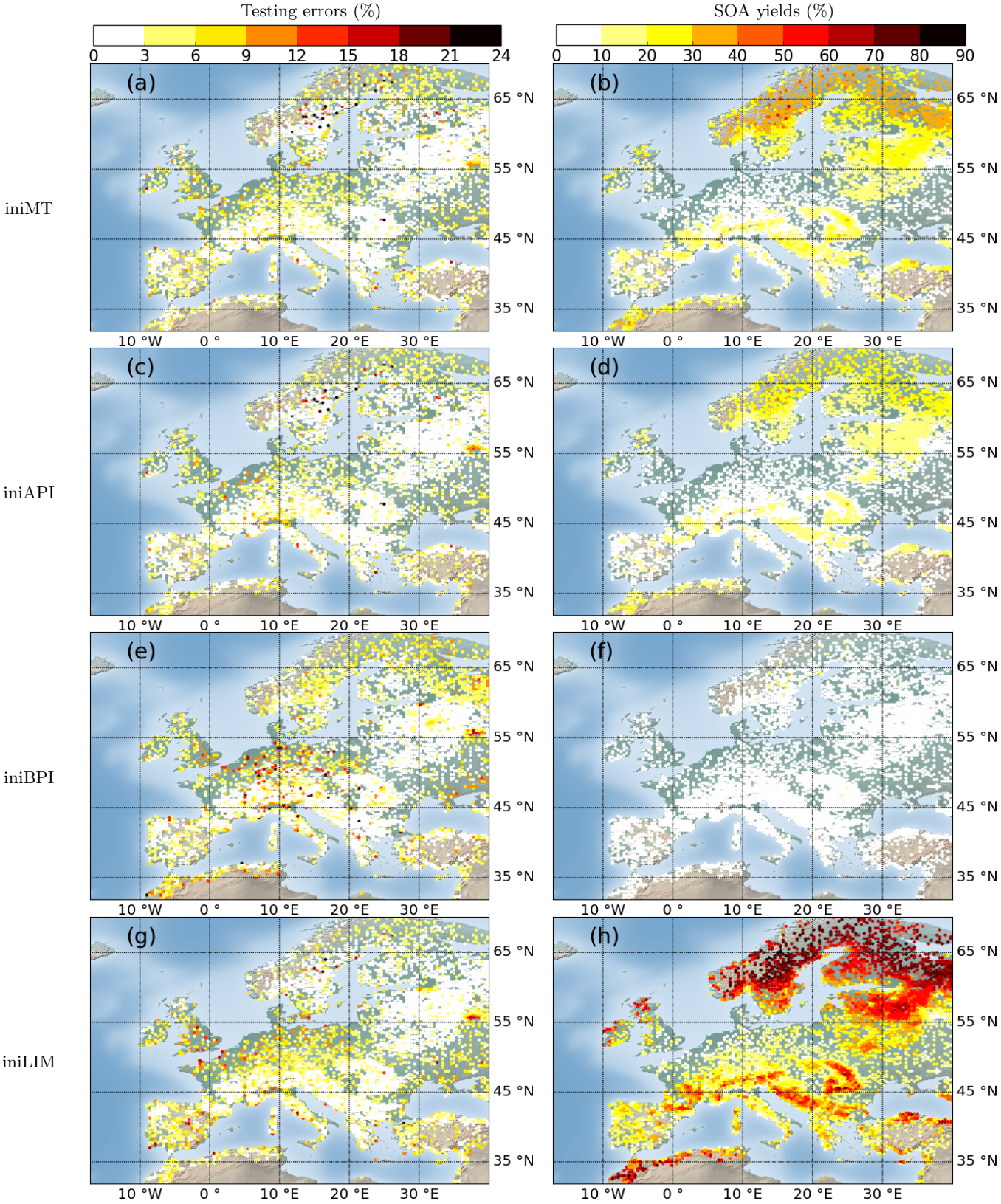


Figure S3: Map distributions of MT-rdc testing errors (left panels) and SOA yields (right panels) simulated with different initial precursor conditions under all testing conditions (9 818 conditions).

S3 Reaction pathways of other reduced mechanisms: case I to III

S3.1 Case I mechanism

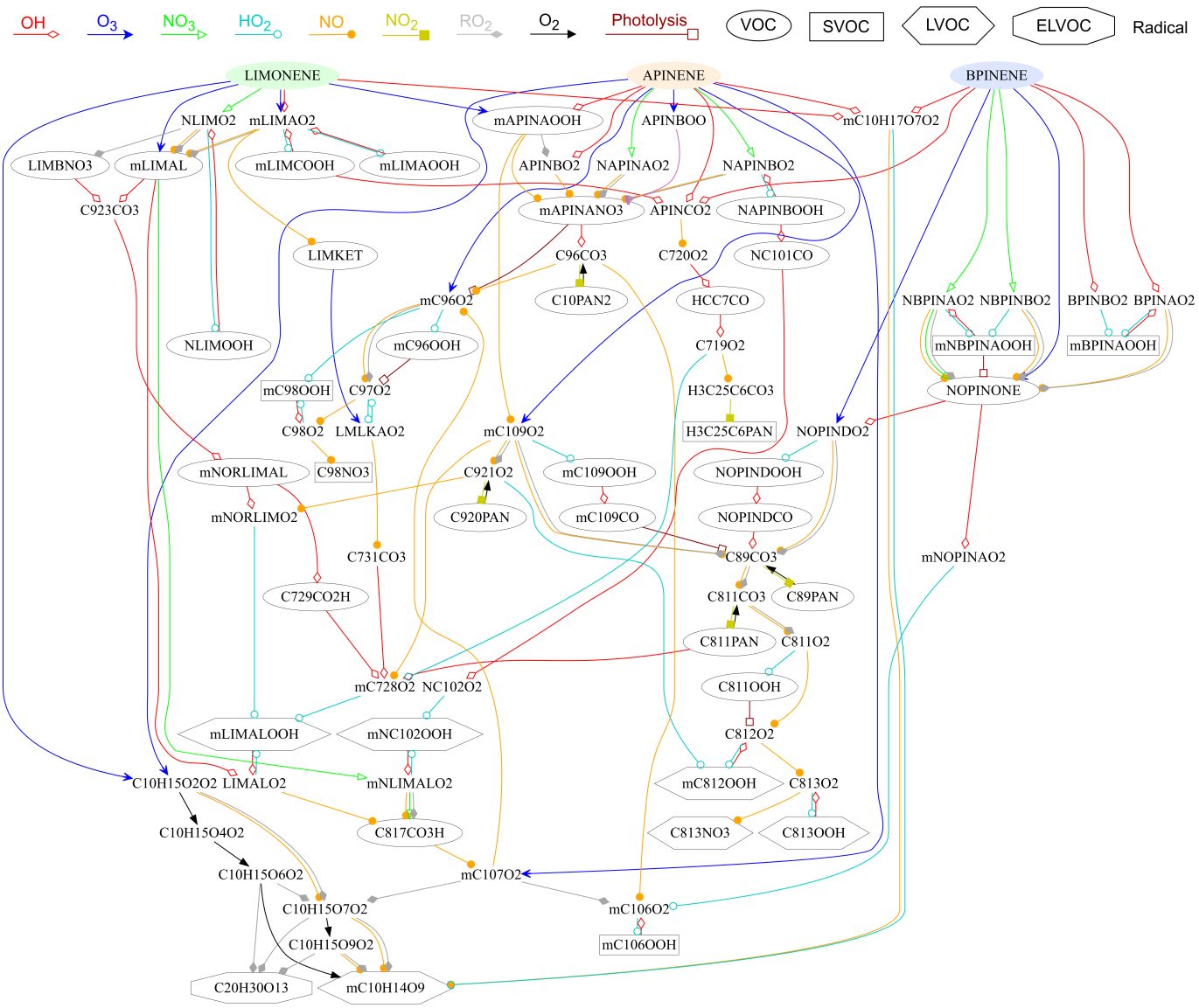


Figure S4: Reaction pathway related to SOA formation in the case I mechanism.

S3.2 Case II mechanism

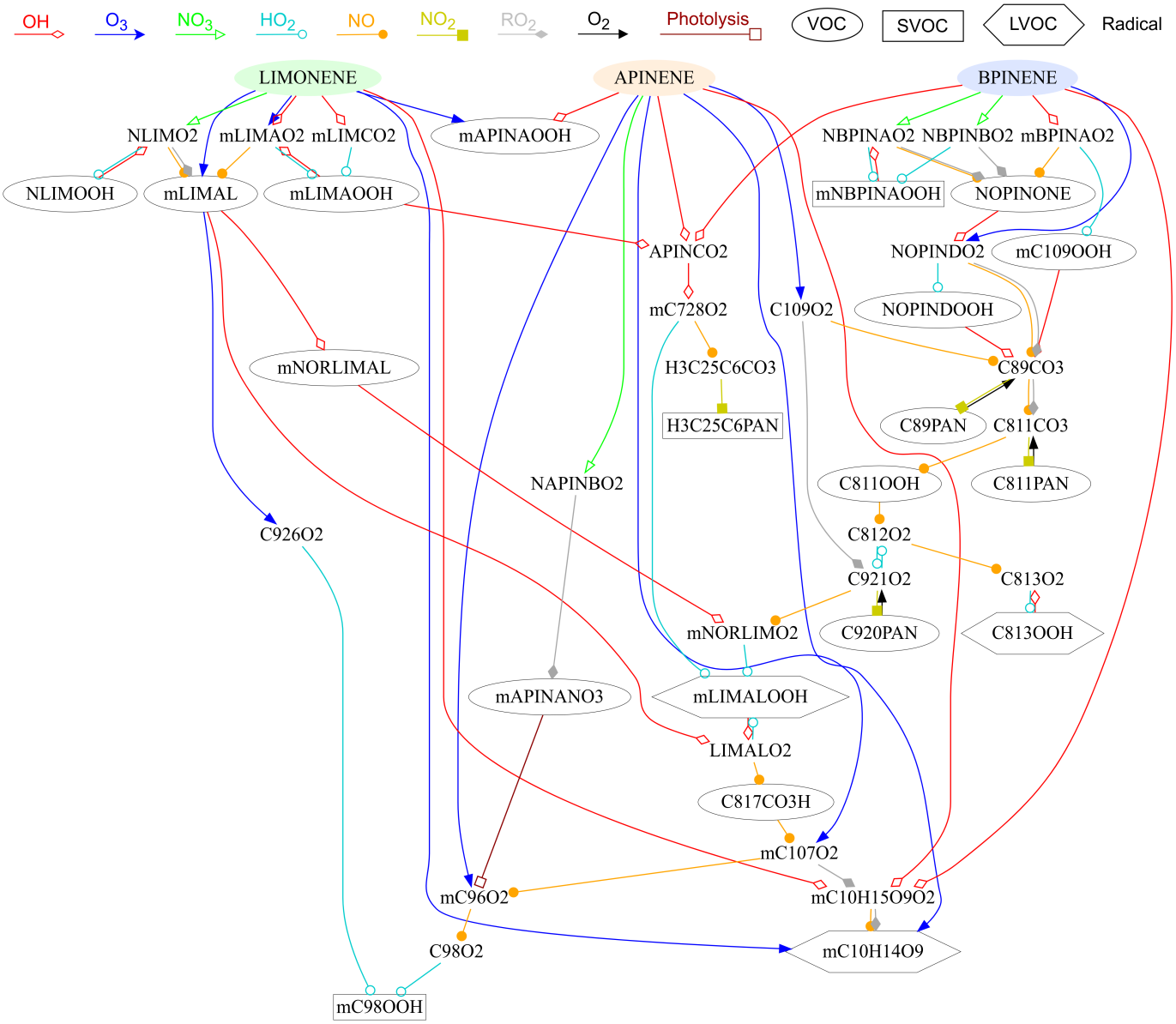


Figure S5: Reaction pathway related to SOA formation in the case II mechanism.

S3.3 Case III mechanism

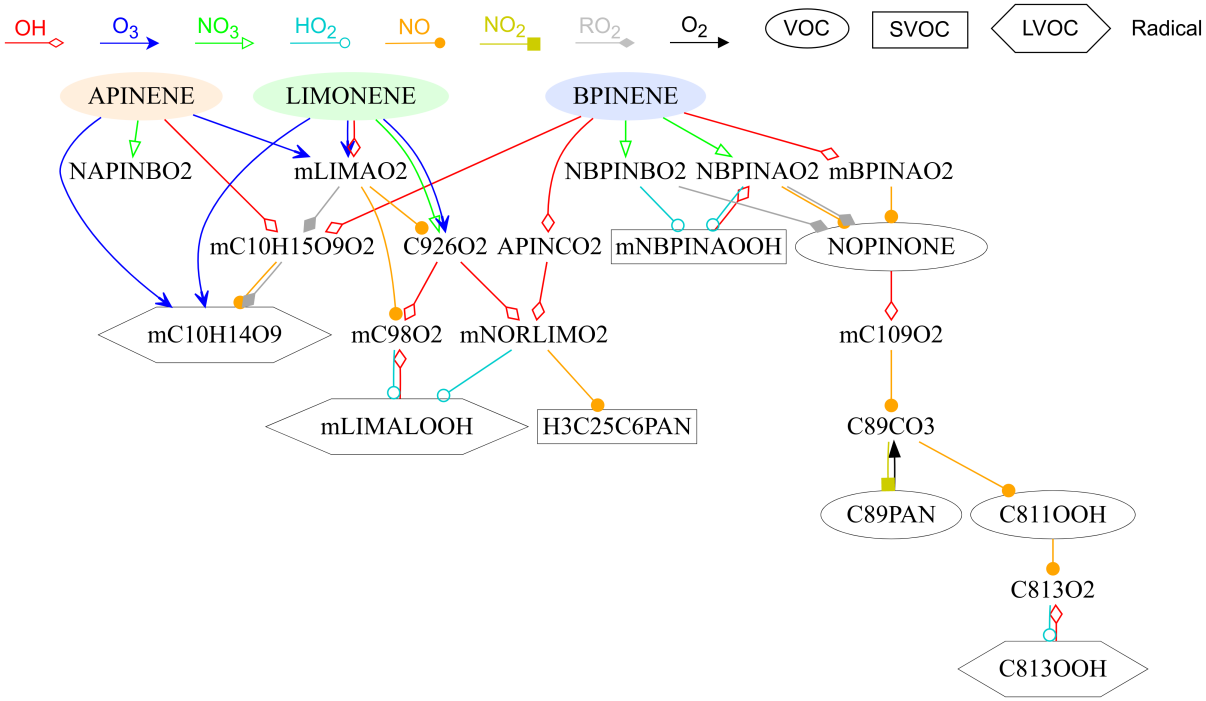


Figure S6: Reaction pathway related to SOA formation in the case III mechanism.

References

[Myrdal and Yalkowsky, 1997] Myrdal, P. B. and Yalkowsky, S. H. (1997). Estimating pure component vapor pressures of complex organic molecules. *Ind. Eng. Chem. Res.*, 36(6):2494–2499.

[Nannoolal et al., 2004] Nannoolal, Y., Rarey, J., Ramjugernath, D., and 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 Equilibr.*, 226:45–63.

[Topping et al., 2016] Topping, D., Barley, M., Bane, M. K., Higham, N., Aumont, B., Dingle, N., and McFiggans, G. (2016). Umansysprop v1.0: an online and open-source facility for molecular property prediction and atmospheric aerosol calculations. *Geosci. Model Dev*, 9(2):899–914.

[Xavier et al., 2019] Xavier, C., Rusanen, A., Zhou, P., Dean, C., Pichelstorfer, L., Roldin, P., and Boy, M. (2019). Aerosol mass yields of selected biogenic volatile organic compounds–a theoretical study with nearly explicit gas-phase chemistry. *Atmos. Chem. Phys.*, 19(22):13741–13758.

Chapter 4

Modeling the response of biogenic secondary organic aerosol formation to anthropogenic NOx emission reduction: a comparison of implicit and semi-explicit aerosol mechanisms

Anthropogenic nitrogen oxide (NOx) emissions from human activities contribute significantly to the formation and evolution of secondary organic aerosols (SOA). An investigation of the influences of NOx emission reduction on biogenic SOA formation using the three-dimensional (3-D) Chemical Transport Model (CTM) CHIMERE is presented in this chapter. The semi-explicit GENOA-generated biogenic SOA mechanism (GBM) and the implicit Hydrophobic/Hydrophilic Organic mechanism (H^2O) are adopted to simulate SOA concentrations and compositions over Europe for the 2018 summer (from June 1 to August 31). The GBM mechanism incorporates multi-generation oxidation pathways from near-explicit mechanisms reduced by GENOA v2.0, involving 228 reactions and 128 species describing the SOA formation from monoterpene (MT) and sesquiterpene (SQT). In contrast, the H^2O mechanism describes the MT and SQT SOA formation with a few model species and reactions organized from experimental data, involving 25 reactions and 18 species.

The use of the detailed semi-explicit GBM mechanism improves the performance of SOA simulation in current 3-D modeling, leading to closer agreement with measurements compared to the H^2O mechanism. The simulated OA concentrations with GBM and H^2O mechanisms are similar in magnitude, with biogenic MT and SQT SOAs dominating the concentrations, particularly in central Europe. The detailed GBM mechanism considers multi-generation oxidation, including the formation of highly oxidized HOMs due to auto-oxidation for MT SOAs, leading to more oxidized aerosols than those estimated by H^2O, which lacks this level of detail.

Furthermore, the influences of NOx emission reduction (i.e., NO, NO_2, and HONO) on biogenic SOA formation have been investigated with GBM and H^2O mechanisms. In response to NOx emission reduction, aerosol concentrations simulated with GBM and H^2O mechanisms increase and become more oxidized. When the anthropogenic

NOx emission is reduced by 50 %, most places in Europe are under the low-NOx regime. Due to the absence of NO, the oxidized products from MT degradation react preferentially with organic peroxy radicals (RO_2) and HO_2, leading to changes in SOA concentration and composition. For MT SOAs, condensables with lower volatility are formed from reactions with RO_2 and HO_2 than those from reactions with NO, favoring auto-oxidation for the formation of highly oxygenated molecules (HOMs). This change, which significantly contributes to SOA formation and aging, can be included in 3-D simulations with the GBM mechanism.

These results suggest that using a semi-explicit SOA mechanism in 3-D modeling can capture more variability in SOA concentrations based on the physical-chemical conditions, compared to using an implicit SOA mechanism. The importance of integrating a detailed SOA mechanism that considers the multi-generation aging process and the formation of highly oxidized molecules (HOMs) is also highlighted for accurate prediction of SOA concentration and composition in air quality modeling.

Contents

Modeling the response of biogenic secondary organic aerosol formation to anthropogenic NOx emission reduction: a comparison of implicit and semi-explicit aerosol mechanisms

Abstract

This study investigates the impacts of NOx emission reduction on secondary organic aerosol (SOA) formation using the three-dimensional (3-D) Chemical Transport Model (CTM) CHIMERE and two SOA mechanisms: a semi-explicit GENOA-generated biogenic SOA mechanism (GBM) and an implicit Hydrophobic/Hydrophilic Organic mechanism (H^2O). The GBM mechanism is generated by the GENerator of Reduced Organic Aerosol Mechanisms (GENOA) version 2.0. It considers detailed SOA formation pathways and aging with multi-generation oxidation related to monoterpene and sesquiterpene SOA formation. The use of a semi-explicit mechanism leads to a significant improvement in model performance compared to measurements. Additionally, the study shows that SOA aging contributes significantly to SOA formation. In response to the emission reduction of nitrogen oxides (NOx), aerosol concentrations simulated with GBM and H^2O mechanisms increase and become more oxidized. The changes are due to enhanced HO_2 and RO_2 reactions with oxidized products from biogenic VOC degradation, which is favorable to auto-oxidation for the formation of highly oxygenated molecules (HOMs) from monoterpene. The results suggest that detailed semi-explicit SOA mechanisms are potentially valuable for accurate predictions of air quality and informed policy decisions. Overall, this study provides insights into the complexity of gas-phase chemistry on SOA formation and highlights the importance of incorporating semi-explicit SOA mechanisms in CTMs.

1 Introduction

Atmospheric aerosols play a crucial role in influencing air quality, climate, and human health ([Breysse, 2013; Seinfeld, 2016; McNeill, 2017]). Among all types of aerosols, secondary organic aerosols (SOAs) have received much attention in air quality studies (e.g., [Kanakidou, 2005; Hallquist, 2009; Couvidat, 2013; Huang, 2014]). The formation processes of SOAs in the atmosphere are complex, involve multiphase physicochemical transformations, and are not yet well understood ([Hodzic, 2016]). After being emitted into the troposphere, volatile organic compounds (VOCs) undergo multi-generation gas-phase chemistry, forming oxidation products with low volatilities that may condense on existing particles to form SOA through gas-particle partitioning under favorable atmospheric conditions ([Hallquist, 2009]). When SOAs are formed, SOA chemical aging occurs due to successive oxidation of the first-generation oxidation products ([Donahue, 2006; Wang, 2018]). Due to a wide variety of VOC origins and SOA formation conditions, the SOA composition and concentrations vary spatially and temporally. Generally, VOCs can come from many sources, including biogenic emissions from vegetation, anthropogenic emissions from human activities (such as transportation and manufacturing), and biomass burning. [Guenther, 2012] reported a yearly production of approximately 1 000 Tg of biogenic VOCs, including 50 % from isoprene, 15 % from monoterpene, and 3 %

from sesquiterpene. Although emissions of monoterpene and sesquiterpene are lower than those of isoprene, they strongly influence SOA formation because of their higher yields ([Seigneur, 2019]).

In contrast to biogenic VOCs, anthropogenic VOCs are linked directly to human activity. Besides VOCs, anthropogenic emissions also include a range of other pollutants, such as nitrogen oxides (NOx) and sulfur dioxide (SO_2), which can alter aerosol formation. Particularly, nitrogen oxides (NOx) are one of the most important pollutants influencing SOA formation [Ng, 2007; Porter, 2021]. In addition to playing a role in the formation and destruction of tropospheric ozone concentration and therefore affecting the oxidation of precursors, NOx can react with the organic radicals (RO_2) formed by the oxidation of precursors and therefore affect the chemical pathways and SOA formation.

Current regulation efforts, e.g., on traffic, are leading to significant reductions in NOx emissions [André, 2020]. It is generally considered that biogenic organic aerosol concentrations drop as a response to current emission regulations and decrease of anthropogenic emissions, particularly in rural and peri-urban areas where oxidant concentrations (e.g., ozone) are predicted to decrease ([Sartelet, 2012; Shrivastava, 2019]). However, some studies suggested that reducing anthropogenic emissions could potentially result in less significant reductions in biogenic organic aerosol concentrations or even an increase (e.g., [Huang, 2020]). These non-linear effects could be attributed to competition between low-NOx and high-NOx conditions or even to non-linear effects due to the interactions of products formed from individual VOCs [Takeuchi, 2022]. This highlights the need for further investigation into the influences of emission reduction on SOA formation. To effectively regulate and mitigate the impact of anthropogenic emissions on the environment, it is necessary to accurately predict the influences of those emissions on aerosol formation. This requires a comprehensive understanding of accurate VOC chemistry, which involves the interactions between emissions and aerosol formation, including the influences of environmental conditions and atmospheric pollutants.

Explicit chemical mechanisms, such as the Generator for Explicit Chemistry and Kinetics of Organics in the Atmosphere (GECKO-A), which is fully explicit, and the Master Chemical Mechanism (MCM), which lumps reactions after first and second generations, contains up-to-date knowledge on VOC chemistry derived from theoretical and experimental studies. As a complement to MCM, the Peroxy Radical Autoxidation Mechanism (PRAM) models auto-oxidation and the formation of extremely low-volatility organic compounds (ELVOCs) from monoterpenes that are missing in MCM ([Roldin, 2019]). However, the direct use of explicit chemical mechanisms in regional-scale modeling is scarce due to the computational consideration ([Li, 2015]). Currently, three-dimensional (3-D) Chemistry-Transport Models (CTM) often use implicit chemical mechanisms based on a few model species and reactions to simulate the formation of organic aerosols. These implicit mechanisms are usually built from chamber measurements, potentially lacking some chemical pathways to depict SOA formation over contrasted regions adequately. Widely used approaches to derive these implicit mechanisms of SOA formation are based on the two-product Odum approach ([Odum, 1996]), the volatility basis set (VBS) approach [Donahue, 2006]), or the surrogate approach (e.g., [Couvidat, 2012]). In the one-dimensional (1-D) VBS approach, organic compounds are divided into logarithmically-spaced bins of similar saturation concentration. In the two-product Odum approach, the oxidation of a VOC precursor is approximated by a reaction forming two lumped prod-

ucts, which are semi-volatile and can condense onto the particle phase. In the surrogate approach, these products are affected by surrogate molecules with representative physicochemical properties for gas/particle partitioning. As detailed in the description of the H^2O mechanism ([Couvidat, 2012]), the surrogate approach allows taking into account the hydrophilic properties of SOA and interactions between compounds. However, these interactions are represented in a very simple way, especially as the surrogate molecules are chosen arbitrarily, based on a few measurements only. To represent auto-oxidation from monoterpenes, a simple chemical scheme built from the measurements of [Ehn, 2014] was implemented in the implicit mechanism H^2O ([Chrit, 2017]). 3-D simulations showed the potentially large influence of auto-oxidation for SOA formation over the Mediterranean ([Chrit, 2017]) and for ultrafine particle formation over the city of Paris ([Sartelet, 2022a]). However, the assumptions used in implicit mechanisms could lead to uncertainties in the model evaluation of the response to regulation assessments.It is, therefore, necessary to evaluate the models and their responses, taking into account the different formation pathways, the non-linear effects due to the interaction of oxidation products, and the dependency on environmental conditions.

To address this issue, the GENerator of Reduced Organic Aerosol Mechanisms (GENOA) was developed by [Wang, 2022; Wang, 2023], generating concise semi-explicit SOA mechanisms from detailed chemical mechanisms. The generated SOA mechanisms preserve the complexity of explicit SOA formation mechanisms within a size suitable for 3-D regional simulations. Version 2.0 of GENOA ([Wang, 2023]) employs a parallel reduction approach, allowing the processing of mechanisms from multiple SOA precursors simultaneously, and it has been applied to monoterpene and sesquiterpene reduction.

In this work, to investigate SOA formation response to anthropogenic emission reduction, the GENOA-generated SOA mechanisms from monoterpene and sesquiterpene are integrated into the 3-D CTM model CHIMERE coupled to the state-of-the-art aerosol module SSH-aerosol and applied to simulations over Europe. With 3-D CHIMERE simulations, the pollutant concentrations simulated with the GENOA-reduced SOA mechanisms and the implicit H^2O mechanism are compared, as well as differences in SOA formation for a NOx emission reduction scenario. The model and the simulation set-up are detailed in section 2. The CHIMERE concentrations simulated with implicit and semi-explicit chemical mechanisms are compared with measurements in section 3.1 and inter-compared in section 3.2. Finally, the influence of the mechanism on concentrations simulated with the NOx emission reduction scenario is investigated in section 3.3. By exploring the relationship between anthropogenic emissions, aerosol formation, and environmental conditions, this study aims at contributing to our understanding of the complex interplay between these factors and support the development of effective strategies for reducing the negative impacts of anthropogenic emissions on air quality.

2 Method

2.1 Model overview

The 3D simulations are carried out with the CHIMERE CTM model ([Menut, 2021]). The model was modified by coupling CHIMERE with the aerosol formation SSH-aerosol ([Sartelet, 2020]) based on a splitting approach, where the different processes (e.g., trans-

port, chemistry, and aerosol dynamics) are solved sequentially. The model first solves together all processes related to transport, deposition, and emissions. Then the model calculates the evolution of concentrations due to chemical reactions. In the last step, CHIMERE launches SSH-aerosol to solve all processes related to aerosol dynamics (condensation/evaporation of semi-volatile compounds and coagulation).

In SSH-aerosol, the gas-particle partitioning is computed with the thermodynamic module SOAP ([Couvidat, 2015]) accounting for interactions between organic and inorganic compounds based on the molecular structure of the molecules (i.e., non-ideality of aerosol). In the current study, thermodynamic equilibrium is assumed for gas-particle partitioning. For the particle size discretization, a sectional approach is used with 10 sections and diameters ranging from 10 nm to 10 μm.

2.2 SOA mechanisms

The SOA formation in the gas phase is simulated with two different SOA chemical mechanisms: the Hydrophilic/Hydrophobic Organic (H^2O) mechanism and the GENOA-reduced SOA mechanism. The H^2O mechanism uses an implicit surrogate approach based on experimental data ([Couvidat, 2012; Majdi, 2019]). Each key SOA precursor is oxidized into several model surrogate species with attached molecular structures representing lumped aerosol species. Its latest version [Sartelet, 2020], which includes a simple chemical scheme to represent the auto-oxidation of monoterpenes, is adopted in this work.

As a mechanism reduction algorithm, GENOA v2.0 generates semi-explicit SOA mechanisms by reducing near-explicit chemical mechanisms, preserving their complexity of VOC chemistry for SOA formation. The GENOA-reduced SOA mechanism, hereafter referred to as "GBM" for GENOA Biogenic Mechanism, was trained by the GENOA v2.0 algorithm ([Wang, 2023]), for the degradation schemes of biogenic SOA precursors, including one C_{15} surrogate (i.e., humulene) representing all sesquiterpenes and three C_{10} surrogates (i.e., α-pinene, β-pinene, limonene) representing all monoterpenes. As done in H^2O and following [Pun, 2006], the α-pinene and β-pinene mechanisms were adopted to represent the SOA formation from sabinene and carene respectively. The limonene mechanism was used to represent the SOA formation from the rest of the MT precursors. The degradation schemes of limonene and ocimene are identical except for the initial degradation rates. For other SOA precursors (i.e., isoprene and aromatics), GBM uses the same SOA schemes as in H^2O. For simplicity and clarity, abbreviations are adopted for the repeatedly mentioned organic names listed in Table 4.1.

Full name	Abbr.	Full name	Abbr.
sesquiterpene	SQT	monoterpene	MT
humulene	HUM	α-pinene	API
β-pinene	BPI	limonene	LIM
ocimene	OCI		

Table 4.1: Abbreviations for organic species mentioned in this work.

The MT SOA scheme of GBM is trained from MCM and the Peroxy Radical Autoxidation Mechanism (PRAM, developed by [Roldin, 2019]). MCM describes the detailed degradation schemes of three MT species (i.e., API, BPI, and LIM) with O_3, OH, and

NO_3 radicals, while PRAM completes the MCM schemes with the formation of highly oxygenated molecules (HOMs) from MT. The GBM's MT SOA scheme contains 198 reactions and 107 species (including 24 condensables), reduced from 3 001 reactions and 1 227 species (including 738 condensables). As for HOMs, 30 reactions and 14 HOM species, including five condensables (i.e., three C_{10} monomers and two C_{20} dimers), are preserved in GBM from PRAM. The detailed training processes have been reported by [Wang, 2023]. The reaction pathways leading to the formation of MT HOM SOAs and all MT SOAs are shown in Figs. 4.1 and 4.22, respectively. The SQT SOA scheme of GBM is also obtained from GENOA v2.0, where the VOC degradation scheme of SQT in MCM (as β-caryophyllene) was used as the reference mechanism. After the reduction in GENOA v2.0, the scheme includes 30 reactions and 21 species (including 8 condensables), down from 1 626 reactions and 579 species (including 356 condensables). The SQT SOA scheme in GBM is presented in Fig. 4.4, and the scheme in H^2O can be presented by Eq. 4.1, where oxidants include O_3, OH, and NO_3 radicals.

$$HUM + oxidants \rightarrow 0.74BiBmP + 0.26BiBlP + oxidants \tag{4.1}$$

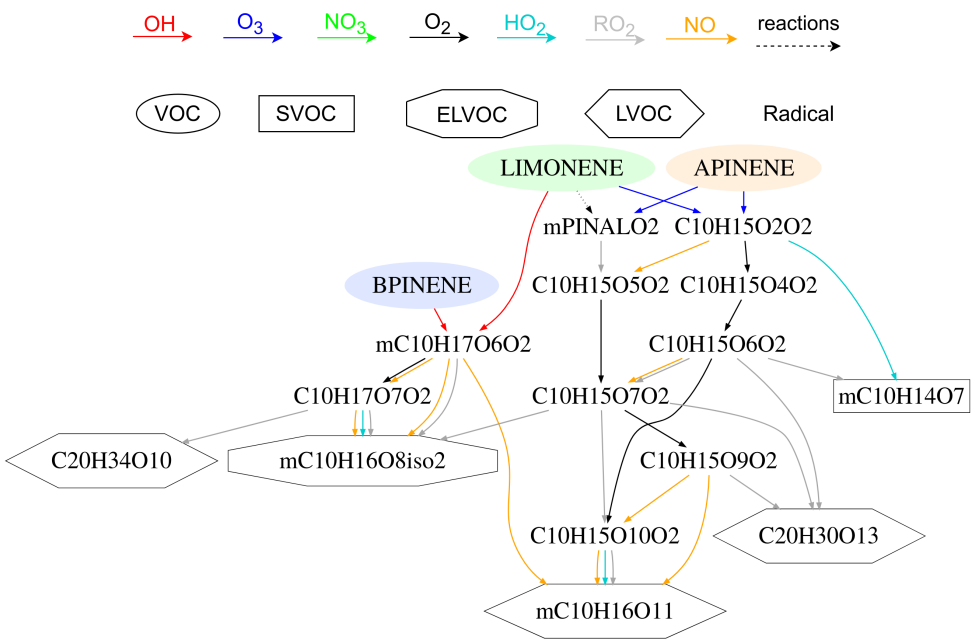


Figure 4.1: Reaction pathways of the MT SOA scheme illustrating HOM formation in the GBM mechanism trained from the PRAM mechanism.

A comparison of the size of the H^2O and GBM mechanisms derived from MT and SQT precursors can be found in Table 4.2, in terms of the numbers of reactions, gas-phase species, gas-phase condensable species, highly oxygenated molecules (HOMs), and organic peroxy radicals (RO_2s). Compared to H^2O, GBM preserves more species and reactions, which are indicative of the greater complexity of gas-phase chemistry in SOA formation.

For the MT scheme in H^2O, non-HOM species "BiA0D", "BiA1D", "BiA2D" and "BiNIT" are formed by oxidation of monoterpenes, representing hydrophilic non-dissociative, hydrophilic monoacid, hydrophilic diacid, and hydrophobic nitrate, respectively ([Couvi-

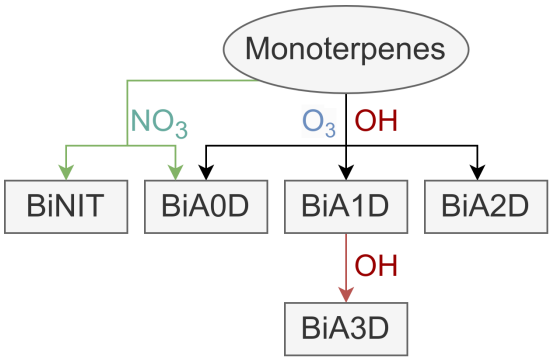


Figure 4.2: Reaction pathways of the monoterpene SOA scheme in the H^2O mechanism regarding the formation of non-HOM species.

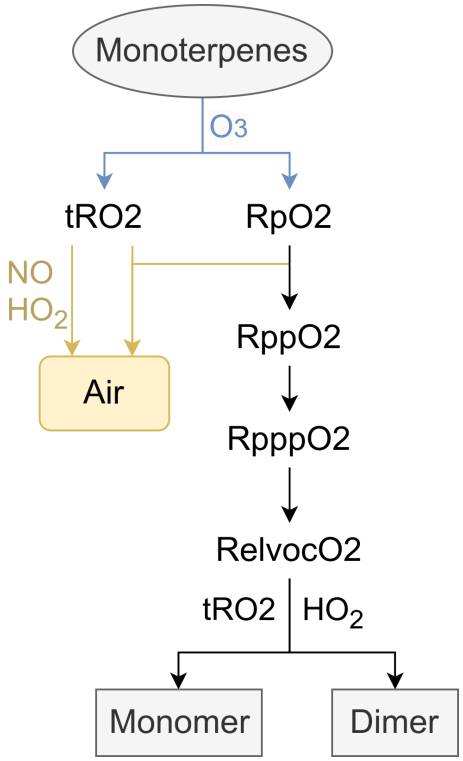


Figure 4.3: Reaction pathways of the monoterpene SOA scheme in the H^2O mechanism regarding the formation of HOM species, i.e., Monomer and Dimer. Species tRO2, RpO2, RppO2, RpppO2, RelvocO2 are peroxy radicals derived from monoterpene oxidation with Ozone, while "Air" indicates the destruction of tRO2 and RpO2 concentrations.

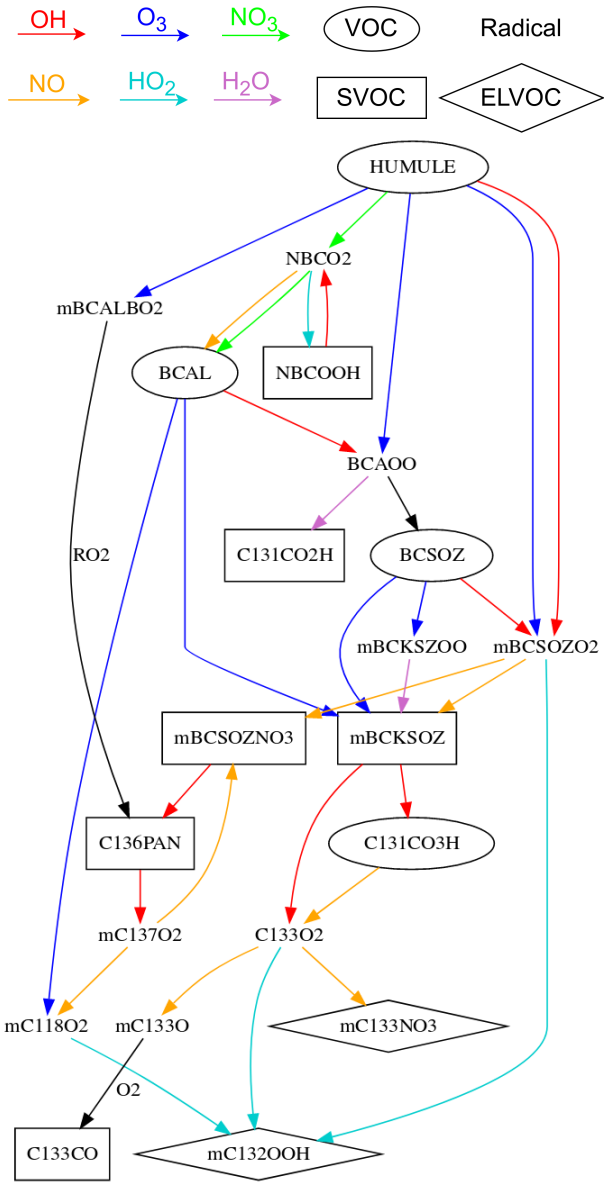


Figure 4.4: Reaction pathways of the sesquiterpene SOA scheme in the GBM mechanism illustrating the SOA formation. Degradation pathways that do not form SOA are not shown.

Precursor	MT		SQT	
Mechanism	H^2O	GBM	H^2O	GBM
Reaction [a]	22	198	3	30
Species	15	107	3	21
Condensables	6	24	2	8
HOMs	2	5	0	0
RO_2s	5	40	0	6

Table 4.2: Number of components in the SOA mechanisms H^2O and GBM for MT and SQT oxidations

[a] same degradation kinetic rate of API, BPI, and LIM in H^2O and GBM as in MCM.

dat, 2012]). Meanwhile, two hydrophobic surrogates "Monomer" and "Dimer" with 5 RO_2 species represent the SOA chemical aging and HOM formation from RO_2-RO_2 reactions and HO_2 reactions, as shown in Fig. 4.3. The stoichiometric ratios and partitioning coefficients in the H^2O MT SOA scheme are derived from experimental data. As shown by the reaction pathways in Fig. 4.2, where only first-generation degradation is considered for non-HOM SOA formation in H^2O, and reactions with inorganics (e.g., NO, HO_2) are not taken into account.

The MT scheme in GBM is more complex, as it is reduced from the near-explicit VOC mechanisms. The scheme attempts to preserve the responses of multi-generation oxidations (up to 13 generations) on SOA formation and aging for the degradation of API, BPI, and LIM. As a result, GBM contains reaction pathways common to two or three MTs, as well as pathways specific to one MT. The influences of inorganics on MT SOA formation are considered, as GBM includes 40 reactions with NO, 6 with NO2, 35 with HO_2, and 2 with H^2O, respectively. Furthermore, for the 24 condensables, their condensations on both the organic and aqueous aerosol phases are computed during partitioning (to be assumed as both hydrophilic and hydrophobic aerosols).

The RO_2-RO_2 reactions of GBM between two organic peroxy radicals (RO_2s) are performed using a "RO_2" pool that sums the concentrations of RO_2 species derived from the targeted SOA precursors and the background RO_2s. Background RO_2 concentrations are the sum of the RO_2 concentrations from inorganic reactions and the RO_2 concentrations from the other SOA precursors. In GBM, the MT RO_2 pool contains 22 MT RO_2 species and contributes to 30 reactions. For the MT HOM formation, as shown in Fig. 4.1, all 5 MT HOMs can be directly formed through RO_2-RO_2 reactions, while for non-HOM, the RO_2-RO_2 reactions products can react with NO or HO_2 to form condensables.

As for the SQT degradation, two hydrophobic organics with low and medium saturation vapor pressure, i.e., BiBlP and BiBmP, are adopted in the H^2O mechanism. In the GBM mechanism, there are 8 condensables from 30 multi-generational oxidations, including 8 reactions with NO, 4 with HO_2, and 2 with H^2O. As for RO_2-RO_2 reactions, only one reaction is preserved from GENOA reduction (see Fig. 4.4). A SQT RO_2 pool consisting of the background RO_2 and the SQT RO_2 species "mBCALBO2", is used to compute the formation of the condensable "C136PAN".

For comparison, the initial degradation rates of VOC oxidation in H^2O and GBM mechanisms are set to be consistent with each other. Consequently, the gas-phase VOC degradation and inorganic variations simulated with the two mechanisms are identical.

In 3-D simulations using the H^2O mechanism, the parameterization of [Pun, 2007] is usually used to account for intraparticle reactions (such as oligomerization) over acidic particles, increasing the partitioning of pinonaldehyde. For example, [Lemaire, 2016] showed that SOA formed via this parameterization could represent 50 % of the simulated concentrations of total biogenic SOA. While theoretically possible, the reactive uptake of pinonaldehyde on an acidic particle was shown to be too slow to be significant under atmospheric conditions ([Couvidat, 2018a]). Therefore, in this study, simulations with H^2O were performed with and without activating this parameterization.

2.3 Configuration of simulations

Simulations are performed over Europe (latitudes from 32 °N to 70 °N and longitude from 17 °W to 39.8 °E), with a horizontal resolution of 0.25 °×0.4 °. The OA concentrations and compositions from 1 June to 31 August 2018 are investigated, considering that biogenic aerosol formation is expected to be significant during the summer. All simulations started 15 days before (on 15 May) to minimize the influence of initial conditions. Boundary conditions are taken from CAMS CIFS global model simulations ([Flentje, 2021]). Meteorology was obtained from the operational analysis of the Integrated Forecasting System (IFS) model of the European Centre for Medium-Range Weather Forecasts (ECMWF) ([Flentje, 2021]). The anthropologic emissions of gas and particles were taken from the CAMS-REG-AP inventory (version v5.1_REF2.1) ([Kuenen, 2022]).

2.4 Computation of biogenic emissions

Biogenic emissions for CHIMERE are calculated in CHIMERE with the MEGAN v2.1 algorithm ([Guenther, 2012]), as described in [Couvidat, 2018b; Menut, 2021]. However, recent studies indicate that there is considerable uncertainty in the estimation of biogenic emissions, up to a factor of two to three at the global level, and even higher values at the regional level ([Messina, 2016; Sindelarova, 2022]). Studies also reported that biogenic emissions computed with MEGAN over Europe might approximately be significantly overestimated for isoprene by a factor of 3, while underestimated for MT by a factor of 3 ([Jiang, 2019; Ciccioli, 2023]). With higher emissions of MT, [Jiang, 2019] shows better model performance in simulated SOA concentrations.

Therefore, several simulations were performed in this study with either the default MEGAN emissions or with the MT and SQT emissions increased by a factor of 3 but with emissions of isoprene decreased by a factor of 3. While the underestimation of emissions may vary spatially and temporally, the use of these simple factors should provide a good estimation of the uncertainties related to biogenic emissions.

2.5 Observation

The observation data is extracted from the EBAS website (https://ebas.nilu.no/, last access: 2023/01/01). EBAS is a dataset that hosts atmospheric measurement data from various national and international programs. As this study is focused on the organic aerosol formation, the comparisons with observations concentrate on available data during the simulation period for $PM_{2.5}$, PM_{10} total concentrations as well as organic concentrations, i.e., organic carbon mass in $PM_{2.5}$ ($OC_{PM_{2.5}}$), organic carbon mass in PM_1 (OC_{PM_1}), and organic mass in PM_1 (OM_{PM_1}. The model to measurement comparisons for inorganic gaseous and other particulate pollutants (i.e., O_3, NO_2, particulate nitrate (pNO_3^-), particulate ammonium (pNH_4^+), particulate chloride (pCl^-), particulate sodium (PNa^+), and particulate sulfate (pSO_4^{2-})) are presented in the Fig. 4.21.

Statistical indicators, i.e., Mean Fractional Errors (MFE) and Mean Fractional Bias (MFB), are adopted for analysis. MFE and MFB are calculated by Eqs. 4.2 and 4.3, respectively, where C_i^{mod} and C_i^{mea} are the simulated and the measured concentrations of the targeted compound at time i (N is the total number of time steps). The following two criteria reported by [Boylan, 2006] have been used in many studies for evaluating the

model performance (e.g., [Couvidat, 2018b; Lannuque, 2020]): For a model performance to be considered acceptable, MFE and MFB should be in the range of 75 % and ± 60 %, respectively. The performance should fall within the range of 50 % and ± 30 % for MFE and MFB to be considered close to the best values a model can achieve. These evaluation criteria together with other statistical (i.e., correlation, the root mean square error (RMSE)) indicators are used in this study.

$$MFE = \frac{1}{N}\sum_{i=1}^{N}\frac{2|C_i^{mod} - C_i^{mea}|}{C_i^{mod} + C_i^{mea}} \tag{4.2}$$

$$MFB = \frac{1}{N}\sum_{i=1}^{N}\frac{2(C_i^{mod} - C_i^{mea})}{C_i^{mod} + C_i^{mea}} \tag{4.3}$$

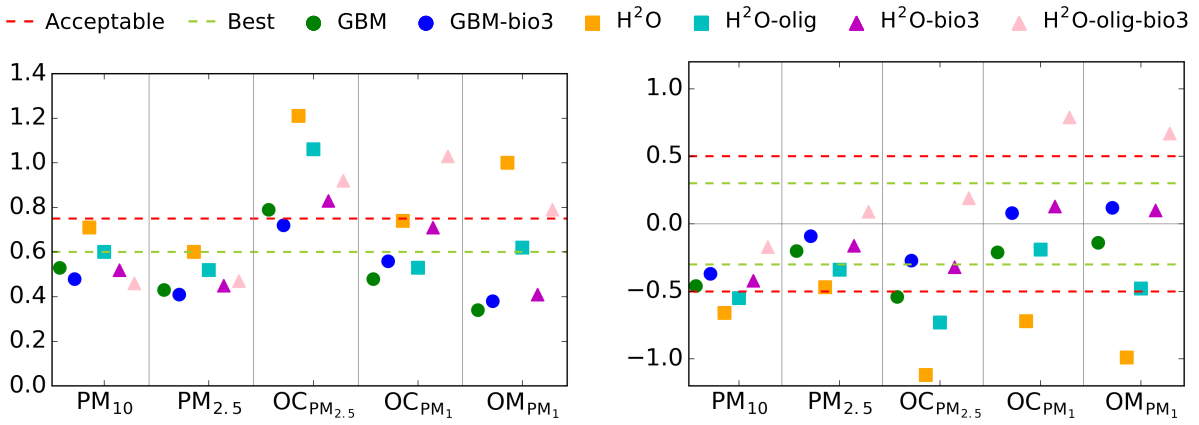
3 Results and discussion

3.1 Comparison between simulation results and measurements

In this section, the 3-D simulation results modeled with different SOA mechanisms are compared with the measurements. In order to determine the best configurations, a total of six simulations are conducted with the GBM and H^2O mechanisms, considering or not the modified biogenic emissions and oligomerization. The details of the simulations are listed below with the corresponding identifiers used in the analysis:

- "GBM": Simulation with the GBM mechanism.
- "GBM-bio3": Simulation with the GBM mechanism and modified biogenic emissions (triple for MTs and SQTs, and one-third for isoprene).
- "H^2O": Simulation with the H^2O mechanism.
- "H^2O-oligo": Simulation with the H^2O mechanism and activated oligomerization for monoterpene oxidation according to the parameterization of [Pun, 2007].
- "H^2O-bio3": Simulation with the H^2O mechanism and modified biogenic emissions (same emission as for the GBM-bio3 simulation).
- "H^2O-oligo-bio3": Simulation with the H^2O mechanism, activated oligomerization, and modified biogenic emissions.

The comparisons for inorganic gaseous and particulate pollutants between simulations and measurements can be found in Fig. 4.21, in terms of MEGAN and MFB. For inorganic pollutants, the results show that all simulations estimated similar inorganic concentrations, suggesting that modifications in SOA mechanisms and biogenic emissions have negligible effects on inorganic gaseous and particulate concentrations. The similarity is predictable since the modifications do not directly influence inorganic chemistry and gas-to-particle partitioning. Compared to the measurements, most of the inorganic concentrations estimations are acceptable, except for particulate sulfate (pSO_4^-), whose MFB value is about 70 %.



(a) MFE (b) MFB

Figure 4.5: MFE and MFB of the different investigated simulations compared to measurements for concentrations of $PM_{2.5}$, PM_{10}, $OC_{PM_{2.5}}$, OC_{PM_1}, and OM_{PM_1}.

Figure 4.5 presents the comparisons between simulation results and the measurements in terms of MFE (Fig. 4.5a) and MFB (Fig. 4.5b) for concentrations of $PM_{2.5}$, PM_{10}, $OC_{PM_{2.5}}$, OC_{PM_1}, and OM_{PM_1}. The numbers of selected observatory stations and compared measurements are listed in Table 4.3. While data are available for a significant number of stations for $OC_{PM_{2.5}}$, data for only a few stations are available for OC_{PM_1} and OM_{PM_1}. For OC_{PM_1}, only two stations with available data are found for the simulation domain during the simulation period. For OM_{PM_1}, 4 stations are initially found. However, two stations are not taken into account as they introduce significant uncertainties in the comparison: one is the Puy de Dôme observatory station in France (EMEP station code: FR0030R) on top of an extinct volcano at an altitude of 1 471 m that differs significantly of the average altitude of the corresponding cell (16.9 m), indicating that the resolution of the model is not sufficient to reproduce the local topography in the vicinity of this station. The other station is the only urban station of the database and is located near Lille (EMEP station code: FR0027U). Therefore, a very high resolution would be necessary to perform a suitable comparison at this station.

According to MFB values recorded in Fig. 4.5b, all simulations with non-modified biogenic emissions (GBM in green circle, H^2O in orange triangle, H^2O-olig in cyan triangle) underestimate the aerosol concentrations compared to measurements. This underestimation could be explained by an underestimation of biogenic emissions over Europe by MEGAN 2.1 as explained previously (see Sect.2.4).

For simulations with modified biogenic emissions, the simulation results of GBM-bio3 (blue circle in Fig. 4.5) and H^2O-bio2 (purple triangle in Fig. 4.5) are significantly improved compared to the results simulated with the original biogenic emissions in both MFE and FMB values for organic aerosols (i.e., $OC_{PM_{2.5}}$, OC_{PM_1}, and OM_{PM_1}) as well as $PM_{2.5}$ and PM_{10}. The results are consistent with the results of [Jiang, 2019; Ciccioli, 2023] who reported that underestimations on simulated SOA concentrations could be explained by an underestimation of biogenic emissions by MEGAN2.1.

With non-modified biogenic emissions, the H^2O-olig simulation improves the quality of the results in terms of MFB and MFE (e.g., MFB reduced from -0.99 to -0.48 for OM_{PM_1}).

However, this parameterization may artificially increase the SOA mass and compensate for errors in biogenic emissions. The performance of H^2O-olig (MFB = -0.73, MFE = 1.06 on $OC_{PM_{2.5}}$) is lower than the performance of H^2O (MFB = -1.12, MFE = 1.21, on $OC_{PM_{2.5}}$). Moreover, when accounted for the modified biogenic emission, H^2O-olig-bio3 leads to a clear overestimation of concentrations (MFB = 0.19, MFE = 0.92, on $OC_{PM_{2.5}}$). Based on all these results and the fact that the parameterization of [Pun, 2007] leads to an overestimation of pinonaldehyde condensation, we considered that the GBM-bio3 and H^2O-bio3 are the most reliable configurations.

Moreover, the statistical analysis shows that the results simulated with the GBM mechanism are closer to the measurement (with lower MFE and MFB values) than those simulated with the H^2O mechanism, suggesting that simulations with the GBM mechanism have better performances in SOA concentrations than those with the H^2O mechanism. The GBM mechanisms may simulate SOA concentrations closer to the measurements than those simulated with the H^2O mechanism, indicating that the detailed SOA mechanism may improve the performance of SOA simulation in current 3-D modeling.

Overall, the comparison with the measurements suggests that the simulations with GBM-bio3 and H^2O-bio3 have better performances than other simulations. As shown in Fig. 4.5, the two simulations have MFE and MFB mostly in acceptable ranges for particulate concentrations except for $OC_{PM_{2.5}}$, where H^2O-bio3 has an MFE of 83 % above the acceptable threshold of 75 % defined by [Boylan, 2006]. Table 4.3 provides more information about the measurement comparison of the GBM-bio3 and H^2O-bio3 simulations, including the number of measurements adopted for evaluation, measured and simulated mean concentrations, correlation, and RMSE. The results show that the GBM-bio3 simulation has higher correlation and lower RMSE values than the H^2O-bio simulation, indicating a better performance simulated with the GBM-bio3 simulation than the H^2O simulation. For the purpose of further investigation, the configurations of the GBM-bio3 and H^2O-bio3 simulations are applied, as these two simulations have the best performances compared to the measurements.

3.2 Comparison between the implicit and detailed SOA mechanisms

3.2.1 Comparison of organic aerosol concentrations

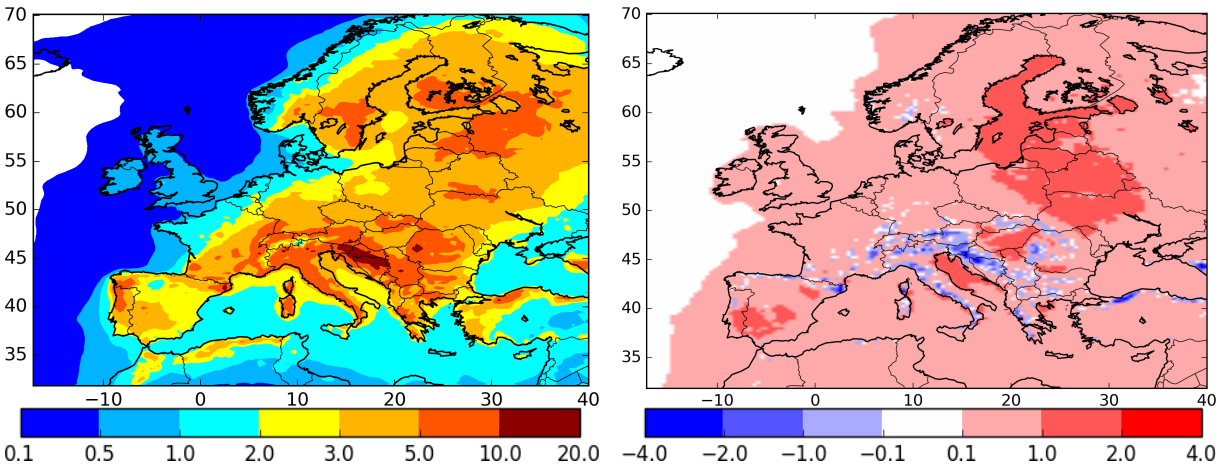
The 3-D results simulated using the GBM and H^2O mechanisms with modified biogenic emissions (one-third of isoprene emissions and triple of MT and SQT emissions estimated with MEGAN v2.1) are compared in this section. As no noticeable differences are observed in the concentrations of oxidants and inorganic gas-phase and particle pollutants simulated with the two mechanisms, the comparison between the two simulations focuses mainly on organic aerosol (OA) concentrations. For simplicity, the simulations are referred to by their SOA mechanism name, i.e., GBM and H^2O, for the following discussion.

Figure 4.6 shows the mass distribution of the total OA concentrations simulated with GBM (Fig. 4.6a), as well as the absolute concentration differences between GBM and H^2O simulations (Fig. 4.6b). Over the whole simulation period (from June to August 2018), the GBM mechanism simulates an average OA concentration over the domain of 2.1 $\mu g/m^3$ (maximum of 15.8 $\mu g/m^3$), while H^2O simulates a concentration of 1.7 $\mu g/m^3$ (maximum of 17.3 $\mu g/m^3$). With the GBM simulation, 95 % of OA average concentrations are

Table 4.3: Statistics analysis for comparing with measurements the simulated daily concentrations of simulations with GBM-bio3 and H^2O-bio3.

	PM_{10}	$PM_{2.5}$	$OC_{PM_{2.5}}$	OC_{PM_1}	OM_{PM_1}
No. station	80	61	25	2	2
No. measurement [b]	92	89	16	17	25
Measurement mean	13.5	7.9	2.5	2.3	4.2
Simulation with GBM - bio3					
Mean	9.3	7.2	2.4	3.1	5.7
RMSE	7.7	4.5	2.3	2.4	3.3
Correlation	0.62	0.64	0.72	0.39	0.84
MFB	-0.37	-0.09	-0.27	0.08	0.12
MFE	0.48	0.41	0.72	0.56	0.38
Simulation with H^2O - bio3					
Mean	8.9	6.7	2.6	3.6	5.9
RMSE	8.1	4.8	3.0	3.4	3.9
Correlation	0.59	0.62	0.69	0.41	0.82
MFB	-0.42	-0.16	-0.32	0.13	0.10
MFE	0.52	0.45	0.83	0.71	0.41

[b] the sum of all stations. Noted that the measurements have been converted to the daily average for comparison.



(a) GBM Conc. ($\mu g/m^3$) (b) GBM. - H^2O Conc.($\mu g/m^3$)

Figure 4.6: Average organic aerosol concentrations simulated with the GBM mechanism (a) and differences with the H^2O mechanism (b) during June-August 2018.

simulated in the range of 0 $\mu g/m^3$ to 5.5 $\mu g/m^3$, and 99 % of those lower than 8.0 $\mu g/m^3$. As shown in Fig. 4.6a, high OA concentrations (≥ 5 $\mu g/m^3$) are observed in central Europe between 35 °N and 50 °N, and in northern Europe, between 55 °N to 65 °N, while extreme high OA concentrations above 10 $\mu g/m^3$ are observed in the western Balkan Peninsula, near Bosnia. Those locations with high OA concentrations are consistent with the places with high biogenic emissions.Compared to H^2O, GBM generally simulates higher OA concentration over most continental areas of Europe. In Eastern Europe, near the Baltic Sea, Southern Spain, and north of the Mediterranean Sea, the simulated OA concentration differences are in the range of 1 $\mu g/m^3$ to 1.6 $\mu g/m^3$ (The highest increase of 1.6 $\mu g/m^3$ is found near the Baltic Sea). However, as shown in Fig. 4.6b, H^2O produces more SOA than GBM in central Europe (up to 3.2 $\mu g/m^3$ difference is found near Bosnia), at the place corresponding to high SOA concentrations and biogenic emissions.

3.2.2 Comparison of OM/OC ratios

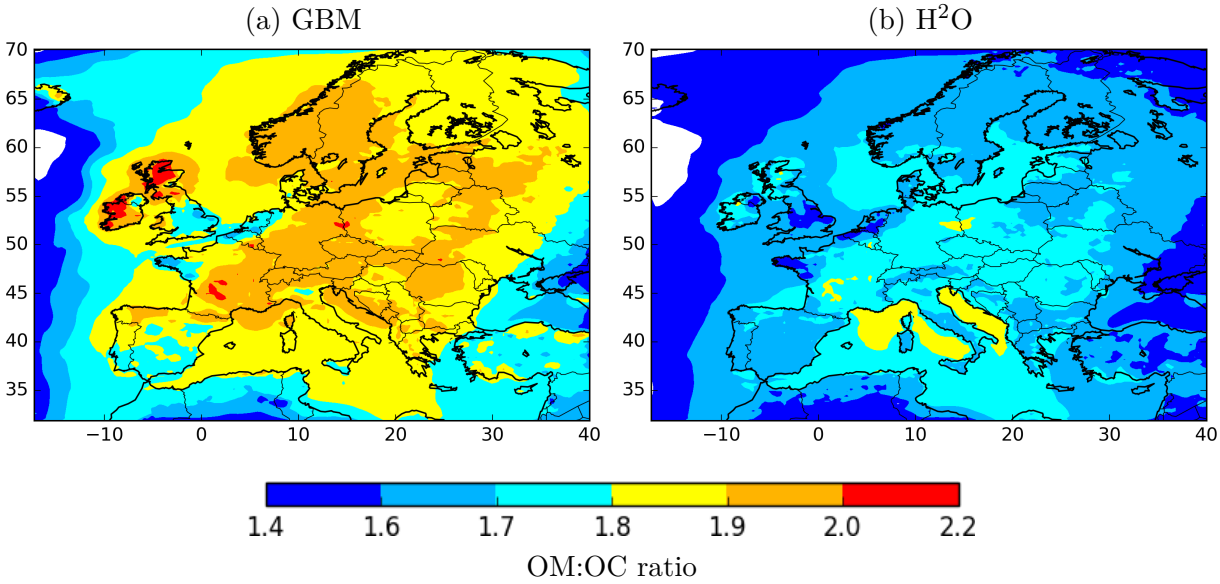


Figure 4.7: Average organic mass to organic carbon (OM:OC) ratios simulated with GBM (a) and H^2O mechanisms (b) during June-August 2018.

The map distributions of OM:OC ratios simulated with GBM and H^2O are presented in Fig. 4.7. A higher OM:OC ratio is simulated with the GBM mechanism with an average of 1.81 (ranging from 1.34 to 2.14), compared to the ratio simulated with H^2O with an average of 1.64 (ranging from 1.32 to 1.85). The OM/OC ratio distribution indicates that more oxidized products are simulated with GBM than with H^2O, which is consistent with the reaction pathways as more oxidized organics derived from high-generation oxidation are included in GBM. High-generation oxidation is referred to the succession oxidation steps after the first and the second generations.

However, due to the low amount of reported data on the OM/OC ratio, it is difficult to determine which simulated ratios are the most realistic ones. [Turpin, 2001] reported a ratio of 2.1 in rural sites in the US. Over Europe, [Poulain, 2011] reported an OM/OC ratio of 1.75 at Melpitz (in eastern Germany) during the summer of 2008. A ratio of 1.95

and 1.75 was simulated during the summer of 2018 with the GBM and H^2O mechanisms, respectively, suggesting a potential overestimation of the ratio by the GBM mechanism. However, the CHARMEX summertime measurements over the Mediterranean indicate a ratio of 2.34 at Ersa in Corsica and 1.97 at Mallorca in the summer of 2013 ([Chrit, 2017; Sartelet, 2022b]). The OM/OC ratio could then be underestimated at these two locations by the GBM mechanism (1.95 at Ersa and 1.85 at Mallorca).

3.2.3 Comparison of MT and SQT SOA

In the total organic mass simulated with GBM, MT SOAs account for approximately 45 %, SQT SOAs account for 41 %, and isoprene SOAs account for 0.3 %. Similar proportions with slight decreases are also observed with the H^2O mechanism, where MT, SQT, and isoprene account for approximately 43 %, 41 % and 0.2 %, respectively. The results show that biogenic SOAs, particularly the SOAs from MTs and SQTs, dominate the total OA concentrations over Europe during the 2019 summertime for both simulations, with a contribution of roughly 85 % with both mechanisms. The result is consistent with several studies demonstrating the dominance of biogenic proportions in SOA concentrations, particularly in rural and suburban areas (e.g., [Kelly, 2018; Hong, 2022]).

To further investigate the variations in aerosol formation, the total OA concentrations are divided into three categories: MT SOAs, SQT SOAs, and other OAs. Figure 4.8 presents their concentration distributions simulated with the GBM mechanism (left panels), as well as the absolute concentration differences between GBM and H^2O. Average concentrations of 0.96 $\mu g/m^3$, 0.88 $\mu g/m^3$, and 0.29 $\mu g/m^3$ of MT, SQT, and other SOA were simulated with GBM over the domain, respectively, against 0.73 $\mu g/m^3$, 0.70 $\mu g/m^3$, and 0.27 $\mu g/m^3$ simulated with H^2O. High MT and SQT SOA concentrations ($\geq$ 2 $\mu g/m^3$)are simulated with GBM in central and northeast Europe (See Fig. 4.8a and Fig. 4.8c). The highest concentrations (7.3 $\mu g/m^3$ for MT and 6.2 $\mu g/m^3$ for SQT) are found near Bosnia. For other OAs (Fig. 4.8e), average concentrations are generally lower than 1.2 $\mu g/m^3$, except in Northern Italy where concentrations of other SOAs locally reach 3.8 $\mu g/m^3$ due to high concentrations of aged primary organic aerosols (3.5 $\mu g/m^3$).

As GBM uses the same mechanism as H^2O for other SOA, no clear differences were found between the two simulations for other SOA concentrations (Fig. 4.8f). The differences in total SOA concentrations shown in Fig. 4.6 can therefore be explained by differences in MT and SQT SOAs. Both MT and SQT SOAs simulated with GBM decreased in part of Central Europe compared to the H^2O results, with a decrease down to 1.6 $\mu g/m^3$ for MT and 1.8 $\mu g/m^3$ for SQT. In northern Europe, GBM simulates more MT SOAs (up to 1.1 $\mu g/m^3$) but fewer SQT SOAs (-0.5 $\mu g/m^3$ to -0.1 $\mu g/m^3$). These results indicate that H^2O simulates higher SOA concentrations than GBM for regions with high OA concentrations (e.g., Central Europe with high biogenic emissions). However, for the other regions of Europe, higher concentrations of MT-SOA and SQT-SOA were simulated with the GBM mechanism.

3.2.4 Comparison of HOM and non-HOM concentrations

Depending on the aerosol properties, MT SOAs can be further divided into non-HOM and HOM SOAs. While some compounds in the MCM (and therefore some compounds formed by aging in the GBM mechanism) can be considered as HOM (because of a high

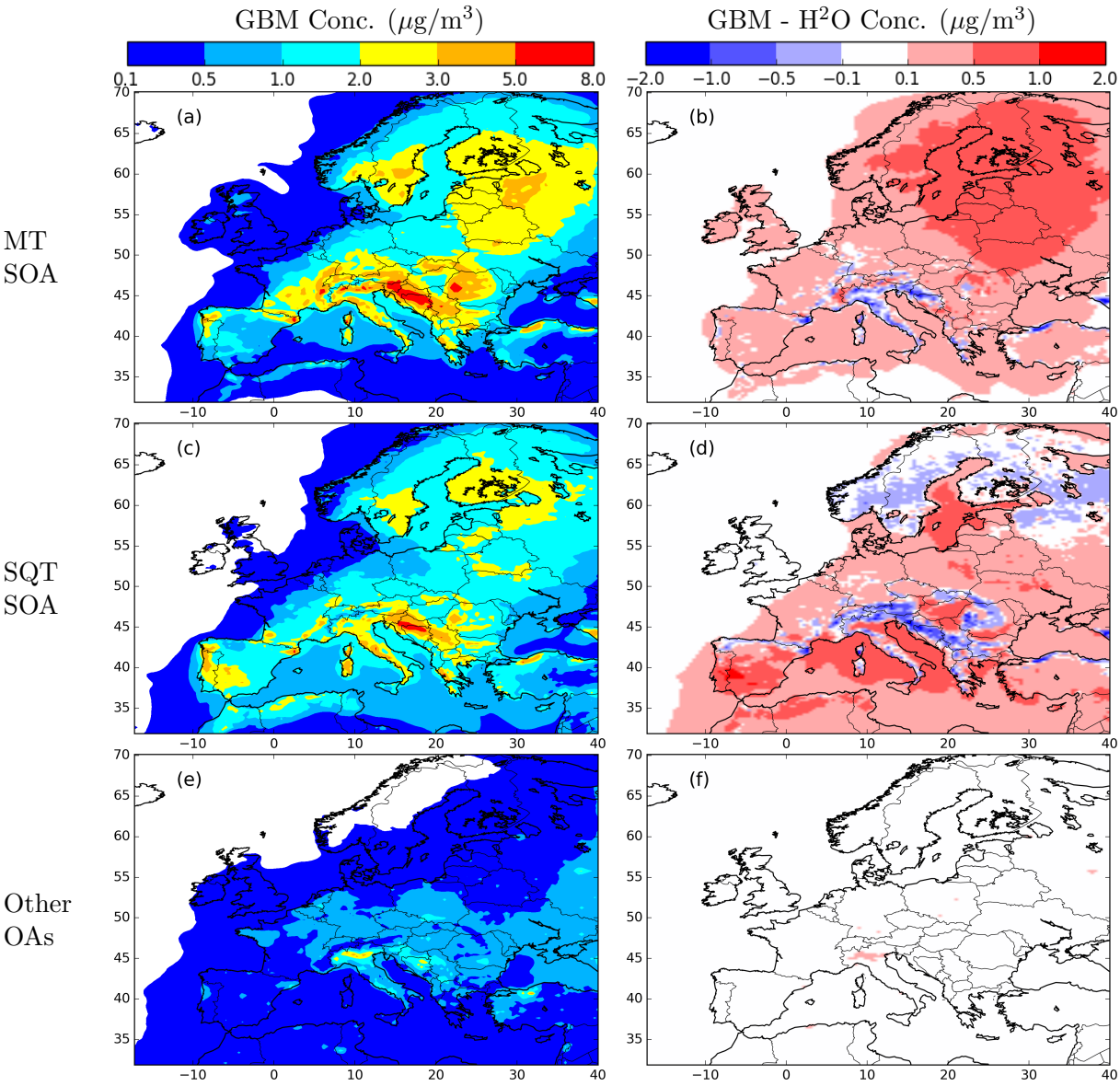


Figure 4.8: Maps of SOA concentrations derived from monoterpene, sesquiterpene, and other SOA precursors simulated with the GBM (left panels) and differences with the H^2O mechanism (right panels).

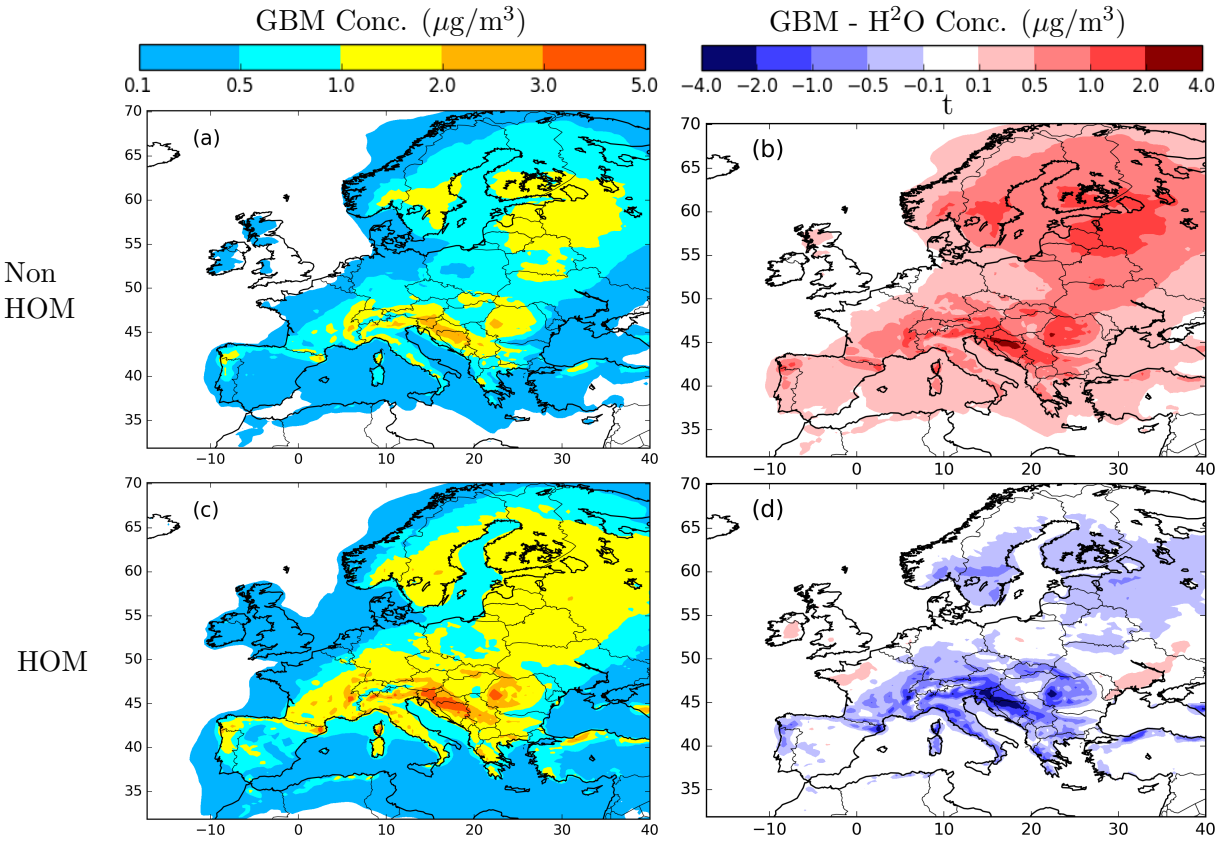
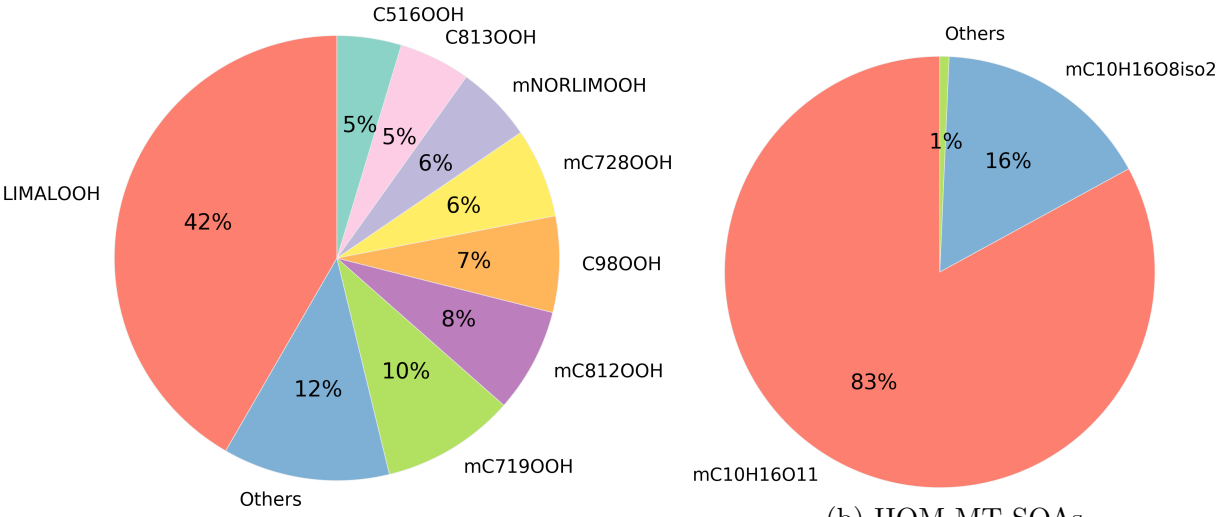


Figure 4.9: Average concentrations of non-HOM (top panels) and HOM MT SOAs (bottom panels) simulated with the GBM mechanism (left panels) and concentration differences with the H^2O mechanism (right panels) during June-August 2018.

O/C ratio), for the purpose of the comparison, we only considered as HOM compounds formed by auto-oxidation and RO_2 dimerization in both mechanisms (species originating from PRAM in GBM; Monomer and Dimer species in H^2O).

Figure 4.9 presents the concentration distributions of non-HOM and HOM MT-SOAs simulated with GBM, as well as the absolute concentration differences between results simulated with GBM and H^2O. Overall, HOMs dominate the total MT SOAs, accounting for 62 % (0.60 $\mu g/m^3$ on average) and 93 % (0.68 $\mu g/m^3$) of MT SOAs simulated with GBM and H^2O, respectively. High HOM concentrations up to 4.4 $\mu g/m^3$ and 7.7 $\mu g/m^3$ and high non-HOM concentrations up to 3.2 $\mu g/m^3$ and 1.3 $\mu g/m^3$ are found in central Europe simulated with GBM and H^2O, respectively.



(a) non-HOM MT SOAs

(b) HOM MT SOAs

Figure 4.10: Composition of non-HOM (a) and HOM (b) MT SOAs simulated with the GBM mechanism during June-August 2018. The fraction "Others" represents the sum of SOA species with a concentration contribution of less than 5 %.

As mentioned previously in Sect. 2.2, both GBM and H^2O involve the formation of non-HOMs and HOMs from MTs, with the GBM mechanism containing more details than the H^2O mechanism involving multi-generation oxidation and a more complex dependency on inorganic radicals.

Figure 4.10 shows the compositions of non-HOM and HOM MT SOAs simulated with GBM. As shown in Fig. 4.10a, species greatly contribute to non-HOM SOAs formed under low NOx conditions (compounds formed from RO_2 reacting with HO_2). Among them, the first-generation oxidant products derived from ozonolysis of LIM, pLIMALOOH, contribute to 42 % of non-MT SOAs, indicating LIM has higher SOA yields than API and BPI. Other non-HOM species having contributions above 5 % shown in Fig. 4.10a are all second-generation or higher products. Together these second-generation or higher species account for 59 % of total non-HOM SOAs, showing that a significant fraction of non-HOM MT SOAs is due to SOA aging. As reaction pathways of multi-generation oxidation reflecting SOA aging for non-HOM species are absent in H^2O, the non-HOM concentrations simulated with H^2O are much lower than those simulated with GBM. These results are consistent with the higher OM:OC ratio with GBM shown in Fig. 4.7 indicating more oxidized SOA.

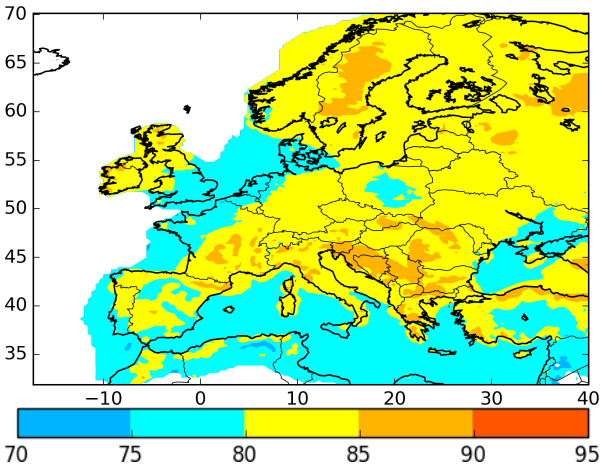


Figure 4.11: Distribution of mC10H16O11 concentration in HOM MT SOA concentrations (%).

As shown in Fig. 4.10b, out of a total of five HOM species in GBM, mC10H16O11 contributes 82.9 % of HOM SOAs, followed by mC10H16O8iso2 contributing 16.4 %. Both mC10H16O11 and mC10H16O8iso2 are terminal species resulting from multiple auto-oxidation steps. The two dimer species resulting from RO_2 dimerization and the remaining HOM species account for 0.7 % of the total HOMs, indicating that auto-oxidation is the major HOM formation pathway. The map distribution of the dominant HOM species, mC10H16O11, is shown in Fig. 4.11. In Eastern Europe, mC10H16O11 can account for up to 90.8 % of HOMs at places observed with high HOM concentrations. Compared to GBM, the highly simplified mechanism of H^2O resulted in high HOM SOA concentrations. A total of 46 % of MT SOAs simulated with H^2O are pMonomer, 47 % are pDimer, 6 % are pBiNITs (SOA formed by MT + NO_3), and the remaining species (including BiA0D, Bi1D, BiA2D, BiA3D) counts for 1 %.

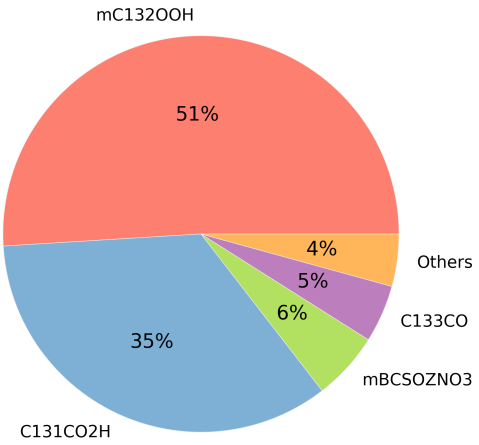


Figure 4.12: Composition of SQT SOAs simulated with the GBM mechanism.

As for SQT SOA concentrations, GBM adopts 8 condensables involved in 30 reactions (shown in Fig. 4.4) to describe the SOA formation and aging process. Figure 4.12

presents the composition of SQT SOAs simulated with GBM. It shows that two condensables mC132OOH (51 %) and C131CO2H (35 %) dominate the SQT SOA concentrations, followed by mBCSOZNO3 accounting for 6 % and C133CO for 5 %. mC131CO2H is the main first-generation oxidation product from the ozonolysis of sesquiterpenes, while pBCSOZNO3 is the first-generation product from the OH-initiated reaction and inorganic reaction with NO. mC132OOH is a lump species representing third and all higher-generation oxidation products. As this last species is the dominant SQT-SOA species, these results suggest that higher-generation products from SQT oxidation make significant contributions to atmospheric SOA formation.

Figure 4.12 shows the map distribution of C131CO2H and mC132OOH concentrations. It is shown that high C131CO2H contributes more SQT SOAs in central Europe, while mC132OOH is produced more in northern Europe. As shown in Eq. 4.1, H^2O adopts only a single oxidation step and two SOA species to simulate the SQT SOA formation process. BiBmP (SVOC compounds from sesquiterpene with medium volatility) represents 48 % of SQT-SOA and BiBlP (SVOC compounds from sesquiterpene with low volatility) is the remaining fraction. Due to the single oxidation in H^2O, the mechanism may be unable to track SOA aging contrary to GBM.

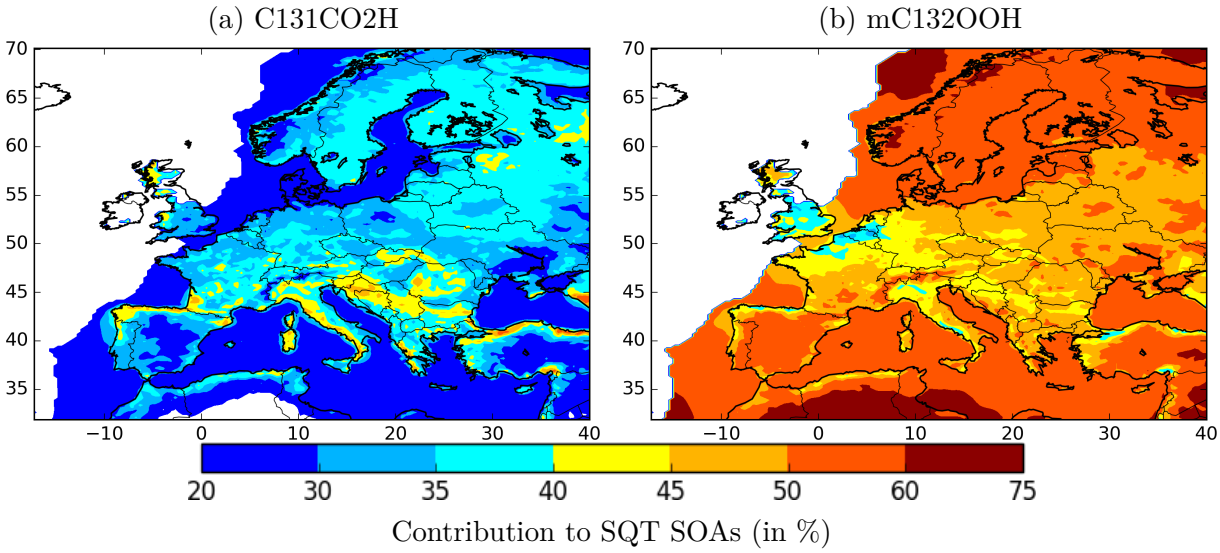


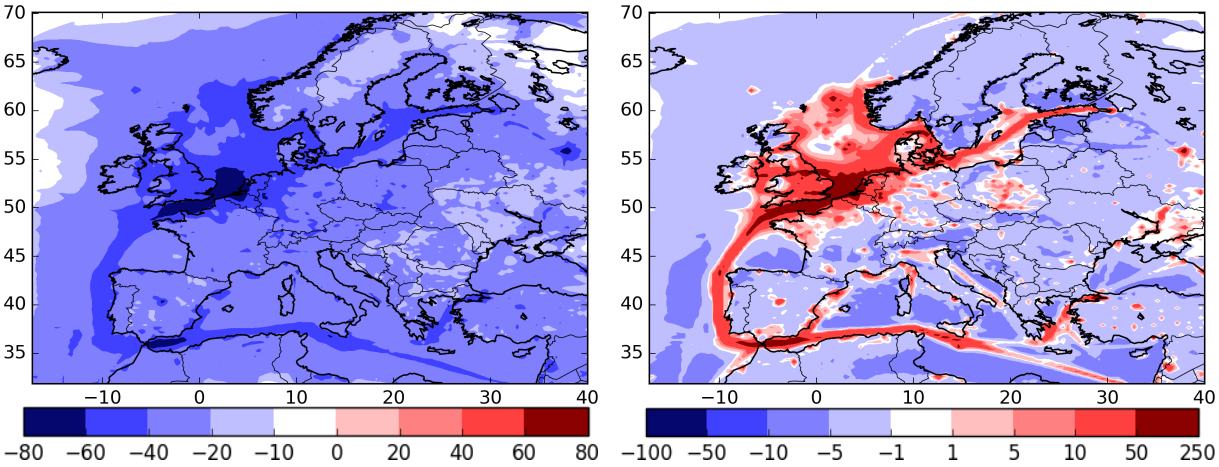
Figure 4.13: Distribution of dominant SQT SOA species in total SQT SOA concentrations simulated with the GBM mechanism. Their contributions to SQT SOA can be found in the GBM SQT scheme in Fig. 4.4.

3.3 Response of biogenic SOA concentrations to NOx emission reduction

Since SOA formation involves multigenerational reactions and multiphase transformations, its potential reaction to anthropogenic reduction is highly nonlinear. NOx (i,e, NO, NO_2, HONO) profoundly affects SOA concentration and composition, either by participating in the oxidation pathways or by forming secondary inorganic aerosols that may interact with the gas-particle partitioning of organic compounds. In this section, both

the GBM and H^2O mechanisms are used to evaluate changes in SOA concentrations due to NOx emission reduction. An anthropogenic NOx emission reduction of 50 % over the simulation domain is considered. The conditions before and after NOx emission reduction are referred to as the "REF" and "NOx50" scenarios, respectively. The simulations are conducted in June-August 2018 over Europe.

3.3.1 Effect of NOx reduction on concentrations of oxidants and radicals



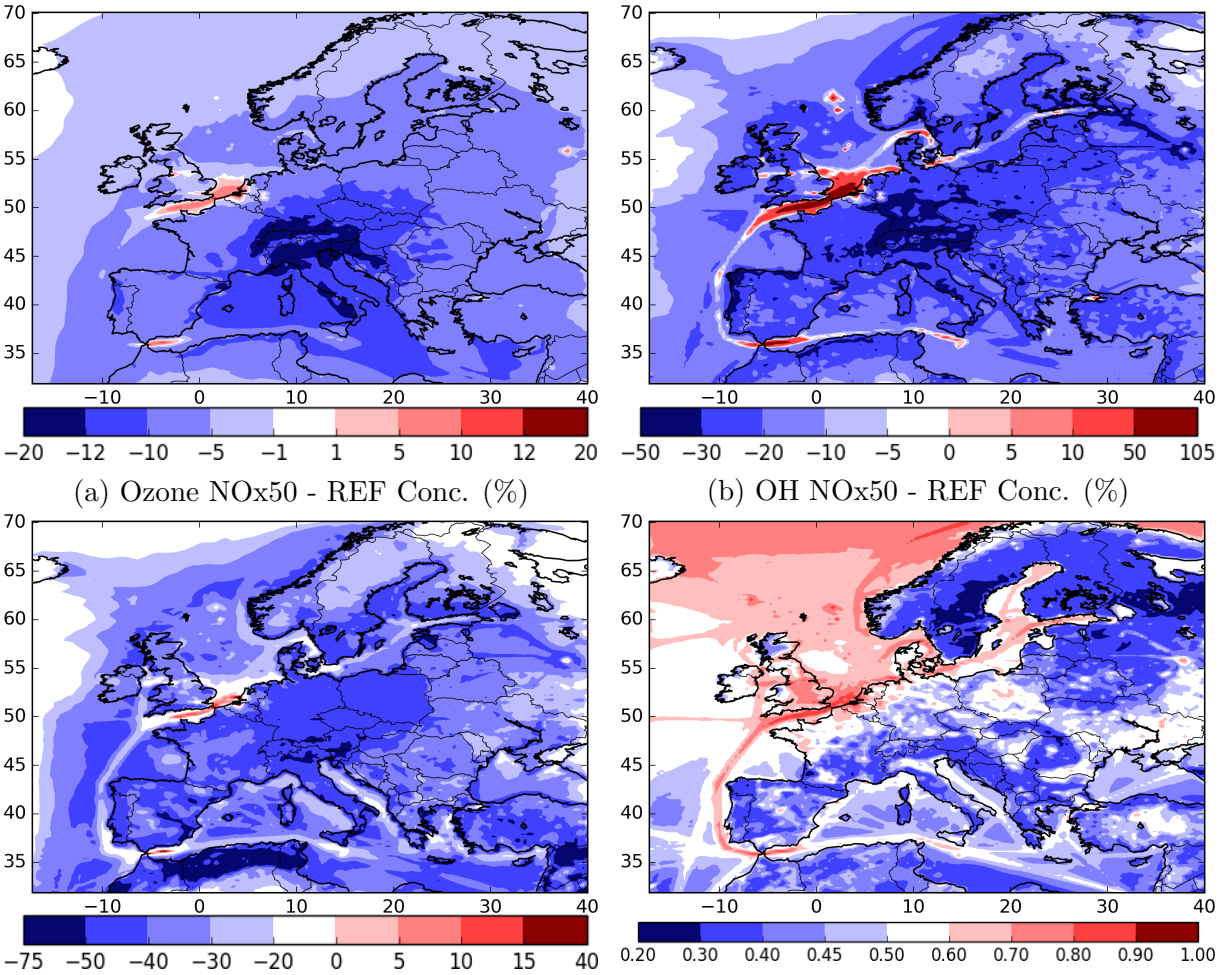
(a) NO NOx50 - REF Conc. (%)

(b) HO2 NOx50 - REF Conc. (%)

Figure 4.14: Relative differences in NO (a) and HO2 (b) concentrations between the NOx50 and REF scenarios. Simulations are performed with the GBM mechanism.

To comprehend the effect of reducing NOx emissions on the formation of secondary organic aerosols (SOAs), it is crucial to investigate how the reduction of NOx affects the levels of oxidants and radicals involved in the formation of SOAs. The map of Fig. 4.14a shows the distribution of relative differences in NO concentrations under the NOx and REF scenarios. In response to a 50 % reduction in anthropogenic NOx emissions, the simulated NO concentrations are reduced between 2 % and 80 % (30 % on average) compared to the REF scenario. Similar decreases are also noted for NO_2 concentration changes (ranging from 3 % to 58 % with an average of 31 %), due to emission reduction and the interdependence between NO and NO_2.

Figure 4.15a, b, and c show the average relative changes in oxidant concentrations (i.e., O_3, OH, and NO_3) resulting from NOx reduction simulated with the GBM mechanism. In most areas of Europe, decreases in concentrations were simulated for all oxidants but to different extents, with an average reduction ratio of -6 % (up to 18 %) for O_3, -34 % (up to 72 %) for NO_3, and -17 % (up to 41 %) for OH. However, oxidant concentrations decrease less or even increase in some areas with high NOx emissions (along the shipping routes in the Baltic Sea, North Sea, and English Channel, as well as in a few urban areas near Moscow). The highest relative increases are simulated up to 19 % for O_3, 104 % for OH, and 38 % for NO_3 in the English Channel near Dunkirk. These results may be due to a reduction of O_3 by the titration reaction in these regions, or to a reduction in the rate of HNO_3 production due to the termination reaction of NO_2 with OH. After NOx



(a) Ozone NOx50 - REF Conc. (%)

(b) OH NOx50 - REF Conc. (%)

(c) NO_3 NOx50 - REF Conc. (%)

(d) Chemcial regime ratio in NOx50 [a]

Figure 4.15: Relative differences in oxidant concentration, i.e., Ozone (a), OH radical (b), and NO_3 radical (c) between the NOx50 and REF scenarios, as well as the chemical regime ratio of NOx50 scenario. Simulations are performed using the GBM mechanism.
[a] The parameterization described in the CHIMERE documentation ([Menut, 2021]) is utilized to compute the chemical regime ratio, which calculates the ratio of the reaction rate of RO2 radicals with NO (high-NOx regime) to the sum of reaction rates of the reactions with HO2 and RO2 (low-NOx regime).

emission reduction, these areas remain under the high-NOx regime (see Fig. 4.15d), while most places in mainland Europe are under the low-NOx regime.

As explained by [Sillman, 1999], under the low-NOx regime, NOx reduction limits O_3 formation by NO_2 photolysis, leading to a decrease in O_3, NO_3, OH, and HO_2, as shown in the mainland Europe due to NOx reduction. Conversely, under the high-NOx regime, a decrease in NOx concentrations results in an antagonistic effect that increases O_3 concentrations, as simulated in places near Moscow, the English Channel, and the Strait of Gibraltar. Accompanying changes in Ozone concentrations were increases in NO_3, OH, and HO_2 under the high-NOx regime and decreases under the low-NOx regime. Moreover, for HO_2, concentrations decrease slightly under the low-NOx regime (≤ 8.7 %) and increase strongly under the high-NOx regime (up to 239 %), resulting in a slight increase (0.4 %) on average.

Similar changes in radical and oxidant concentrations were observed for those simulated with the H^2O mechanism, as the H^2O mechanism shares common inorganic reactions and VOC degradation rates with the GBM mechanism.

3.3.2 Comparison of total organic aerosols and OC:OM ratio

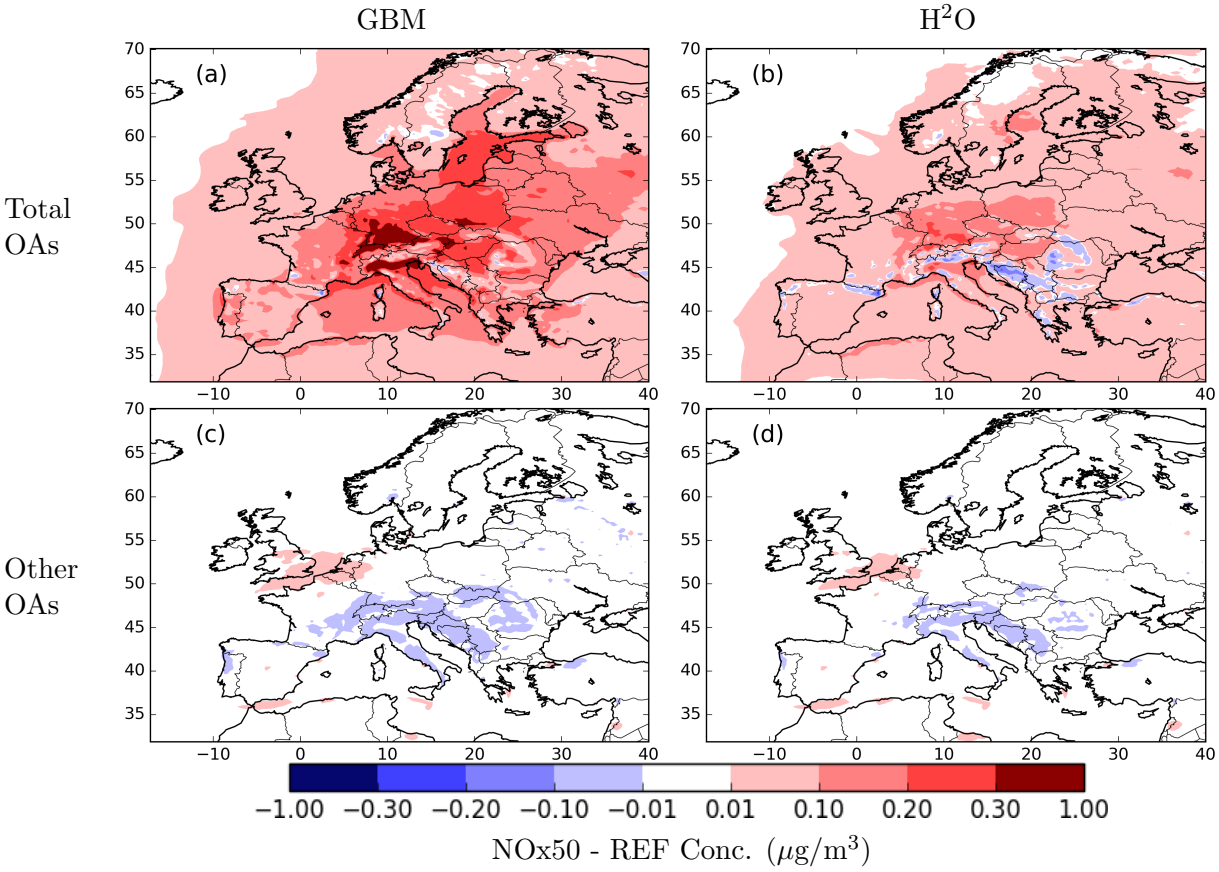


Figure 4.16: Absolute concentration differences in total OA (top panels) and OAs besides MT and SQT SOAs (noted as "Other OAs", shown in the bottom panels) between NOx50 and REF scenarios simulated with the GBM (left panels) and H^2O (right panels) mechanisms.

The concentration differences in total SOAs as well as other OAs besides MT and SQT SOAs due to NOx emission reduction are presented in Fig. 4.16. As a response to NOx emissions, simulations using the detailed SOA mechanism suggest an overall increase in organic aerosol concentrations (up to 0.68 $\mu g/m^3$ in central Europe) with some decreases in areas with high biogenic emissions (down to -0.31 $\mu g/m^3$), whereas the simulations with the implicit H^2O mechanism estimated a lower increase over mainland Europe (up to 0.27 $\mu g/m^3$) with more decrease in central Europe (down to -0.31 $\mu g/m^3$). On average, the total OA concentrations increase by 0.08 $\mu g/m^3$ and by 0.04 $\mu g/m^3$ with the GBM mechanism and H^2O mechanism, respectively. These increases are consistent with the changes in the sum of MT and SQT SOAs as a result of NOx reduction. For other OAs presented in Figs. 4.16c and 4.16d, simulations with both mechanisms estimate slight increases under the high-NOx regime and decreases under the low-NOx regime, correspondent to the changes in oxidant concentrations, particularly O_3 and OH.

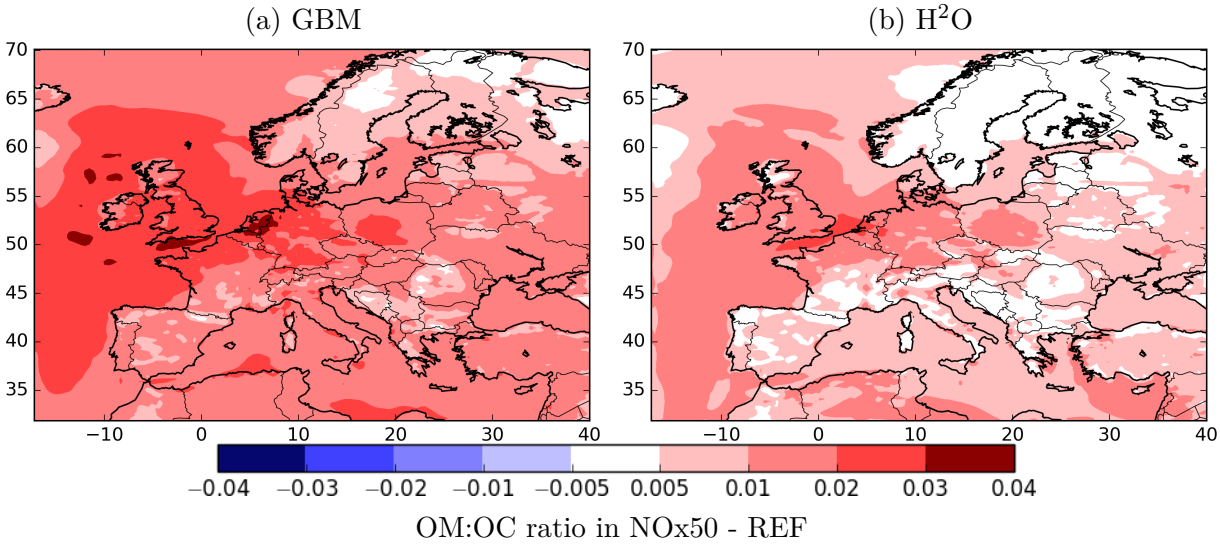


Figure 4.17: Absolute differences in organic mass to organic carbon (OM:OC) ratios between NOx50 and REF scenarios simulated with the GBM (a) and H^2O (b) mechanisms.

Figure 4.17 illustrates the organic mass to organic carbon (OM:OC) ratio changes due to NOx reduction simulated with the GBM and H^2O mechanisms. As a response to NOx emission reduction, increases in OM:OC ratios are simulated with the GBM and H^2O mechanisms, with an average increase of the ratios by 0.02 and 0.01, respectively. This evolution suggests that slightly more oxidized SOAs are formed when NOx emissions are reduced, probably as a result of enhanced HOM formation of MT SOAs. Compared to those simulated with the H^2O mechanism, OM:OC ratios simulated with the GBM mechanism show larger increases (up to 0.032 for GBM and 0.026 for H^2O) due to more oxidized SOAs produced from auto-oxidation. Some reactions with RO_2 and HO_2 may also contribute to the increase in OM:OC ratios.

3.3.3 Comparison of MT SOAs

The impacts of NOx emission reductions on the average concentration of MT SOAs simulated using two mechanisms, GBM and H^2O, were investigated. Figure 4.18 illustrates

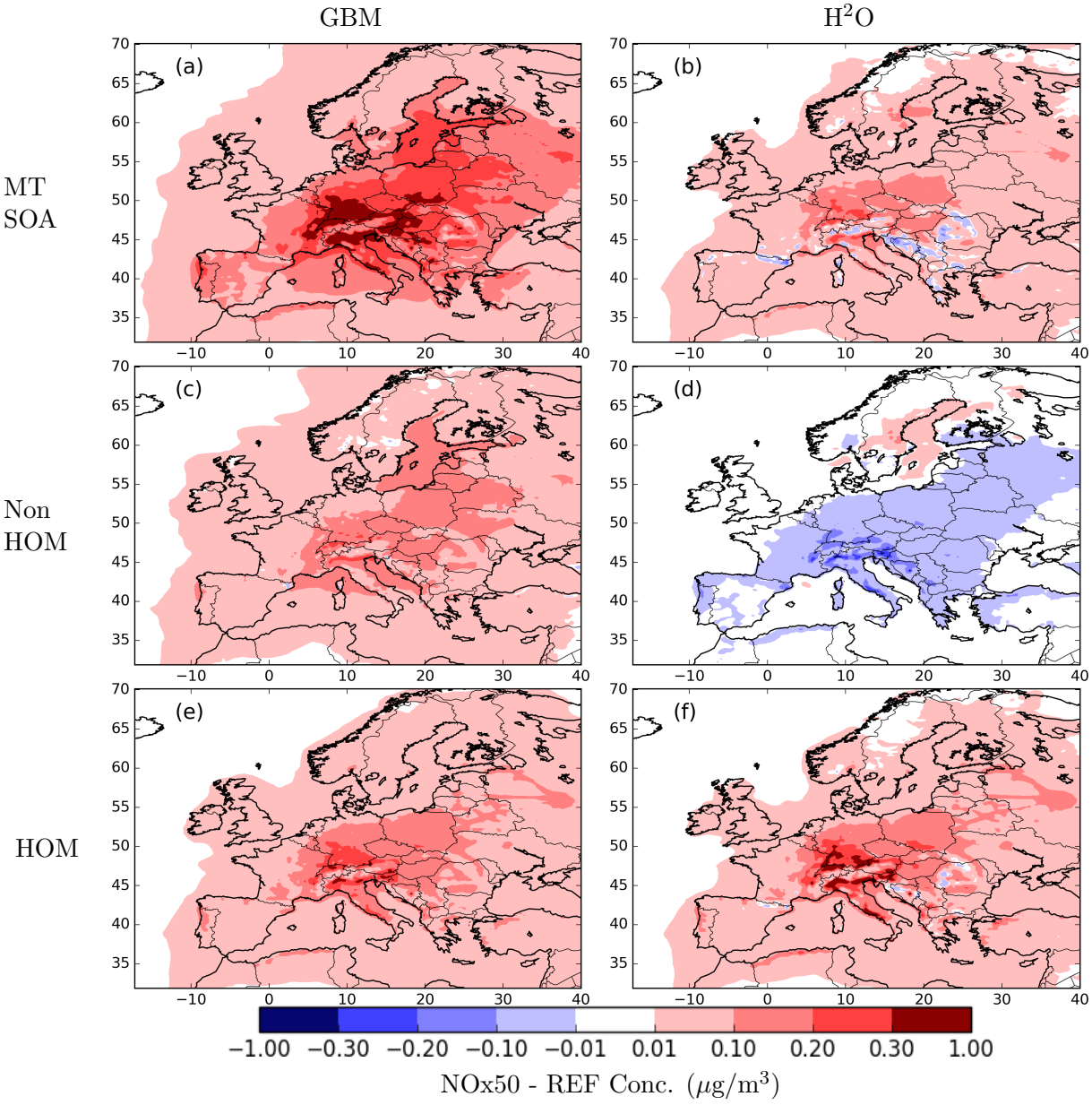


Figure 4.18: Absolute differences in MT SOA concentrations (top), HOM MT SOA (middle), and non-HOM MT SOA (bottom) between NOx50 and REF scenarios, simulated with the GBM (left panels) and H^2O (right panels) mechanisms.

the differences in total MT SOA, MT non-HOMs, and MT concentrations between the NOx50 and REF scenarios. Notably, the simulated response of MT SOA to NOx reduction differs between the two mechanisms in response due to different responses of the SOA mechanisms to the changes in oxidants and radical concentrations.

On average over the simulated domain, both mechanisms predict an increase in MT SOAs due to NOx emission reduction, but the GBM mechanism predicts a larger increase. The average increase with GBM is around 0.09 $\mu g/m^3$, corresponding to a 15 % relative difference, while the increase with H^2O is around 0.03 $\mu g/m^3$ (10 %). When separating the total MT SOAs into non-HOM and HOM concentrations, it is observed that the GBM mechanism estimates increases of SOAs in both non-HOM (0.04 $\mu g/m^3$ corresponding to 24 % relative difference) and HOM (0.046 $\mu g/m^3$ or 10 %) species as a response to NOx emission reduction. In contrast, the H^2O mechanism estimates an increase for HOM species (0.043 $\mu g/m^3$ or 10 %) but a decrease for non-HOM MT species (-0.01 $\mu g/m^3$ or -2 %). It is noted that for HOM MT concentrations, the simulations of the two mechanisms estimate a similar amount of increases on average, while higher peaks of MT HOM concentrations are found with H^2O at 0.72 $\mu g/m^3$ against 0.46 $\mu g/m^3$ simulated with GBM.

The differences in the SOA mechanisms can explain the changes in SOA concentrations. For non-HOM formation, the GBM mechanism preserves multi-generation oxidation and the reactions of oxidized products (i.e., RO_2 species) with NO, HO_2, or another RO_2 species. Therefore, when anthropogenic NOx emission is reduced, RO_2 radicals form during the oxidation and will react preferentially with RO_2 and HO_2 in most places of Europe as lower concentrations of NO are simulated. Depending on the chemical pathways that are favored, a decrease in NOx may either promote or impede the formation of less volatile semi-volatile organic compounds (SVOCs). The reaction between RO_2 and NO results in the production of more volatile products such as aldehydes, ketones, and organic nitrates in MT SOA formation ([Presto, 2005; Ng, 2007]), in contrast to the RO_2 + HO_2 reactions that produce less volatile hydroxyperoxides. Therefore, a decrease in NOx may lead to an increase in MT SOAs. Both [Xavier, 2019] and [Yu, 2021] reported findings, demonstrating a negative correlation between the concentration of monoterpene SOAs and NOx levels modeled with the MCM + PRAM mechanism. Furthermore, since NOx reduction leads to a greater decrease in NO_3 concentrations (34 %) than O_3 (6 %) and OH (17 %), there may be a reduction in the initial oxidation of precursors that results from decreasing NOx.

The implicit H^2O mechanism contains only a single oxidation step (as shown in Fig. 4.2) which is only influenced by the degradation of precursors with the different oxidants and does not account for the reaction of RO_2 with the other radicals. Therefore, the changes in non-HOM concentrations simulated with H^2O directly reflect the changes in oxidant concentrations (i.e., O_3, OH, NO_3), presenting a decrease in mainland Europe while a small increase is simulated covering the Baltic Sea (at high-NOx regime due to shipping pollution).

For HOM formation, the GBM mechanism preserves auto-oxidation and RO_2 reactions with NO, HO_2, and RO_2 that partially favor the termination of the auto-oxidation process. The RO_2 reactions with HO_2 and RO_2 are enhanced as a result of NOx emission reduction, which further encourages auto-oxidation and leads to an increase in HOM concentration.

In the H^2O mechanism, the RO_2 - RO_2 reactions and influences of NO and HO_2 on

HOM formation are also preserved but with a very simplified mechanism. As illustrated in Fig. 4.3, in the H^2O mechanism, NO does not directly participate in the HOM formation, it only contributes to the destruction of tRO2 species. On the contrary, HO_2 not only contributes to the destruction of tRO2 but also directly participates in the HOM formation in the H^2O mechanism, i.e., the formation of Monomer and Dimer. Therefore, when reactions with HO_2 and other RO_2 species are favorable due to NOx reduction, considerably more HOMs are produced with H^2O.

The MT SOA compositions simulated with the GBM mechanism under the NOx50 scenario are presented in Fig. 4.23. No significant composition differences are observed compared to the results under the REF scenario. Since most MT SOAs can be formed from multiple reactions with NO, HO_2, and RO_2, changes in oxidants and NOx do not result in notable changes in MT SOA composition. Overall, the results indicate that the detailed GBM mechanism can preserve more complexity in the gas-phase chemistry of MT SOA formation, particularly the pathways involving auto-oxidation, as well as RO_2 reactions with HO_2 and RO_2. Those are critical for understanding the pathways and factors affecting SOA concentrations in response to NOx emission reductions.

3.3.4 Comparison of SQT SOAs

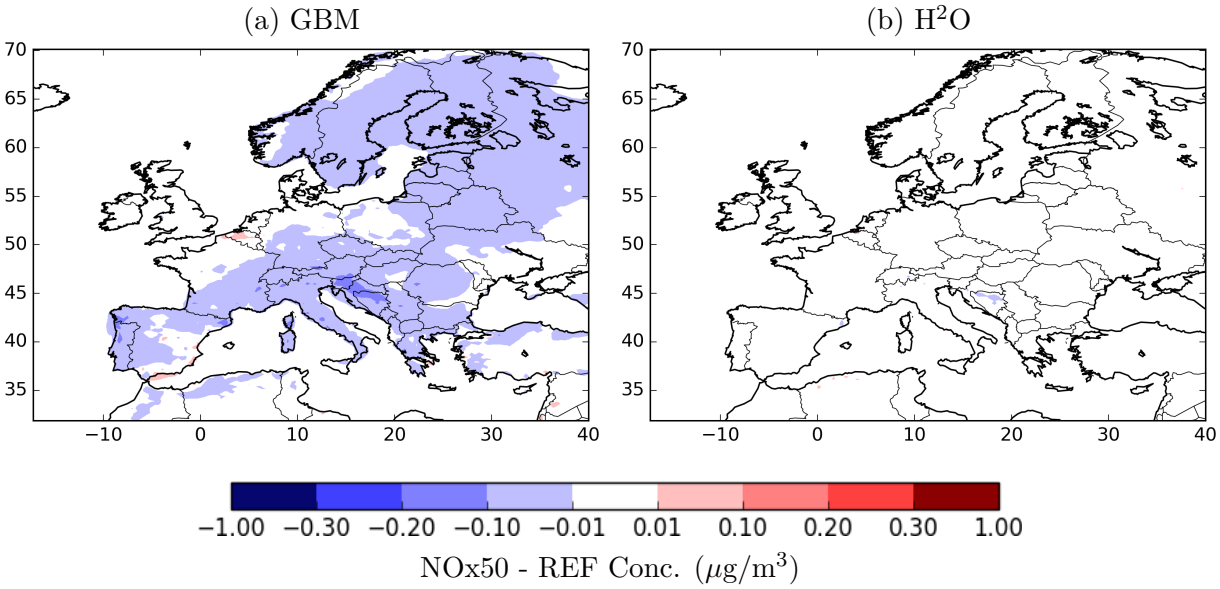


Figure 4.19: Effect on NOx reduction by 50% on SQT SOA concentrations simulated with the GBM (a) and H^2O (b) mechanisms.

The concentration differences in SQT SOAs between the NOx50 and REF scenarios simulated with the GBM and H^2O mechanism are also compared. Figure 4.19 presents the simulation results for SQT SOA, where slight decreases (up to 0.22 $\mu g/m^3$) in SQT SOA concentrations over mainland Europe as a response to NOx emission reduction are simulated with the GBM mechanism and no significant change (inferior to 0.022 $\mu g/m^3$) is simulated with the H^2O mechanism. The changes in the total SQT SOA concentration simulated with the GBM mechanism can be explained by a compensation between increased SOAs from $RO_2 + HO_2$ reactions (i.e., C132OOH) and decreased SOAs from RO_2

+ NO reactions (i.e., mBCSOZNO3, mBCKSOZ, mC133NO3, and C133CO). The RO_2 + RO_2 reaction in the GBM SQT SOA scheme does not lead to noticeable SOA formation ($\leq 1\%$) and therefore does not contribute to concentration changes. Figure 4.20 presents the SQT SOA composition simulated with the GBM mechanism under the NOx50 scenario. Notably, the composition of SQT SOA species and their distribution in total SQT SOAs is highly akin to that simulated under the REF scenario (See Fig. 4.12). Compared to the result simulated under the REF scenario shown in Fig. 4.12, the contribution of mC132OOH in SQT-SOA is increased from 51 % to 53 %. The contribution of C133CO in SQT-SOA is decreased from 5 % to 4 %. This confirms that SQT SOAs from RO_2 + HO_2 reactions increase in response to reduced NOx emissions. In the case of simulations with the H^2O mechanism, NOx emissions are not expected to affect the SQT SOA concentrations, since the H^2O mechanism adopts a highly simplified SQT scheme shown in Eq. 4.1, disregarding the dependence of the NOx regime on SQT SOA formation.

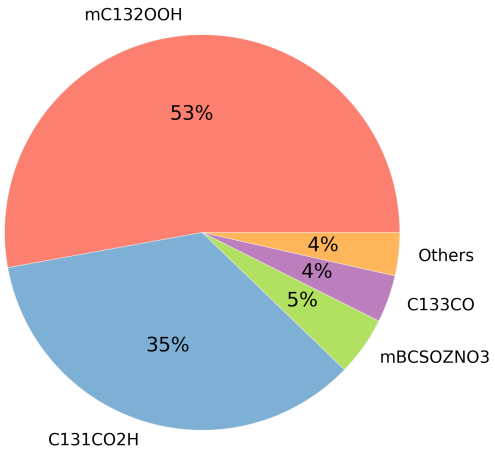


Figure 4.20: Composition of SQT SOAs simulated with the GBM mechanism and NOx50 emission reduction scenario during June-August 2018.

4 Conclusion

In this study, the impacts of NOx emission reduction on SOA formation are investigated using the three-dimensional (3-D) chemical transport model CHIMERE model and a semi-explicit GENOA-generated biogenic SOA mechanism (GBM). The GBM mechanism was generated by the GENerator of reduced Organic Aerosol mechanisms (GENOA) version 2.0 and contains detailed degradation schemes of monoterpene (MT) and sesquiterpene (SQT) accounting SOA formation mechanisms reduced from the Master Chemical Mechanism (MCM) for both MT and SQT and Peroxy Radical Autoxidation Mechanism (PRAM) for MT. Aerosol concentrations and compositions are simulated over Europe for the 2018 summer (June-August) using the GBM mechanism and the implicit Hydrophobic/Hydrophilic Organic mechanism (H^2O) for comparison.

The results simulated with the two mechanisms (GBM and H^2O) and different estimations of biogenic emissions (regular biogenic emission estimated with MEGAN v2.1,

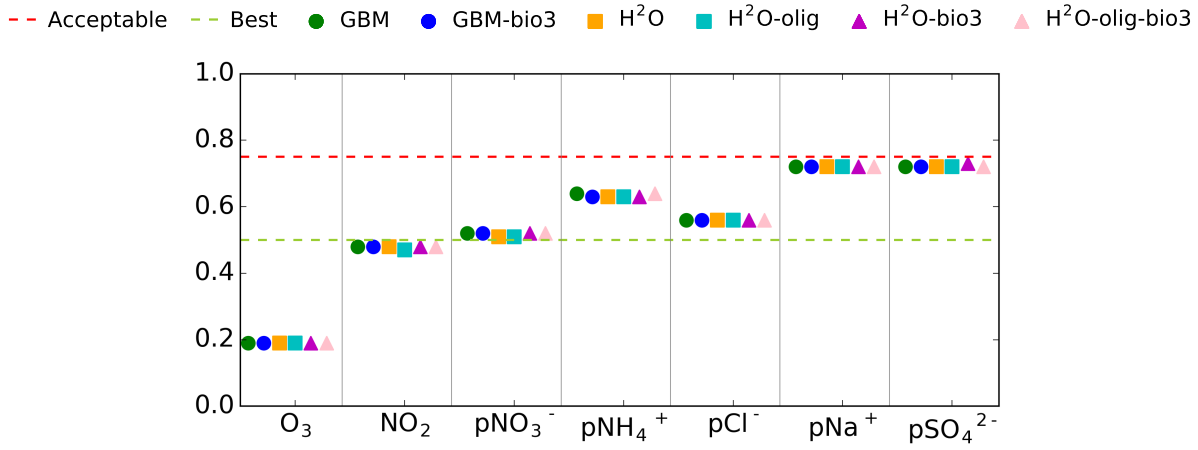
and the modified biogenic emissions with one-third of isoprene and triple of MT and SQT of the regular emission to account for the potential errors in biogenic emissions suggested by the literature review) are first compared to the measurements. The results showed that the 3-D simulations with regular MEGAN biogenic emissions tend to underestimate aerosol concentrations compared to the measurements. However, after modifying biogenic emissions, the organic concentrations estimated with both mechanisms meet the model performance standard reported by [Boylan, 2006]. This suggests that biogenic VOC emissions may have a major impact on SOA formation and it may be necessary to improve the calculation of biogenic emissions over Europe and to reduce the uncertainties. Using the GBM mechanism leads to an enhancement in the model's performance than that simulated with the H^2O mechanism in terms of correlation, bias, and error when compared to the measurements. This feature emphasizes the need to implement more detailed SOA mechanisms in 3D air quality models in order to improve the performance in simulating SOA concentration.

The simulations performed with both the GBM and H^2O mechanisms showed that biogenic SOA concentrations, mainly the MT and SQT SOAs, dominate the OA concentrations, with high SOA concentrations observed in central Europe. This is because the GBM considers more detailed pathways regarding the SOA formation and aging with multi-generation oxidation than H^2O, including the formation of HOMs via auto-oxidation. Accounting for those high-generation oxidation products simulated with GBM also leads to higher OM:OC ratios than those simulated with the H^2O. The results indicate that the products formed over several generations of oxidation estimated with GBM contribute significantly to SOA formation.

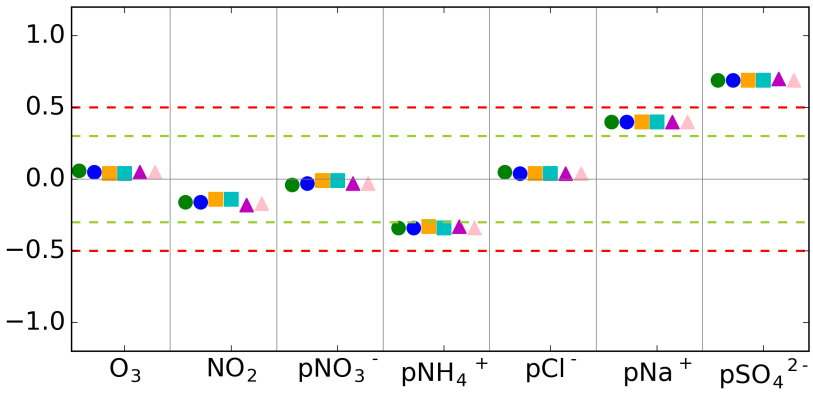
Finally, the study investigated the influences of a Nitrogen oxides (NOx) emission reduction scenario on biogenic SOA formation with semi-explicit and implicit SOA mechanisms. Both mechanisms indicate that SOA concentrations should increase in concentration and becomes more oxidized when reducing NOx emissions. With the semi-explicit GBM mechanism, simulated MT non-HOM and HOM concentrations increase as a response to NOx reduction. The H^2O mechanism also shows raised aerosol concentrations due to MT HOM concentration growths.

In conclusion, this study demonstrates the importance of using detailed SOA mechanisms (such as the semi-explicit mechanisms generated with the GENOA mechanisms) in order to preserve the dependency of SOA formation on environmental conditions and accurately determine the impact of emission reduction scenarios. The current regulation assessments may be insufficient due to simplified SOA mechanisms, which may miss essential pathways in SOA modeling. Therefore, detailed semi-explicit SOA mechanisms are potentially valuable to evaluate the models and their responses, taking into account the different formation pathways, the nonlinear effects caused by the interaction of oxidation products, and the dependency on environmental conditions. Overall, our findings provide insights into the complexity of gas-phase chemistry on SOA formation and the importance of incorporating semi-explicit SOA mechanisms in CTMs for accurate predictions of air quality and informed policy decisions.

Appendix: Additional information related to 3-D simulation



(a) MFE



(b) MFB

Figure 4.21: MFE (a) and MFB (b) between simulations and measurements on different inorganic pollutants. From left to right, measurements of O_3, NO_2, particulate nitrate, ammonium, chloride, sodium, and sulfate are compared. The numbers of stations and measurements compared with are listed in Table 4.3.

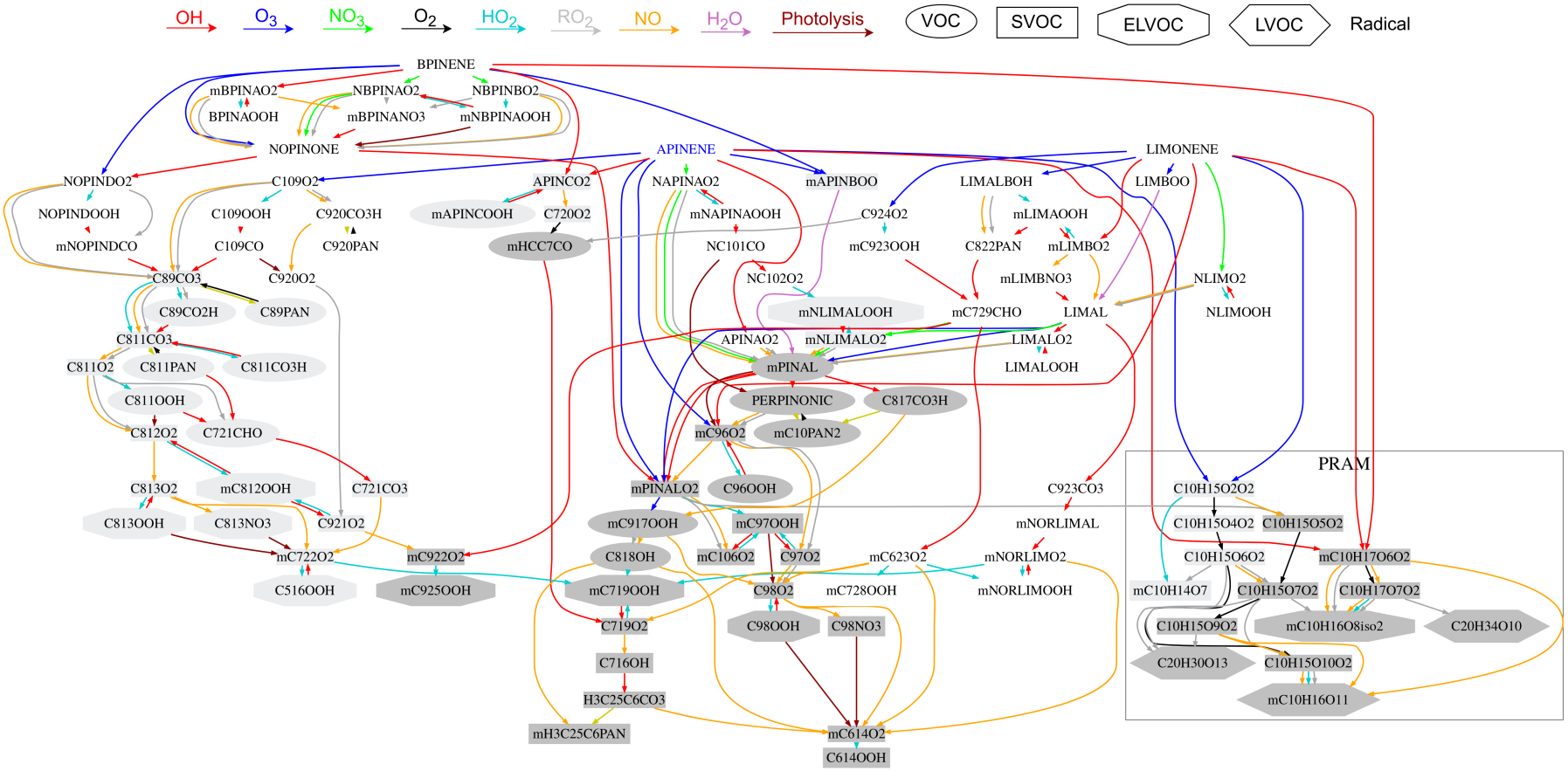


Figure 4.22: Reaction pathways that lead to monoterpene SOA formation in GBM mechanism. The reactions are shown by lines, where different colors indicate different types of reactions. The oxidation reactions of VOCs with various oxidants are displayed in different colors: red for Ozone, blue for OH radical, and green for NO_3 radical. Reactions with NO, HO_2, RO_2 pool, H_2O, O_2, and photolysis are depicted in orange, cyan, gray, brown, black, and orchid colors, respectively. The species are illustrated with distinct shadow colors: no color, light gray, and deep gray, representing the number of SOA precursors it can be derived from (one, two, or three precursors). Additionally, species with light and dark gray shadows are distinguished by different shapes, including no shape, ellipse, box, hexagon, and octagon, which signify their properties, such as radicals, VOCs, semi-volatile organic compounds (SVOCs with P_{sat} lower than 10^{-9} atm), low-volatile organic compounds (LVOCs with P_{sat} between 10^{-9} atm and 10^{-13} atm), and ELVOCs with P_{sat} lower than 10^{-13} atm at 298 K, respectively. The same legends are applied to Fig. 4.1.

Table 4.4: Monoterpene aerosol properties related to the GBM and H^2O mechanisms. [a]

Name	MW	P_{sat}	Henry's	γ_{inf}	H_{vap}	Precursors
GBM mechanism						
mNAPINAOOH	231.2	7.5×10^{-8}	1.5×10^{4}	4.9×10^{4}	95.2	API
mNBPINAOOH	231.2	3.6×10^{-8}	4.0×10^{4}	3.9×10^{4}	98.3	BPI
BPINAOOH	186.2	4.6×10^{-8}	2.5×10^{5}	4.8×10^{3}	97.4	BPI
mC97OOH	190.6	2.1×10^{-8}	1.7×10^{7}	1.6×10^{2}	101.0	API, BPI, LIM
mNLIMALOOH	262.4	1.9×10^{-10}	6.9×10^{7}	4.2×10^{3}	121.8	API, LIM
mC719OOH	178.8	7.5×10^{-10}	1.1×10^{10}	6.7	117.5	API, BPI, LIM
C716OH	158.2	1.2×10^{-6}	1.3×10^{7}	3.6	83.1	API, BPI, LIM
LIMALOOH	218.2	1.5×10^{-10}	2.7×10^{9}	1.4×10^{2}	123.6	LIM
mNORLIMOOH	202.6	4.9×10^{-10}	2.0×10^{9}	54.8	118.3	LIM
mC728OOH	171.8	1.3×10^{-10}	6.2×10^{10}	7.1	126.0	LIM
mC812OOH	194.8	3.2×10^{-11}	4.2×10^{10}	40.9	132.1	API, BPI
C98OOH	204.2	9.2×10^{-10}	1.5×10^{9}	39.4	114.6	API, BPI, LIM
C98NO3	233.2	8.0×10^{-9}	3.4×10^{7}	2.0×10^{2}	104.7	API, BPI, LIM
mH3C25C6PAN	238.5	1.4×10^{-8}	5.9×10^{8}	6.8	103.2	API, BPI, LIM
mC925OOH	217.5	8.6×10^{-13}	1.5×10^{13}	4.4	147.7	API, BPI, LIM
C614OOH	162.1	9.1×10^{-9}	2.5×10^{9}	2.5	104.6	API, BPI, LIM
C813OOH	206.2	8.8×10^{-13}	7.0×10^{12}	9.0	148.4	API, BPI
C813NO3	235.2	1.6×10^{-11}	1.1×10^{11}	30.8	134.8	API, BPI
C516OOH	164.1	7.1×10^{-12}	1.1×10^{13}	7.2×10^{-1}	139.1	API, BPI
mC10H16O11	318.3	2.6×10^{-16}	1.9×10^{16}	11.0	95.1	API, BPI, LIM
mC10H14O7	234.0	1.2×10^{-9}	4.2×10^{8}	1.1×10^{2}	68.5	API, LIM
C20H30O13	478.0	2.0×10^{-20}	5.8×10^{17}	4.9×10^{3}	105.5	API, BPI, LIM
mC10H16O8iso2	266.4	1.8×10^{-12}	2.1×10^{11}	1.5×10^{2}	84.6	API, BPI, LIM
C20H34O10	434.0	3.1×10^{-16}	3.3×10^{12}	5.4×10^{4}	103.9	API, BPI, LIM
H^2O mechanism						
BiA0D	168.0	3.6×10^{-7}	1.4×10^{3}	2.0×10^{6}	50.0	API, BPI, LIM
BiA1D	170.0	2.9×10^{-10}	1.1×10^{8}	126.5	50.0	API, BPI, LIM
BiA2D	186.0	1.9×10^{-10}	2.7×10^{8}	37.0	109.0	API, BPI, LIM
BiA3D	204.0	4.3×10^{-10}	ELVOC[b]	ELVOC	109.0	API, BPI, LIM
BiNIT	215.0	3.3×10^{-9}	HPHO[c]	HPHO	50.0	API, BPI, LIM
Monomer	278.0	1.3×10^{-17}	HPHO	HPHO	50.0	API, BPI, LIM
Dimer	432.0	1.3×10^{-17}	HPHO	HPHO	50.0	API, BPI, LIM

[a] Columns from left to right are the name of the condensable species, molar mass (MW in mol/g), the saturation vapor pressure at 298 K (P_{sat} in atm), Henry's law constant (Henry's in mol L^{-1} atm^{-1}), the activity coefficient at infinite dilution in water (γ_{inf}), the enthalpy of vaporization (H_{vap} in kJ/mol), and the corresponding SOA precursors (API: α-pinene, BPI: β-pinene, and LIM: limonene).

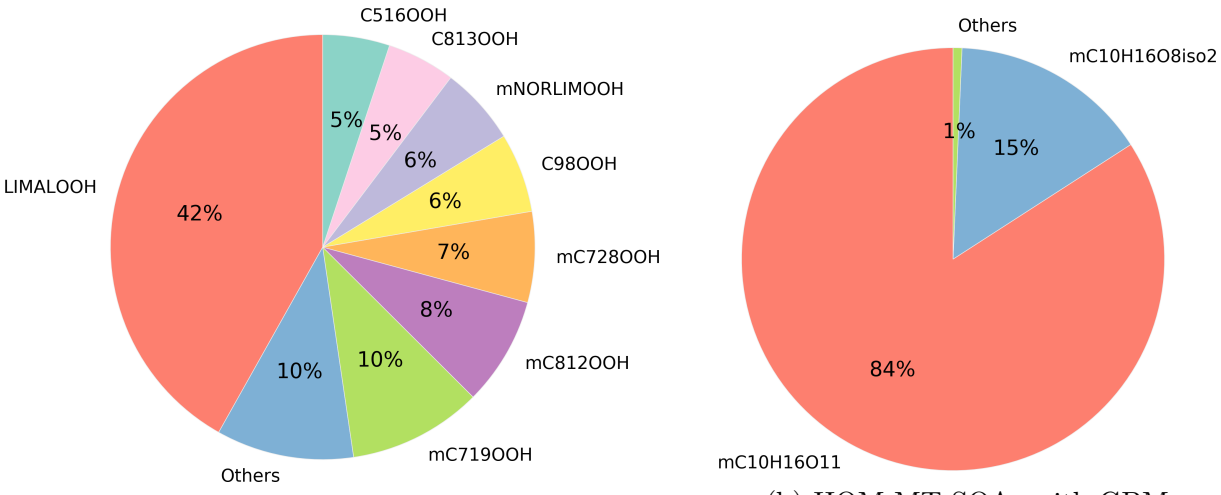
[b] BiA3D is considered as a non volatile species.

[c] HPHO: Hydrophilic species (i.e., BiNIT, Monomer, and Dimer) condense only on organic phases in the H^2O mechanism, so Henry's and γ_{inf} are not computed for those species.

Name	MW	P_{sat}	Henry's	γ	H_{vap}
GBM mechanism					
NBCOOH	299.4	4.0×10^{-12}	1.8×10^{6}	7.6×10^{6}	118.6
C131CO2H	254.3	5.8×10^{-12}	5.0×10^{9}	1.9×10^{3}	120.3
mBCKSOZ	252.8	2.0×10^{-8}	1.4×10^{4}	1.9×10^{5}	91.1
C136PAN	299.3	1.8×10^{-9}	1.2×10^{5}	2.6×10^{5}	101.0
mC132OOH	280.3	4.9×10^{-14}	1.9×10^{11}	6.0×10^{3}	135.3
mBCSOZNO3	330.0	8.5×10^{-13}	9.8×10^{7}	6.7×10^{5}	123.9
mC133NO3	302.5	1.8×10^{-14}	2.4×10^{12}	1.3×10^{3}	139.2
C133CO	256.3	6.6×10^{-13}	1.3×10^{12}	64.0	128.1
H^2O mechanism					
BiBmP	236.0	3.9×10^{-10}	HPHO	HPHO	175.0
BiBlP	298.0	7.9×10^{-13}	HPHO	HPHO	175.0

Table 4.5: Sesquiterpene SOA properties related to the GBM and H^2O mechanisms. The explanation for "HPHO" sees 4.4.

Columns from left to right are the name of the condensable species, molar mass (MW in mol/g), the saturation vapor pressure at 298 K (P_{sat} in atm), Henry's law constant (Henry's in mol L^{-1} atm^{-1}), the activity coefficient at infinite dilution in water, and the enthalpy of vaporization (H_{vap} in kJ/mol).



(a) non-HOM MT with GBM

(b) HOM MT SOAs with GBM

Figure 4.23: Composition of non-HOM (a) and HOM (b) MT SOAs simulated with the GBM mechanism and NOx50 emission reduction scenario during June-August 2018.

Conclusions and perspectives

1 Conclusions

This thesis presents a comprehensive investigation of the formation and evolution of secondary organic aerosol (SOA) in regional-scale Chemical Transport Models (CTMs), with a particular focus on developing effective and accurate mechanisms for simulating SOA formation. The work can be divided into two main parts: the development and application of the GENerator of reduced organic Aerosol Mechanisms (GENOA), a mechanism reduction algorithm that produces semi-explicit mechanisms for simulating SOA formation and aging; and the impact investigation of reducing the emission of anthropogenic nitrogen oxides (NOx) on biogenic aerosol formation using the CTM model CHIMERE and the detailed biogenic SOA mechanisms generated by GENOA.

In the first part of the work, detailed in chapters 2 and 3, the GENOA algorithm is developed and applied to reduce mechanisms describing SOA formation from different SOA precursors. The obtained semi-explicit mechanisms preserve the accuracy of near-explicit chemical mechanisms for SOA formation to an extent demanded by the user while requiring significantly fewer reactions and species, which are more computationally efficient than explicit mechanisms and suitable for large-scale air quality modeling. The first version of GENOA (GENOA v1.0) described in chapter 2, is shown to be effective in reducing mechanisms describing sesquiterpene (SQT) from the Master Chemical Mechanism (MCM), resulting in a reduced mechanism that is 2 % of the size of the original mechanism, while inducing an average error of 2.7 % in a testing process under 12 159 conditions over Europe. To further improve the reduction efficiency and allow the simultaneous reduction of mechanisms from several SOA precursors, a second version of the GENOA algorithm (GENOA v2.0) is developed. As presented in chapter 3, GENOA v2.0 adopts a parallel reduction structure with updates on the reduction strategy and the evaluation method. When applied to reduce monoterpene (MT) SOA mechanisms from MCM and the Peroxy Radical Autoxidation Mechanism (PRAM), GENOA v2.0 results in a reduced MT SOA mechanism with 197 reactions and 110 species (23 of them being condensables) that is about 7 % of the size of the original mechanism, while inducing an average error of less than 3 % under 9 818 testing conditions over Europe.

In the second part of the work, detailed in chapter 4, the impacts of anthropogenic emission reduction on biogenic SOA formation are investigated using 3-D CHIMERE over Europe for the 2018 summer. Simulations are conducted using the semi-explicit GENOA-generated biogenic SOA mechanism (GBM), containing MT and SQT SOA schemes trained by GENOA v2.0, compared to those simulated with the implicit Hy-

drophobic/Hydrophilic Organic mechanism (H^2O). The results show that the use of a semi-explicit SOA mechanism leads to a significant improvement in model performance compared to measurements. The biogenic aerosols simulated with GBM are more oxidized than those simulated with H^2O as a result of GBM taking into account SOA aging due to multi-generational oxidation.

Reducing anthropogenic NOx emissions by 50 % was estimated to increase MT SOA yields using both GBM and H^2O mechanisms, with a greater increase estimated using GBM. As a result of the absence of NO under the low-NOx regime, peroxy radicals (RO_2) from MT oxidation react preferentially with other RO_2 and HO_2, forming condensables with lower volatility that increase aerosol concentration. The formation of highly oxygenated molecules (HOMs) of MT SOAs from auto-oxidation is also favorable in response to NOx reduction. These results underscore the importance of considering the complex interactions between anthropogenic NOx emissions and biogenic organic aerosol formation and highlight the need for implementing semi-explicit SOA mechanisms to improve SOA predictions in 3-D air quality modeling.

In summary, the work presented in this thesis demonstrates the effectiveness of the GENOA algorithm in producing semi-explicit mechanisms for simulating SOA formation and aging, emphasizing the importance of incorporating such detailed mechanisms in regional-scale CTM modeling. The research suggests that incorporating semi-explicit SOA mechanisms can help preserve more SOA variations in 3-D air quality modeling, depending on the physical-chemical environment, leading to better predictions of air quality and assessment of the impact of emission reduction scenarios. These insights can inform policymakers in making informed decisions concerning air quality management.

2 Perspectives

The GENOA algorithm, developed in this study, has the potential to enhance the accuracy of 3-D air quality models by improving the representation of SOA formation processes. This novel algorithm, moreover, holds great potential for advancing our understanding of the explicit VOC mechanisms that underlie SOA formation across different scales, including urban and regional contexts. Further application and development can be applied to GENOA.

2.1 Model development

Moving forward, there are several avenues for improving the accuracy and efficiency of the GENOA algorithm in reducing complex atmospheric chemical mechanisms.

To enhance the accuracy of the GENOA algorithm, improvements can be made in the selection of evaluation datasets, evaluation criteria, and selection of reduced mechanisms. Currently, the evaluation datasets used for training the reduction are selected randomly or manually from 3-D simulation results. The selection can be time-consuming and may not cover all conditions effective on SOA formation. To address this, an effective method for obtaining representative conditions is needed to ensure the representability of environmental conditions. For example,[Porter, 2021] quantified atmospheric parameter ranges for ambient secondary organic aerosol formation, which can be employed to build evaluation datasets. In addition, cluster analysis may also be helpful in sorting and categorizing

representative environmental conditions for SOA formation.

The current evaluation criteria are based on a reduction error calculated from the total SOA concentration simulated with the 0-D aerosol model. As a result, the current GENOA does not preserve the concentration variations of other compounds besides SOA (e.g., organic peroxy radicals, oxidants, and inorganics) on SOA formation. To improve the evaluation, an "error matrix" can be introduced, where the performance of reduction candidates can be estimated based on the variations of multiple parameters (e.g., concentrations of crucial organic condensables and radicals) on SOA formation. Additionally, the updated error matrix may allow GENOA to extend to other atmospheric pollutants besides SOAs.

To further improve the accuracy of mechanism reduction, an alternative strategy is to exhaustively explore all potential late-stage reductions and identify those with both high accuracy and small size. This can be achieved through the implementation of a "reduction tree" framework, in which each node represents a reduced mechanism with a potential reduction candidate and branches indicate different reduction paths. By employing this approach, the future GENOA will be able to systematically traverse all possible late-stage reductions, thereby minimizing the impact of reduction parameters (e.g., user-defined parameters) on the training results. This strategy, however, may have significant computational requirements and should be reduced to some extent by combining it with other approaches.

In addition to improving reduction accuracy, enhancing reduction efficiency is essential for applying the GENOA algorithm to more complex mechanisms, particularly in its application to full explicit mechanisms with millions of reactions and species. Current versions of GENOA can be applied to near-explicit MCM mechanisms or mechanisms with similar levels of complexity, but reductions for full explicit mechanisms such as GECKO-A may take months or longer. To improve the efficiency of GENOA, the reduction structure can be reorganized to group reduction candidates and accept multiple reduction attempts at a time, reducing the total number of reduction steps required while preserving the combination effects ($1+1 > 2$) of reductions. For example, instead of testing many reduction candidates and accepting one at a time (as currently applied to GENOA v2.0), an advanced evaluation system can be developed, which groups all those reduction candidates into several sets, and then directly investigate the performance of the group reductions. The evaluation system may be developed using machine learning algorithms, which can automate the selection of reduction candidates and evaluate the impact of reductions on the overall mechanism. This could potentially speed up the reduction process and improve the accuracy of the final reduced mechanism.

Overall, those further developments could help to expand the applicability and utility of the GENOA algorithm in advancing our understanding of atmospheric chemistry and air quality.

2.2 Model application

GENOA is a powerful and flexible tool for generating reduced organic aerosol mechanisms with a balance between accuracy and size. The mechanisms produced by GENOA can be implemented in various air quality models for different case studies related to SOA formation. In-depth investigations using detailed GENOA-generated SOA mechanisms can

provide valuable insights into the spatial and temporal variations of SOA concentration and composition, which are often challenging to capture with implicit mechanisms.

In addition to NOx, other inorganic emissions such as SO_2 and NH_3 have been reported to impact SOA formation ([Xu, 2021b]), and their effects can be explored using GENOA-generated SOA mechanisms. Another possible application of GENOA is to generate semi-explicit SOA mechanisms for anthropogenic precursors, including alkanes and aromatics (e.g., toluene and xylene). Detailed SOA mechanisms for these compounds are expected to provide a better understanding of urban air pollution dynamics. Likewise, the semi-explicit SOA mechanism of isoprene degradation can also be investigated with GENOA.

Moreover, GENOA-generated SOA mechanisms can be used to investigate the interactions in multi-precursor systems, where the presence of one VOC may affect the SOA formation process of another VOC as reported by [McFiggans, 2019]. By considering all main SOA precursors, both anthropogenic and biogenic, the semi-explicit SOA mechanism can contribute to exploring how different emissions interact and influence SOA production. Furthermore, GENOA-generated SOA mechanisms may have further applicability in predicting SOA pollution from various environmental events, including shipping emissions, wildlife, and industrial accidents. Such investigations may enhance our comprehension of the impacts of SOA formation on human health and the environment.

As the capabilities of the GENOA model continue to develop, it may become feasible to apply it to fully explicit chemical mechanisms such as GECKO-A. This would enable a comparison between the semi-explicit SOA mechanisms generated from different explicit chemical mechanisms, including those reduced from MCM and GECKO-A. Additionally, if the inorganic and SOA concentrations can be preserved, GENOA could also produce a single, semi-explicit SOA mechanism from all VOC precursors, preserving the complexity of the gas-phase chemistry involved in both SOA formation and inorganic reactions.

References

[André, 2020] M. André, K. Sartelet, S. Moukhtar, J.M. André, and M. Redaelli. "Diesel, petrol or electric vehicles: What choices to improve urban air quality in the Ile-de-France region? A simulation platform and case study". *Atmos. Environ.* 241 (2020), p. 117752 (cit. on pp. 3, 105).

[Atkinson, 2000] Roger Atkinson. "Atmospheric chemistry of VOCs and NOx". *Atmos. Environ.* 34.12-14 (2000), pp. 2063–2101 (cit. on pp. 5, 61).

[Atkinson, 2003] Roger Atkinson and Janet Arey. "Atmospheric Degradation of Volatile Organic Compounds". *Chem. Rev.* 103.12 (2003), pp. 4605–4638 (cit. on pp. 5, 14, 60).

[Aumont, 2005] B. Aumont, S. Szopa, and S. Madronich. "Modelling the evolution of organic carbon during its gas-phase tropospheric oxidation: development of an explicit model based on a self generating approach". *Atmos. Chem. Phys.* 5.9 (2005), pp. 2497–2517 (cit. on pp. 15, 23, 61).

[Bianchi, 2019] Federico Bianchi, Theo Kurtén, Matthieu Riva, Claudia Mohr, Matti P. Rissanen, Pontus Roldin, et al. "Highly Oxygenated Organic Molecules (HOM) from Gas-Phase Autoxidation Involving Peroxy Radicals: A Key Contributor to Atmospheric Aerosol". *Chem. Rev.* 119.6 (2019), pp. 3472–3509 (cit. on pp. 16, 61).

[Boylan, 2006] James W. Boylan and Armistead G. Russell. "PM and light extinction model performance metrics, goals, and criteria for three-dimensional air quality models". *Atmos. Environ.* 40.26 (2006), pp. 4946–4959 (cit. on pp. 112, 115, 132).

[Breysse, 2013] Patrick N Breysse, Ralph J Delfino, Francesca Dominici, Alison CP Elder, Mark W Frampton, John R Froines, et al. "US EPA particulate matter research centers: summary of research results for 2005–2011". *Air Qual. Atmos. Health* 6.2 (2013), pp. 333–355 (cit. on pp. 11, 104).

[Camredon, 2007] M Camredon, B Aumont, J Lee-Taylor, and S Madronich. "The SOA/VOC/NO x system: an explicit model of secondary organic aerosol formation". *Atmos. Chem. Phys.* 7.21 (2007), pp. 5599–5610 (cit. on p. 61).

[Carter, 2010] William P.L. Carter. "Development of the SAPRC-07 chemical mechanism". *Atmos. Environ.* 44.40 (2010), pp. 5324–5335 (cit. on p. 22).

[Carter, 2020] WPL Carter. "The SAPRC-18 Atmospheric Chemical Mechanism". *web site at http://www.cert. ucr. edu/˜ carter/SAPRC/18* (2020) (cit. on p. 23).

[Chen, 2022] Tianzeng Chen, Peng Zhang, Biwu Chu, Qingxin Ma, Yanli Ge, Jun Liu, et al. "Secondary organic aerosol formation from mixed volatile organic compounds: Effect of RO2 chemistry and precursor concentration". *npj Clim. Atmos. Sci.* 5.1 (2022), pp. 1–8 (cit. on p. 61).

[Chrit, 2017] M. Chrit, K. Sartelet, J. Sciare, J. Pey, N. Marchand, F. Couvidat, et al. "Modelling organic aerosol concentrations and properties during ChArMEx summer campaigns of 2012 and 2013 in the western Mediterranean region". *Atmos. Chem. Phys.* 17 (2017), pp. 12509–12531 (cit. on pp. 106, 118).

[Ciccioli, 2023] Piero Ciccioli, Camillo Silibello, Sandro Finardi, Nicola Pepe, Paolo Ciccioli, Francesca Rapparini, et al. "The potential impact of biogenic volatile organic compounds (BVOCs) from terrestrial vegetation on a Mediterranean area using two different emission models". *Agric. For. Meteor.* 328 (2023), p. 109255 (cit. on pp. 112, 114).

[Coggon, 2019] M. M. Coggon, C. Y. Lim, A. R. Koss, K. Sekimoto, B. Yuan, J. B. Gilman, et al. "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". *Atmos. Chem. Phys.* 19.23 (2019), pp. 14875–14899 (cit. on p. 61).

[Couvidat, 2013] F. Couvidat, Y. Kim, K. Sartelet, C. Seigneur, N. Marchand, and J. Sciare. "Modeling secondary organic aerosol in an urban area: application to Paris, France". *Atmos. Chem. Phys.* 13.2 (2013), pp. 983–996 (cit. on p. 104).

[Couvidat, 2018a] F. Couvidat, M. G. Vivanco, and B. Bessagnet. "Simulating secondary organic aerosol from anthropogenic and biogenic precursors: comparison to outdoor chamber experiments, effect of oligomerization on SOA formation and reactive uptake of aldehydes". *Atmos. Chem. Phys.* 18.21 (2018), pp. 15743–15766 (cit. on p. 111).

[Couvidat, 2018b] Florian Couvidat, Bertrand Bessagnet, Marta Garcia-Vivanco, Elsa Real, Laurent Menut, and Augustin Colette. "Development of an inorganic and organic aerosol model (CHIMERE 2017β v1. 0): seasonal and spatial evaluation over Europe". *Geosci. Model Dev* 11.1 (2018), pp. 165–194 (cit. on pp. 112, 113).

[Couvidat, 2012] Florian Couvidat, Edouard Debry, Karine Sartelet, and Christian Seigneur. "A hydrophilic/hydrophobic organic (H2O) aerosol model: Development, evaluation and sensitivity analysis". *J. Geophys. Res.-Atmos.* 117.D10 (2012) (cit. on pp. 24, 26, 61, 105–108).

[Couvidat, 2015] Florian Couvidat and Karine Sartelet. "The Secondary Organic Aerosol Processor (SOAP v1. 0) model: a unified model with different ranges of complexity based on the molecular surrogate approach". *Geosci. Model Dev* 8.4 (2015), pp. 1111–1138 (cit. on pp. 18, 26, 28, 107).

[Donahue, 2011] N. M. Donahue, S. A. Epstein, S. N. Pandis, and A. L. Robinson. "A two-dimensional volatility basis set: 1. organic-aerosol mixing thermodynamics". *Atmospheric Chemistry and Physics* 11.7 (2011), pp. 3303–3318 (cit. on pp. 24, 61).

[Donahue, 2006] N. M. Donahue, A. L. Robinson, C. O. Stanier, and S. N. Pandis. "Coupled Partitioning, Dilution, and Chemical Aging of Semivolatile Organics". *Env. Sc. and Tech.* 40.8 (2006), pp. 2635–2643 (cit. on pp. 7, 15, 24, 61, 104, 105).

[Ehn, 2014] Mikael Ehn, Joel A Thornton, Einhard Kleist, Mikko Sipilä, Heikki Junninen, Iida Pullinen, et al. "A large source of low-volatility secondary organic aerosol". *Nature* 506.7489 (2014), pp. 476–479 (cit. on pp. 61, 106).

[Flentje, 2021] H. Flentje, I. Mattis, Z. Kipling, S. Rémy, and W. Thomas. "Evaluation of ECMWF IFS-AER (CAMS) operational forecasts during cycle 41r1–46r1 with calibrated ceilometer profiles over Germany". *Geosci. Model Dev* 14.3 (2021), pp. 1721–1751 (cit. on p. 112).

[Fry, 2009] J. L. Fry, A. Kiendler-Scharr, A. W. Rollins, P. J. Wooldridge, S. S. Brown, H. Fuchs, et al. "Organic nitrate and secondary organic aerosol yield from NO_3 oxidation of β-pinene evaluated using a gas-phase kinetics/aerosol partitioning model". *Atmos. Chem. Phys.* 9.4 (2009), pp. 1431–1449 (cit. on p. 61).

[Gelencsér, 2007] András Gelencsér, Barbara May, David Simpson, Asunción Sánchez-Ochoa, Anne Kasper-Giebl, Hans Puxbaum, et al. "Source apportionment of PM2.5 organic aerosol over Europe: Primary/secondary, natural/anthropogenic, and fossil/biogenic origin". *J. Geophys. Res.-Atmos.* 112.D23 (2007) (cit. on p. 60).

[Goldstein, 2007] Allen H Goldstein and Ian E Galbally. "Known and unexplored organic constituents in the earth's atmosphere". *Env. Sc. and Tech.* 41.5 (2007), pp. 1514–1521 (cit. on p. 60).

[Goliff, 2013] W.S. Goliff, W.R. Stockwell, and C.V. Lawson. "The Regional Atmospheric Chemistry Mechanism, version 2". *Atmos. Environ.* 68 (2013), pp. 174–185 (cit. on pp. 22, 28, 64).

[Goshua, 2022] Anna Goshua, Cezmi A Akdis, and Kari C Nadeau. "World Health Organization global air quality guideline recommendations: Executive summary". *Allergy* 77.7 (2022), pp. 1955–1960 (cit. on pp. 11–13).

[Granier, 2019] C. Granier, S. Darras, H. Denier van der Gon, J. Doubalova, N. Elguindi, B. Galle, et al. "The Copernicus Atmosphere Monitoring Service global and regional emissions (April 2019 version)" (2019) (cit. on p. 12).

[Griffin, 2003] Robert J. Griffin, Khoi Nguyen, Donald Dabdub, and John H. Seinfeld. "A Coupled Hydrophobic-Hydrophilic Model for Predicting Secondary Organic Aerosol Formation". *J. Atmos. Chem.* 44.2 (2003), pp. 171–190 (cit. on p. 61).

[Guenther, 2012] A. B. Guenther, X. Jiang, C. L. Heald, T. Sakulyanontvittaya, T. Duhl, L. K. Emmons, et al. "The Model of Emissions of Gases and Aerosols from Nature version 2.1 (MEGAN2.1): an extended and updated framework for modeling biogenic emissions". *Geosci. Model Dev* 5.6 (2012), pp. 1471–1492 (cit. on pp. 60, 73, 104, 112).

[Hallquist, 2009] M. Hallquist, J. C. Wenger, U. Baltensperger, Y. Rudich, D. Simpson, M. Claeys, et al. "The formation, properties and impact of secondary organic aerosol: current and emerging issues". *Atmos. Chem. Phys.* 9.14 (2009), pp. 5155–5236 (cit. on pp. 60, 61, 104).

[Han, 2023] S. Han and M. Jang. "Modeling daytime and nighttime secondary organic aerosol formation via multiphase reactions of biogenic hydrocarbons". *Atmos. Chem. Phys.* 23.2 (2023), pp. 1209–1226 (cit. on p. 61).

[Hodzic, 2016] A. Hodzic, P. S. Kasibhatla, D. S. Jo, C. D. Cappa, J. L. Jimenez, S. Madronich, et al. "Rethinking the global secondary organic aerosol (SOA) budget: stronger production, faster removal, shorter lifetime". *Atmos. Chem. Phys.* 16.12 (2016), pp. 7917–7941 (cit. on p. 104).

[Hong, 2022] Youwei Hong, Xinbei Xu, Dan Liao, Taotao Liu, Xiaoting Ji, Ke Xu, et al. "Measurement report: Effects of anthropogenic emissions and environmental factors on the formation of biogenic secondary organic aerosol (BSOA) in a coastal city of southeastern China". *Atmos. Chem. Phys.* 22.11 (2022), pp. 7827–7841 (cit. on p. 118).

[Huang, 2014] Ru-Jin Huang, Yanlin Zhang, Carlo Bozzetti, Kin-Fai Ho, Jun-Ji Cao, Yongming Han, et al. "High secondary aerosol contribution to particulate pollution during haze events in China". *Nature* 514.7521 (2014), pp. 218–222 (cit. on p. 104).

[Huang, 2020] Xin Huang, Aijun Ding, Jian Gao, Bo Zheng, Derong Zhou, Ximeng Qi, et al. "Enhanced secondary pollution offset reduction of primary emissions during COVID-19 lockdown in China". *Natl. Sci. Rev.* 8.2 (2020) (cit. on pp. 21, 105).

[Isaacman-VanWertz, 2021] G. Isaacman-VanWertz and B. Aumont. "Impact of organic molecular structure on the estimation of atmospherically relevant physicochemical parameters". *Atmos. Chem. Phys.* 21.8 (2021), pp. 6541–6563 (cit. on p. 25).

[Jenkin, 2020] M. E. Jenkin, R. Valorso, B. Aumont, M. J. Newland, and A. R. Rickard. "Estimation of rate coefficients for the reactions of O_3 with unsaturated organic compounds for use in automated mechanism construction". *Atmos. Chem. Phys.* 20.21 (2020), pp. 12921–12937 (cit. on p. 61).

[Jenkin, 2008] ME Jenkin, LA Watson, SR Utembe, and DE Shallcross. "A Common Representative Intermediates (CRI) mechanism for VOC degradation. Part 1: Gas phase mechanism development". *Atmos. Environ.* 42.31 (2008), pp. 7185–7195 (cit. on pp. 25, 62).

[Jenkin, 2012] ME Jenkin, KP Wyche, CJ Evans, T Carr, PS Monks, MR Alfarra, et al. "Development and chamber evaluation of the MCM v3. 2 degradation scheme for β-caryophyllene". *Atmos. Chem. Phys.* 12.11 (2012), pp. 5275–5308 (cit. on pp. 61, 64).

[Jenkin, 1997] Michael E Jenkin, Sandra M Saunders, and Michael J Pilling. "The tropospheric degradation of volatile organic compounds: a protocol for mechanism development". *Atmos. Environ.* 31.1 (1997), pp. 81–104 (cit. on pp. 23, 61).

[Jiang, 2019] J. Jiang, S. Aksoyoglu, G. Ciarelli, E. Oikonomakis, I. El-Haddad, F. Canonaco, et al. "Effects of two different biogenic emission models on modelled ozone and aerosol concentrations in Europe". *Atmos. Chem. Phys.* 19.6 (2019), pp. 3747–3768 (cit. on pp. 112, 114).

[Kaduwela, 2015] Ajith Kaduwela, Deborah Luecken, William Carter, and Richard Derwent. "New directions: Atmospheric chemical mechanisms for the future". *Atmos. Environ.* 122 (2015), pp. 609–610 (cit. on p. 62).

[Kampa, 2008] Marilena Kampa and Elias Castanas. "Human health effects of air pollution". *Environ. Pollut.* 151.2 (2008), pp. 362–367 (cit. on p. 11).

[Kanakidou, 2005] Maria Kanakidou, JH Seinfeld, SN Pandis, Ian Barnes, Franciscus Johannes Dentener, Maria Cristina Facchini, et al. "Organic aerosol and global climate modelling: a review". *Atmos. Chem. Phys.* 5.4 (2005), pp. 1053–1123 (cit. on pp. 60, 104).

[Kelly, 2018] J. M. Kelly, R. M. Doherty, F. M. O'Connor, and G. W. Mann. "The impact of biogenic, anthropogenic, and biomass burning volatile organic compound emissions on regional and seasonal variations in secondary organic aerosol". *Atmos. Chem. Phys.* 18.10 (2018), pp. 7393–7422 (cit. on p. 118).

[Kelp, 2022] Makoto M Kelp, Daniel J Jacob, Haipeng Lin, and Melissa P Sulprizio. "An online-learned neural network chemical solver for stable long-term global simulations of atmospheric chemistry". *J. Adv. Model. Earth Syst.* 14.6 (2022), e2021MS002926 (cit. on p. 62).

[Kesselmeier, 1999] John Kesselmeier and Michael Staudt. "Biogenic volatile organic compounds (VOC): an overview on emission, physiology and ecology". *J. Atmos. Chem.* 33 (1999), pp. 23–88 (cit. on p. 4).

[Khan, 2017] MAH Khan, ME Jenkin, A Foulds, RG Derwent, CJ Percival, and DE Shallcross. "A modeling study of secondary organic aerosol formation from sesquiterpenes using the STOCHEM global chemistry and transport model". *J. Geophys. Res.-Atmos.* 122.8 (2017), pp. 4426–4439 (cit. on p. 25).

[Kuenen, 2022] J. Kuenen, S. Dellaert, A. Visschedijk, J.-P. Jalkanen, I. Super, and H. Denier van der Gon. "CAMS-REG-v4: a state-of-the-art high-resolution European emission inventory for air quality modelling". *Earth Syst. Sci. Data* 14.2 (2022), pp. 491–515 (cit. on p. 112).

[Lannuque, 2018] V. Lannuque, M. Camredon, F. Couvidat, A. Hodzic, R. Valorso, S. Madronich, et al. "Exploration of the influence of environmental conditions on secondary organic aerosol formation and organic species properties using explicit simulations: development of the VBS-GECKO parameterization". *Atmos. Chem. Phys.* 18.18 (2018), pp. 13411–13428 (cit. on pp. 25, 62, 84).

[Lannuque, 2020] V. Lannuque, F. Couvidat, M. Camredon, B. Aumont, and B. Bessagnet. "Modeling organic aerosol over Europe in summer conditions with the VBS-GECKO parameterization: sensitivity to secondary organic compound properties and IVOC (intermediate-volatility organic compound) emissions". *Atmos. Chem. Phys.* 20.8 (2020), pp. 4905–4931 (cit. on pp. 25, 113).

[Lemaire, 2016] V. Lemaire, I. Coll, F. Couvidat, C. Mouchel-Vallon, C. Seigneur, and G. Siour. "Oligomer formation in the troposphere: from experimental knowledge to 3-D modeling". *Geosci. Model Dev* 9.4 (2016), pp. 1361–1382 (cit. on p. 111).

[Li, 2015] Jingyi Li, Meredith Cleveland, Luke D. Ziemba, Robert J. Griffin, Kelley C. Barsanti, James F. Pankow, et al. "Modeling regional secondary organic aerosol using the Master Chemical Mechanism". *Atmos. Environ.* 102 (2015), pp. 52–61 (cit. on pp. 61, 105).

[Li, 2022] Qi Li, Jia Jiang, Isaac K Afreh, Kelley C Barsanti, and David R Cocker III. "Secondary organic aerosol formation from camphene oxidation: measurements and modeling". *Atmos. Chem. Phys.* 22.5 (2022), pp. 3131–3147 (cit. on p. 61).

[Lugon, 2021] Lya Lugon, Karine Sartelet, Youngseob Kim, Jéremy Vigneron, and Olivier Chrétien. "Simulation of primary and secondary particles in the streets of Paris using MUNICH". *Faraday Discussions* 226 (2021), pp. 432–456 (cit. on p. 8).

[Majdi, 2019] M. Majdi, K. Sartelet, G. M. Lanzafame, F. Couvidat, Y. Kim, M. Chrit, et al. "Precursors and formation of secondary organic aerosols from wildfires in the Euro-Mediterranean region". *Atmos. Chem. Phys.* 19.8 (2019), pp. 5543–5569 (cit. on p. 107).

[Manisalidis, 2020] Ioannis Manisalidis, Elisavet Stavropoulou, Agathangelos Stavropoulos, and Eugenia Bezirtzoglou. "Environmental and Health Impacts of Air Pollution: A Review". *Public Health Front.* 8 (2020) (cit. on pp. 9, 11).

[McFiggans, 2019] Gordon McFiggans, Thomas F Mentel, Juergen Wildt, Iida Pullinen, Sungah Kang, Einhard Kleist, et al. "Secondary organic aerosol reduced by mixture of atmospheric vapours". *Nature* 565.7741 (2019), pp. 587–593 (cit. on pp. 61, 140).

[McNeill, 2017] V Faye McNeill. "Atmospheric Aerosols: Clouds, Chemistry, and Climate". *Annual Rev. Chem. Biomol. Eng.* 8 (2017), pp. 427–444 (cit. on pp. 6, 104).

[Menut, 2021] L. Menut, B. Bessagnet, R. Briant, A. Cholakian, F. Couvidat, S. Mailler, et al. "The CHIMERE v2020r1 online chemistry-transport model". *Geoscientific Model Development* 14.11 (2021), pp. 6781–6811 (cit. on pp. 26, 106, 112, 125).

[Messina, 2016] P. Messina, J. Lathière, K. Sindelarova, N. Vuichard, C. Granier, J. Ghattas, et al. "Global biogenic volatile organic compound emissions in the ORCHIDEE and MEGAN models and sensitivity to key parameters". *Atmospheric Chemistry and Physics* 16.22 (2016), pp. 14169–14202 (cit. on p. 112).

[Mouchel-Vallon, 2022] Camille Mouchel-Vallon and Alma Hodzic. "Towards emulating an explicit organic chemistry mechanism with random forest models". *J. Geophys. Res.* (2022), e2022JD038227 (cit. on p. 62).

[Mouchel-Vallon, 2020] Camille Mouchel-Vallon, Julia Lee-Taylor, Alma Hodzic, Paulo Artaxo, Bernard Aumont, Marie Camredon, et al. "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". *Atmos. Chem. Phys.* 20.10 (2020), pp. 5995–6014 (cit. on pp. 23, 61).

[Nannoolal, 2008] Yash Nannoolal, Jürgen Rarey, and Deresh Ramjugernath. "Estimation of pure component properties: Part 3. Estimation of the vapor pressure of non-electrolyte organic compounds via group contributions and group interactions". *Fluid Phase Equilibr.* 269.1 (2008), pp. 117–133 (cit. on p. 71).

[Nannoolal, 2004] Yash Nannoolal, Jürgen Rarey, Deresh Ramjugernath, and Wilfried Cordes. "Estimation of pure component properties: Part 1. Estimation of the normal boiling point of non-electrolyte organic compounds via group contributions and group interactions". *Fluid Phase Equilibr.* 226 (2004), pp. 45–63 (cit. on p. 71).

[Newland, 2022] M. J. Newland, C. Mouchel-Vallon, R. Valorso, B. Aumont, L. Vereecken, M. E. Jenkin, et al. "Estimation of mechanistic parameters in the gas-phase reactions of ozone with alkenes for use in automated mechanism construction". *Atmos. Chem. Phys.* 22.9 (2022), pp. 6167–6195 (cit. on p. 61).

[Ng, 2007] N. L. Ng, P. S. Chhabra, A. W. H. Chan, J. D. Surratt, J. H. Kroll, A. J. Kwan, et al. "Effect of NO_x level on secondary organic aerosol (SOA) formation from the photooxidation of terpenes". *Atmos. Chem. Phys.* 7.19 (2007), pp. 5159–5174 (cit. on pp. 105, 129).

[Odum, 1996] Jay R. Odum, Thorsten Hoffmann, Frank Bowman, Don Collins, Richard C. Flagan, and John H. Seinfeld. "Gas/Particle Partitioning and Secondary Organic Aerosol Yields". *Env. Sc. and Tech.* 30.8 (1996), pp. 2580–2585 (cit. on pp. 23, 24, 61, 105).

[Pankow, 2008] J. F. Pankow and W. E. Asher. "SIMPOL.1: a simple group contribution method for predicting vapor pressures and enthalpies of vaporization of multifunctional organic compounds". *Atmos. Chem. Phys.* 8.10 (2008), pp. 2773–2796 (cit. on p. 71).

[Pankow, 1994] James F. Pankow. "An absorption model of gas/particle partitioning of organic compounds in the atmosphere". *Atmos. Environ.* 28.2 (1994), pp. 185–188 (cit. on p. 17).

[Porter, 2021] William C Porter, Jose L Jimenez, and Kelley C Barsanti. "Quantifying atmospheric parameter ranges for ambient secondary organic aerosol formation". *ACS Earth Space Chem.* 5.9 (2021), pp. 2380–2397 (cit. on pp. 61, 73, 105, 138).

[Poulain, 2011] L. Poulain, G. Spindler, W. Birmili, C. Plass-Dülmer, A. Wiedensohler, and H. Herrmann. "Seasonal and diurnal variations of particulate nitrate and organic matter at the IfT research station Melpitz". *Atmos. Chem. Phys.* 11.24 (2011), pp. 12579–12599 (cit. on p. 117).

[Presto, 2005] Albert A. Presto, Kara E. Huff Hartz, and Neil M. Donahue. "Secondary Organic Aerosol Production from Terpene Ozonolysis. 2. Effect of NOx Concentration". *Env. Sc. and Tech.* 39.18 (2005), pp. 7046–7054 (cit. on p. 129).

[Pun, 2007] B. K. Pun and C. Seigneur. "Investigative modeling of new pathways for secondary organic aerosol formation". *Atmos. Chem. Phys.* 7.9 (2007), pp. 2199–2216 (cit. on pp. 24, 111, 113, 115).

[Pun, 2006] Betty K. Pun, Christian Seigneur, and Kristen Lohman. "Modeling Secondary Organic Aerosol Formation via Multiphase Partitioning with Molecular Data". *Env. Sc. and Tech.* 40.15 (2006), pp. 4722–4731 (cit. on pp. 24, 61, 107).

[Raes, 2000] Frank Raes, Rita Van Dingenen, Elisabetta Vignati, Julian Wilson, Jean-Philippe Putaud, John H Seinfeld, et al. "Formation and cycling of aerosols in the global troposphere". *Atmos. Environ.* 34.25 (2000), pp. 4215–4240 (cit. on p. 19).

[Roldin, 2019] Pontus Roldin, Mikael Ehn, Theo Kurtén, Tinja Olenius, Matti P Rissanen, Nina Sarnela, et al. "The role of highly oxygenated organic molecules in the Boreal aerosol-cloud-climate system". *Nat. Commun.* 10.1 (2019), pp. 1–15 (cit. on pp. 15, 61, 69, 71, 105, 107).

[Sartelet, 2012] K. Sartelet, F. Couvidat, C. Seigneur, and Y. Roustan. "Impact of biogenic emissions on air quality over Europe and North America." *Atmos. Environ.* 53 (2012), pp. 131–141 (cit. on pp. 4, 21, 105).

[Sartelet, 2022a] K. Sartelet, Y. Kim, F. Couvidat, M. Merkel, T. Petäjä, J. Sciare, et al. "Influence of emission size distribution and nucleation on number concentrations over Greater Paris". *Atmos. Chem. Phys.* 22.13 (2022), pp. 8579–8596 (cit. on p. 106).

[Sartelet, 2022b] Karine Sartelet. "Secondary Aerosol Formation and Their Modeling". *Atmospheric Chemistry in the Mediterranean Region: Volume 2 - From Air Pollutant Sources to Impacts*. Ed. by François Dulac, Stéphane Sauvage, and Eric Hamonou. Cham: Springer International Publishing, 2022, pp. 165–183 (cit. on p. 118).

[Sartelet, 2020] Karine Sartelet, Florian Couvidat, Zhizhao Wang, Cédric Flageul, and Youngseob Kim. "SSH-Aerosol v1. 1: A Modular Box Model to Simulate the Evolution of Primary and Secondary Aerosols". *Atmosphere* 11.5 (2020), p. 525 (cit. on pp. 26, 64, 106, 107).

[Sarwar, 2008] Golam Sarwar, Deborah Luecken, Greg Yarwood, Gary Z. Whitten, and William P. L. Carter. "Impact of an Updated Carbon Bond Mechanism on Predictions from the CMAQ Modeling System: Preliminary Assessment". *J. Applied Meteor.* 47.1 (2008), pp. 3–14 (cit. on pp. 22, 28, 64).

[Saunders, 2003] Sandra M Saunders, Michael E Jenkin, RG Derwent, and MJ Pilling. "Protocol for the development of the Master Chemical Mechanism, MCM v3 (Part A): tropospheric degradation of non-aromatic volatile organic compounds". *Atmos. Chem. Phys.* 3.1 (2003), pp. 161–180 (cit. on p. 61).

[Schraufnagel, 2019] Dean E Schraufnagel, John R Balmes, Clayton T Cowl, Sara De Matteis, Soon-Hee Jung, Kevin Mortimer, et al. "Air pollution and noncommunicable diseases: A review by the Forum of International Respiratory Societies' Environmental Committee, Part 2: Air pollution and organ systems". *Chest* 155.2 (2019), pp. 417–426 (cit. on p. 11).

[Schreck, 2022] John S Schreck, Charles Becker, David John Gagne, Keely Lawrence, Siyuan Wang, Camille Mouchel-Vallon, et al. "Neural Network Emulation of the Formation of Organic Aerosols Based on the Explicit GECKO-A Chemistry Model". *J. Adv. Model. Earth Syst.* 14.10 (2022), e2021MS002974 (cit. on pp. 25, 62).

[Seigneur, 2019] Christian Seigneur. *Air Pollution: Concepts, Theory, and Applications*. Cambridge University Press, 2019 (cit. on pp. 5, 7, 13, 20, 105).

[Seinfeld, 2016] John H Seinfeld, Christopher Bretherton, Kenneth S Carslaw, Hugh Coe, Paul J DeMott, Edward J Dunlea, et al. "Improving our fundamental understanding of the role of aerosol-cloud interactions in the climate system". *Proc. Nat. Acad. Sci.* 113.21 (2016), pp. 5781–5790 (cit. on pp. 9, 104).

[Shen, 2022] L. Shen, D. J. Jacob, M. Santillana, K. Bates, J. Zhuang, and W. Chen. "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". *Geosci. Model Dev* 15.4 (2022), pp. 1677–1687 (cit. on p. 62).

[Shrivastava, 2019] M. Shrivastava, M.O. Andreae, P. Artaxo, H.M.J. Barbosa, L.K. Berg, J. Brito, et al. "Urban pollution greatly enhances formation of natural aerosols over the Amazon rainforest". *Nat. Commun.* 10 (2019), p. 1046 (cit. on p. 105).

[Shrivastava, 2017] Manish Shrivastava, Christopher D. Cappa, Jiwen Fan, Allen H. Goldstein, Alex B. Guenther, Jose L. Jimenez, et al. "Recent advances in understanding secondary organic aerosol: Implications for global climate forcing". *Rev. Geophys* 55.2 (2017), pp. 509–559 (cit. on pp. 10, 61).

[Sillman, 1999] Sanford Sillman. "The relation between ozone, NOx and hydrocarbons in urban and polluted rural environments". *Atmos. Environ.* 33.12 (1999), pp. 1821–1845 (cit. on p. 126).

[Sindelarova, 2022] K. Sindelarova, J. Markova, D. Simpson, P. Huszar, J. Karlicky, S. Darras, et al. "High-resolution biogenic global emission inventory for the time period 2000–2019 for air quality modelling". *Earth Syst. Sci. Data* 14.1 (2022), pp. 251–270 (cit. on p. 112).

[Stockwell, 2020] William R Stockwell, Emily Saunders, Wendy S Goliff, and Rosa M Fitzgerald. "A perspective on the development of gas-phase chemical mechanisms for Eulerian air quality models". *J. Air Waste Manag. Assoc.* 70.1 (2020), pp. 44–70 (cit. on p. 61).

[Stolzenburg, 2022] Dominik Stolzenburg, Mingyi Wang, Meredith Schervish, and Neil M. Donahue. "Tutorial: Dynamic organic growth modeling with a volatility basis set". *J. Atmos. Sci.* 166 (2022), p. 106063 (cit. on p. 61).

[Sturm, 2023] Patrick Obin Sturm, Astrid Manders, Ruud Janssen, Arjo Segers, Anthony S Wexler, and Hai Xiang Lin. "Advecting Superspecies: Efficiently Modeling Transport of Organic Aerosol With a Mass-Conserving Dimensionality Reduction Method". *J. Adv. Model. Earth Syst.* 15.3 (2023), e2022MS003235 (cit. on p. 62).

[Szopa, 2005] S. Szopa, B. Aumont, and S. Madronich. "Assessment of the reduction methods used to develop chemical schemes: building of a new chemical scheme for VOC oxidation suited to three-dimensional multiscale HO_x-NO_x-VOC chemistry simulations". *Atmos. Chem. Phys.* 5.9 (2005), pp. 2519–2538 (cit. on p. 62).

[Takeuchi, 2022] M. Takeuchi, T. Berkemeier, G. Eris, and N.L. Ng. "Non-linear effects of secondary organic aerosol formation and properties in multi-precursor systems". *Nat. Commun.* 13 (2022), p. 7883 (cit. on p. 105).

[Topping, 2016] D. Topping, M. Barley, M. K. Bane, N. Higham, B. Aumont, N. Dingle, et al. "UManSysProp v1.0: an online and open-source facility for molecular property prediction and atmospheric aerosol calculations". *Geosci. Model Dev* 9.2 (2016), pp. 899–914 (cit. on p. 71).

[Turner, 2020] Michelle C Turner, Zorana J Andersen, Andrea Baccarelli, W Ryan Diver, Susan M Gapstur, C Arden Pope III, et al. "Outdoor air pollution and cancer: An overview of the current evidence and public health recommendations". *CA: Cancer J. Clin.* 70.6 (2020), pp. 460–479 (cit. on p. 4).

[Turpin, 2001] B. J. Turpin and H.-J. Lim. "Species contributions to PM2.5 mass concentrations: Revisiting common assumptions for estimating organic mass". *Aerosol Sc. and Tech.* 35 (2001), pp. 602–610 (cit. on p. 117).

[Utembe, 2011] S.R. Utembe, M.C. Cooke, A.T. Archibald, D.E. Shallcross, R.G. Derwent, and M.E. Jenkin. "Simulating secondary organic aerosol in a 3-D Lagrangian chemistry transport model using the reduced Common Representative Intermediates mechanism (CRI v2-R5)". *Atmos. Environ.* 45.8 (2011), pp. 1604–1614 (cit. on p. 25).

[Utembe, 2009] S.R. Utembe, L.A. Watson, D.E. Shallcross, and M.E. Jenkin. "A Common Representative Intermediates (CRI) mechanism for VOC degradation. Part 3: Development of a secondary organic aerosol module". *Atmos. Environ.* 43.12 (2009), pp. 1982–1990 (cit. on p. 25).

[Wang, 2018] N. Wang, E. Kostenidou, N. M. Donahue, and S. N. Pandis. "Multi-generation chemical aging of α-pinene ozonolysis products by reactions with OH". *Atmos. Chem. Phys.* 18.5 (2018), pp. 3589–3601 (cit. on p. 104).

[Wang, 2022] Zhizhao Wang, Florian Couvidat, and Karine Sartelet. "GENerator of reduced Organic Aerosol mechanism (GENOA v1.0): an automatic generation tool of semi-explicit mechanisms". *Geosci. Model Dev* 15.24 (2022), pp. 8957–8982 (cit. on pp. 62–65, 68, 72, 106).

[Wang, 2023] Zhizhao Wang, Florian Couvidat, and Karine Sartelet. "Implementation of a Parallel Reduction Algorithm in the Generator of Reduced Organic Aerosol Mechanisms (Genoa V2.0): Application to Multiple Monoterpene Aerosol Precursors" (2023) (cit. on pp. 106–108).

[Watson, 2008] L.A. Watson, D.E. Shallcross, S.R. Utembe, and M.E. Jenkin. "A Common Representative Intermediates (CRI) mechanism for VOC degradation. Part 2: Gas phase mechanism reduction". *Atmos. Environ.* 42.31 (2008), pp. 7196–7204 (cit. on pp. 25, 62).

[Weber, 2020] J. Weber, S. Archer-Nicholls, P. Griffiths, T. Berndt, M. Jenkin, H. Gordon, et al. "CRI-HOM: A novel chemical mechanism for simulating highly oxygenated organic molecules (HOMs) in global chemistry–aerosol–climate models". *Atmos. Chem. Phys.* 20.18 (2020), pp. 10889–10910 (cit. on pp. 25, 62).

[Whitehouse, 2004a] L. E. Whitehouse, A. S. Tomlin, and M. J. Pilling. "Systematic reduction of complex tropospheric chemical mechanisms, Part I: sensitivity and time-scale analyses". *Atmos. Chem. Phys.* 4.7 (2004), pp. 2025–2056 (cit. on p. 62).

[Whitehouse, 2004b] L. E. Whitehouse, A. S. Tomlin, and M. J. Pilling. "Systematic reduction of complex tropospheric chemical mechanisms, Part II: Lumping using a time-scale based approach". *Atmos. Chem. Phys.* 4.7 (2004), pp. 2057–2081 (cit. on p. 62).

[Wiser, 2023] F. Wiser, B. K. Place, S. Sen, H. O. T. Pye, B. Yang, D. M. Westervelt, et al. "AMORE-Isoprene v1.0: a new reduced mechanism for gas-phase isoprene oxidation". *Geosci. Model Dev* 16.6 (2023), pp. 1801–1821 (cit. on p. 62).

[Wolf, 2021] Sarah Wolf, Jonas Teitge, Jahel Mielke, Franziska Schütze, and Carlo Jaeger. "The European Green Deal—more than climate neutrality". *Intereconomics* 56 (2021), pp. 99–107 (cit. on p. 12).

[Wu, 1994] Pei-Ming Wu and Kikuo Okada. "Nature of coarse nitrate particles in the atmosphere—A single particle approach". *Atmospheric Environment* 28.12 (1994), pp. 2053–2060 (cit. on p. 7).

[Xavier, 2019] Carlton Xavier, Anton Rusanen, Putian Zhou, Chen Dean, Lukas Pichelstorfer, Pontus Roldin, et al. "Aerosol mass yields of selected biogenic volatile organic compounds–a theoretical study with nearly explicit gas-phase chemistry". *Atmos. Chem. Phys.* 19.22 (2019), pp. 13741–13758 (cit. on pp. 71, 76, 129).

[Xia, 2009] A. G. Xia, D. V. Michelangeli, and P. A. Makar. "Mechanism reduction for the formation of secondary organic aerosol for integration into a 3-dimensional regional air quality model: α-pinene oxidation system". *Atmos. Chem. Phys.* 9.13 (2009), pp. 4341–4362 (cit. on p. 62).

[Xu, 2021a] L. Xu, X. Liu, H. Gao, X. Yao, D. Zhang, L. Bi, et al. "Long-range transport of anthropogenic air pollutants into the marine air: insight into fine particle transport and chloride depletion on sea salts". *Atmos. Chem. Phys.* 21.23 (2021), pp. 17715–17726 (cit. on p. 4).

[Xu, 2021b] Li Xu, Lin Du, Narcisse T Tsona, and Maofa Ge. "Anthropogenic effects on biogenic secondary organic aerosol formation". *Adv Atmos Sci* . 38 (2021), pp. 1053–1084 (cit. on p. 140).

[Ying, 2011] Qi Ying and Jingyi Li. "Implementation and initial application of the near-explicit Master Chemical Mechanism in the 3D Community Multiscale Air Quality (CMAQ) model". *Atmos. Environ.* 45.19 (2011), pp. 3244–3256 (cit. on p. 61).

[Yu, 2021] Zechen Yu, Myoseon Jang, Tianyu Zhang, Azad Madhu, and Sanghee Han. "Simulation of Monoterpene SOA Formation by Multiphase Reactions Using Explicit Mechanisms". *ACS Earth Space Chem.* 5.6 (2021), pp. 1455–1467 (cit. on p. 129).

[Zhang, 2007] Qi Zhang, Jose L Jimenez, MR Canagaratna, J David Allan, H Coe, I Ulbrich, et al. "Ubiquity and dominance of oxygenated species in organic aerosols in anthropogenically-influenced Northern Hemisphere midlatitudes". *Geophys. Res. Lett.* 34.13 (2007) (cit. on p. 7).

[Zhu, 2015] S. Zhu, K. N. Sartelet, and C. Seigneur. "A size-composition resolved aerosol model for simulating the dynamics of externally mixed particles: SCRAM (v 1.0)". *Geoscientific Model Development* 8.6 (2015), pp. 1595–1612 (cit. on p. 26).