

Development of a simple unified volatility-based scheme (SUVS) for secondary organic aerosol formation using genetic algorithms

A. G. Xia, C. A. Stroud, and P. A. Makar

Air Quality Research Division, Science and Technology Branch, Environment Canada, Toronto, Ontario, Canada

Received: 10 January 2011 – Published in Atmos. Chem. Phys. Discuss.: 3 February 2011

Revised: 19 June 2011 – Accepted: 21 June 2011 – Published: 1 July 2011

Abstract. A new method is proposed to simplify complex atmospheric chemistry reaction schemes, while preserving SOA formation properties, using genetic algorithms. The method is first applied in this study to the gas-phase α -pinene oxidation scheme. The simple unified volatility-based scheme (SUVS) reflects the multi-generation evolution of chemical species from a near-explicit master chemical mechanism (MCM) and, at the same time, uses the volatility-basis set speciation for condensable products. The SUVS also unifies reactions between SOA precursors with different oxidants under different atmospheric conditions. A total of 412 unknown parameters (product yields of parameterized products, reaction rates, etc.) from the SUVS are estimated by using genetic algorithms operating on the detailed mechanism. The number of organic species was reduced from 310 in the detailed mechanism to 31 in the SUVS. Output species profiles, obtained from the original subset of the MCM reaction scheme for α -pinene oxidation, are reproduced with maximum fractional error at 0.10 for scenarios under a wide range of ambient HC/NO_x conditions. Ultimately, the same SUVS with updated parameters could be used to describe the SOA formation from different precursors.

1 Introduction

Aerosols play an important role in atmospheric chemistry, with impacts on local, regional and global air quality and human health exposure. Sulfate has been known to be a dominant inorganic component, sometimes 50% or more, in particles with aerodynamic diameter less than 2.5 μm (PM_{2.5}) in the eastern United States (Seinfeld, 2004). Recent particle mass spectrum measurements on both aircrafts and at ground level monitoring sites (Zhang et al., 2007; Jimenez

et al., 2009) indicate that organic to sulfate ratio is about 1:1 in tropospheric fine particles (Hallquist et al., 2009). Organic aerosols are classified as primary organic aerosols (POA) and secondary organic aerosols (SOA). The POA is formed directly from primary emissions, such as fossil fuel combustion, biomass burning, and biogenic materials (Sun and Ariya, 2006). The SOA is formed via gas/particle transfer (Pankow, 1994) or heterogeneous reactions (Jang et al., 2002; Tolocka et al., 2004; Kalberer et al., 2004; Liggi et al., 2005) of the organic products from the atmospheric oxidation of biogenic and/or anthropogenic volatile organic compounds (VOCs). According to the gas/particle partitioning theory (Pankow, 1994), the products with low volatility remain in the aerosol phase and they are also called condensable products.

There are two major approaches used to model SOA formation which have been summarized in a recent review (Hallquist et al., 2009). The first approach is a two-product gas/particle partitioning model (Odum et al., 1996), in which two surrogate species are used to describe hundreds of oxidation products from each VOC precursor. In this model, two parameters for mass-based stoichiometric yields (α_1 and α_2) and two for partitioning coefficients ($K_{p,1}$ and $K_{p,2}$) are obtained by least square fitting of smog chamber measurements. In this empirical model, the two surrogate products in the gas phase are described by a one-step chemical reaction. This simple model is easy to implement in a chemical transport model, and it has been widely used to study SOA formation in regional (Andersson-Skold and Simpson, 2001; Schell et al., 2001; Pun et al., 2003; Slowik et al., 2010) and global models (Chung and Seinfeld, 2002; Tsigaridis and Kanakidou, 2003).

The second approach (Hallquist et al., 2009) is to use chemical mechanisms, such as the Master Chemical Mechanism (MCM) (Jenkin et al., 1997; Saunders et al., 2003; Jenkin et al., 2003) or the self-generating mechanism (Aumont et al., 2005), to describe the gas phase



Correspondence to: A. G. Xia
(airquality@gmail.com)

multi-generational chemical evolutions of condensable organic products. The total numbers of species and reactions from the explicit chemical mechanisms are often on the order of thousands. These explicit gas-phase chemical mechanisms are coupled with the gas/particle partitioning model (Pankow, 1994) to describe SOA formation (Stroud et al., 2004; Jenkin, 2004; Johnson et al., 2006; Camredon et al., 2007; Capouet et al., 2008; Xia et al., 2008; Kelly et al., 2010) from the oxidation of a variety of VOCs, such as toluene, α - and β -pinene, and others. The major difference between the simple model and explicit model lies in the number of chemical species and reactions involved. Although some modifications were made for the one-step, two-product model approach to account for some simple interactions between different products in both gas and aerosol phases (Tsigaridis and Kanakidou, 2003; Chan et al., 2007), the explicit mechanism is more precise in describing complex chemistry for the multi-generation products in a SOA system. Nevertheless, explicit chemical mechanisms have been used to study SOA formation using zero-dimensional box models rather than three-dimensional regional models, due to the computational burden for solving the set of differential equations for thousands of species and reactions.

Explicit chemical mechanisms also have limitations: the complexity of the reaction mechanism may be based on estimates, extrapolations and mechanistic analogies, rather than individually laboratory-verified chemical reaction pathways, products and reaction rates. The explicit mechanisms do, however, attempt to approximate the complexity of the ambient atmosphere. The explicit mechanisms therefore provide a good starting point for a mathematical analysis of that complexity and for the development of methodologies and parameterizations for its simplification.

The purpose for this work is to develop a methodology for creating new simple chemical reaction mechanisms suitable for studying SOA formation in a 3-dimensional chemical transport model. This simplified chemical mechanism should also simulate the same evolution of chemical species as provided by a more explicit chemical mechanism. Two requirements are needed to meet this goal: (1) the new chemical reaction mechanism must have a small number of species; and (2) the new chemical mechanism should mimic a detailed chemical reaction mechanism to describe chemical interactions among different species.

For the first requirement, the volatility-basis set (VBS) approach (Donahue et al., 2006) provides a good framework to use a few lumped species to represent a large number of chemicals in the ambient atmosphere. Donahue et al. (2006) first extended the simple two-product model into a multiple (up to 9) lumped product model. The effective saturation concentrations C_i^* for each lumped product is assumed to be fixed, rather than data-fitted varied values in the two-product model (Odum et al., 1996). Specifically, the fixed C_i^* of the multiple lumped products are separated by one order of magnitude each. This means a large number of chem-

ical species from smog chamber experiments could be represented by multiple lumped products. As an improved method over the two-product model for organic aerosol formation, the VBS approach unifies condensable products from different oxidation systems under the framework of the same fixed volatility distribution. For example, the same lumped products could be used to represent the oxidation products between α -pinene + O₃ system and toluene + OH system (Lane et al., 2008; Donahue et al., 2009).

However, to address the issue regarding the second requirement to obtain a simple chemical system that mimics a detailed chemical mechanism, the original chemical reaction scheme (Donahue et al., 2006) via the VBS approach is not sufficient for such a purpose. This is because the reaction system in the VBS approach, similar to those from the two-product approach, is basically a one-step reaction system, although some simple chemical aging process has been taken into account (Farina et al., 2010) to improve this method. In this paper, a new chemical reaction scheme using the VBS concept is proposed. Essentially, the new simple unified volatility-based scheme (SUVS) is constructed by following chemical reaction protocols (Atkinson and Arey, 2003; Saunders et al., 2003; Kroll and Seinfeld, 2008) for the organic compounds in the ambient atmosphere.

Equally important in developing the new reaction scheme, we also need to determine chemical parameters, such as chemical reaction rate coefficients and product yields. First, for this purpose, the MCM is used as a benchmark to provide concentration profiles for detailed chemical species under a wide range of conditions. Next, the output concentrations for the detailed MCM chemical species are grouped, according to their volatility distributions, into a small number of lumped species in the new chemical reaction scheme. Finally, the computed profiles for the lumped species are treated as “experimental data” for data fitting of a large number of unknown chemical parameters (product yields and reaction rate constants) in the newly derived simple reaction scheme. This is accomplished here using genetic algorithms, rather than using traditional least square fitting methods for a small number (less than 12) of unknowns in the original one-step VBS reaction scheme (Stanier et al., 2008).

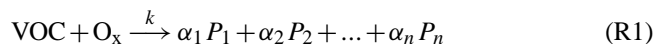
Our primary focus was to develop a simple scheme for organic aerosol formation, but not one which is fully coupled with an existing chemical mechanism. In that respect, our work is similar the idea of a yield approach, in which yields of SOA mass are calculated based on the amount of precursor that is oxidized, but there is no attempt to conserve the gas-phase products formed. We envisage the SUVS as an ancillary mechanism in the context of its use in a regional air pollution model: the gas-phase reaction mechanism of the regional model being used to provide both precursor oxidation and HO_x/NO_x/O_x levels that are in turn used to drive the SUVS solution. So our intention, at this stage, is not to have the SUVS modify the inorganic chemistry, but rather, to be driven by both it and a regional model’s organic chemistry.

This paper is organized as follows. First, we give a brief introduction on the VBS approach in Sect. 2. In Sect. 3, we propose a new simple reaction scheme for SOA formation. In Sect. 4, we run the MCM v3.1 under a wide range of conditions to obtain the “experimental data” for the new reaction scheme. In Sect. 5, we introduce the genetic algorithms method to derive the fitted chemical parameters for the new simple reaction scheme. Finally, we present model results and discussions on this new chemical mechanism framework for SOA formation.

2 Volatility-based set (VBS) approach for organic aerosol formation

In this section, we give a brief introduction on the two-product model for SOA formation, which is followed by a concise description of the volatility-based set approach. Next, four issues about the VBS approach will be raised, and these four issues motivate our work to develop a new chemical scheme in this paper.

Pankow (1994) proposed a theoretical absorptive model to study organic aerosols phase partitioning, and this absorptive model was applied by Odum et al. (1996) to describe SOA formation using smog chamber experimental data. According to Odum et al. (1996), a chemical oxidation system for SOA formation can be expressed as:



where VOC represents a volatile organic compound precursor, such as α -pinene, toluene, isoprene etc. The O_x represents any of the oxidants, such as O_3 , OH, NO_3 , and even Cl (Cai and Griffin, 2006). P_1, P_2, \dots, P_n are n number of condensable products; $\alpha_1, \alpha_2, \dots, \alpha_n$ are mass-based stoichiometric yields for the n products; $K_{p,1}, K_{p,2}, \dots, K_{p,n}$, are the corresponding gas/particle partitioning coefficients ($\text{m}^3 \mu\text{g}^{-1}$) in which

$$K_{p,i} = \frac{C_i^p / C_i^g}{C_{\text{OA}}} \quad (1)$$

where C_i^p and C_i^g are aerosol phase and gas phase concentrations ($\mu\text{g m}^{-3}$) for the i -th product, C_{OA} is the total aerosol mass concentration ($\mu\text{g m}^{-3}$). By using this framework, the SOA yield (Y) can be expressed as:

$$Y = \frac{\Delta M_{\text{OA}}}{\Delta \text{HC}} = C_{\text{OA}} \sum_{i=1}^n \frac{\alpha_i K_{p,i}}{1 + C_{\text{OA}} K_{p,i}} = \sum_{i=1}^n \frac{\alpha_i}{1 + C_i^* / C_{\text{OA}}} \quad (2)$$

In order to fit smog chamber experimental data, Odum et al. (1996) found that two surrogate products are enough to describe the shape of the SOA yield curve for one oxidation system, and a total of four parameters ($\alpha_1, \alpha_2, K_{p,1}$, and $K_{p,2}$) are obtained through a least square fitting method. This is the so called “two-product model” for SOA formation. But the two-product model has difficulty describing

the wide range of organic compounds in the ambient atmosphere, especially for those highly volatile organic compounds (Donahue et al., 2006).

To address this issue, Donahue et al. (2006) proposed the use of the VBS approach to study organic aerosol formation from React. (R1), in which the volatilities C_i^* for all n (up to 9) lumped products are assumed to be fixed, rather than variables derived through the least square fitting method in the two-product model. Still, the mass-based stoichiometric yields ($\alpha_1, \alpha_2, \dots, \alpha_n$) are obtained by using linear least square fitting of the smog chamber experimental SOA yield curves (Stanier et al., 2008).

The VBS approach can cover the wide range of volatilities for the organic products found in the atmosphere. In addition, the VBS approach can also be used to map all oxidation products from different oxidation systems into one framework because volatilities for all organic products are lumped into a fixed framework. Chemical transformations between different lumped products, such as aging through OH oxidation, could be described by this scheme (Donahue et al., 2006). This approach was used to study organic aerosol formation from both primary and secondary sources (Robinson et al., 2007; Lane et al., 2008).

We suggest here four possible improvements to the existing VBS approach. First, as mentioned earlier in the introduction, the reaction scheme in Reaction (R1) for the VBS approach is still a one-step reaction system. The complex chemical reactions for the products, especially for those multi-generational products, would not be captured well by using a one-step reaction scheme in a period longer than the typical duration (a few hours) for a smog chamber experiment for SOA formation. Second, large uncertainties still exist for the mass-based stoichiometric yields derived via least square fitting, especially for those yields corresponding to the high volatile products (Presto and Donahue, 2006; Pathak et al., 2007). The uncertainties are mainly caused by using total SOA yield as the only one important constraint for the data fitting process. Third, the set of parameters for a single reaction system change when the driving conditions change. For example, five different sets of VBS parameterizations were developed for ozonolysis of α -pinene (Pathak et al., 2007) under five different conditions. Fourth, the reaction scheme between a VOC precursor, say α -pinene, with different oxidants, such as OH, NO_3 , or O_3 , are not coupled in the fitting process, while the actual precursor VOC may react with both O_3 and OH during the day and NO_3 and O_3 at night. The original least square fitting tries to isolate parameters for each reaction separately, but they are intricately coupled and require higher order fitting techniques.

Motivated by the four points mentioned above, we develop a new approach to write and parameterize the reaction scheme under which α -pinene oxidation forms condensable products. This new VBS scheme is a multi-step reaction system, which is constructed by following standard organic reaction protocols. The chemical profiles of the new reaction

scheme mimic those from a detailed chemical mechanism, and demonstrate the power of this approach.

3 Development of a simple unified volatility-based scheme (SUVS)

In this section, we give an overview of the general features for ambient atmospheric oxidation. Then, based on these general features, we construct a Simple Unified Volatility-based Scheme (SUVS) for organic aerosol formation.

Oxidation reactions of ambient organic species alter their volatility, either to increase it through cleavage of carbon-carbon bond, or to decrease it through addition of other functional groups. In recent years, exploration of particle-phase organic chemical reactions, including heterogeneous reaction and multiphase reactions, has been very active (Jang et al., 2002; Tolocka et al., 2004; Kalberer et al., 2004; Gao et al., 2004; Iinuma et al., 2004; Liggió et al., 2005). A review on current understanding of chemical kinetics and reactions in the gas and aerosol phases for SOA formation is given by Kroll and Seinfeld (2008). In this work, as a first step to explore the new chemical reaction scheme and fitting method, we focus our study on atmospheric gas phase reactions only.

3.1 Brief review of atmospheric oxidation

Gas phase chemical reaction pathways vary with individual SOA precursors because of their unique molecular structures. Only a small fraction of reactions for a few ambient VOCs have been studied extensively. A number of generic structure-activity relationships derived from the experimental studies are used as protocols to infer chemical parameters and subsequent products for those larger VOCs during construction of an explicit chemical mechanism such as the MCM (Jenkin et al., 1997; Saunders et al., 2003; Aumont et al., 2005).

An excellent review on atmospheric oxidation of a wide range of VOCs can also be found in Atkinson and Arey (2003). Here, we only give a brief overview of those aspects of oxidation mechanisms relevant to SOA formation. First, the initial gas phase chemical reaction of a VOC species with primary oxidants, such as OH, NO₃, and O₃, leads to the formation of alkyl radicals (R) or Criegee intermediates and then rapidly to form organic peroxy radicals (RO₂). Subsequent reactions for RO₂ radicals and alkoxy radicals (RO) are very important for SOA formation, since these may lead to lower volatility products. In the troposphere, RO₂ radicals react with (1) NO, (2) RO₂, (3) NO₃, (4) HO₂, and (5) NO₂.

1. RO₂ + NO: The RO₂ radicals react with NO to form an RO radical or an organic nitrate (RONO₂). In the atmosphere, the branching of RO₂ with NO competes with the reaction of RO₂ + HO₂ to form ROOH. For most SOA reaction systems, like those for light aromatics

and monoterpenes (Song et al., 2005), the volatility of RONO₂ is comparatively higher than that of the ROOH. As a result, for α-pinene, higher SOA mass is expected to be formed under lower NO_x condition.

2. RO₂ + RO₂: Under ambient conditions, the self reaction and cross reactions of RO₂ + RO₂ form RO and OH via chain propagation or carbonyl, alcohol, or organic peroxides (ROOR) via chain-termination. The carbonyl and alcohol are more volatile than the ROOR (Ziemann, 2002) and those formed ROOH from RO₂ + HO₂.
3. RO₂ + NO₃: The reaction of RO₂ + NO₃ is believed to be an important loss process at night time (Kirchner and Stockwell, 1996). Based on available information (Lightfoot et al., 1992; Biggs et al., 1994, 1995; Daele et al., 1995), the reaction of RO₂ with NO₃ is assumed to produce alkoxy radicals (RO) and NO₂ (Saunders et al., 2003).
4. RO₂ + HO₂: The RO₂ radicals may react with HO₂ to form a low-volatility product of organic peroxide (ROOH). Organic peroxides have been predicted to be an important fraction of total SOA formed in the MCM (Bonn et al., 2004; Xia et al., 2008).
5. RO₂ + NO₂: The reaction of RO₂ radicals with NO₂ forms peroxy nitrates (ROONO₂), which are temporary reservoirs for RO₂ and NO₂, because of lifetimes on the order of seconds at 298 K.

Meanwhile, the reaction for the RO radical has two major chemical pathways: (1) dissociation through cleavage of carbon-carbon bond or H-abstraction with both pathways forming a carbonyl and an alkyl radical (R), which is rapidly converted to RO₂, or HO₂, (2) isomerization to form another alkyl radical (R') via a 1,5-hydrogen shift.

The reaction pathways described above lead to the formation of a wide range of oxygenated products, such as carbonyl compounds, organic nitrates (RONO₂), hydroperoxides (ROOH), alcohols (ROH), carboxylic acid (RC(O)OH). As reviewed in Atkinson and Arey (2003), the degradation for these newly formed compounds include photolysis and further reactions with OH and NO₃. Specifically, the photolysis of the oxygenated compounds leads to the formation of RO, and the reactions of the oxygenated compounds with OH and NO₃ generally result in the formation of RO₂.

3.2 A simple unified volatility-based scheme (SUVS)

A wide variety of starting organic compounds and possible complex chemical reaction pathways outlined in the preceding subsection lead to a myriad of compounds formed in the ambient atmosphere. As mentioned earlier, "volatility" is an important parameter to describe organic aerosol formation. Volatility moderates a compound's existence in

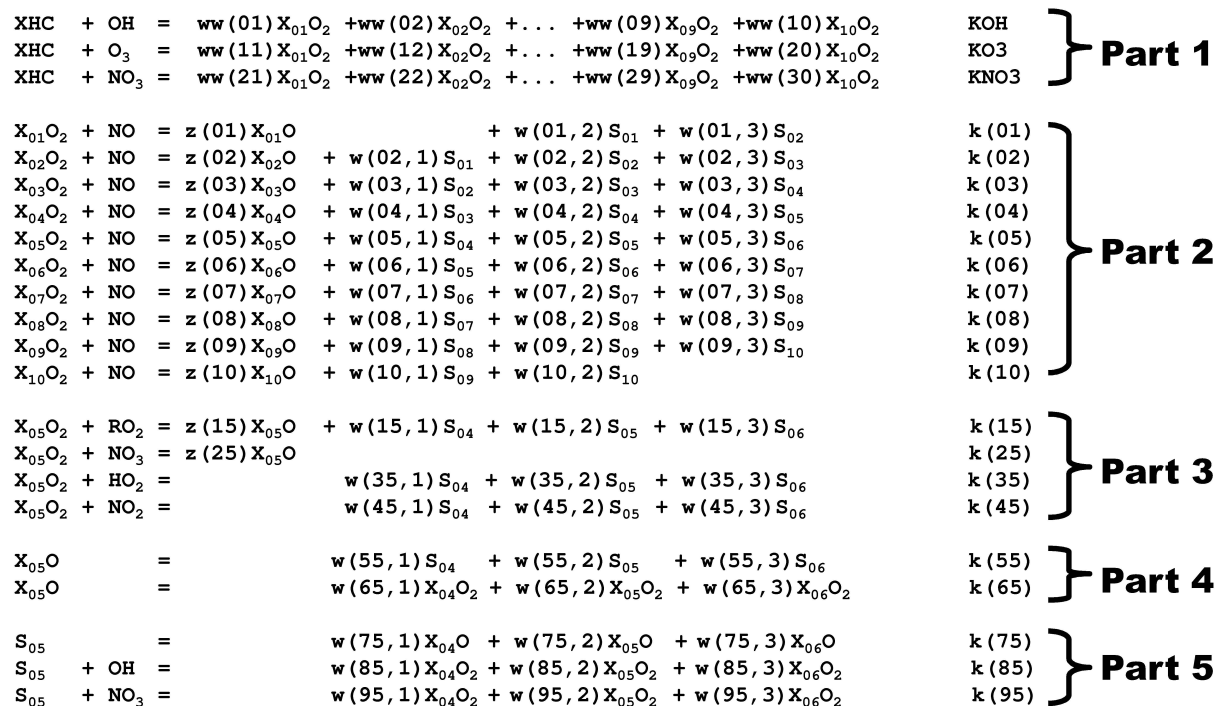


Fig. 1. A simple unified volatility-based scheme (SUVS) for SOA formation.

gas- and aerosol phases. To model evolution of a chemical system, a lumping method based on volatility is an effective way to reduce the number of organic species (Bian and Bowman, 2005).

In this work, we first lump organic compounds in a chemical system into 10 volatility-based surrogate condensable products termed volatility bins: S_{01} , S_{02} , ..., S_{10} . The volatilities for the 10 compounds span from 10^{-2} to $10^7 \mu\text{g m}^{-3}$ at the reference temperature of 298 K: S_{01} is in volatility bin 1 (10^{-2} to $10^{-1} \mu\text{g m}^{-3}$), S_{02} is in volatility bin 2 (10^{-1} to $10^0 \mu\text{g m}^{-3}$), ... and S_{10} is in volatility bin 10 (10^6 to $10^7 \mu\text{g m}^{-3}$). Note that the organic compounds, such as the formaldehyde, acetaldehyde, with volatility higher than $10^7 \mu\text{g m}^{-3}$ are excluded for the lumping process. Also note that the use of “volatility bin” terminology is taken directly from Donahue et al. (2006), and the nomenclature of the compounds in this paper is similar to the one used in their works, in which up to 9 products were assumed.

As mentioned in Sect. 2, a chemical mechanism without RO_2 and RO is insufficient to describe chemical evolution of a complex chemical system over time lasting longer than the few hours duration for a typical smog chamber experiment for SOA formation. Therefore, we assume that a large number of RO_2 in the chemical system are grouped into 10 surrogates: $X_{01}O_2$, $X_{02}O_2$, ..., $X_{10}O_2$. This same lumping approach is also applied to the RO for 10 surrogates: $X_{01}O$, $X_{02}O$, ..., $X_{10}O$. For convenience, we use $X_{ii}O_2$ to represent the 10 RO_2 surrogates and $X_{ii}O$ for the 10 RO surrogates.

As a result, hundreds and/or thousands of chemical species in a detailed chemical mechanism are reduced to only 30 organic compounds in 10 volatility bins in a simple reaction scheme, except the starting organics. Next, we will describe the simple unified volatility-based scheme (SUVS), which is constructed based on the protocols outlined in Sect. 3.1.

In this work, the SUVS is constructed for one single starting organic compound of α -pinene which undergoes atmospheric oxidation with three oxidants: OH , O_3 , and NO_3 . We assume the oxidation of the starting organic, represented by XHC , leads to the formation of 10 surrogate products: $X_{01}O_2$, $X_{02}O_2$, ..., $X_{10}O_2$. The reaction rate coefficients for the three initial reactions can be obtained from the literature or any reliable detailed chemical mechanism. A total of 30 stoichiometric coefficients from $ww(01)$ to $ww(30)$ for the 10 surrogates are assigned, as shown in reaction scheme Part 1 in Fig. 1.

Next, each $X_{ii}O_2$ radical can react with NO (as outlined in the generic reaction mechanism description in Sect. 3.1) to form an alkoxy (RO) radical or an organic nitrate ($RONO_2$). As shown in reaction scheme Part 2 in Fig. 1, we assume the newly formed product of RO belongs to the same volatility bin of $X_{ii}O_2$, i.e., $X_{ii}O_2 + NO \rightarrow X_{ii}O$. When the organic nitrates are formed from the reaction of $X_{ii}O_2 + NO$, we assume the volatility for organic nitrate are distributed into three neighboring volatility bins: $(ii-1)$ -th, ii -th, and $(ii+1)$ -th volatility bins. i.e., $X_{ii}O_2 + NO \rightarrow w(ii,1)S_{ii-1} + w(ii,2)S_{ii} + w(ii,3)S_{ii+1}$, where $w(ii,1)$, $w(ii,2)$,

and $w(ii,3)$ are the stoichiometric coefficients for the corresponding three products. Note that, information for compounds with different functional groups, such as hydroperoxy (ROOH), nitrate (RONO₂), peroxy nitrates (ROONO₂), alcohols (ROH), and carboxylic acid (RC(O)OH), are replaced by volatility indexes. The reaction rate coefficients for the 10 reactions are expressed as $k(01)$, $k(02)$, ..., $k(10)$ in Fig. 1. Also note that in this work, for simplicity, we use only three neighboring volatility products, rather than the whole suite of 10 products.

For demonstration purpose, we use compounds in the 5th volatility bin as reactants in the following description. Similarly, X₀₅O₂ reacts with RO₂. Because the permutation and cross reactions of the RO₂ lead to alkoxy radicals (RO) and carbonyls and alcohols, the products are then represented by X₀₅O radical and three condensable species: S₀₄, S₀₅, and S₀₆. This is expressed by the 15th reaction in the SUVS, and reaction coefficient for this reaction is shown as $k(15)$ in Part 3 in Fig. 1. The stoichiometric coefficients for the four products are $z(15)$, $w(15, 1)$, $w(15, 2)$, and $w(15, 3)$. RO₂ represents the total organic peroxide radicals in the chemical system, i.e., RO₂ = R₀₁O₂ + R₀₂O₂ + ... + R₁₀O₂. Meanwhile, the reaction between X₀₅O₂ and NO₃ leads to the formation of the corresponding alkoxy radical of X₀₅O with the reaction coefficient of $k(25)$ listed in Part 3 in Fig. 1.

X₀₅O₂ could also react with HO₂ and NO₂ to generate ROOH and RO₂NO₂, respectively. Both products (ROOH versus RO₂NO₂) are represented by condensable products of S₀₄, S₀₅, and S₀₆ with different set of stoichiometric coefficients: $w(35, 1)$, $w(35, 2)$, and $w(35, 3)$ for the ROOH; $w(45, 1)$, $w(45, 2)$, and $w(45, 3)$ for the RO₂NO₂.

Next, we discuss the chemical reactions of the alkoxy radicals (RO) in Part 4 in Fig. 1. As mentioned in Sect. 3.1, the degradation of RO leads to the formation of carbonyl and RO₂. In our simple chemical reaction scheme, the products of the carbonyls are represented by three condensable species of S₀₄, S₀₅, and S₀₆, and the corresponding stoichiometric coefficients are $w(55,1)$, $w(55,2)$, and $w(55,3)$. When the RO undergoes dissociation and isomerization to form peroxy radicals in the 65th reaction, we assume the corresponding products would be in three volatility bins: X₀₄O₂, X₀₅O₂, and X₀₆O₂. This treatment is different from the first 30th reactions in which RO₂ reacts with NO, RO₂, and NO₃ to form only one RO. This is because the dissociation and isomerization of the RO leads to significant changes of the chemical structures, which impacts the corresponding volatility bins.

Finally, we discuss the oxidation of the condensable products in the simple chemical reaction scheme. On the one hand, the condensable species, S₀₅, would undergo thermal decomposition or photolysis (the 75th reaction in the SUVS), and the products are the X₀₄O, X₀₅O, and X₀₆O. Note that the photolysis reaction rate coefficient ($k(75)$) for the 75th reaction is constrained by normalized sun intensity of SUN, which is 0 at night time and 1 at noon. On the other hand, when the condensable species S₀₅ reacts with OH and NO₃,

the products are organic peroxy radicals. In this case, they are X₀₄O₂, X₀₅O₂, and X₀₆O₂ in the 85th and 95th reaction in the SUVS. The above SUVS is constructed to describe the parameterized degradation of gas phase chemical reactions for a starting organic compound via the VBS speciation. In order to apply this simple scheme, equally important, we need to determine the values for 312 (=30 (w) + 30 (z) + 28 × 9 (w)) stoichiometric coefficients and 100 ($k(1)$ to $k(100)$) reaction rate coefficients.

Before we discuss how to determine the 412 unknown parameters, we briefly explain a simple treatment of inorganic chemistry and the HO_x/NO_x/O_x budget issue in the SUVS. Theoretically, the inorganic gas-phase chemistry could be coupled with the SUVS; however, our work here investigates the use of the SUVS as an ancillary mechanism that is affected by, but does not modify, the inorganic gas reactions. Our primary focus in this work is to develop a simple scheme for organic aerosol formation, we assume the concentrations profiles of the inorganic species (HO_x/NO_x/O_x) are known and obtained either from experimental observations or prior model simulations with explicit coupled inorganic-organic mechanisms. This part is explained in detail in the end of Sect. 4.2. In summary, SUVS does not modify the HO_x/NO_x/O_x budget.

Now two challenges face us before we apply this newly proposed SUVS in a regional air quality model. The first challenge is how to obtain time series of different volatility-based species under a wide range of experimental conditions for use in parameter estimation. Here, the detailed chemical mechanism of MCM v3.1 is used to generate time series of individual organic species. When the condensable organic species in the MCM v3.1 are lumped into 10 volatility-based compounds, we obtain the concentration profiles for the 10 volatility-based compounds in the SUVS system. The resulting accuracy of the parameterization is thus limited to that of the MCM.

The second challenge is on how to determine 412 unknown chemical parameters. Traditionally, a simple linear least square fitting method would be applied to determine reaction rate coefficients and stoichiometric coefficients for a one-step reaction scheme, like that in Stanier et al. (2008). Part of this challenge requires the construction of a set of test conditions sufficiently large in order to span the space of all atmospheric conditions. For such a large number of unknown parameters in the SUVS, a non-traditional method of genetic algorithms is utilized in this task (Sect. 5). In Sect. 4, we will introduce the settings for the MCM v3.1 under a wide range of conditions.

4 Master Chemical Mechanisms and volatility distribution

4.1 Settings for the Master Chemical Mechanism: creating a set of conditions for testing

The near-explicit Master Chemical Mechanism (MCM) v3.1 describes the degradation of α -pinene in a very detailed level, and includes the entire reaction sequence from the initial oxidation step to the final products, CO_2 and H_2O .

In this first application of the new SUVS for SOA formation, for simplicity, we focus our study on the oxidation of one single SOA precursor, the well-studied α -pinene, rather than all species in the entire oxidation system. Therefore, a subset of the MCM v3.1 describing α -pinene oxidation is used in a box model to study the chemical evolutions of 331 compounds with 976 reactions. Detailed information regarding the introduction and settings of the α -pinene oxidation system for a zero-dimensional box-model study can also be found in our previous studies (Xia et al., 2008; Slowik et al., 2010). In this work, the major input parameters for the box model include temperature, relative humidity, α -pinene emission rate, and initial concentrations of NO_x . It should be noted that photolysis rates for the core reactions in the MCM correspond to clear sky conditions on equinox at a latitude of 45°N and the midday $J(\text{NO}_2)$ is 0.073 s^{-1} .

In this study, temperature and relative humidity are constant values. The temperature is assumed to be fixed at a reference temperature of 298 K. The selection of the reference temperature is consistent with the reference volatilities, as used by Donahue et al. (2006). The relative humidity is also set to be at 10% all the time. This is considered to be a dry condition (Pathak et al., 2007). The third difference from our previous study is the α -pinene emission rate. A total of 21 logarithmly evenly spaced α -pinene emission rates are set from 0.5×10^7 to 8×10^7 molecule $\text{cm}^{-3}\text{ s}^{-1}$. Finally, a total of 27 logarithmly evenly spaced initial mixing ratios of NO are set from 2.5 ppbv to 1000 ppbv. The initial mixing ratios of NO_2 are set to be 1/3 of the NO mixing ratios for each scenario. Thus, a total of 567 ($=21 \times 27$) scenarios have been generated for the study of α -pinene oxidation. Each scenario is run for 3 days (72 h) starting from midnight. During the daytime, some compounds undergo photolysis reactions under natural sunlight. As a result, the concentration profiles for all 331 compounds can be calculated from the 3-day box model simulation.

Note that only gas phase chemical reactions are considered for this simulation, because we focus on the application of our new SUVS system on the formation of lumped condensable species in the gas phase chemistry. The gas/particle partitioning process and aerosol heterogeneous reactions are important but not included in this study. To obtain the concentration profiles for the 10 lumped condensable species from S_{01} to S_{10} in the SUVS, we first need to know the volatility of each stable organic species. Then, all stable

species are lumped into the 10 species according to their individual volatilities.

4.2 Volatility distribution

In this subsection, we first briefly introduce the concept of effective saturation concentration and its estimation method via saturation vapor pressure (SVP). Then, we describe volatility distributions for 10 lumped species from a typical α -pinene oxidation scenario.

Rather than the traditional term of partitioning coefficient, Donahue et al. (2006) proposed to use an alternative term C_i^* ($\mu\text{g m}^{-3}$), effective saturation concentration, to study the gas/particle partitioning behavior for the i -th product. The C_i^* can be expressed theoretically via Pankow's (1994) partitioning theory as:

$$C_i^* = \frac{1}{K_{p,i}} = \frac{10^6 M_i \zeta_i' P_{L,i}^o}{760RT} \quad (3)$$

where R ($8.206 \times 10^{-5} \text{ m}^3 \text{ atm K}^{-1} \text{ mol}^{-1}$) is the gas constant, T (K) is the temperature, M_i (g mol^{-1}) is the molecular weight for the i -th compound, $P_{L,i}^o$ (torr) is the saturation vapor pressure (SVP) of the i -th pure compound at temperature T , and ζ_i' is a mole-fraction-scale-based activity coefficient for the i -th compound in the condensed phase.

In general, the activity coefficients of the compounds in the organic aerosols are in the range of 0.3 to 3.0 (Seinfeld and Pankow, 2003). However, as adopted by many researchers (Kamens et al., 1999; Jenkin, 2004; Xia et al., 2008) as a first approximation, the activity coefficients are also assumed to be unity in this work for the calculation of the C_i^* for each stable species. Once the SVP and M_i from Eq. (3) are known at a given temperature T , we can determine the C_i^* , an intrinsic property of compound i for an ideal solution. Next, we will describe the estimation method for the SVP outlined in Schwarzenbach et al. (2003).

By making an assumption of the linear relationship for the vaporization enthalpy at different temperatures, integrating the Clausius-Clapeyron equation from T_b (normal boiling point) to T , and using Trouton's rule, the SVP can be expressed as

$$\ln \frac{P_L^o}{760} = -K_F(4.4 + \ln T_b)[1.8\left(\frac{T_b}{T} - 1\right) - 0.8\ln\left(\frac{T_b}{T}\right)] \quad (4)$$

where T_b (K) denotes the normal boiling point, T (K) designates the temperature of interest, K_F (no unit) is a Fishtine factor (Fishtine, 1963), and the P_L^o (torr) represents the SVP. The T_b is estimated by using a group contribution method (Joback and Reid's (1987) fragmentation method, modified and extended by Stein and Brown (1994), and the K_F factor is estimated by the method given in Fishtine (1963) or Sage and Sage (2000). Overall, this SVP estimation method by using Eq. (4) is comparable to the universal functional activity coefficient (UNIFAC P_L^o) method (Asher et al., 2002)

in estimation of the SVP against measured data (Xia et al., 2008). The Eq. (4) was also used to estimate the SVP for the condensable species from our previous study (Xia et al., 2008).

Only 180 out of the 331 compounds in the chemical system are stable organic species, the remaining compounds are either inorganic compounds or organic intermediates. As mentioned early, the volatility for each stable organic compound is estimated by Eq. (3) at the reference temperature of 298 K. In this part, our goal is to group the 180 stable organic species into 10 lumped ones.

To represent ambient organic aerosol, the volatilities for the 10 lumped species at the reference temperature of 298 K are separated by one order of magnitude, spanning from 10^{-2} to $10^7 \mu\text{g m}^{-3}$. In this work, we have pre-defined 10 volatility bins. The n -th volatility bin covers the volatility range from $10^{(n-3)-0.5} \sim 10^{(n-3)+0.5} \mu\text{g m}^{-3}$, where $n = 2, 3, \dots, 10$. When $n = 1$, the first volatility bin covers the volatility range no more than $10^{-1.5} \mu\text{g m}^{-3}$. With the given setting, a linear interpolation within the bin is used to determine the volatility bin location for each of the 180 species.

By using the linear interpolation lumping method, all 180 condensable species in the subset of MCM for α -pinene oxidation are grouped into the 10 lumped species (gas phase). Figure 2 shows diurnal profiles for the 10 lumped species from one scenario lasting for 72 h when there is constant α -pinene emission.

Starting from midnight, the number concentrations of 10 lumped compounds start to accumulate from zero and there are different patterns for the diurnal changes. For example, in the daytime starting from sunrise at 30th hour, the concentrations for the 5th, 8th, and 9th lumped species undergo large increases. But the concentrations for the 6th and 7th lumped species start to decline until sunset at the 42nd hour. This kind of diurnal change could be different for different scenarios. In Sect. 6, we will examine these distinctive diurnal change patterns mainly due to photolysis changes.

In this section, we have obtained diurnal profiles, termed “experimental data”, for the 10 lumped species. Additionally, inorganic species are vital for the SUVS, and the hourly profiles for these species are obtained directly from the MCM model simulations. A linearly interpolated method is applied to the hourly profiles of the 10 lumped species and the inorganic species within each our-hour period. Next, we will introduce genetic algorithms for estimation of the 412 unknown parameters in the SUVS system.

5 Genetic algorithms (GAs)

5.1 Brief review of genetic algorithms

The underlying ideas of genetic algorithms (GAs) were inspired by the mechanism of natural selection. A complete description of GAs techniques can be found in Goldberg (1989)

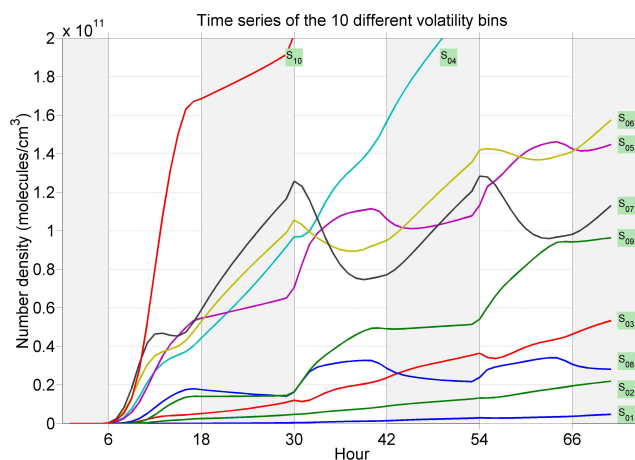


Fig. 2. Diurnal profiles for the 10 lumped species from one α -pinene oxidation scenario lasting for 72 h. The shaded areas indicate night time.

and Michalewicz (1999). A simple and practical introduction on how to incorporate these ideas in a computational setting can be found in Haupt and Haupt (2004). GAs has been applied mostly in mathematics and physics. As a new tool, GAs have also been recently used in the atmospheric sciences with various applications.

Specifically, GAs were applied to determine aerosol size distributions (Lienert et al., 2001, 2003) and aerosol refractive index (Barkey et al., 2007) from polar nephelometer data. In another study, the GAs were coupled with an atmospheric dispersion model to characterize pollutant emission locations, times, and quantities (Haupt, 2005; Allen et al., 2007). GAs were applied to optimize a set of physical and computational parameters in both the 5th-generation PSU/NCAR Mesoscale Model (MM5) and the Regional Atmospheric Modeling System (RAMS) for improved agreement between model simulations and observational or synthetic data (Lee et al., 2006; O’Steen and Werth, 2009). This technique was utilized to calibrate seasonal PM_{10} emission inventories in Beijing to improve the Models-3/Community Multiscale Air Quality (CMAQ) model performance (Li et al., 2010). Similar to our task in this paper, GAs were employed in chemical kinetics for determining parameters in various chemical systems (Carroll, 1996; Polifke et al., 1998; Harris et al., 2000; Tsuchiya and Ross, 2001; Elliott et al., 2004).

Fundamentally, GAs are used to solve an optimization problem by evolving the best solution from an initial set of completely random guesses. In other word, GAs are an adaptive heuristic search method. A typical GAs requires:

1. A fitness function f to reflect the objective optimization problem (minimum or maximum) at hand. A fitness function includes one or several variables that a system needs to determine. The variables are called

“parameters” in the GAs. The fitness function is used to distinguish between “bad” and “good” solutions. We shall use a simple example to explain concepts and methodologies used in the GAs field. An example could involve finding the maximum value of $f = 200 - (x - 5)^2 - (y - 10)^2$, and the corresponding two unknown parameters are (x, y) , where $0 \leq x, y \leq 15$.

2. An encoding method to represent the solution domain. There are several different encoding methods, such as binary encoding, gray encoding, and float-point encoding. Each parameter is encoded as a string (or termed a gene). For example, (8, 12) could be described by two tentative 4-digit binary genes as (1000, 1100). In GAs, one individual represents a set of all unknown parameters, and it is expressed as a chain where bits standing for all parameters are lined up to form a binary string. The chain is called genetic sequencing, or chromosome in GAs. As a result, (8, 12) is encoded as (10001100).

After the fitness function and the encoding method are determined for a GAs. A 7-step procedure can be used to apply the GAs for a given task.

1. Initialize possible solutions. Individual possible solutions (termed Individuals) are given from random guesses to form an initial population. A population is an array of individuals. The size of the population depends on the number of unknown parameters in the problem. In this simple example, the population size is tentatively set to be four, and four initial randomly generated individuals are: (1, 2), (2, 13), (12, 4), and (15, 15).
2. Evaluate the fitness value of each individual in this population. The fitness values for the four initial individuals are: $f(1, 2) = 120$; $f(2, 13) = 182$; $f(12, 4) = 115$; and $f(15, 15) = 75$.
3. Select best-fit individuals for reproduction of new individuals. In this example, two individuals (2, 13) and (1, 2) are singled out as “Parent 1” and “Parent 2” for mating because of their high fitness values of 182 and 120.
4. Generate new individuals via two genetic operators: crossover and mutation. The two operations can be performed only when all individuals are encoded as genes.
 - (a) Crossover: One-point crossover is the simplest crossover method, in which one same point is selected on both parents and the data beyond that point are swapped from two parents. In this example, the fifth point is selected for crossover, which leads to

the generation of two new children (2, 10) and (1, 5).

Parent 1	(2, 13)	(00101 101)
Parent 2	(1, 2)	(00010 010)
		↓
Child 1	(2, 10)	(00101 010)
Child 2	(1, 5)	(00010 101)

- (b) Mutation: Mutation is applied to a single parent. Similar to the biogenic mutation, an arbitrary bit in the genetic sequencing will be changed from its original state. For example, the first bit from “Parent 1” undergoes mutation and a new child (offspring) (10, 13) is generated.

Parent 1	(2, 13)	(00101101)
		↓
Child 1	(10, 13)	(10101101)

5. Evaluate the fitness of the new individuals.
6. Replace the least-fit individuals with the new individuals to form a new population.
7. Repeat the procedures from step 2 to 6 until termination.

At each iteration, the GAs perform crossover and mutation operations with two different probabilities ($P_{\text{crossover}}$ and P_{mutation}) to produce a new population. Each successive population is called a new generation. The GAs usually uses some of the following conditions to determine when to stop: (1) maximum number of generation; (2) time limit; (3) sufficient fitness achieved.

In GAs, elitism is a commonly used technique where one or more of highest fitness individuals are copied, unchanged, from one generation to the next. Elitism could rapidly increase the GAs performance, because it prevents losing the best found solution to date. Figure 3 shows a schematic representation of the procedures outlined above for a basic GAs. Note that the number of parameters for the example shown above is only two. It could vary from one to hundreds (412 for our SUVS case), and even to billions (Goldberg et al., 2007) in actual applications.

As Elliot et al. (2004) summarized that when contrasted with traditional gradient-based search methods, GAs do not require knowledge of the gradient of the fitness functions. This makes GAs particular suitable for problems, such as this SUVS reaction scheme, where the analytic expression of the fitness function is not known.

5.2 Settings of the GAs for SUVS

5.2.1 Objective functions of the GAs

Fan et al. (2004) demonstrated that the design of the fitness function is instrumental in performance improvement for the GAs. In our case, 10 ambient-like scenarios were selected

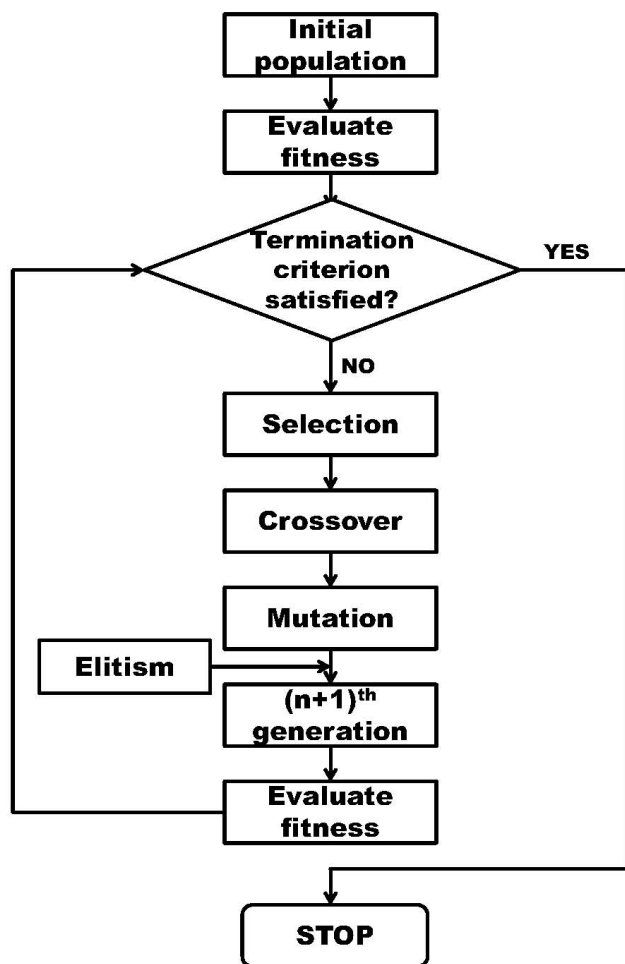


Fig. 3. Schematic representation of procedures used for a simple GAs application.

for testing the SUVS. Each scenario ran for 72 h starting from midnight. The profiles for 11 species obtained from the SUVS were compared with the profiles for the 11 species from the MCM v3.1. The 11 species included the 10 lumped species (S_{01} , S_{02} , ..., and S_{10}) and the total RO_2 species. To avoid comparison of very low (near zero) concentrations of the species in the first few hours of model runs, the timing used for the comparison are selected from a tentative starting time at 10th hour to the end time at 72nd hour. The fitness function defined for this study is:

$$f = \frac{1}{10^{-8} + \frac{1}{N_{SC}} \frac{1}{N_{SP}} \sum_{i=1}^{N_{SC}} \sum_{j=1}^{N_{SP}} \sqrt{\frac{\frac{1}{N_h} \sum_{k=starthour}^{endhour} (x_{i,j,k}^{calc} - x_{i,j,k}^{orig})^2}{\frac{1}{N_h} \sum_{k=starthour}^{endhour} x_{i,j,k}^{orig}}} \quad (5)$$

where

- $x_{i,j,k}^{calc}$ represents calculated number density of the j -th lumped species from i -th scenario at k -th hour by using the SUVS system;
- $x_{i,j,k}^{orig}$ is the corresponding original number density of the lumped j -th species from i -th scenario at k -th hour by using the full chemical mechanism of the MCM v3.1;
- N_{SC} (=10) represents total number of selected scenarios for testing; N_{SP} (=11) is the total number of species used for comparison; and N_h (=endhour–starthour + 1) denotes the total hours for this comparison, in which the starthour is 10 and the endhour is 72.
- The maximum value of the function f is 10^8 , which corresponds to a perfect fit. The constant is added into the term to avoid numerical overflow.

- The term $\sqrt{\frac{\frac{1}{N_h} \sum_{k=starthour}^{endhour} (x_{i,j,k}^{calc} - x_{i,j,k}^{orig})^2}{\frac{1}{N_h} \sum_{k=starthour}^{endhour} x_{i,j,k}^{orig}}}$ in the fitness function (Eq. 5) is conventionally called the coefficient of variation for the root mean square deviation – CV(RMSD). For simplicity, we define this term as $ERROR_{i,j}$, i.e.,

$$ERROR_{i,j} = \frac{\sqrt{\frac{1}{N_h} \sum_{k=starthour}^{endhour} (x_{i,j,k}^{calc} - x_{i,j,k}^{orig})^2}}{\frac{1}{N_h} \sum_{k=starthour}^{endhour} x_{i,j,k}^{orig}} \quad (6)$$

Three additional errors can be formed via $ERROR_{i,j}$ as:

$$g_{SP,j} = \frac{1}{N_{SC}} \sum_{i=1}^{N_{SC}} ERROR_{i,j} \quad (7)$$

$$g_{SC,i} = \frac{1}{N_{SP}} \sum_{j=1}^{N_{SP}} ERROR_{i,j} \quad (8)$$

$$g = \frac{1}{N_{SC}} \frac{1}{N_{SP}} \sum_{i=1}^{N_{SC}} \sum_{j=1}^{N_{SP}} ERROR_{i,j} = \frac{1}{f} - 10^{-8} \quad (9)$$

where $g_{SP,j}$ is the mean CV(RMSD) for the j -th species from all N_{SC} scenarios, $g_{SC,i}$ is the mean CV(RMSD) for all N_{SP} species in the i -th scenario, and g is the mean CV(RMSD) for all 10 scenarios and 11 species. The simple relation between the mean error of g and the fitness function f is also indicated in Eq. (9). We will discuss the relationship between $g_{SC,i}$, $g_{SP,j}$, and g in more details in Sect. 6.

5.2.2 Parameters for the GAs

As mentioned in Sect. 5.1, the unknown variables could be encoded with a binary representation or floating point representation. In general, the floating point approach is better than the binary representation in that the floating point approach is faster and it can cover larger domains at a higher resolution than the binary approach (Elliott et al., 2004). In this work, the floating point approach is applied and each variable is represented by a 6-digit long number from 0.000000 to 0.999999. Among the total of 412 unknown variables, 312 of them are the stoichiometric coefficients, which vary from 0.0 to 1.0. It is straightforward to use the floating point encoding method for them. Note that a sensitivity test showed that the computation cost is far too expensive if the 312 stoichiometric coefficients are assumed to be in the wider range from 0.0 to 10.0. In this work, they are assumed in the range from 0.0 to 1.0. The remaining 100 variables are for reaction rate constants (from $k(1)$ to $k(100)$).

In the MCM, organic reaction rate constants differ for different functional groups. Overall, they vary from 10^{-16} to 10^{-10} , except for those in reaction of $\text{RO} \rightarrow \text{RO}_2$. In the SUVS, for simplicity, 90 out of the 100 reaction rate constants are consistently set in the same range of those from the MCM. Only $k(61) \sim k(70)$ for the type of $\text{RO} \rightarrow \text{RO}_2$ reaction have fast reaction rate constants in a range of 10^5 to 10^7 . To effectively represent reaction rate constant by a 6-digit long number, the reaction rate constant $k(i)$ has been converted to be $kk(i)$ as follows,

$$kk(i) = \frac{1}{(-10)^{-(-16)}} (\log_{10} k(i) - (-16)) \quad (10)$$

$$= \frac{1}{6} (16 + \log_{10} k(i)) \quad (i = 1, \dots, 60, 71, \dots, 100)$$

and each converted $kk(i)$ is in the range between 0.0 to 1.0. Similar conversion can be performed for the remaining 10 parameters from $k(61)$ to $k(70)$.

In summary, as demonstrated by Elliott et al. (2003), incorporation of realistic physical bounds on the 312 stoichiometric coefficients and 100 reaction rate constant ensures minimization of non-physical solutions using GAs. Note that reaction rate coefficients in the SUVS might be different from those in the MCM even for the same reaction type, because the species from the MCM are lumped into the SUVS. The reaction rate coefficients ranges, however, are physically reasonable. The genetic operators and the parameters used in this work were taken to be as follows:

- population size $N_{\text{population}} = 5000$
- crossover probability $P_{\text{crossover}} = 0.80$
- mutation probability $P_{\text{mutation}} = 0.004$
- elitism, elitism parameter $n_{\text{elitism}} = 2$

The population size affects the GAs performance. The minimum requirement is that the population size be set at least

the value of the number of variables, so that the individuals in each population span the space being searched. In general, increasing the population size enables the GAs to search more points and thereby obtain a better result. However, the larger the population size, the longer the GAs takes to compute each generation. In this work, the number of variables is 412; the population size is 5000, which is more than a factor of 12 larger than the number of variables.

We would like to mention that the computational cost is very expensive by setting the population size at 5000, because within each generation, or called iteration, there are 5000 different sets of 412 parameters, each set of data running 10 selected scenarios for 72 h. In particular, the computational cost is especially heavy when some set of 412 parameters leads to highly singular ordinary differential equations (ODEs) for the gas phase chemistry of the SUVS system. We will discuss more about computational cost in Sect. 6.4.

5.3 Numerical methods

Three numerical methods are involved for this work: (1) the MCM gas phase chemistry for α -pinene oxidation; (2) the SUVS in Fig. 1; and (3) the implementation of GAs.

The integration method used for the MCM gas phase chemistry is a Gear-type solver of ODEs via the FACSIMILE 3.0 integrator. A Rosenbrock ODE solver is used for the integration of the SUVS via a Kinetic PreProcessor (KPP) package (Damian et al., 2002; Sandu and Sander, 2006). Finally, to deal with the huge computational burden for the GAs, a Message Passing Interface (MPI) enabled the parallel genetic algorithm package (Charbonneau, 1995; Metcalfe and Charbonneau, 2003), named as MPI-PAKAIA, to be selected and implemented with an IBM power4 supercomputer system for solving this heavy computational cost problem.

6 Model results and discussions

As outlined in Fig. 5, the GAs give 5000 different set of randomized initial solutions to our problem. The fitness function for each of 5000 populations is then evaluated. Based on the performance, “good” candidates with higher fitness are selected for generating new populations for the next generation. The populations in the new generation undergo the same evaluation. This process is repeated until a termination criterion is met for a stop. In this work, the calculation is stopped at the 21,549th generation because no further improvement was found after a huge computational cost. For convenience, we call the 21,549th generation as the final generation in this paper.

6.1 Evolution of the best individual

At each generation, we can find the worst, the median, and the best individuals. Figure 4a shows the evolutions for the three individuals until the final 21,549th generation. The best

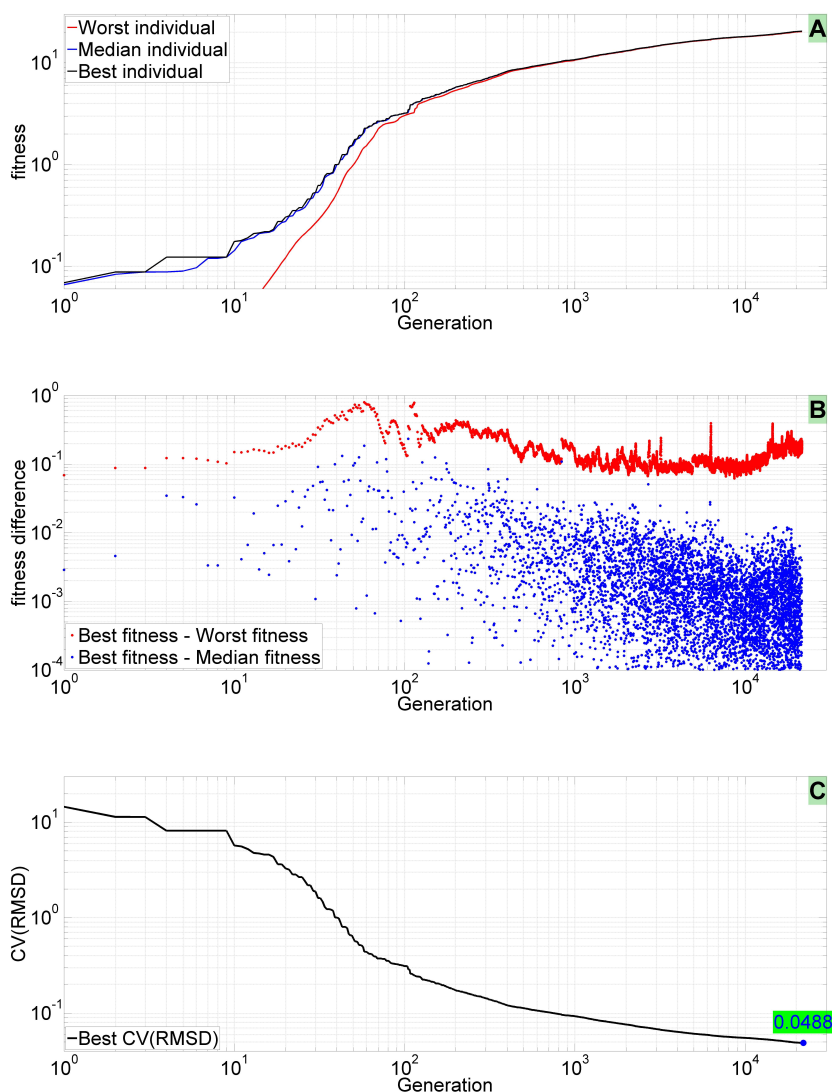


Fig. 4. (A) Evolution of the worst, median, and the best individuals from the 1st generation to the final 21,549th generation; (B) The fitness differences between the best and the worst individuals, and between the best and median individuals; (C) Evolution of the best CV(RMSD) from 14.514 at the 1st generation to 0.0488 in the final 21,549th generation.

individual increased monotonically from 0.0689 in the first generation to 20.5099 in the final generation (larger numbers mean a more successful fitness for the given generation). Meanwhile, the fitness for the median individual and the worst individual increased from 0.0660 and 1.0907×10^{-10} (not shown) in the first generation to 20.5099 and 20.3247 in the final generation. Generally, Fig. 4a shows that larger fitness differences among the three individuals exist in the first 100 generations. Further, Fig. 4b shows the evolutions of two fitness differences: the fitness difference between the best individual and the worst individual; and the fitness difference between the best individual and the median individual. As expected, the fitness difference between the best and the worst individuals is always higher than the fitness differ-

ence between the best and the median individuals. Overall, these differences are small and less than 1.0, and all individuals within each generation converge. Note that the fitness function is defined in Eq. (5) and it is related but still different from the CV(RMSD) in Eq. (9).

Intuitively, it is meaningful to check the errors, rather than the fitness function. Figure 4c shows the evolution for the mean error of g , which is shown in Eq. (9) as the mean CV(RMSD) for all 11 species over 10 scenarios, for the best individual from 14.519 (fraction) in the first generation to 0.0488 (fraction) in the final generation. This means the average error for the species from all scenarios is less than 5% by using the best set of data from the final generation.

6.2 Comparison of concentration profiles of individual species

Next, we are going to compare individual species number concentration profiles between the full mechanism (the MCM v3.1) and the SUVS. The profiles from the later one are obtained by using the GAs generated best set of data from the final generation. For a full list of the final SUVS reactions, including the GAs-derived 412 parameters and mean molecular weights of the 10 lumped species (S_{01} – S_{10}), please refer to Supplement.

Due to the settings outlined in Sect. 4.1 for different constant α -pinene emissions and different initial concentrations of the NO_x , the diurnal changes for the 11 lumped species from the full mechanism, as shown in Fig. 5, can be grouped into four types:

1. Concentrations are always increasing with little fluctuation. These are the low volatility species from S_{01} to S_{04} . Black solid lines in Fig. 5a show this pattern for a representative species of S_{02} over 10 different scenarios.
2. Strong diurnal fluctuations for species with intermediate volatility. Overall, species concentrations are increasing on the daily basis. This corresponds to the species from S_{05} to S_{08} . Solid lines in Fig. 5b unfold this kind of diurnal changes for a representative species of S_{07} . For example, in scenario 5 and 6, peak concentrations of S_{07} on the second day are comparable with those peaks on the third day. Note that diurnal variations in this work are driven mainly by the photolysis changes.
3. Concentrations are always increasing, but the increment is more intense at daytime than that at night time. S_{09} and S_{10} are the two species following this pattern. Figure 5c shows the profiles for S_{09} .
4. Strong diurnal changes with high concentrations at daytime and very low concentrations at nighttime. This pattern is slightly different from type 2, in which species concentrations are increasing on the daily basis. This pattern, as shown in Fig. 5d, is only for the total RO_2 , and the concentrations are extremely low (near zero) at night time.

Figure 5a–d also shows the comparison of the diurnal profiles for the four types of representative species between the full mechanism and the final generation SUVS. In order to qualitatively describe the differences between the two mechanisms, the CV(RMSD), i.e. the $g_{\text{SP},j}$ in Eq. (7), for the j -th species are calculated over 10 scenarios. For example, the $g_{\text{SP},2}$, a measure of the fractional error for the S_{02} , is 0.017. This suggested that the SUVS is able to accurately describe the diurnal changes and concentration profiles for the S_{02} over 10 different scenarios with an average error of 1.7%. Figure 5a displays the S_{02} hourly profiles from the two mechanisms, full mechanism versus the

SUVS, over the 10 scenarios. There is very good agreement with each other because of a low error at 1.7%. Furthermore, the SUVS is also capable of capturing the strong diurnal changes for the second type of representative species of the S_{07} . The calculated error of $g_{\text{SP},7}$ is 0.046. Although Fig. 5b shows that the calculated small error at 4.6% stems mainly from the peaks and troughs of the diurnal cycles for a few scenarios, the timings of the turning points for the diurnal changes are very accurate. Likewise, the error of $g_{\text{SP},9}$ for the S_{09} is 0.049. Figure 5c illustrates that the SUVS is able to describe big increases at daytime for the S_{09} , and the timings for the small errors are coming from night time. Finally, the total RO_2 from 10 different scenarios are compared. In the SUVS, we have a relationship of total $\text{RO}_2 = X_{01}\text{O}_2 + X_{02}\text{O}_2 + \dots + X_{09}\text{O}_2 + X_{10}\text{O}_2$. Each $X_{ii}\text{O}_2$ corresponds to a summation of several individual organic peroxy radicals from the full chemical mechanism of the MCM v3.1. One straightforward method is to compare the individual $X_{ii}\text{O}_2$ between the full mechanism and the SUVS. But in this work by using GAs, we simply compare the total RO_2 only, rather than 10 individual $X_{ii}\text{O}_2$, ($ii = 1, \dots, 10$), from the two mechanisms. The GAs is able to find a good solution for the SUVS to describe the strong diurnal changes of the total RO_2 shown in Fig. 5d.

The errors for each of 11 species from 10 scenarios are summarized in Table 1. Because the fitness function in Eq. (5) includes the errors contributed from all 11 species and 10 scenarios, to achieve an overall good performance at the error of 0.0488, the errors for some species are higher than the average values. For example, the maximum error of 0.092 is for the S_{08} . An examination of the model performance for this species (not shown here) indicates that improvement for the GAs is still needed, although the SUVS is able to capture diurnal changes and peak values for S_{08} among the 10 different scenarios. On the whole, the fractional errors for the 6 out of the 11 species are less than 0.05, in particular, the errors for the three species (S_{02} , S_{06} , and S_{10}) are less than 0.03.

Overall, the new SUVS with the best set of 412 parameters from the final generation of GAs is able to describe the evolution of the 10 lumped species and the total RO_2 . Figure 6a presents the distribution of a total of 7920 ($= 72\text{ h} \times 11\text{ species} \times 10\text{ scenarios}$) point-to-point errors $(x_{i,j,k}^{\text{calc}} - x_{i,j,k}^{\text{orig}})/x_{i,j,k}^{\text{orig}}$ from all 11 species over the 10 scenarios. The x -axis in Fig. 6a is the normalized number density, which is calculated via normalization of one species concentration against the maximum concentration for the same species from the same scenario within the 72-hour period. Figure 6a shows that the point-to-point errors are higher at lower normalized number density. Most of the point-to-point errors lie within the ± 0.10 (fraction) lines. The pattern for the point-to-point errors against the normalized number density looks like an overall shape (\triangleright). Larger errors occur for low concentrations and smaller errors for

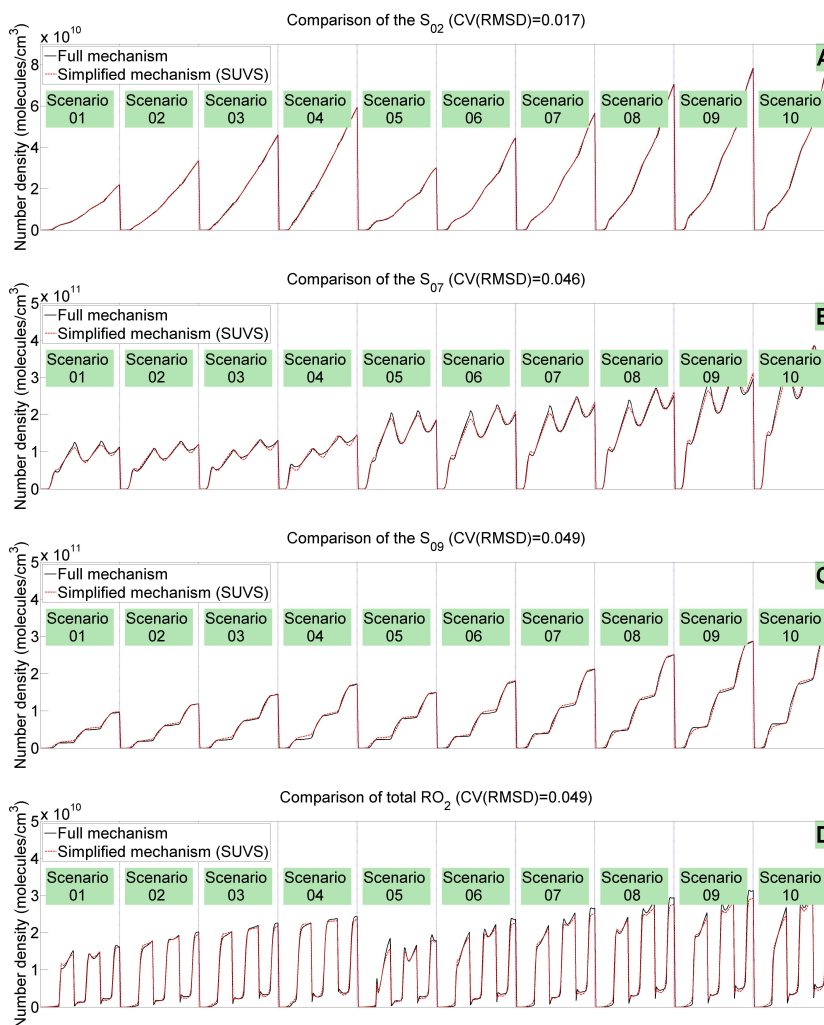


Fig. 5. Comparison of number density profiles for four representative species (S_{02} , S_{07} , S_{09} , and RO_2) between the full mechanism of the MCM v3.1 and the simplified chemical mechanism (SUVS), which is obtained by using the genetic algorithms generated best set of data from the 21,549th generation. The calculated CV(RMSD) for the four species are 0.017, 0.046, 0.049, and 0.049.

Table 1. The CV(RMSD) ($g_{SP,j}$ ($j = 1, \dots, 11$)) for the 11 species from 10 different scenarios.

Species	S_{01}	S_{02}	S_{03}	S_{04}	S_{05}	S_{06}	S_{07}	S_{08}	S_{09}	S_{10}	RO_2
$g_{SP,j}$	0.066	0.017	0.063	0.056	0.059	0.022	0.046	0.092	0.049	0.029	0.049

higher concentrations. This is consistent with the use of the CV(RMSD) in the fitness function, because CM(RMSD) is computed through normalization of RMSD against the mean concentration of the same species from one entire scenario.

In order to qualitatively describe the point-to-point distributions, we first sum the number of points located within each grid cell in Fig. 6a. Here, the grid cells are defined with 0.1×0.1 grid configuration. Next, when the number for the

points in each grid cell is normalized by the total number of points (7920), a point density is obtained. Figure 6b exhibits the distributions of the point density for the point-to-point errors in each grid cell. For example, it shows that 80.4 % of the errors are bounded within the ± 0.10 (fraction). Furthermore, it demonstrates that the majority of the normalized number densities lie in the range between 0.2 and 0.3.

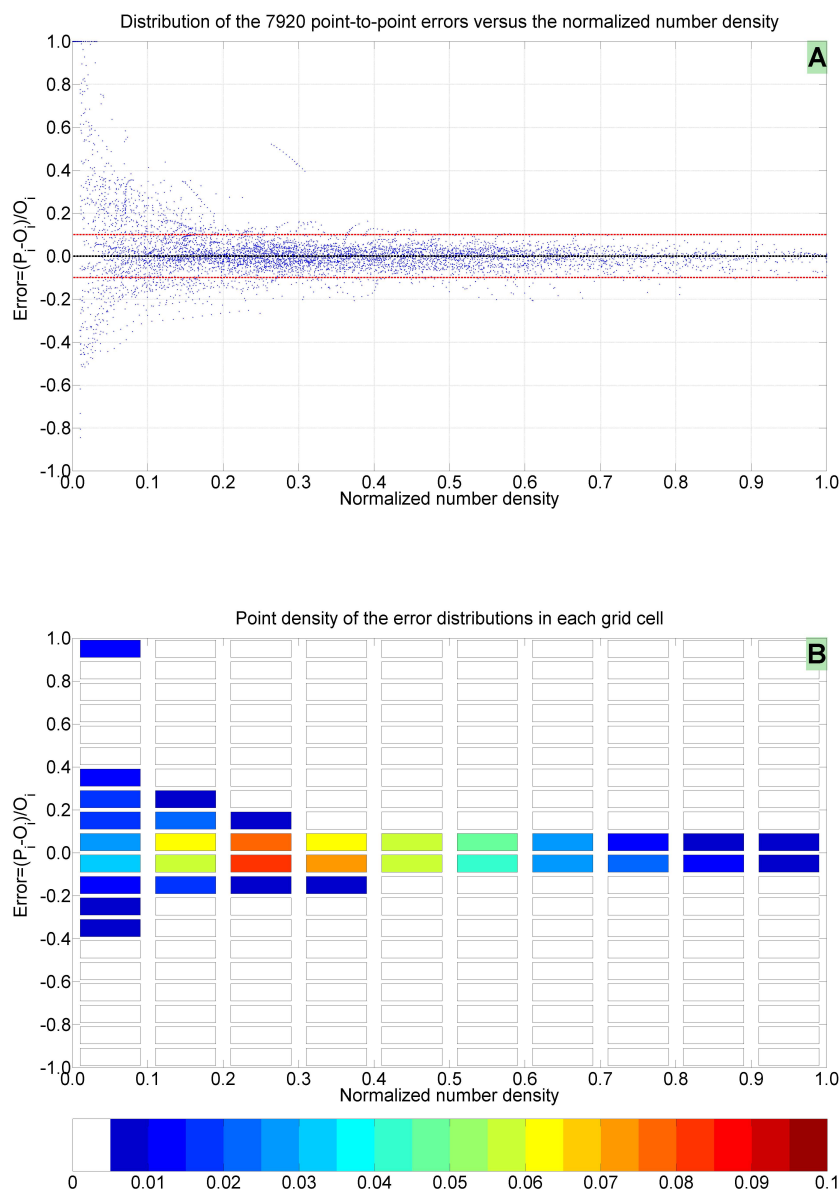


Fig. 6. (A) The distribution of the 7920 point-to-point errors between the SUVS (P_i) and the full mechanism (O_i) against the normalized number density for the 11 species over 10 scenarios. The dashed lines correspond to the fractional errors at 0.00 (black), ± 0.10 (red). (B) The point density of the error distribution in each grid cell. The point density is defined as a ratio between the number of points located in each grid cell and the total number of points.

A sensitivity test demonstrated that when a point-to-point error $(x_{i,j,k}^{\text{calc}} - x_{i,j,k}^{\text{orig}})/x_{i,j,k}^{\text{orig}}$, rather than the CV(RMSD), is directly included in the design of fitness function, the performance for the derived SUVS cannot be comparable with the CV(RMSD). During the test, a large computational cost is associated with the excessive time the GAs require in fine-tuning the parameters, in order for the point-to-point error to be reduced even when the number density is already very low. The point-to-point error design method thus treats low concentrations and high concentrations equally. In this work,

however, we focus heavily on accurate prediction of higher concentrations, rather than low concentrations. Our objective in this work favors the selection of CV(RMSD) over the point-to-point error. This reinforces the importance of a good design of the fitness function, which sometimes is considered more of “an art than a science”.

6.3 Performance of the SUVS for additional tested scenarios

It is a common practice to check the validity of the new system or mechanism against a wider range of conditions. For example, Asher et al. (2002) first exercised 10 % of observational data to determine parameters in their new formula to calculate vapor pressure of organic compounds. Later on, the dataset from the remaining 90 % were evaluated against this new formula. Similarly, Xia et al. (2009) derived a reduced chemical mechanism for α -pinene oxidation system via five traditional mechanism reduction techniques with data from 108 selected scenarios. This reduced mechanism was later evaluated against an additional 135 scenarios from a wider range of conditions.

Likewise, the SUVS is evaluated against an additional 557 scenarios. In the previous section, we focused our analysis on the individual species via $g_{SP,j}$ from Eq. (7). In this part, another metric, $g_{SC,i}$ from Eq. (8), is employed to evaluate model performance for each individual scenario. Fig. 7a shows the performance of $g_{SC,i}$ from a total of 567 scenarios. Although a maximum of one is set for the color scale in Fig. 7a, the maximum and minimum fractional errors among all 567 scenarios are 4.5951 and 0.0399, respectively. The large errors are caused when the SUVS is run under atmospherically unrealistic conditions. For example, the maximum error of 4.5951 corresponds to an extremely high NO_x condition, in which the average NO_x concentration is as high as 1,322.7 ppbv. Meanwhile, the calculated error of $g_{SC,i}$ would be larger than 1.00 when the average α -pinene concentrations are at least 200 ppbv and the average NO_x is less than 0.7 ppbv. These large errors contrast to smaller errors for 10 selected scenarios highlighted with larger black circles. In general, fractional errors decrease gradually towards small values with more blue colors around 0.10 when the corresponding average concentrations of the α -pinene and NO_x come to close to those of the 10 selected scenarios.

In order to identify SUVS's range of influence, Fig. 7b shows the additional 165 scenarios with fractional errors at most 0.10. Although the 10 selected scenarios represent only limited range of conditions, Figure 7b indicates that the SUVS is still valid over a wider range of ambient conditions. For example, when the average α -pinene is 10 ppbv and the average NO_x is 20 ppbv, the calculated fractional error for this scenario is 0.084. Overall, the colors for the errors in Fig. 7b change gradually when the errors increase from the smallest 0.0399 to 0.10.

In summary, 10 randomly selected ambient-like scenarios are used for the GAs analysis so that the SUVS could reproduce the full mechanism of MCM v3.1. By using the best parameters obtained from the GAs, the fractional errors for not only the original 10 selected scenarios, but also an additional 165 scenarios are at most 0.10. Here, a threshold of 0.10 is tentatively used as a criterion for judging whether the error is acceptable. If a different set of 10 scenarios are se-

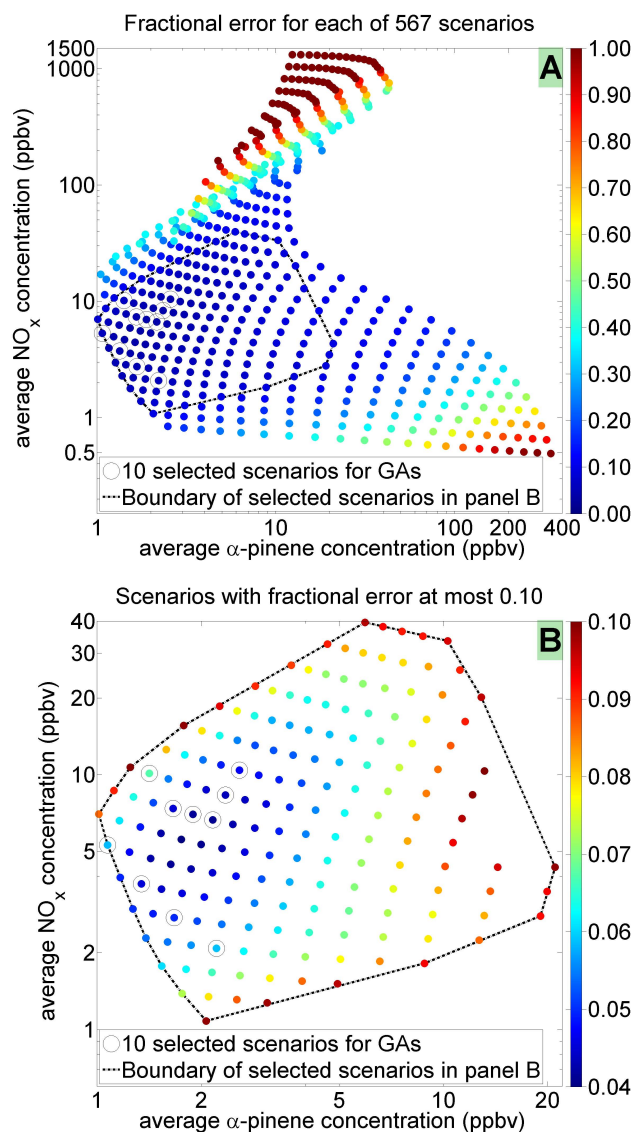


Fig. 7. (A) Errors for each of 567 scenarios under a wide range of conditions. The 10 randomly selected ambient scenarios are used in the GAs to derive a best set of parameters for the SUVS; (B) a total of 175 scenarios are identified with the fractional error at most 0.10. Note that color scales for the fractional errors differ between the two panels, and the smaller numbers (more blue colors) correspond to more accurate simulations.

lected for the GAs, the pattern for the error map shown in Fig. 7a might be different. This suggests that the SUVS's range of influence depends strongly on the scenarios chosen for the GAs.

6.4 Computational cost

Finally, we will discuss the computational cost for implementing the GAs for this work. The computational cost is expensive due to three reasons.

1. First, the chemical mechanism (SUVS) is run over 3 days for 10 different scenarios. The method is not a simple computation, rather we need to solve the ODEs when a chemical system is integrated over three days for 10 scenarios. It is equivalent to integrate a chemical system for 30 days when only one set of parameters are used. Sometimes, when a set of parameters is unrealistic, the ODEs for the SUVS become singular and that tremendously increases the computational burden.
2. The second factor is due to as large as 412 unknown variables for determination. When the number of variables is large, a high value of population size is required for the GAs. The rule of thumb is that the population size should be at least 10 times of the number of variables. In our case, we choose 5000, and we used an IBM power4 supercomputer to perform the calculation.
3. The third reason is related to the second one in that the GAs would not be able to find a convergent solution with a good fit within 1,000 generations because of the 412 unknown variables. In our case, the GAs was terminated at the 21,549th generation when little improvement was found with additional computational cost.

As mentioned in Sect. 5.2, the total computation for this work is equivalent to an integration of 7.7576×10^{10} h (=21,549 generations \times 5000 sets \times 10 scenarios \times 72) for a chemical system. This is equivalent to simulating 30 years gas phase chemistry for a typical air quality model with a grid configuration of 100 (grids in x-axis) \times 100 (grids in y-axis) \times 30 (vertical layers).

Due to the facts outlined above, we performed this task via MPI parallel computation in an IBM power4 supercomputer with the use of 64 nodes. For comparison purpose, the computational cost is converted to equivalent time if the calculation is performed by using one node only. Two conclusions can be drawn from the analysis of the computation cost. First, the tracking of computational cost indicates that it would take 369 days to obtain the results demonstrated above if one single node from the same IBM supercomputer machine (with a clock speed of 1.9 GHz) is used. Second, average computational cost for each generation is equivalent to 5.6 h in the first 18 generations, and the computational cost drops to an average 0.403 hour per generation. The high cost in the first 18 generation is caused by singular chemical ODEs from some randomized unrealistic initial populations and their offspring.

7 Summary and conclusions

Based on protocols for a detailed chemical mechanism, a simple unified volatility-based scheme (SUVS) is proposed for describing the gas phase chemistry of secondary organic aerosol formation. This SUVS has three distinctive features.

First, this new SUVS is able to unify chemical reactions for the SOA precursors with different oxidants. This coupled feature is an improvement over previous commonly used methods, in which different sets of parameters are used when oxidation conditions are changed.

Second, the SUVS is able to describe multi-generation reactions. This is also an improvement over the simple one-step reaction scheme, because the multi-generation method would be able to describe chemical evolutions which last longer than a smog chamber experiments that typically last no more than five hours.

Third, aside from the initial organic species, the species in the SUVS is volatility based and the number of organic species in the chemical system is only 30, which is a factor of 10 smaller than a detailed chemical mechanism. Moreover, the SUVS is suitable for describing oxidation of a general organic species. In this work, the SUVS was used for describing α -pinene oxidation, and it has been compared with a detailed chemical mechanism of MCM v3.1 for α -pinene oxidation. In order to derive the 412 unknown parameters in the SUVS, we resolved two issues: experimental data and fitting method.

To get experimental data, the MCM v3.1 was run with 10 selected scenarios. Then, the hourly profiles from a total of 180 condensable organic species in the MCM v3.1 were then grouped into 10 lumped species. The resulting profiles for 11 species (the 10 lumped species and a species for the total RO₂) are used as “experimental data” for the SUVS.

A genetic algorithms method was used to determine the 412 unknown parameters from the SUVS. A CV(RMSD) was included in the design of a fitness function for the GAs. This is an essential part in the application of GAs. Meanwhile, a MPI-version parallel genetic algorithm package, named as MPI-PIKAIA, was used to speed up the calculation. The GAs method used a 6-digit-long floating point encoding method with a population size of 5000. The size of the population was thus more than 12 times larger than the number of unknown variables. Most importantly, physical bounds on the 312 stoichiometric coefficients and the 100 reaction rate constants ensure minimization of non-physical solutions. In the GAs, the fitness value for the best solution of the SUVS is always monotonically increasing and the corresponding CV(RMSD) is decreasing to 0.0488 (fraction) in the final 21,549th generation. The best solution from the 21,549th generation was used to evaluate the SUVS.

The SUVS combined with the best set of parameters is evaluated by using profiles of the “experimental data” for 11 species from 10 different scenarios. We evaluated the errors based on species and scenario, respectively.

In term of species, the patterns of the diurnal changes for the 11 species can be grouped into four types, and the SUVS is able to capture the diurnal changes for each type. The errors for some species are as low as 0.017, but large error at 0.092 for a species still exists. This highlights the importance for further model improvement by using various

GAs techniques. For example, multi-objective genetic algorithms could be applied to avoid early convergence to a local maximum, rather than global maximum. In addition, the island model or the cellular genetic algorithms (Alba and Dorronsoro, 2008) could be used to explore the solutions from a global prospective.

Because all unknown parameters in the SUVS are bounded to certain realistic constraints, they should give physical interpretations, but not a unique solution. This is because GAs, similar to the neural network method, sometimes produce many different sets of parameters with similarly good fits. One possible way to get an identical “best” solution is to use a hybrid optimization technique (Renders and Flasse, 1996; Rodriguez-Fernandez et al., 2006). This means other optimization methods are applied to post-process candidate solutions when no further improvement was found using the GAs. The use of suitable optimization methods to post-process a large multi-dimensional system, such as the SUVS, merit further investigation.

Finally, we evaluated the performance of the SUVS over a wide range of conditions. The evaluation indicates that the SUVS is able to reproduce the full mechanism with fractional error at most 0.10 for not only the 10 selected scenarios, but also the additional 165 scenarios under a wider range of conditions.

Overall, compared with traditional mechanism reduction method (Xia et al. 2009; Utembe et al., 2009) for SOA formation, the SUVS is able to reduce the number of species and reactions by a factor of 10 and reproduce the species profiles from a detailed chemical mechanism. An examination of the final SUVS reactions, given in the Supplement, indicates that some species could be removed from product list when corresponding stoichiometric coefficients in certain reactions are close to zero. However, systematic mechanism reduction techniques, such as those traditionally applied in Xia et al. (2009), could be employed to further reduce the SUVS.

Essentially, the validity of SUVS does depend strongly on the accuracy of the “experimental data” used by the GAs for optimization. In this work, the “experimental data” were obtained through simulating the chemistry in the MCM v3.1. Thus, the uncertainties for the SUVS stem from structure activity relationship, which was used to construct the MCM. If the benchmark mechanism of the MCM v3.1 or other similar explicit chemical mechanisms (Aumont et al, 2005), is revised, the 412 parameters in the SUVS should be updated to reflect new changes.

In addition, the lumping method and estimation methods for boiling point and saturation vapor pressure also play an important role for the SUVS. Moreover, the inclusion of activity coefficients could potentially have a significant impact, relative to our initial work in which unity is used. A value of unity for the activity coefficients was assumed in order to simplify the calculation and avoid the need for group method estimation of activities. However, it should be noted that under conditions of high relative humidity non-unity activity

coefficients may significantly affect the result. Our methodology would be applicable for cases in which activity coefficients are explicitly calculated, and the underlying methodology would remain unchanged in such a case. This should be examined in future research.

We would like to point out that the “experimental data” used for optimization of the SUVS are not limited only to the explicit chemical mechanisms, such as the MCM. We could also use real smog chamber experimental data. In the later case, the unity and/or non-unity activity coefficients for all compounds are embedded into the term of “effective saturation coefficients” (Donahue et al., 2006). The accuracy of the resulting SUVS, coupled with a gas/particle partitioning model, would depend on the accuracies of instruments to measure concentrations and volatilities and the representativeness of the laboratory-generated particle composition to ambient particle composition.

Moreover, a single SUVS could be used to describe various mixtures of primary species. For example, to develop a revised SUVS for the α - and β -pinene mixed system, 30 additional unknown parameters for β -pinene initial oxidation with three oxidants (OH, O₃, and NO₃) would be integrated into the SUVS system. These 30 new unknown parameters are assigned similar to those for α -pinene system in Part 1 in Fig. 1. In other words, similar to the original VBS proposed by Donahue et al. (2006), the framework of this simple chemical mechanism is sustainable to describe different SOA precursor oxidation systems.

Supplementary material related to this article is available online at:

<http://www.atmos-chem-phys.net/11/6185/2011/acp-11-6185-2011-supplement.pdf>.

Acknowledgements. This research is financially supported by Natural Sciences and Engineering Research Council of Canada (NSERC) and the PERD (Program for Energy Research and Development) Particles and Related Emissions Program, Project C12.007. Many thanks to Yan Yang and the AURAMS team at Environmental Canada for fruitful discussion of this project. This work was computed with CMC IBM supercomputers from Environment Canada. The authors also would like to thank Balbir Pabla and Sylvain Ménard for their help with MPI settings. This paper is dedicated to late Diane Michelangeli from York University, Canada for her unlimited support in passing many milestones in our academic careers.

Edited by: J. H. Seinfeld

References

- Alba, E. and Dorronsoro, B.: Cellular genetic algorithms, Springer, Berlin, xiii, 245 p, 2008.
- Allen, C. T., Haupt, S. E., and Young, G. S.: Source characterization with a genetic algorithm-coupled dispersion-backward

- model incorporating scipuff, *J. Appl. Meteorol. Climatol.*, 46, 273–287, doi:10.1175/jam2459.1, 2007.
- Andersson-Skold, Y. and Simpson, D.: Secondary organic aerosol formation in northern Europe: A model study, *J. Geophys. Res.-Atmos.*, 106, 7357–7374, 2001.
- Asher, W. E., Pankow, J. F., Erdakos, G. B., and Seinfeld, J. H.: Estimating the vapor pressures of multi-functional oxygen-containing organic compounds using group contribution methods, *Atmos. Environ.*, 36, 1483–1498, 2002.
- Atkinson, R. and Arey, J.: Atmospheric degradation of volatile organic compounds, *Chemical Reviews* (Washington, DC, United States), 103, 4605–4638, 2003.
- 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, doi:10.5194/acp-5-2497-2005, 2005.
- Barkey, B., Paulson, S. E., and Chung, A.: Genetic algorithm inversion of dual polarization polar nephelometer data to determine aerosol refractive index, *Aerosol Sci. Tech.*, 41, 751–760, doi:10.1080/02786820701432640, 2007.
- Bian, F. and Bowman, F. M.: A lumping model for composition- and temperature-dependent partitioning of secondary organic aerosols, *Atmos. Environ.*, 39, 1263–1274, 2005.
- Biggs, P., Canosamas, C. E., Fracheboud, J. M., Shallcross, D. E., and Wayne, R. P.: Investigation into the kinetics and mechanism of the reaction of NO_3 with CH_3O_2 at 298-K and 2.5-torr – a potential source of OH in the nighttime troposphere, *J. Chem. Soc. Faraday. T*, 90, 1205–1210, 1994.
- Biggs, P., Canosamas, C. E., Fracheboud, J. M., Shallcross, D. E., and Wayne, R. P.: Rate constants for the reactions of C_2H_5 , $\text{C}_2\text{H}_5\text{O}$ and $\text{C}_2\text{H}_5\text{O}$ radicals with NO_3 at 298 K and 2.2 torr, *J. Chem. Soc. Faraday. T*, 91, 817–825, 1995.
- Bonn, B., von Kuhlmann, R., and Lawrence, M. G.: High contribution of biogenic hydroperoxides to secondary organic aerosol formation, *Geophys. Res. Lett.*, 31, L10108, doi:10.1029/2003GL019172, 2004.
- Cai, X. Y. and Griffin, R. J.: Secondary aerosol formation from the oxidation of biogenic hydrocarbons by chlorine atoms, *J. Geophys. Res.-Atmos.*, 111, D14206, doi:10.1029/2005JD006857, 2006.
- Camredon, M., Aumont, B., Lee-Taylor, J., and Madronich, S.: The SOA/VOC/ NO_x system: an explicit model of secondary organic aerosol formation, *Atmos. Chem. Phys.*, 7, 5599–5610, doi:10.5194/acp-7-5599-2007, 2007.
- Capouet, M., Muller, J.-F., Ceulemans, K., Compernelle, S., Vereecken, L., and Peeters, J.: Modeling aerosol formation in alpha-pinene photo-oxidation experiments, *J. Geophys. Res.-Atmos.*, 113, D02308, doi:10.1029/2007JD008995, 2008.
- Carroll, D. L.: Chemical laser modeling with genetic algorithms, *Aiaa Journal*, 34, 338–346, 1996.
- Chan, A. W. H., Kroll, J. H., Ng, N. L., and Seinfeld, J. H.: Kinetic modeling of secondary organic aerosol formation: effects of particle- and gas-phase reactions of semivolatile products, *Atmos. Chem. Phys.*, 7, 4135–4147, doi:10.5194/acp-7-4135-2007, 2007.
- Charbonneau, P.: Genetic algorithms in astronomy and astrophysics, *Astrophys. J. Suppl. Ser.*, 101, 309–334, 1995.
- Chung, S. H. and Seinfeld, J. H.: Global distribution and climate forcing of carbonaceous aerosols, *J. Geophys. Res.-Atmos.*, 107, 4407, doi:10.1029/2001JD001397, 2002.
- Daele, V., Laverdet, G., Lebras, G., and Poulet, G.: Kinetics of the reactions $\text{CH}_3\text{O} + \text{NO}$, $\text{CH}_3\text{O} + \text{NO}_3$, and $\text{CH}_3\text{O}_2 + \text{NO}_3$, *J. Phys. Chem.*, 99, 1470–1477, 1995.
- Damian, V., Sandu, A., Damian, M., Potra, F., and Carmichael, G. R.: The kinetic preprocessor kpp – a software environment for solving chemical kinetics, *Comput. Chem. Eng.*, 26, 1567–1579, 2002.
- 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, 2006.
- Donahue, N. M., Robinson, A. L., and Pandis, S. N.: Atmospheric organic particulate matter: From smoke to secondary organic aerosol, *Atmos. Environ.*, 43, 94–106, doi:10.1016/j.atmosenv.2008.09.055, 2009.
- Elliott, L., Ingham, D. B., Kyne, A. G., Mera, N. S., Pourkashanian, M., and Wilson, C. W.: Incorporation of physical bounds on rate parameters for reaction mechanism optimization using genetic algorithms, *Combust. Sci. Technol.*, 175, 619–648, doi:10.1080/00102200390196377, 2003.
- Elliott, L., Ingham, D. B., Kyne, A. G., Mera, N. S., Pourkashanian, M., and Wilson, C. W.: Genetic algorithms for optimisation of chemical kinetics reaction mechanisms, *Prog. Energ. Combust.*, 30, 297–328, 2004.
- Fan, W. G., Fox, E. A., Pathak, P., and Wu, H.: The effects of fitness functions on genetic programming-based ranking discovery for web search, *J. Am. Soc. Inf. Sci. Technol.*, 55, 628–636, 2004.
- Farina, S. C., Adams, P. J., and Pandis, S. N.: Modeling global secondary organic aerosol formation and processing with the volatility basis set: Implications for anthropogenic secondary organic aerosol, *J. Geophys. Res.-Atmos.*, 115, D09202 doi:10.1029/2009jd013046, 2010.
- Fishtine, S. H.: Reliable latent heats of vaporization, *Ind. Eng. Chem.*, 55, 20–28, 1963.
- Gao, S., Keywood, M., Ng, N. L., Surratt, J., Varutbangkul, V., Bahreini, R., Flagan, R. C., and Seinfeld, J. H.: Low-molecular-weight and oligomeric components in secondary organic aerosol from the ozonolysis of cycloalkenes and alpha-pinene, *J. Phys. Chem. A*, 108, 10147–10164, 2004.
- Goldberg, D. E.: Genetic algorithms in search, optimization, and machine learning, Addison-Wesley Pub. Co., Reading, Mass., xiii, 412 p., 1989.
- Goldberg, D. E., Sastry, K., and Llorca, X.: Toward routine billion-variable optimization using genetic algorithms, *Complexity*, 12, 27–29, doi:10.1002/cplx.20168, 2007.
- 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 impact of secondary organic aerosol: current and emerging issues, *Atmos. Chem. Phys.*, 9, 5155–5236, doi:10.5194/acp-9-5155-2009, 2009.
- Harris, S. D., Elliott, L., Ingham, D. B., Pourkashanian, M., and Wilson, C. W.: The optimisation of reaction rate parameters for

- chemical kinetic modelling of combustion using genetic algorithms, *Comput. Method. Appl. M.*, 190, 1065–1090, 2000.
- Haupt, S. E.: A demonstration of coupled receptor/dispersion modeling with a genetic algorithm, *Atmos. Environ.*, 39, 7181–7189, doi:10.1016/j.atmosenv.2005.08.027, 2005.
- Haupt, R. L. and Haupt, S. E.: *Practical genetic algorithms*, 2nd ed., John Wiley, Hoboken, N.J., xvii, 253 p., 2004.
- Iinuma, Y., Boge, O., Gnauk, T., and Herrmann, H.: Aerosol-chamber study of the alpha-pinene/O₃ reaction: Influence of particle acidity on aerosol yields and products, *Atmos. Environ.*, 38, 761–773, 2004.
- Jang, M. S., Czoschke, N. M., Lee, S., and Kamens, R. M.: Heterogeneous atmospheric aerosol production by acid-catalyzed particle-phase reactions, *Science*, 298, 814–817, 2002.
- Jenkin, M. E.: Modelling the formation and composition of secondary organic aerosol from α - and β -pinene ozonolysis using MCM v3, *Atmos. Chem. Phys.*, 4, 1741–1757, doi:10.5194/acp-4-1741-2004, 2004.
- 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, 1997.
- Jenkin, M. E., Saunders, S. M., Wagner, V., and Pilling, M. J.: Protocol for the development of the Master Chemical Mechanism, MCM v3 (Part B): tropospheric degradation of aromatic volatile organic compounds, *Atmos. Chem. Phys.*, 3, 181–193, doi:10.5194/acp-3-181-2003, 2003.
- Jimenez, J. L., Canagaratna, M. R., Donahue, N. M., Prevot, A. S. H., Zhang, Q., Kroll, J. H., DeCarlo, P. F., Allan, J. D., Coe, H., Ng, N. L., Aiken, A. C., Docherty, K. S., Ulbrich, I. M., Grieshop, A. P., Robinson, A. L., Duplissy, J., Smith, J. D., Wilson, K. R., Lanz, V. A., Hueglin, C., Sun, Y. L., Tian, J., Laaksonen, A., Raatikainen, T., Rautiainen, J., Vaattovaara, P., Ehn, M., Kulmala, M., Tomlinson, J. M., Collins, D. R., Cubison, M. J., Dunlea, E. J., Huffman, J. A., Onasch, T. B., Alfarra, M. R., Williams, P. I., Bower, K., Kondo, Y., Schneider, J., Drewnick, F., Borrmann, S., Weimer, S., Demerjian, K., Salcedo, D., Cottrell, L., Griffin, R., Takami, A., Miyoshi, T., Hatakeyama, S., Shimono, A., Sun, J. Y., Zhang, Y. M., Dzepina, K., Kimmel, J. R., Sueper, D., Jayne, J. T., Herndon, S. C., Trimborn, A. M., Williams, L. R., Wood, E. C., Middlebrook, A. M., Kolb, C. E., Baltensperger, U., and Worsnop, D. R.: Evolution of organic aerosols in the atmosphere, *Science*, 326, 1525–1529, doi:10.1126/science.1180353, 2009.
- Joback, K. G. and Reid, R. C.: Estimation of pure-component properties from group-contributions, *Chem. Eng. Commun.*, 57, 233–243, 1987.
- Johnson, D., Utembe, S. R., and Jenkin, M. E.: Simulating the detailed chemical composition of secondary organic aerosol formed on a regional scale during the TORCH 2003 campaign in the southern UK, *Atmos. Chem. Phys.*, 6, 419–431, doi:10.5194/acp-6-419-2006, 2006.
- Kalberer, M., Paulsen, D., Sax, M., Steinbacher, M., Dommen, J., Prevot, A. S. H., Fisseha, R., Weingartner, E., Frankevich, V., Zenobi, R., and Baltensperger, U.: Identification of polymers as major components of atmospheric organic aerosols, *Science*, 303, 1659–1662, 2004.
- Kamens, R., Jang, M., Chien, C. J., and Leach, K.: Aerosol formation from the reaction of alpha-pinene and ozone using a gas-phase kinetics aerosol partitioning model, *Environ. Sci. Technol.*, 33, 1430–1438, 1999.
- Kelly, J. L., Michelangeli, D. V., Makar, P. A., Hastie, D. R., Mozurkewich, M., and Auld, J.: Aerosol speciation and mass prediction from toluene oxidation under high NO_x conditions, *Atmos. Environ.*, 44, 361–369, doi:10.1016/j.atmosenv.2009.10.035, 2010.
- Kirchner, F. and Stockwell, W. R.: Effect of peroxy radical reactions on the predicted concentrations of ozone, nitrogenous compounds, and radicals, *J. Geophys. Res.-Atmos.*, 101, 21007–21022, 1996.
- Kroll, J. H. and Seinfeld, J. H.: Chemistry of secondary organic aerosol: Formation and evolution of low-volatility organics in the atmosphere, *Atmos. Environ.*, 42, 3593–3624, doi:10.1016/j.atmosenv.2008.01.003, 2008.
- Lane, T. E., Donahue, N. M., and Pandis, S. N.: Simulating secondary organic aerosol formation using the volatility basis-set approach in a chemical transport model, *Atmos. Environ.*, 42, 7439–7451, doi:10.1016/j.atmosenv.2008.06.026, 2008.
- Lee, Y. H., Park, S. K., and Chang, D.-E.: Parameter estimation using the genetic algorithm and its impact on quantitative precipitation forecast, *Ann. Geophys.*, 24, 3185–3189, doi:10.5194/angeo-24-3185-2006, 2006.
- Li, M. J., Chen, D. S., Cheng, S. Y., Wang, F., Li, Y., Zhou, Y., and Lang, J. L.: Optimizing emission inventory for chemical transport models by using genetic algorithm, *Atmos. Environ.*, 44, 3926–3934, doi:10.1016/j.atmosenv.2010.07.010, 2010.
- Lienert, B. R., Porter, J. N., and Sharma, S. K.: Repetitive genetic inversion of optical extinction data, *Appl. Optics*, 40, 3476–3482, 2001.
- Lienert, B. R., Porter, J. N., and Sharma, S. K.: Aerosol size distributions from genetic inversion of polar nephelometer data, *J. Atmos. Ocean. Tech.*, 20, 1403–1410, 2003.
- Liggio, J., Li, S. M., and McLaren, R.: Reactive uptake of glyoxal by particulate matter, *J. Geophys. Res.-Atmos.*, 110, D10304, doi:10.1029/2004JD005113, 2005.
- Lightfoot, P. D., Cox, R. A., Crowley, J. N., Destriau, M., Hayman, G. D., Jenkin, M. E., Moortgat, G. K., and Zabel, F.: Organic peroxy-radicals - kinetics, spectroscopy and tropospheric chemistry, *Atmos. Environ.*, 26, 1805–1961, 1992.
- Metcalf, T. S. and Charbonneau, P.: Stellar structure modeling using a parallel genetic algorithm for objective global optimization, *J. Comput. Phys.*, 185, 176–193, doi:10.1016/s0021-9991(02)00053-0, 2003.
- Michalewicz, Z.: *Genetic algorithms + data structures = evolution programs*, 3rd rev. and extended edn., Springer-Verlag, Berlin, New York, xx, 387 p., 1999.
- O'Steen, L. and Werth, D.: The application of an evolutionary algorithm to the optimization of a mesoscale meteorological model, *J. Appl. Meteorol. Climatol.*, 48, 317–329, doi:10.1175/2008jamc1967.1, 2009.
- 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, 1996.
- Pankow, J. F.: An absorption-model of gas-particle partitioning of organic-compounds in the atmosphere, *Atmos. Environ.*, 28, 185–188, 1994.
- Pathak, R. K., Presto, A. A., Lane, T. E., Stanier, C. O., Donahue, N. M., and Pandis, S. N.: Ozonolysis of α -pinene: parameterization

- of secondary organic aerosol mass fraction, *Atmos. Chem. Phys.*, 7, 3811–3821, doi:10.5194/acp-7-3811-2007, 2007.
- Polifke, W. Geng, W. Q., and Dobbeling, K.: Optimization of rate coefficients for simplified reaction mechanisms with genetic algorithms, *Combust Flame*, 113, 119–134, 1998.
- Presto, A. A. and Donahue, N. M.: Investigation of alpha-pinene plus ozone secondary organic aerosol formation at low total aerosol mass, *Environ. Sci. Technol.*, 40, 3536–3543, doi:10.1021/es052203z, 2006.
- Pun, B. K., Wu, S. Y., Seigneur, C., Seinfeld, J. H., Griffin, R. J., and Pandis, S. N.: Uncertainties in modeling secondary organic aerosols: Three-dimensional modeling studies in nashville/western tennessee, *Environ. Sci. Technol.*, 37, 3647–3661, 2003.
- Renders, J. M. and Flasse, S. P.: Hybrid methods using genetic algorithms for global optimization, *IEEE Trans. Syst. Man Cybern. Part B-Cybern.*, 26, 243–258, 1996.
- Robinson, A. L., Donahue, N. M., Shrivastava, M. K., Weitkamp, E. A., Sage, A. M., Grieshop, A. P., Lane, T. E., Pierce, J. R., and Pandis, S. N.: Rethinking organic aerosols: Semivolatile emissions and photochemical aging, *Science*, 315, 1259–1262, 2007.
- Rodriguez-Fernandez, M., Mendes, P., and Banga, J. R.: A hybrid approach for efficient and robust parameter estimation in biochemical pathways, *Biosystems*, 83, 248–265, doi:10.1016/j.biosystems.2005.06.016, 2006.
- Sage, M. L. and Sage, G. W.: Vapor pressure, in: *Handbook of property estimation methods for environmental chemicals: Environmental and health sciences*, edited by: Boethling, R. S., and Mackay, D., Lewis Publishers, New York, 53–67, 2000.
- Sandu, A. and Sander, R.: Technical note: Simulating chemical systems in Fortran90 and Matlab with the Kinetic PreProcessor KPP-2.1, *Atmos. Chem. Phys.*, 6, 187–195, doi:10.5194/acp-6-187-2006, 2006.
- 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, doi:10.5194/acp-3-161-2003, 2003.
- Schell, B., Ackermann, I. J., Hass, H., Binkowski, F. S., and Ebel, A.: Modeling the formation of secondary organic aerosol within a comprehensive air quality model system, *J. Geophys. Res.-Atmos.*, 106, 28275–28293, 2001.
- Schwarzenbach, R. P., Gschwend, P. M., and Imboden, D. M.: *Environmental organic chemistry*, 2nd ed., Wiley, Hoboken, N.J., xiii, 1313p., 2003.
- Seinfeld, J. H.: Air pollution: A half century of progress, *Aiche J.*, 50, 1096–1108, 2004.
- Seinfeld, J. H. and Pankow, J. F.: Organic atmospheric particulate material, *Annu. Rev. Phys. Chem.*, 54, 121–140, 2003.
- Slowik, J. G., Stroud, C., Bottenheim, J. W., Brickell, P. C., Chang, R. Y.-W., Liggio, J., Makar, P. A., Martin, R. V., Moran, M. D., Shantz, N. C., Sjostedt, S. J., van Donkelaar, A., Vlasenko, A., Wiebe, H. A., Xia, A. G., Zhang, J., Leitch, W. R., and Abbatt, J. P. D.: Characterization of a large biogenic secondary organic aerosol event from eastern Canadian forests, *Atmos. Chem. Phys.*, 10, 2825–2845, doi:10.5194/acp-10-2825-2010, 2010.
- Song, C., Na, K. S., and Cocker, D. R.: Impact of the hydrocarbon to NO_x ratio on secondary organic aerosol formation, *Environ. Sci. Technol.*, 39, 3143–3149, 2005.
- Stanier, C. O., Donahue, N. M., and Pandis, S. N.: Parameterization of secondary organic aerosol mass fractions from smog chamber data, *Atmos. Environ.*, 42, 2276–2299, doi:10.1016/j.atmosenv.2007.12.042, 2008.
- Stein, S. E. and Brown, R. L.: Estimation of normal boiling points from group contributions, *J. Chem. Inf. Comp. Sci.*, 34, 581–587, 1994.
- Stroud, C. A., Makar, P. A., Michelangeli, D. V., Mozurkewich, M., Hastie, D. R., Barbu, A., and Humble, J.: Simulating organic aerosol formation during the photooxidation of toluene/NO_x mixtures: Comparing the equilibrium and kinetic assumption, *Environ. Sci. Technol.*, 38, 1471–1479, 2004.
- Sun, J. and Ariya, P. A.: Atmospheric organic and bio-aerosols as cloud condensation nuclei (ccn): A review, *Atmos. Environ.*, 40, 795–820, 2006.
- Tolocka, M. P., Jang, M., Ginter, J. M., Cox, F. J., Kamens, R. M., and Johnston, M. V.: Formation of oligomers in secondary organic aerosol, *Environ. Sci. Technol.*, 38, 1428–1434, 2004.
- Tsigaridis, K. and Kanakidou, M.: Global modelling of secondary organic aerosol in the troposphere: a sensitivity analysis, *Atmos. Chem. Phys.*, 3, 1849–1869, doi:10.5194/acp-3-1849-2003, 2003.
- Tsuchiya, M. and Ross, J.: Application of genetic algorithm to chemical kinetics: Systematic determination of reaction mechanism and rate coefficients for a complex reaction network, *J. Phys. Chem. A*, 105, 4052–4058, 2001.
- Utembe, S. R., Watson, L. A., Shallcross, D. E., and Jenkin M. E.: A Common Representative Intermediates (CRI) mechanism for VOC degradation. Part 3: Development of a secondary organic aerosol module, *Atmos. Environ.*, 43, 1982–1990, doi:10.1016/j.atmosenv.2009.01.008, 2009.
- Xia, A. G., Michelangeli, D. V., and Makar, P. A.: Box model studies of the secondary organic aerosol formation under different hc/NO_x conditions using the subset of the master chemical mechanism for alpha-pinene oxidation, *J. Geophys. Res.-Atmos.*, 113, D10301, doi:10.1029/2007jd008726, 2008.
- Xia, A. G., Michelangeli, D. V., and Makar, P. A.: 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, 4341–4362, doi:10.5194/acp-9-4341-2009, 2009.
- Zhang, Q., Jimenez, J. L., Canagaratna, M. R., Allan, J. D., Coe, H., Ulbrich, I., Alfarra, M. R., Takami, A., Middlebrook, A. M., Sun, Y. L., Dzepina, K., Dunlea, E., Docherty, K., Decarlo, P. F., Salcedo, D., Onasch, T., Jayne, J. T., Miyoshi, T., Shimojo, A., Hatakeyama, S., Takegawa, N., Kondo, Y., Schneider, J., Drewnick, F., Borrmann, S., Weimer, S., Demerjian, K., Williams, P., Bower, K., Bahreini, R., Cottrell, L., Griffin, R. J., Rautiainen, J., Sun, J. Y., Zhang, Y. M., and Worsnop, D. R.: Ubiquity and dominance of oxygenated species in organic aerosols in anthropogenically-influenced northern hemisphere midlatitudes, *Geophys. Res. Lett.*, 34, L13801, doi:10.1029/2007gl029979, 2007.
- Ziemann, P. J.: Evidence for low-volatility diacyl peroxides as a nucleating agent and major component of aerosol formed from reactions of O₃ with cyclohexene and homologous compounds, *J. Phys. Chem. A*, 106, 4390–4402, 2002.