Mechanism reduction for the formation of secondary organic aerosol for integration into a 3-dimensional regional air quality model: \(\alpha\)-pinene oxidation system

A. G. Xia\(^1,*\), D. V. Michelangeli\(^1\), and P. A. Makar\(^2\)

\(^1\)Department of Earth and Space Science and Engineering, York University, Toronto, Ontario, Canada
\(^2\)Air Quality Research Division, Environment Canada, Toronto, Ontario, Canada

*now at: Air Quality Research Division, Environment Canada, Toronto, Ontario, Canada

Received: 4 June 2008 – Published in Atmos. Chem. Phys. Discuss.: 14 July 2008
Revised: 9 June 2009 – Accepted: 9 June 2009 – Published: 7 July 2009

Abstract. A detailed \(\alpha\)-pinene oxidation mechanism was reduced systematically through the successive application of five mechanism reduction techniques. The resulting reduced mechanism preserves the ozone- and organic aerosol-forming properties of the original mechanism, while using less species. The methodologies employed included a directed relation graph method with error propagation (DRGEP, which removed a large number of redundant species and reactions), principal component analysis of the rate sensitivity matrix (PCA, used to remove unnecessary reactions), the quasi-steady-state approximation (QSSA, used to remove some QSS species), an iterative screening method (ISSA, which removes redundant species and reactions simultaneously), and a new lumping approach dependent on the hydrocarbon to NO\(_x\) ratio (which reduced the number of species in mechanism subsets for specific hydrocarbon to NO\(_x\) ranges).

This multistage methodology results in a reduction ratio of 2.5 for the number of both species and reactions compared with the full mechanism. The simplified mechanism reproduces the important gas and aerosol phase species (the latter are examined in detail by individual condensing species as well as in classes according to four functional groups: PANs, nitrates, organic peroxides, and organic acids). The total SOA mass is also well represented in the condensed mechanism, to within 16\% of the detailed mechanism under a wide range of conditions. The methodology described here is general, and may be used in general mechanism reduction problems.

1 Introduction

Most widely used gas phase chemical reaction mechanisms, including those used in 3-dimensional grid-based Eulerian regional air quality models, were developed primarily for ozone formation under a wide range of conditions. Examples of such popularly used mechanisms include ADOM-II (Stockwell and Lurmann, 1989), CBM-IV (Gery et al., 1989), RADM2 (Stockwell et al., 1990), RACM (Stockwell et al., 1997), and EMEP (Simpson, 1995; Simpson et al., 2003).

On the other hand, detailed gas phase chemical mechanisms, such as the Master Chemical Mechanism (MCM) (Jenkin et al., 1997; Jenkin et al., 2003; Saunders et al., 2003) and a newly developed fully-explicit chemical mechanism based on a self-generating approach (Aumont et al., 2005), could be used to describe the oxidation of a large number of volatile organic compounds (VOCs) in the ambient atmosphere. The detailed mechanisms, rather than the simplified mechanisms, are capable of depicting the formation of multifunctional organic compounds. This property is of importance in that many of these multifunctional compounds may partition between gas and aerosol phases to form secondary organic aerosol (SOA).

For example, the MCM has been applied to describe SOA formation from the oxidation of \(\alpha\)-pinene under a wide range of conditions (Xia et al., 2008), the ozonolysis of the \(\alpha\)- and \(\beta\)-pinene (Jenkin, 2004), the oxidation of toluene (Johnson et al., 2004; Stroud et al., 2004) and other volatile organic compounds (VOCs) (Johnson et al., 2006a, b) by using a trajectory model. The self-generated detailed mechanism by Aumont et al. (2005) was used to simulate SOA formation.
formation from the oxidation of 1-octene under atmospheric relevant conditions (Camredon et al., 2007). Another quasi-explicit model solely developed for \( \alpha \)-pinene oxidation was evaluated against 28 smog chamber experiments (Capouet et al., 2008).

However, the number of species and reactions from a detailed gas phase chemical mechanism for SOA formation are much larger than those commonly used within 3-dimensional Eulerian regional air quality models for the purpose of predicting ozone or particulate matter concentrations. The detailed mechanisms represent an extremely large computational burden in memory and CPU time in this context, hence the need to reduce these mechanisms to allow an accurate yet simple mechanism capable of describing both ozone and SOA formation. Many reduction techniques, especially in the fields of atmospheric chemistry and combustion chemistry, have been developed to condense chemical mechanisms. Generally, the techniques can be categorized as reducing either species, or reactions, or both. Techniques used in the mechanism reduction prior to the year 2000 can be found in three excellent reviews (Griffiths, 1995; Tomlin et al., 1997; Okino and Mavrovouniotis, 1998).

Several simple chemical mechanisms for SOA formation from \( \alpha \)-pinene oxidation are available in the literature. For example, based on smog chamber observations, Kamens et al. (1999) derived a mechanism with simple gas phase chemistry coupled with gas/particle partitioning model. This mechanism was extended by Andersson-Skold and Simpson (2001) to model SOA formation in northern Europe. Other simple \( \alpha \)-pinene oxidation mechanisms include one that was extracted from RACM (Barthelmie and Pyror, 1999), and a mechanism by Chen and Griffin (2005) that takes an intermediate approach between the non-specific SAPRC mechanism (Carter, 2000) and near-explicit MCM.

Some simplified chemical mechanisms for \( \alpha \)-pinene oxidation have been developed based on subsets of the detailed MCM, such as the “Mainz Alpha-pinene Mechanism” (MAM) scheme compiled recently by Bonn et al. (2004, 2005) to simulate SOA formation in a global CTM. The mechanism was evaluated against other reference mechanisms, such as RACM and MCM, for the formation of ozone by comparing the results from a few selected scenarios (Bonn et al., 2005). The performance of this mechanism versus the original under a large variety of conditions for ozone and SOA formation have not been established, and the methodology used to create the reduced mechanism was not described in detail.

In previous work (Xia et al., 2008), the subset of the detailed MCM describing \( \alpha \)-pinene oxidation was used to examine SOA formation under a wide range of hydrocarbon/NO\(_x\) conditions. In the current work, the same subset of the MCM is reduced by using a sequence of successively applied mechanism reduction techniques. The objective of the mechanism reduction process is to remove insignificant species and reactions from the chemical system, while to maintaining the accuracy of predictions of the dominant species for SOA formation and ozone. The methodology employed here is general, and could be applied to other mechanism reduction problems.

This paper is organized as follows. First, we introduce five mechanism reduction techniques in Sect. 2. Later, we present the results for the mechanism reduction of the subset of \( \alpha \)-pinene oxidation system from the five different stages. Finally, the implications of the mechanism reduction for the \( \alpha \)-pinene oxidation system are discussed.

## 2 Mechanism reduction techniques

In this section we review some of the methods typically used to reduce species and reactions in chemical mechanisms.

### 2.1 Methods to eliminate species

Time-scale analysis, lumping schemes, and compound contribution methods are the three main techniques used to eliminate species.

Time-scale analysis methods have been used to identify and then eliminate some short life-time species through the application of quasi-steady-state assumption (QSSA) and partial equilibrium assumption (McRae et al., 1982; Chen, 1988; Turanyi et al., 1993). Intrinsic low dimensional manifold (ILDM) (Maas and Pope, 1992) and computational singular perturbation (CSP) methods (Lam and Goussis, 1988, 1994; Massias et al., 1999a) have been also used to construct a reduced mechanism, through separation of fast and slow dynamic processes, in combustion chemistry (Massias et al., 1999b; Lu et al., 2001) and atmospheric chemistry (Neophytou et al., 2004; Lovas et al., 2006).

Lumping approaches reduce the number of species through combining reacting species to reduce the number of reactants. The transformation can be operated with rigorous mathematical approaches, such as exact lumping methods (Wei and Kuo, 1969; Li and Rabitz, 1989; Li et al., 1994a), and approximate lumping (Kuo and Wei, 1969; Li and Rabitz, 1990; Li et al., 1994b; Tomlin et al., 1994; Wang et al., 1998). For example, during the development of RADM2 (Stockwell et al., 1990), the primary VOCs were lumped based on aggregation factors of their reactivity with OH radicals. The reactivity lumping approach was updated via a hybrid reactivity weighting (Makar et al., 1996), integrated pseudo-species (Makar and Polavarapu, 1997), and a transient lumping method (Makar, 2000). The approximate lumping methods above depend on knowledge of “typical” relative concentrations of primary VOCs in a one-time decision made during mechanism construction (Stockwell et al., 1990; Makar et al., 1996; Makar and Polavarapu, 1997), or on an ongoing basis, with the aim of temporary reduction of the mechanism during gas-phase integration in a chemical transport model (Makar, 2000).
Other lumping strategies have been developed, such as lumping of similar chemical structures or functional groups for high temperature combustion systems (Fournet et al., 2000; Ranzi et al., 2001), lumping multiple semi-volatile organic aerosol components into one or a few groups with similar saturation vapor pressures, activity coefficients in water, or chemical structures (Bian and Bowman, 2002, 2005).

Whitehouse et al. (2004a) developed a lumping scheme applied to chemical reactions with similar oxidation pathways in version 2.0 of MCM. The disadvantage of their lumping method is that the ratios between the species in the lumped group depend not only on the reaction rate coefficients, but also on concentrations of the species at each time point. In particular, pre-calculated tables of the lumping coefficients for each time point must be provided before lumping.

A third method to eliminate unnecessary species is based on the study of compounds contribution towards species of interest. For example, the directed related graph (DRG) method, recently proposed and developed by Lu and Law (2005, 2006), is used to remove those species which always contribute a very small fraction towards the concentration of any species deemed important for the process under study.

2.2 Methods to eliminate reactions

Redundant reactions can be removed via principal component analysis (PCA) (Vajda et al., 1985; Turanyi et al., 1989) of a “sensitivity matrix” (defined later in this work). This method has been used to reduce chemical reactions for the subset of butane oxidation (Zeng et al., 1997), OH/HO and H2O/RO2 system (Carslaw et al., 1999) of the MCM, the CBM-EX (Heard et al., 1998), and combustion chemistry (Tomlin et al., 1992; Brown et al., 1997).

2.3 Other methods to reduce chemical mechanisms

Optimization methods based on integer programming (Petzold and Zhu, 1999; Androulakis, 2000; Banerjee and Jerapetritou, 2003; Bhattacharjee et al., 2003) and genetic algorithms (Edwards et al., 1998; Elliott et al., 2005) have been recently introduced to reduce chemical mechanism in the fields of biology (Maurya et al., 2005) and combustion (Banerjee and Jerapetritou, 2006).

Another two reduced methods should be mentioned for reducing tropospheric chemistry; “common representative intermediates” (CRI) method proposed by Jenkin et al. (2008) with an essential assumption that the ozone formation depends on the number of reactive carbon bonds only, and the “chemical operator” method (Gery et al., 1989; Carter, 1990, 2000) developed to decrease the number of the peroxy radicals (RO2) into a small number of reactivity groups ( < 10) (Madronich and Calvert, 1990). For the latter method, artificial species, called “chemical operators”, are used to represent net effects of peroxy radical reactions on NO consumption, NO2 formation, HO2 consumption and formation, and nitrate or hydroperoxide formation. The operator method was recently applied by Szopa et al. (2005) to reduce the self-generated mechanism (Aumont et al., 2005) with 350 000 species and 2 million reactions into a much smaller mechanism with 147 species and 472 reactions. Although the operator method is efficient and involves no approximation when O3 formation occurs, it remains unsatisfactory for the prediction of SOA formation (Carter, 2008), being unable to represent the formation of some condensable organic hydroperoxides under low NOx conditions.

2.4 Mechanism reduction to the subset of α-pinene oxidation

In order to obtain a simple mechanism for SOA formation in this work, five systematic mechanism reduction techniques were applied, in sequence to reduce the subset of α-pinene oxidation of the MCM, under a wide range of conditions. These techniques include the application of:

1. the directed relation graph (DRG) method, and its offshoot, the Directed Relation Graph method with Error Propagation (DRGEP) (Lu and Law, 2005, 2006; Pepiot-Desjardins and Pituch, 2008) to resolve species coupling and to remove redundant species;
2. PCA of the rate sensitivity matrix to remove redundant reactions (Turanyi et al., 1989);
3. QSSA to identify and remove of some quasi-steady-state species (Turanyi et al., 1993);
4. the Iterative Screening and Structure Analysis method (Mauersberger, 2005) to remove unimportant species and reactions simultaneously;
5. a new automatic linear lumping approach in which some species are combined and tested at different HC/NOx ratios.

Before the five methods are briefly introduced, we first categorize the chemical species, according to Turanyi (1990), into three groups in mechanism reduction: target species, necessary species, and redundant species. Target species are the species of interest; the concentrations of the target species should be reproduced as accurately as possible in the reduced mechanism. Necessary species must be kept in the reduced system to reproduce target species concentrations. Redundant species are sufficiently unimportant that they may be eliminated from the mechanism without compromising the accuracy of the target species. In contrast to redundant species, target species and necessary species remain in the reduced system. For convenience, the latter two groups are jointly termed as “major species” in this paper.

2.4.1 Directed relation graph method and directed relation graph method with error propagation

Directed relation graph: The “directed relation graph” (DRG) method, a fast computational method, for mechanism
Mechanism reduction was very recently developed (Lu and Law, 2005) and applied to a detailed n-Heptane and iso-octane mechanism (Lu and Law, 2006) in combustion chemistry. The DRG method is used to resolve the extent to which species are coupled, and the redundant species, based on a selected threshold value, are then removed.

The details about the DRG method can be found in Lu and Law (2005, 2006). A brief summary of a slightly modified version of this method is given here. In a chemical system with different species, if species $A$ is a target species, the relationship between another species, say $B$, and the target species $A$ can be established through a normalized contribution coefficient, defined as:

$$ r_{AB} \equiv \frac{\sum_{i=1}^{N} v_{Ai} R_i \delta_{Bi}}{\sum_{i=1}^{N} v_{Ai} R_i}, $$

$$ \delta_{Bi} = \begin{cases} 1, & \text{if species } B \text{ is a reactant in the } i^{\text{th}} \text{ reaction,} \\ 0, & \text{otherwise} \end{cases} \quad (1) $$

where $r_{AB}$ is the normalized contribution coefficient from species $B$ to species $A$, $R_i$ denotes the reaction rate for the $i^{\text{th}}$ reaction, $v_{Ai}$ is the stoichiometric coefficient for species $A$ in the $i^{\text{th}}$ reaction (the stoichiometric coefficient is positive if the compound is a product and negative if it is a reactant), and $N$ is the total number of reactions in the system.

The normalized contribution coefficient between each pair of species is an element of a matrix. Each element satisfies relationship of $0 \leq r_{AB} \leq 1$. If both $r_{AB}$ and $r_{BA}$ are large values, that means species $A$ and species $B$ are coupled. In this situation, species $A$ and $B$ should be either kept or removed together.

However, due to the existence of the reversible reactions in combustion chemistry, the definition of $r_{AB}$ from the original work (Lu and Law, 2005, 2006) is slightly different from the one used in this paper. According to Lu and Law (2005, 2006), $\delta_{Bi}$ is 1 as long as species $B$ is involved, either as a reactant or a product, in the $i^{\text{th}}$ reaction. However, the (detailed) mechanisms for gas-phase reactions resolve reversible reactions into forward and reverse reactions, and $\delta_{Bi}$ is defined therefore to be 1 only if species $B$ is a reactant in this work. Secondly, the reaction rate $R_i$ is the net reaction rate between forward and backward reaction rates in the original work (Lu and Law, 2005, 2006), yet it only refers to forward reaction rate here.

The calculated normalized contribution coefficients between two species and the matrix composed of the coefficients vary with time, because the concentrations and reaction rates change with time and scenario. Conceptually, the mechanism reduction procedure could be conducted at each time point by using the matrix of the normalized contribution coefficients obtained from that time point. The computational cost of this approach would be prohibitive.

To avoid coefficient recomputation at each time period, the DRG method is used here for a wide range of conditions as a screening tool, with a maximum value of each normalized contribution coefficient $r_{AB}$ from all sampled points (under a wide range of conditions) being calculated to create the elements of a maximum coefficient matrix. This maximum coefficient matrix is then used only once in the entire mechanism reduction process to identify redundant species. Our work demonstrated that this use of the DRG method as a screening tool is more efficient than the conceptual way to reduce the mechanism.

Mechanism reduction using the DRG method is an iterative process. For example, if species $A$ is a target species, species $B$ will remain in the system as a major species only when $r_{AB}$ is larger than a given threshold $\varepsilon_1$. If $B$ has been selected as a major species, the focus turns to reactions creating $B$, and species $I$ with $r_{BI} > \varepsilon_1$ must also be retained in the reduced mechanism, and so on. The iterative process is applied initially to each target species and it stops when the total number of “major species” converges. The remaining species at this point are redundant with respect to the target species, and can be safely removed from the product list of chemical reactions. Subsequently, all reactions that consume redundant species can be removed as well. A simplified reaction scheme is thus obtained for a given threshold $\varepsilon$. Different simplified mechanisms may be generated using different thresholds, though all should be evaluated against the full mechanism.

**Directed relation graph method with error propagation** In DRG method, all major species are considered equally important, no matter whether the species are directly or indirectly related. However, the effect of removing one species will consequently diminish when the length of the propagation path (the pathway for a sequence of two directed-related species, such as $A \rightarrow B$, $B \rightarrow C$, and $C \rightarrow D$, is described as “$A \rightarrow B \rightarrow C \rightarrow D$”) increases. To address this issue, Pepiot-Desjardins and Pitsch (2008) extended the original
DRG method (Lu and Law, 2005) into a method called “directed relation graph with error propagation (DRGEP)” to study the indirect effect among the species.

As illustrated in Fig. 1, species A depends on species B and C directly with the contribution coefficients of \( r_{AB} \) and \( r_{AC} \), respectively. Species B depends on species C at the contribution coefficients of \( r_{BC} \). As a result, A depends on both directly and indirectly through species B.

To quantify the complex coupling between A and C, a path dependent interaction coefficient \( r_{AC,i} \) on path \( i \) for the products encountered on the path is defined. For example, in Fig. 1, path #1 is \( A \rightarrow B \rightarrow C \), the coupling coefficient between A and C on path #1 is defined as:

\[
r_{AC,i} = r_{AB} \times r_{BC}
\]

where \( r_{AB} \) and \( r_{BC} \) are the maximum normalized contribution coefficients used by the DRG method.

From the above analysis, many different paths exist from one species to another in a chemical system. A generalized coupling coefficient \( R_{AC} \) between species A and C is defined as:

\[
R_{AC} = \max_{all \ path \ i} \{ r_{AC,i} \}
\]

In the above example,

\[
R_{AC} = \max \{ r_{AB} \times r_{BC}, r_{AC} \}
\]

Each species is associated with other species at different generalized contribution coefficients. Any species X will be selected as a major species to a given target species A if

\[
R_{AX} \geq \varepsilon_2
\]

where \( \varepsilon_2 \) is a particular threshold.

In the end, the final resulting species of the simplified mechanism are the union of the species from all subsets of each target species. The simplified reaction scheme is generated similarly to the DRG method once the redundant species are identified.

2.4.2 Principal component analysis of the rate sensitivity matrix

A spatially homogeneous reaction system can be described by a set of ODEs as:

\[
\frac{dc_i}{dt} = f_i(c, k) = \sum_j v_{ij} R_j \quad i = 1, N
\]

where \( N \) is the total number of species, \( c_i \) is the concentration for species \( i \) at time \( t \), \( f_i \) is the overall reaction rate for species \( i \), \( v_{ij} \) is the stoichiometric coefficient for species \( i \) in the \( j \)th reaction, \( R_j \) is the reaction rate of the \( j \)th reaction, \( c \) is the \( N \)-vector of species concentrations, and \( k \) is the \( M \)-vector of reaction rate coefficients.

In order to study the effect of perturbations to the reaction rate coefficients from an initial set \( k^0 \) to a perturbed set \( k \), Turanyi et al. (1989) first introduced an objective function \( Q \), which is the square of the normalized deviation of overall reaction rate \( f_i \), summed for all species.

\[
Q(c, \alpha) = \sum_{i=1}^{N} \left[ \frac{f_i(c, \alpha) - f_i(c, \alpha^0)}{f_i(c, \alpha^0)} \right]^2
\]

where \( \alpha = \ln k \) and \( \alpha^0 = \ln k^0 \).

\( Q \) is a measure of the sensitivity of the overall reaction rates to perturbations in the rate coefficients. As Turanyi et al. (1989) pointed out that in the neighbourhood of \( \alpha^0 \) (i.e. for small perturbations close to the initial state), the objective function \( Q \) can be approximated by the following expression:

\[
Q \approx (\Delta \alpha)^T F^T F (\Delta \alpha)
\]

where \( \Delta \alpha = \alpha - \alpha^0 \), and \( F \) is a \( N \times M \) matrix, where \( N \) is the total number of species and \( M \) is the total number of reactions) is the log-normalized rate sensitivity matrix (Turanyi et al., 1989). The element of \( F \) is computed as:

\[
F_{ij} = \frac{\partial \ln f_i}{\partial \ln k_j} = \frac{v_{ij} R_j}{f_i}
\]

where \( k_j \) is the \( j \)th reaction rate coefficient.

Let \( U \) denotes the matrix of normalized eigenvectors \( U_j \), of \( F^T F \) such that \( U_j^T U_j = I_N \), where \( I_N = \text{diag}(1, \ldots, 1) \). The new set of the parameters \( \Psi = U^T \alpha \) are called principal components. Then the objective function can be expressed in terms of the principal components,

\[
Q(\Psi) \approx \sum_{j=1}^{M} \lambda_j (\Delta \Psi_j)^2
\]

where \( \Delta \Psi = U^T \Delta \alpha \) and \( \lambda_1 > \lambda_2 > \ldots > \lambda_M \) denotes the eigenvalues of \( F^T F \) in descending order. The values of the eigenvectors and eigenvalues may be derived using singular value decomposition of the matrix \( F^T F \).

As a result of this principal component analysis (PCA), or eigenvalue/eigenvector decomposition, of the objective function, Eq. (10) reveals that the eigenvalues measure the significance of the principal component to the overall mechanism at a chosen time point, and the elements in an eigenvector of \( F^T F \) represent weights of the reactions within a principal component. In other words, a large eigenvalue means the respective principal component is important, while a large absolute value of an element of the eigenvector suggests the corresponding reaction contributes significantly to that principal component.

Therefore, important reactions could be identified by the PCA method (Turanyi et al., 1989) via two stages at each time point: (1) to select the first \( p \) significant principal components as long as their eigenvalues are larger than a given threshold \( \lambda_1 (\lambda_1 > \lambda_2 > \ldots > \lambda_p > \Delta_1 \), where \( 1 \leq p \leq M \);
(2) within each of the $p$ selected principal components, to find any significant reaction as long as the absolute value of the corresponding element of the eigenvector is larger than a second given threshold $\Delta_2 (|U_{ij}| > \Delta_2$, where $1 \leq q \leq M$). Finally, the procedure is repeated at all sampled time points, and the reactions kept in the reduced mechanism are the union of important reactions identified at all sampled time points. The remaining reactions are the redundant reactions that can be eliminated from the system. The choice of two thresholds ($\Delta_1$ and $\Delta_2$) are a matter of trial and error, and must be made in conjunction with comparisons to the reference reaction mechanism. The detailed procedure on how the PCA method is applied to remove redundant reactions is also demonstrated in Turanyi et al. (1989).

### 2.4.3 Application of quasi-steady-state assumption

The dynamic behavior of a chemical system involves multiple time scales, and the application of quasi-steady-state assumption (QSSA) identifies species with lifetimes sufficiently short that their presence may be removed by assuming a local equilibrium. Once the differential equations for the fastest time-scale QSS species are replaced by algebraic equations (the range of timescales of the ODEs becomes reduced as an added benefit of the reduction method; this decreases computational integration time).

According to Turanyi et al. (1993), an instantaneous error ($\Delta c_i$) introduced by the application of the QSSA to a single species is:

$$\Delta c_i = \frac{1}{J_{ii}} \frac{d c_i}{d t}$$  \hspace{1cm} (11)

where $J_{ii}$ is the diagonal Jacobian element given as

$$J_{ii} = \frac{\partial f_i}{\partial c_i} ; \quad f_i = \frac{c_i}{d t}$$  \hspace{1cm} (12)

A species with fractional errors ($\Delta c_i/c_i$) always less than a given threshold is identified as a QSS species for the reduced mechanism. Once the QSS species have been selected, and if the chemical reactions for the QSS species are in a simple form, the chemical system can be reduced by replacing QSS species with equilibrium expressions or by rewriting the reactions so that the resulting products are produced immediately, without the QSS intermediates being present (Tomlin et al., 1997).

### 2.4.4 Iterative screening and structure analysis method

The iterative screening and structure analysis (ISSA) method (Mauersberger, 2005) removes redundant species and redundant reactions simultaneously. A brief summary of a slightly modified ISSA method (without structure analysis) follows.

First, the production and loss terms of ODEs (Eq. 6) from chemical kinetics are separated. The ODEs are re-written as:

$$P_i(c, t) - L_i(c, t) = \sum_j v_{ij} R_j(c, t)$$  \hspace{1cm} (13)

where $P_i(c, t)$ is the production term for the species $i$, $L_i(c, t)$ is the loss term for species $i$, $v_{ij}$ and $R_j(c, t)$ are the same to those in Eq. (6).

Next, the stoichiometric coefficients are separated for reactants ($v_{ij}^+$) and products ($v_{ij}^-$) respectively, i.e.

$$v_{ij}^+ = \begin{cases} v_{ij} & \text{for } v_{ij} > 0 \\ 0 & \text{for } v_{ij} \leq 0 \end{cases}$$

$$v_{ij}^- = \begin{cases} -v_{ij} & \text{for } v_{ij} < 0 \\ 0 & \text{for } v_{ij} \geq 0 \end{cases}$$  \hspace{1cm} (14)

Now, the production terms and the loss terms can be re-expressed as:

$$P_i(c, t) = \sum_j v_{ij}^+ R_j(c, t)$$

$$L_i(c, t) = \sum_j v_{ij}^- R_j(c, t)$$  \hspace{1cm} (15)

In order to find the relative importance of the $j$th reaction to the species $i$, two normalized valuation coefficients of the ISSA are formulated as:

$$f_{ij} = \frac{v_{ij} R_j}{\sum_v v_{ij} R_j} = \frac{v_{ij} R_j}{P_i}$$

$$g_{ij} = \frac{v_{ij}^- R_j}{\sum_v v_{ij}^- R_j} = \frac{v_{ij}^- R_j}{-L_i}$$  \hspace{1cm} (16)

where $f_{ij}$ is to measure the relative importance of the $j$th reaction as a production reaction to the overall production term ($P_i$) for species $i$, while $g_{ij}$ represents the relative importance of the $j$th reaction as a loss reaction to the overall loss term ($L_i$) for species $i$.

Important reactions and species may be determined as follows, a simplified version of the original methodology of Mauersberger (2005).

First, the valuation coefficients $f_{ij}$ and $g_{ij}$ for the actual group of target species are calculated.

Second, both $f_{ij}$ and $g_{ij}$ are sorted in a descending order, namely, the highest value is the first in the list. Then the first $J$ reactions are selected as important reactions if

$$\sum_{j=1,J} f_{ij} \leq (1 - \varepsilon_3) \quad \text{or} \quad \sum_{j=1,J} g_{ij} \leq (1 - \varepsilon_3)$$  \hspace{1cm} (17)

where $\varepsilon_3$ is a threshold ($0 \leq \varepsilon_3 \leq 1$).

Third, reactants of the selected important reactions identified from previous step are joined into the group of important species.

The iterative procedure repeats from the first step until no more species and reactions are added into the list of important species and reactions. The species and reactions that have not been selected are eliminated from the system.

### 2.4.5 Linear lumping method

Formally, lumping is an effective method to reduce the number of species ($n$) in the original mechanism into $\tilde{n}$ species
where \( \hat{n} < n \). In order to obtain a simple and easily implemented chemical scheme, a mathematically based linear approximation lumping method has been developed in this work for a large chemical system such as the MCM.

Linear lumping requires a strategy for determining when two reactant species, \( A_1 \) and \( A_2 \), have sufficiently similar concentration ratios that they may be treated as a single species. Consider the following example with four reactions:

\[
\begin{align*}
A_1 + Z & \xrightarrow{\kappa_1} Y + X \quad \text{(R1)} \\
A_2 + Z & \xrightarrow{\kappa_2} X \quad \text{(R2)} \\
Y & \xrightarrow{k_3} A_1 \quad \text{(R3)} \\
A_1 + A_2 & \xrightarrow{k_4} X \quad \text{(R4)}
\end{align*}
\]

where species \( A_1 \) and \( A_2 \) are the two (potentially) lump-able species; \( X, Y, \) and \( Z \) are other three species in the system; and \( \kappa_1, \kappa_2, k_3, \) and \( k_4 \) are the reaction rate coefficients for each reaction.

According to Huang et al. (2005), if two species, \( A_1 \) and \( A_2 \), can be lumped, the two species (units: mixing ratio) can be written as

\[
\begin{pmatrix}
A_1 \\
A_2
\end{pmatrix} = (A_1 + A_2) \begin{pmatrix}
\frac{A_1}{A_1 + A_2} \\
\frac{A_2}{A_1 + A_2}
\end{pmatrix}
= A \begin{pmatrix}
x \\
1 - x
\end{pmatrix}
\]

where \( x \) is the mixing ratio fraction of \( A_1 \) in the two species \((A_1+A_2)\) system, and \( A \) is the sum of the two species with the unit of mixing ratio.

A criterion for the linear approximation lumping may be introduced: in a chemical system, if the mean fraction \( x \) is almost a constant (for the set of \( A_1 \) and \( A_2 \) encompassing the range of atmospherically relevant concentrations), then we can lump the two species together. The mean fraction can be easily calculated via the mixing ratios of the two compounds.

Based on the criterion, the Reactions (R1–R4) can be easily re-written as:

\[
\begin{align*}
A + Z & \xrightarrow{x_k_1} Y + X \quad \text{(R5)} \\
A + Z & \xrightarrow{(1-x)k_2} X \quad \text{(R6)} \\
Y & \xrightarrow{k_3} A \quad \text{(R7)} \\
A + A & \xrightarrow{x(1-x)k_4} X \quad \text{(R8)}
\end{align*}
\]

in which the reaction rate coefficients for the (R1), (R2), and (R4) have been modified by a factor, the mixing ratio fraction of the corresponding species. However, the rate coefficient for (R3) does not change because the lumping species (\( A_1 \)) in that reaction is a product only.

While the correspondence between Reactions (R1) through (R4) to Reactions (R5) through (R8) will only be exact when the value of \( x \) is a constant, lumping becomes a viable mechanism reduction strategy when the values of \( x \) vary within a sufficiently small range for atmospheric conditions that the remainder of the mechanism is unaffected by exchanging the original (R1–R4) for the lumped (R5–R8) reactions. This concept will be explored here using an example from the MCM.

**Fig. 2.** One species (NAPINAOOH) in a two species system (NAPINAOOH+NAPINBOOH) in the gas phase: (A) fraction from one selected scenario, in which median value and errors for this scenario are calculated; (B) median fraction from 108 selected scenarios covering different HC/NOx (ppbvC/ppbv) regimes.
Figure 2A shows the relationship of two species (NAPINAOOH and NAPINBOOH) of the MCM system, in which the fractions of NAPINAOOH in the two species system (NAPINAOOH+NAPINBOOH) are displayed during the simulation of one scenario. To avoid large perturbation of the fractions, the top 10% and bottom 10% data points were first removed from the total 48 sampled hourly points. Three parameters are acquired from the 80% selected data: the median value of \( x(0.723) \) and two deviations from the median value. The first deviation \( \text{deviation}_1 = 0.0495 \) is the difference between the maximum and the median value, and the second deviation \( \text{deviation}_2 = -0.0287 \) is the difference between the minimum and the median. The two deviations are labeled and indicated by lines with arrows in Fig. 2A.

Figure 2B shows various median fractions (x-axis) of the species NAPINAOOH in the two species system (NAPINAOOH+NAPINBOOH) with two deviations (y-axis) from the selected 108 scenarios. The two dark circles represent the selected scenario studied in Fig. 2A. The values in the x-axis for the two circles in Fig. 2B is the median value (0.723), and the values in the y-axis correspond to the two deviations (on the opposite side of the zero line in Fig. 2B), respectively. The range of the deviations \[ \text{deviation}_2 \text{ deviation}_1 \] and the range of fractions \[ \text{median} + \text{deviation}_2 \text{ median} + \text{deviation}_1 \] for each of the 108 selected scenarios are reflected in Fig. 2B.

Clearly, a large range [0.70 0.88] of the fraction from the 108 scenarios would not be able to give a good estimation of the mean fraction \( x \) used by the approximate linear lumping method. The median fractions are clustered when the 108 selected scenarios are grouped according to their HC/NO\(_x\) ratios (ppbvC/ppbv) (note that the “HC” refers to \( \alpha \)-pinene only in this work). Particularly, the range of the fraction for the first group (HC/NO\(_x\) < 0.67) from 27 scenarios is in a range of [0.708 0.757] with small variations. The mean fraction \( x \) (0.727) of the 27 scenarios can represent the lowest HC/NO\(_x\) situations. While the range of values of \( x \) suggests that a single mean value across all experiments would not always allow Reactions (R5–R8) to approximate Reactions (R1–R4), the clustering of \( x \) values observed here suggests that the HC/NO\(_x\) ratio may be used to define sub-regions within the space of possible HC/NO\(_x\) ratios where the lumping approximation may be valid.

The mean fraction \( x \) in this instance would vary for different groups of HC/NO\(_x\) conditions, and the reduced mechanism by using the lumping method therefore becomes HC/NO\(_x\) dependent. The sets of reaction rate coefficients modified by the mean fraction, as shown in Reactions (R5–R8), will be different for different HC/NO\(_x\) groups, but the chemical reactions (i.e., the stoichiometric coefficients for the reactants and the products in those reactions) remain the same.

Furthermore, the linear lumping method can be performed for more than two species in a group as long as the fraction between any pair of two species in that group does not have a large variation.

### 3 Application of the methods, and discussion

The five methods, DRG/DRGEP, PCA, QSSA, ISSA, and linear lumping, were applied in sequence to the subset of MCM describing \( \alpha \)-pinene oxidation. The \( \alpha \)-pinene subset contains 928 chemical reactions and 310 chemical compounds. In addition to these organic reactions, 48 inorganic thermal chemical reactions and 21 inorganic compounds were combined with that subset for \( \alpha \)-pinene. The complete chemical mechanism with 976 reactions and 331 species in the gas phase, coupled with Pankow’s (1994) gas/particle partitioning module from 149 identified organic condensable species, is called “original mechanism” here. This original mechanism was introduced in our earlier work (Xia et al., 2008) to study SOA formation from 108 scenarios, which cover a wide range of HC/NO\(_x\) ratios between 0.18 and 8.43 (ppbvC/ppbv). The mechanism was run in a box model, representing different HC/NO\(_x\) ratios, for 72 h under “typical” conditions with diurnal forcing in temperature and radiation. The data points from the last 48 h of each 108 scenarios are used for analysis of mechanism accuracy.

The same number of gas phase chemical reactions from the original mechanism were coupled with Pankow’s (1994) gas/particle partitioning module for the 28 important condensable species from Xia et al. (2008), forming what we will hereafter refer to as the “reference mechanism”. This reference mechanism will be the subject of the five different stages of mechanism reduction, described above, and the reduced mechanisms from each stage will be evaluated against the reference mechanism. Note that all inorganic species and reactions are kept untouched at each stage, and the reduction techniques are applied to the organic part only.

The identification of target species is also critical to the mechanism reduction by using the DRG, DRGEP, and ISSA methods. Prior to the reduction procedure to the reference mechanism, 39 target species are selected: the most important 28 condensable organic species; 10 inorganic species (\( O_2 \), HO, O(\(^1\)D), O(\(^3\)P), NO, NO\(_2\), NO\(_3\), HNO\(_3\), HO\(_2\), and H\(_2\)O\(_2\)); and the precursor \( \alpha \)-pinene.

### 3.1 DRG versus DRGEP

As discussed in Sect. 2.4.1, the DRG method focuses on the direct effect among species, while the DRGEP method studies the indirect effect by considering all paths from one species to another. In this section, the two methods are applied separately from the reference mechanism with the same 39 target species, and the results from the two methods are compared. Both methods need a threshold to reduce the
mechanism, noting that the two thresholds have a different meaning, hence different values. In both methods, when the threshold is very small, the number of major species (target species and the necessary species) is close to that of the reference mechanism, and the extent of mechanism reduction is small. In contrast, when a very large threshold value is chosen, the reduced mechanism from that large cutoff value might not be able to reproduce the reference mechanism very well.

Accordingly, the selection of the optimum threshold value is a balance between mechanism accuracy and reduction intensity. After intensive trials of different thresholds, a cutoff value ($\varepsilon_1$) of 0.06 for the DRG method and a cutoff value ($\varepsilon_2$) of 0.01 for the DRGEP method were carefully chosen to derive the reduced mechanisms, these values giving the best trade-off between mechanism accuracy, relative to the reference mechanism, and species reduction. As a result, 139 species and 387 reactions were removed by the DRG method, and 140 species and 377 reactions were removed by the DRGEP method. A large proportion of the species removed (125 in total) were the same for both methods.

The two reduced mechanisms are evaluated against the reference mechanism for the performance of averaged $O_3$ and $NO_x$ in Fig. 3A and B for the 108 scenarios. Both reduced mechanisms under-predict $O_3$ concentrations and over-predict $NO_x$ concentrations for low HC/NOx scenarios. The
The systematic behavior of the reduced mechanisms is caused by the elimination of the organic species and organic reactions from the reference mechanism. Under low HC/NO_x conditions, NO_x concentrations are relatively high. When the organic species, especially some organic radicals (RO_2), are removed from the mechanism, the branches of the NO_x+RO_2, such as RO_2+NO_2=RO2NO2 and RO2+NO=ROONO, consume less NO_x. As a result, the concentrations of the NO_x are higher in the reduced mechanisms. When NO_x concentrations become higher, the reaction of NO_x+O_3 becomes more effective in consuming ozone, and O_3 concentrations are smaller and under-predicted.

From Fig. 3A and B, there is not much difference between the two methodologies in predicting ozone and NO_x concentrations. Therefore, we cannot judge which method is better purely from the performance of ozone and NO_x. Figure 3C shows the errors of the averaged total SOA mass from the two reduced mechanisms against the reference mechanism by evaluating the 108 selected scenarios. Both reduced mechanisms under-predict averaged total SOA mass, because the over-predicted (higher) NO_x in the reduced mechanism affects the formation of SOA through the competition of the chemical paths between RO_2+HO_2 and RO_2+NO_x (the former reaction leads to more condensable products than the latter, and the NO_x over-prediction in the reduced mechanism therefore results in less aerosol formation). The detailed explanation for the low SOA mass is the same as the one given in our separate paper (Xia et al., 2008) when HC/NO_x ratio decreases. The two reduced mechanisms give similar performances for total SOA at low HC/NOx (=0.167 ppbC/ppb). The DRGEP method always gives better results than the DRG method at higher HC/NO_x conditions. Under worst condition, the DRG method produces an error of -24%, and DRGEP only -20% when the HC/NO_x ratio is around 1.0.

DRG and DRGEP methods are very efficient in reducing large chemical mechanisms. The computational cost to reduce a chemical mechanism is linear to the number of the chemical species. Both methods require a minimum systems-dependent knowledge to reduce chemical mechanisms.

One of the objectives for the model reduction is to have an accurate prediction of the total SOA mass. Since the reduced mechanism from the DRGEP method is better than that from the DRG in this aspect, the results of the DRGEP are adopted for further reduction in later stages, where it is called the “first reduced mechanism”.

At this first stage, 140 out of 310 species and 377 out of 928 reactions from the reference mechanism have been removed with the application of the DRGEP method.

3.2 PCA method

The PCA method is applied to the “first reduced mechanism” (the result of Sect. 3.1) to reduce redundant reactions. During the application of PCA to the reaction rate sensitivity matrix, the cutoff values for the eigenvalue and the eigenvector are critical to determine the importance of the reactions at each time point (Turanyi et al., 1989). Reactions are considered important if they appear as significant elements of the eigenvectors associated with large eigenvalues at any time point. The remaining reactions can be removed from the system.

In the previous studies (Turanyi, 1990; Heard et al., 1998; Carslaw et al., 1999), the threshold criterion for eigenvalue (\(\Delta_1\)) was chosen between 0.001 and 1.0 and the corresponding threshold for eigenvector (\(\Delta_2\)) is in the range of 0.01 to
0.2. The selection of these thresholds was solely based on trial-and-error; no systematic method exists for their selection.

The eigenvalues vary from $10^{-40}$ to $10^{10}$ for the “first reduced mechanism”. In order to find the redundant reactions at all the 5184 time points (=108 scenarios ×48 time points/scenario), systematic tests were conducted by varying the thresholds of the eigenvalue ($0.1 \leq \Delta_1 \leq 200$) and eigenvectors ($0.1 \leq \Delta_2 \leq 0.5$) to the “first reduced mechanism”. Note that the eigenvalues below $10^{-10}$, obtained through numerical analysis at double precision in this case, will not be accurately estimated. However, only the eigenvalues above 0.1, which are larger than $10^{-10}$, are studied here.

The number of redundant reactions changes with different combinations of the two thresholds ($\Delta_1$ and $\Delta_2$). In this work, small thresholds ($\Delta_1$ is 1.0, and $\Delta_2$ is 0.2) chosen similarly to previous published studies led to only 20 redundant reactions being identified. One reason for identifying such a small number of redundant reactions is that the PCA method is applied to a larger number of time points than those in the published papers. Another reason is that the a priori application of the DGREP method means that significant mechanism reduction has already taken place.

Higher thresholds are used in this work; the threshold for a significant eigenvalue was set to be $200.0$ ($\Delta_1$), and the corresponding threshold for the eigenvector was $0.5$ ($\Delta_2$). The selection of these thresholds is based on trial and error to remove the maximum number of reactions without significantly altering the chemistry. Finally, the selection results in the eliminating of 76 out of 551 reactions. The reduced mechanism from this stage is called the “second reduced mechanism”.

The PCA method affects the four functional groups (PANs: peroxynitrates; Nitrates: nitrates; ROOHs: organic peroxides; and Acids: organic acids) for the organic aerosol formation. The group of Acids is underestimated and the other three groups are overestimated when the “second reduced mechanism” is compared with the “first reduced mechanism”. The overestimation of the three groups is caused by elimination of reactions, especially some loss reactions that contribute insignificantly to the major species in the three groups. As a result, the errors of the averaged total SOA mass are increased by about 5% for the “second reduced mechanism” in Fig. 4, due to the positive changes of the major species in the two dominant groups of PANs and ROOHs.

All of the species are considered equally important in the PCA method. If the PCA method were implemented earlier than the DRGEP method, some reactions only important to the redundant species would not be removed. Hence, to get a highly reduced mechanism, the DRGEP method should be applied before the PCA method if the two methods are applied together.

### 3.3 QSSA method

Based on the criterion that the species with a fractional error (Eqs. 11 and 12) always less than 0.05 are the candidate QSS species, 30 species have been identified as QSS removable species following the creation of the second reduced mechanism. The 30 species are all alkoxy radicals (RO) – their removal and replacement by the product species is a common step in mechanism reduction through lumping. Similar RO compounds were identified by Whitehouse et al. (2004b) when the QSSA method was applied to reduce the whole set of MCM v2.0. Here we point out that the systematic approach confirms the feasibility of the assumption used during the traditional mechanism compression – the RO species are the QSS species due to their consistently short lifetimes. This does not mean that short lifetime species are necessarily QSS species.

Comparison between the reduced mechanism by using QSSA and that from stage 2 (after the DRGEP and PCA methods) showed that the maximum error of averaged concentrations of the species in both gas and aerosol phases introduced through the application of QSSA is only 0.1% under different HC/NO$_x$ conditions. Moreover, the maximum error for the averaged total SOA mass is only 0.04%, because the total SOA mass is the summation of all the condensable species, with cancellation of errors for individual condensable species.

Following the removal of the 30 QSS species, further reduction may be possible via reaction lumping. The DRGEP, PCA, and QSSA methods were applied to the “reference mechanism” in sequence; some of the species removed via QSSA alter the relative importance of the remaining species. To clean the mechanism, the DRGEP method was applied again to remove another three species which are only important for some of the 30 removable QSS species. The resulting chemical scheme was reduced to 137 species and 439 reactions, and it is called the “third reduced mechanism”.

### 3.4 ISSA method

The value of $\varepsilon_3$ in Eq. (17) is critical and sensitive to the mechanism reduction intensity; usually $\varepsilon_3$ is less than 0.1. In this work, we compare four different options for the application of ISSA to the “third reduced mechanism”.

When the reduction is conducted at each time point and the $\varepsilon_3$ is 0.01 (stage 4#1 in Table 1), the reduction leads to a removal of 2 species and 43 reactions. This new reduced mechanism is compared with the “third reduced mechanism”, and the comparison shows small differences for individual condensable species ($|\text{errors}|<2\%$) and for the total SOA mass ($|\text{errors}|<0.1\%$).

But when $\varepsilon_3$ increases and the reduction procedure is conducted at each sampled time point, the reduced chemical mechanism tends to become inaccurate, computationally singular, and computationally expensive. The reason for the
system becoming computationally singular is that some significant terms in the ODEs for some species are removed during the reduction.

Mauersberger (2005) suggested that the time-averaged reaction rates used in the ISSA method could result in significantly higher reduction than the established method to conduct the reduction at each time point due to the fundamental difference between daytime and nighttime chemistry in atmospheric chemical mechanisms.

When the reaction rates are averaged over the 48 h and the same $\epsilon_3$ (0.01) is applied to the mechanism (stage 4#2), the ISSA method results in the removal of 79 reactions, nearly twice the number of reactions removed by using the analysis conducted at each time point. The number of reactions that could be removed increases to 91 at the $\epsilon_3$ of 0.02, and to 108 at the $\epsilon_3$ of 0.023.

Figure 5 shows the errors of the total SOA mass from four reduced mechanisms against the reference mechanism with the use of this time averaging procedure. The four mechanisms include the three options used in ISSA method (stage 4#2, #3, #4) and the “third reduced mechanism” from previous stage.

Regarding the averaged total SOA mass, the 3 options (stage 4#2, #3, #4) can describe the formation of the SOA accurately within 15%, especially for the option with $\epsilon_3$ at 0.023. Detailed analysis of individual condensable species showed that errors as large as 150% occur for three important condensable species. Thus, 0.023 is not an optimum value to
**Fig. 6.** Additional errors of the condensable species in the gas phase (panel A) and aerosol phase (panel B) due to the linear lumping method.

**Table 2.** A total of 16 species are lumped into 7 groups by using the linear lumping method.

<table>
<thead>
<tr>
<th>Lumped Groups</th>
<th>Species</th>
<th>2&lt;HC/NO(_x)≤3.33</th>
<th>3.33&lt;HC/NO(_x)≤6.67</th>
<th>HC/NO(_x)&gt;6.67</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>C720NO3</td>
<td>0.323</td>
<td>0.312</td>
<td>0.332</td>
</tr>
<tr>
<td></td>
<td>APINCNO3</td>
<td>0.223</td>
<td>0.246</td>
<td>0.293</td>
</tr>
<tr>
<td></td>
<td>C719NO3</td>
<td>0.454</td>
<td>0.442</td>
<td>0.375</td>
</tr>
<tr>
<td>2</td>
<td>C716OH</td>
<td>0.314</td>
<td>0.337</td>
<td>0.360</td>
</tr>
<tr>
<td></td>
<td>C717NO3</td>
<td>0.490</td>
<td>0.445</td>
<td>0.383</td>
</tr>
<tr>
<td></td>
<td>HCC7CO</td>
<td>0.196</td>
<td>0.219</td>
<td>0.257</td>
</tr>
<tr>
<td>3</td>
<td>C106O2</td>
<td>0.481</td>
<td>0.476</td>
<td>0.463</td>
</tr>
<tr>
<td></td>
<td>PINALO2</td>
<td>0.519</td>
<td>0.524</td>
<td>0.537</td>
</tr>
<tr>
<td>4</td>
<td>C3PAN2</td>
<td>0.324</td>
<td>0.350</td>
<td>0.380</td>
</tr>
<tr>
<td></td>
<td>PAN</td>
<td>0.676</td>
<td>0.650</td>
<td>0.620</td>
</tr>
<tr>
<td>5</td>
<td>C4CODIAL</td>
<td>0.631</td>
<td>0.606</td>
<td>0.578</td>
</tr>
<tr>
<td></td>
<td>CO13C4CHO</td>
<td>0.369</td>
<td>0.394</td>
<td>0.422</td>
</tr>
<tr>
<td>6</td>
<td>CH3CO3</td>
<td>0.662</td>
<td>0.641</td>
<td>0.612</td>
</tr>
<tr>
<td></td>
<td>HCOCH2CO3</td>
<td>0.338</td>
<td>0.359</td>
<td>0.388</td>
</tr>
<tr>
<td>7</td>
<td>NAPINAOOH</td>
<td>0.838</td>
<td>0.858</td>
<td>0.874</td>
</tr>
<tr>
<td></td>
<td>NAPINBOOH</td>
<td>0.162</td>
<td>0.142</td>
<td>0.126</td>
</tr>
</tbody>
</table>
reduce the mechanism, the more “accurate” total SOA mass for ($\varepsilon_3=0.023$) having resulted from significant compensating errors for specific condensable species.

The reduced mechanism from stage 4#3 ($\varepsilon_3=0.02$) is the final reduced mechanism at this stage because the accuracy of the chemical mechanism is preserved; while more reactions have been removed than the reduced mechanism from $\varepsilon_3$ at 0.01. Thus, the reduced mechanism from $\varepsilon_3$ at 0.02 is called the “fourth reduced mechanism”.

To conclude, 3 out of 137 species and 91 out of 439 reactions have been removed by using the ISSA method with time-averaged reaction rates.

Up to the present, only one reduced mechanism can be obtained from each stage. In the next stage with the linear lumping method, the reduced mechanisms are different, in terms of the reaction rate coefficients, under different HC/NO$_x$ regimes.

### 3.5 Linear lumping method

As introduced in Sect. 2.4.5, the mean fraction $x$ of one species in a two-species group is approximately constant within set HC/NO$_x$ ranges. The resulting lumping schemes are then HC/NO$_x$ dependent.

Relationships between any two species from the 134 remaining species have been examined to search for lump-able species. As a result, 16 species are lumped into 7 groups. Table 2 shows the name of the lumped species and the fractions of each species in its group when HC/NO$_x$ ranges are different.

Most of the species in their lumping group have the similar chemical structures, such as the three nitrates in the 1st group (C720NO3, APINCNO3, and C719NO3), two RO$_2$ in the 3rd group (C106O2 and PINALO2), two PANs in the 4th group (C5PAN2 and PAN), and two condensable species in the 7th group (NAPINAOOH and NAPINBOOH). The species with different functional groups can also be lumped together in the 2nd group (C716OH, C717NO3, and HCC7CO).

The mean fraction changes with HC/NO$_x$ ratio, because chemical paths differ under different HC/NO$_x$ condition. For example, in the 2nd group, when HC/NO$_x$ (ppbvC/ppbv) increases from the range of 2.00–3.33 to >6.67, the fraction of nitrate species C717NO3 decreases from 0.49 to 0.38. In contrast, the fraction of the C716OH increases from 0.314 to 0.36.

To apply the linear lumping method, the whole spectra of HC/NO$_x$ ratios were tentatively divided into 5 subranges: (1) HC/NO$_x$≤0.667; (2) 0.667<HC/NO$_x$≤2.00; (3) 2.00<HC/NO$_x$≤3.33; (4) 3.33<HC/NO$_x$≤6.67; and (5) HC/NO$_x$>6.67. The last three of these subranges (with HC/NO$_x$ ratio larger than 2.0) will be used here to demonstrate the effects of lumping. The results are similar for the other two subranges. As mentioned in the beginning of this subsection, three HC/NO$_x$ dependent lumping schemes are then obtained for the corresponding last three subranges; each lumping scheme is applied to the scenarios with HC/NO$_x$ ratios in the respective subrange.

Figure 6 compares each of the three reduced lumping mechanisms to the “fourth reduced mechanism” from previous stage, and illustrates that the errors for most condensable species are less than 2% in both gas and aerosol phases, and only three species (H3C25C6PAN, NC102OOH, and C719OOH) have errors up to 20%. These errors result from the mismatches between the lumped and un-lumped product mass in subsequent chemistry following the lumping stage. However, all the three species (H3C25C6PAN, NC102OOH, and C719OOH) have small contributions to the total SOA mass, because their aerosol mass fractions are less than 0.025. Consequently, the errors of the total SOA mass caused by the lumping are smaller than 1% between the “fourth reduced mechanism” and the reduced lumping mechanisms at the fifth stage. Finally, “five fifth reduced mechanisms” were obtained, one for each HC/NO$_x$ range.

### 3.6 Final evaluation

The five reduced mechanisms from the fifth stage are evaluated thoroughly against the reference mechanism with conditions from a total of 243 scenarios, which included the original selected 108 scenarios and another additional 135 scenarios. More information regarding the setting for the 243 scenarios can be found in Xia et al. (2008).

The evaluations are conducted on four levels: (1) O$_3$ and NO$_x$; (2) individual condensable species in the gas phase and
First of all, Fig. 7 shows the errors of O$_3$ and NO$_x$ between the final five reduced mechanisms and the reference mechanism. The reduced mechanism underpredicts O$_3$ to 20%, and overpredicted NO$_x$ up to 25%. The explanation for this behavior for the two species is given in Sect. 3.1.

Next, Fig. 8 shows the errors of the individual condensable species, in both gas and aerosol phases, from the groups of PANs and Nitrates between the final five reduced lumping mechanisms and the reference mechanism. Generally, the reduced lumping mechanism can describe the dominant condensable species (with aerosol mass fraction larger than 0.10, such as C811PAN and C920PAN) accurately within 12% in both gas phase and aerosol phase. But the absolute errors for H3C25C6PAN and C813NO3 reach 40% when the aerosol mass fractions for the two species are less than 0.08.

Figure 9 illustrates that the majority of errors for the species from the groups of ROOHs and Acids are underestimated in both gas and aerosol phases, and these errors are mainly introduced into the reduced mechanism via the DRGEP method in the first stage, where the underestimation of the ROOHs and Acids are introduced by the removal of the reactions for HO$_2$+RO$_2$. The maximum aerosol mass fraction for the ROOHs and Acids is less than 0.10, so these errors from individual species do not have a big impact on the total SOA mass.

Another interesting result is that the errors for the compounds in the groups of PANs and Nitrates are smaller when their mass fraction is large, but the errors for the compounds in the group of ROOHs and Acids are not. This is because...
the formation of the ROOHs and Acids undergoes multiple paths compared to those for the formation of PAN-like compounds and nitrates, and each path could be affected when the mechanism is reduced.

For the condensable species, the errors in the gas phase and those in the aerosol phase are correlated, but they do not follow a directly proportional relationship. Instead, for most species, there is a shift towards positive errors in the gas phase and negative errors in the aerosol phase in Fig. 10; the reduced mechanism will over-predict the gas-phase, and under-predict the aerosol phase. The error shift is caused by the combination of the mechanism reduction and the gas/particle partitioning process: (1) the removal of HO₂+RO₂ reactions in the mechanism reduction process led to the total SOA mass being under-predicted; (2) the under-predicted total SOA mass led to the underprediction of most condensable species in the aerosol phase (a feedback via the absorption); (3) the concentrations of the gas phase condensable species were therefore over-predicted.

Although the errors for individual condensable species can reach as high as 50%, these individual errors cancel with each other. The errors for the four functional groups are displayed in Fig. 11.

The maximum of the absolute error for the group of ROOHs is 12% in Fig. 11, which is smaller than the
Fig. 11. Errors of the 4 groups between the lumped reduced mechanisms and the reference mechanism from the 243 scenarios. The directions of the arrows in this figure reveal the 243 scenarios with increasing HC/NO\textsubscript{x} ratios.

Fig. 12. Errors of the total SOA mass between the five fifth reduced mechanisms and the reference mechanism from the 243 scenarios.
maximum error for individual species at 50%. In addition, the maximum of the absolute errors for the PANs, Nitrates and Acids are only 4%. Regarding the four functional groups, the reduced mechanism could describe the mechanism for SOA formation accurately within 12%.

Finally, the five fifth reduced lumping mechanisms are evaluated against the reference mechanism in terms of total SOA mass. Figure 12 shows that the reduced lumping mechanisms from each subrange of HC/NO\textsubscript{x} ratios could reproduce the averaged total SOA mass within 16% for both the selected 108 scenarios and the additional 135 scenarios. The maximum errors occur at the HC/NO\textsubscript{x} around 1.0, and the total SOA mass is in the range of 3–7 µg/m\textsuperscript{3}. In addition, the maximum difference between the final five reduced lumping mechanisms and the “fourth reduced mechanism” is 2%.

As demonstrated, the final reduced lumping mechanisms can describe accurately for the dominant condensable species, the four functional groups, and the total SOA mass. But large errors still exist for some insignificant condensable species in the reduced mechanism.

Several chemical mechanisms (Kamens et al., 1999; Colville and Griffin, 2004; Chen and Griffin, 2005; Leungsakul et al., 2005; Hu et al., 2007) have been proposed and evaluated against smog chamber studies for the SOA formation. Most of the comparisons focus on the model performance of the total SOA mass only. From the above analysis, reasonably accurate predictions of the total SOA mass, even from many different situations, do not necessarily mean that the mechanism in use describes the underlying chemistry accurately, the total SOA mass being the summation of different components. Ideally, a mechanism should be able to accurately predict the total SOA mass and important functional groups in both the gas and aerosol phases.

To conclude, the reduced mechanism contains only 125 species, and the number of the reactions for the lump-able species does not change. The number of species and the reactions for the reduced mechanisms at each stage are summarized in Table 3.

4 Summary and conclusions

A large chemical mechanism, describing the α-pinene oxidation for the SOA and ozone formation, has been reduced to a simplified mechanism through five stages of mechanism reduction.

First, the 28 condensable species, 10 inorganic species, and α-pinene were selected as target species under a wide range of conditions from 108 selected scenarios; the main objective of the mechanism reduction is that the reduced mechanism could predict these target species accurately at each stage.

The first stage in the mechanism reduction was to identify necessary species for the 39 target species by the DRGEP method, in which a generalized coupling coefficient was used to measure the contribution of one species to the formation/consumption of another species. A threshold of 0.01 for the generalized coupling coefficients was used to find necessary species for each target species. As a result, a total of 140 out of 310 original species and 377 out of 928 original reactions from the reference mechanism were removed by using the DRGEP method.

We would like to point out that the performance of the reduced mechanism based on the DRG or DRGEP method depends strongly on the chosen threshold. The tests performed here were an example, and the 20% bias found in the tests could be reduced through using a more restrictive threshold criteria in the selection process. The selection of the optimum threshold value is a balance between mechanism accuracy and reduction intensity. A smaller threshold value would lead to a better model performance for ozone and the SOA formation. In the future, a constraint of an upper limit error (say 5% or 10%) for ozone and the SOA could be applied in mechanism reduction procedure.

Moreover, the DRG and DRGEP methods could be improved with further examination of the removed species and reactions. Specifically, some methods are proposed, such as replacing the “unimportant species” with chemically related “major species” and modifying reaction rate coefficient to compensate on the removal of completing reactions. All of these proposed methods merit future investigation.

The second stage in the reduction was to use PCA of the rate sensitivity matrix to remove redundant reactions. A large threshold of 200.0 for the eigenvalue and a high threshold of 0.5 for the eigenvector was used to remove an additional 76 out of 551 chemical reactions.

The third stage involved the application of the QSSA by evaluating instantaneous errors instead of the lifetimes. A total of 30 QSS species were identified and removed from the mechanism. In addition, some species could be expressed through algebraic relationships instead of the ODEs.

The fourth stage in the mechanism reduction was to apply the ISSA method to remove unimportant species and reactions. When the reaction rates are averaged on the daily basis, the application of the ISSA method resulted in an efficient reduction of the mechanism. A carefully chosen threshold (ε\textsubscript{3}) of 0.02 led to the removal of 91 out of 439 reactions and 3 species from the mechanism.

The last stage was to apply a newly developed linear lumping approach, in which the species are combined and tested for different HC/NO\textsubscript{x} ratios under the proposed criteria. A total of 16 species have been lumped into 7 groups, and the reduced lumping mechanisms could capture the features of the SOA and ozone formation.

Note that lumping based on the chemical structure and reactivity is referred to as chemical lumping method. It is a widely used method for developing and/or reducing a chemical mechanism, such as the SAPRC99 (Carter, 2000) mechanism. Many directly emitted VOC species have been lumped...
in this way to describe the formation of ozone. In addition, many products with similar chemical structures, such as the ketones, PAN, hydroperoxide could have been lumped.

Our focus in this paper has been the application of lumping methodologies that make use of objective criteria (such as a threshold cutoff in a diagnostic parameter and the resulting associated bias compared to the original mechanism) to reduce the mechanisms. Essentially, we have examined mathematically “automated” approaches. Mathematically based methodologies have appeared frequently in the literature. Based on rigorous or relaxed mathematic transformation, the number of species could be reduced. As mentioned in Sect. 2.1, many techniques (Li and Rabitz, 1989; Li and Rabitz, 1990; Li et al., 1994a, b; Huang et al., 2005) have been developed to describe the exact and/or approximate lumping since the original mathematically based lumping approaches were proposed (Kuo and Wei, 1969; Wei and Kuo, 1969). The simple linear lumping method developed in this paper is a method that belongs to this class of mathematically based lumping approach.

A key component of the mathematical approaches is that the test sets of initial concentrations span the range of realistically possible species concentrations: if the VOC ratios do not change over the range of reasonable atmospheric conditions, the mathematics will show that the lumping is reasonable in that the impact on the remainder of the mechanism is below the error criterion. If the VOC ratios do vary over different test cases, then the VOCs would not be selected for lumping. For a mechanism with many precursor VOCs, it is thus more likely that the product VOCs from a given precursor may be lumped, rather than the initial precursor VOCs themselves. An exception would be a case in which a pair of precursor VOCs with similar reactivities are emitted in similar ratios by a relative small number of emitting processes. The key point is that the methodology is self-correcting and should prevent the sort of errors as potential problems in a larger-scale test. The user of the methods may set, a priori, an error criterion for any of the methods described here. As we’ve recommended in the paper, any use of these methods must be accompanied by comparison of the resulting mechanism to the original one, and bias calculations must be used to evaluate the new mechanism. If removal of a species results in significant mass loss, for example, this will show up in the resulting biases, and the criterion for cut-off of species may be made more restrictive.

The linear lumping method can be applied complementarily to chemical-based lumping method. Linear lumping is unlikely to lump products from different emitted species (since they won’t fulfill the criterion that their concentrations remain at similar ratios, and this will in turn increase the resulting biases relative to the detailed mechanism to an unacceptable level). The linear lumping method could lump the products and/or intermediate products from individual precursor VOCs as demonstrated in this paper. The linear lumping method’s primary use may therefore be the analysis and reduction of very detailed complex mechanisms, such as demonstrated here. Further reductions may require chemical-based lumping.

Ultimately, the full mechanism for α-pinene oxidation was reduced by a factor of 2.5 in terms of the species and reactions via the sequential application of the five reduction techniques.

All the reduced mechanisms have been evaluated against the reference mechanism or the reduced mechanism from previous stages. The reduced mechanisms from the last stage (the fifth stage) can predict the dominant condensable species, four functional groups, and the total SOA mass accurately within 16% for not only the selected 108 scenarios, but also the additional 135 situations.

Because of the complex chemical paths at different HC/NO_x regimes, the errors for the reduced mechanism are less than 10% when HC/NO_x is larger than 2.0 or less than 0.4 (ppbvC/ppbv).

The simplified mechanism could be applied in a 3-dimensional air quality model to predict α-pinene SOA and ozone formation. The methodology used is generally applicable, and could be used to reduce more extensive reaction mechanisms while preserving ozone and SOA formation properties of those mechanisms.

Acknowledgements. This work was supported by Natural Sciences and Engineering Research Council of Canada (NSERC).

References


A. G. Xia et al.: Mechanism reduction for SOA formation


