**Atmospheric
Chemistry
and Physics**

Systematic reduction of complex tropospheric chemical mechanisms, Part II: Lumping using a time-scale based approach

L. E. Whitehouse[1], A. S. Tomlin[1], and M. J. Pilling[2]

[1]Energy and Resources Research Institute, University of Leeds, Leeds LS2 9JT, UK
[2]School of Chemistry, University of Leeds, Leeds LS2 9JT, UK

Abstract. This paper presents a formal method of species lumping that can be applied automatically to intermediate compounds within detailed and complex tropospheric chemical reaction schemes. The method is based on grouping species with reference to their chemical lifetimes and reactivity structures. A method for determining the forward and reverse transformations between individual and lumped compounds is developed. Preliminary application to the Leeds Master Chemical Mechanism (MCMv2.0) has led to the removal of 734 species and 1777 reactions from the scheme, with minimal degradation of accuracy across a wide range of test trajectories relevant to polluted tropospheric conditions. The lumped groups are seen to relate to groups of peroxy acyl nitrates, nitrates, carbonates, oxepins, substituted phenols, oxeacids and peracids with similar lifetimes and reaction rates with OH. In combination with other reduction techniques, such as sensitivity analysis and the application of the quasi-steady state approximation (QSSA), a reduced mechanism has been developed that contains 35% of the number of species and 40% of the number of reactions compared to the full mechanism. This has led to a speed up of a factor of 8 in terms of computer calculation time within box model simulations.

1 Introduction

Secondary pollutants such as ozone, nitrogen dioxide and peroxy acyl nitrate species may all have detrimental effects on human health and the environment. The formation of these species in the atmosphere needs to be understood in order to develop ways in which their concentration levels can be controlled through the reduction of their precursory emissions. Often numerical models are used to investigate their

formation in the troposphere as a result of emissions of NOx (nitrogen oxides) and a large range of emitted volatile organic compounds (VOCs). Many different chemical mechanisms describing the tropospheric oxidation of VOCs in the atmosphere have been developed with varying degrees of complexity. One of the challenges is to create a chemical model which describes the relative impact of individual emitted compounds, but in a computationally efficient way. Explicit mechanisms such as the Leeds Master Chemical Mechanism (Pilling et al., 1999; Jenkin et al., 1997), are capable of representing the degradation pathways of a large number of individual VOC compounds. Such mechanisms are therefore useful in determining the relative impact of the emissions of individual compounds on the formation of secondary pollutants. Explicit mechanisms are computationally expensive however, due to the large numbers of intermediate species they contain. In order to facilitate easier examination of the impact of a wide range of emissions scenarios, chemical mechanism reduction techniques may be used to generate smaller, less computationally expensive schemes from such explicit mechanisms.

In Whitehouse et al. (2004), the application of sensitivity and QSSA (Quasi-Steady State Analysis) analysis has been shown to lead to a reduction of the MCMv2.0 to a level at which most of the very fast and very slow time-scales have been removed. Their application shows that this can leave a large block of species with intermediate time-scales which all contribute in some way to the formation rates of the important and necessary species chosen. In order to reduce the size of the mechanism any further it is necessary to examine these intermediate time-scale species, and to develop a method by which their numbers can be decreased. An approach commonly used is species lumping, where a group of species can be represented in the mechanism by one variable, see Fig. 1. In this way a smaller set of equations are needed to represent the dynamics of the chemical system.

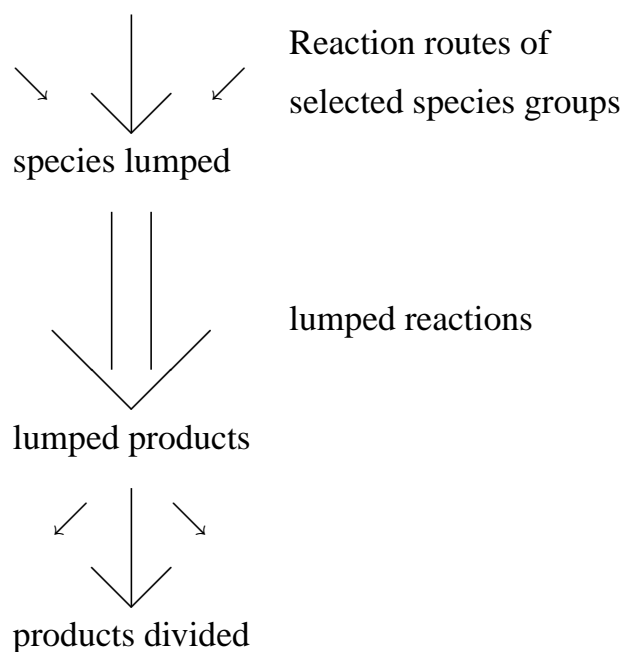*Correspondence to:* A. S.Tomlin
(fueast@leeds.ac.uk)

Fig. 1. An illustration of the theory behind species lumping.

This paper presents a formalised approach to species lumping based on the analysis of the chemical time-scales within the model. The approach is applied to intermediate compounds and is thus relevant to a broad range of emissions scenarios. It is demonstrated through application to a reduced version 2 of the Master Chemical Mechanism (MCMv2.0) as described in Whitehouse et al. (2004). Section 2 of the paper describes previous approaches that have been applied for the lumping of species based on their chemical structures and reactivities. Section 3 describes formal mathematical methods for lumping and introduces the concept of selecting lumping groups based on a time-scale analysis of the dynamics of the chemical system. Section 4 describes the application of a time-scale based lumping approach to the reduced version of the MCMv2.0. Section 5 presents a comparison between the full mechanism and the lumped mechanism for a range of different scenarios and assesses the accuracy of the reduced scheme. Section 6 presents a discussion of the results and Sect. 7 final conclusions.

2 Previous approaches to chemical lumping

In tropospheric chemistry, there are large numbers of volatile organic compounds (VOCs) present in the atmosphere which lead to the large numbers of species within explicit mechanisms. These compounds are traditionally lumped leaving inorganic species expressed explicitly. The mathematical basis for lumping methods has been described in Wang

et al. (1998) and Tomlin et al. (1998). New lumped variables can be created as linear or non-linear combinations of the original species. Formally then, lumping can be described as the reduction of the original set n of equations to an $\hat{n}$−dimensional lumped set where $\hat{n} < n$.

The lumping of VOCs can be achieved in several ways. The simplest approach uses surrogate species not necessarily related to the emitted species in order to represent hydrocarbons released into the atmosphere (Hough, 1988). An example of this simplified lumping can be seen in Leone and Seinfeld (1985), where the total reactive hydrocarbon concentration is represented as 75% n-butene and 25% propene. These, coupled with four aldehyde species, are the only species used to represent organic compounds in the mechanism.

A second approach involves grouping together reactions of species with similar molecular structure, for example splitting VOCs into alkanes, alkenes and aromatic compounds. This type of chemical lumping implies that the lumped compounds are summations of several individual VOCs and is therefore a simple application of linear lumping. Reaction rates and rate constants are determined by analysing kinetic and mechanistic data from various authors as well as empirically fitting smog chamber data. This method was used in Stockwell et al. (1990) with RADM2, where VOCs are grouped together into a manageable set of VOC classes based on their similarity in oxidation reactivity and emission magnitudes (Middleton et al., 1990). Each category of VOC is represented by several model species that span the required reactivity range.

When the organic species are lumped, the principle of reactivity weighting is followed to ensure that the differences in reactivity are taken into account. In Stockwell et al. (1990) reactivity weighting is based on the assumption that the effect of VOC emissions on the simulation results is approximately proportional to the amount of a compound that reacts on a daily basis. Therefore, an emitted compound can be represented by a model species which reacts at a different weight, provided that a weighting factor is applied to the emissions of the compound under consideration. The factor used is the ratio of the fraction of the emitted compound which reacts to the fraction of the model species which reacts:

$$F = \frac{1 - e^{(-k_{OH\ Emit} \times \int [OH] dt)}}{1 - e^{(-k_{OH\ Model} \times \int [OH] dt)}}, \tag{1}$$

where $k_{OH\ Emit}$ is the rate constant for the reaction of OH with the individual compound, $k_{OH\ Model}$ is the rate constant for the reaction of OH with the model species, and the term $\int [OH] dt$ is the daily average integrated OH radical concentration. If the emitted species and the model species are highly reactive this factor tends to 1.

These methods led to the development of the Regional Acid Deposition Model (RADM2) which has been revised with new experimental data to give a reduced mechanism, RACM. In RACM the VOCs are aggregated into 16 anthropogenic and 3 biogenic model species. The rate constants

for the reactions of the model species with OH were calculated as the weighted mean of the rate constants of individual compounds on the basis of the emissions E in units of moles per year taken from the US emissions inventory, (Middleton et al., 1990).

An updated version of the approach was developed in Makar and Polavarapu (1997); Makar et al. (1996) which, instead of using integrated reactivity weighting, uses a dynamic lumping method that directly alters the differential equations at each time-step. The method is shown to reduce the lumping errors when compared with the average approach, since in reality OH shows a strong diurnal profile. The disadvantage of these types of approaches is that they depend on the relative emissions chosen for the different VOC compounds and the lumped mechanism is not therefore guaranteed to apply to future emissions scenarios or emissions scenarios for different countries or regions.

A third type of chemical lumping uses a structural approach where organic species are grouped according to bond type. Reactions of similar carbon bonds are assumed to have similar reactivities. This method has an advantage over the previous two as fewer surrogate species are required to represent a wide range of organic compounds in the atmosphere. One example of this approach is the Carbon Bond Mechanism, CBM-EX, discussed in Gery et al. (1989), which contains 204 reactions between 90 species. While inorganic species remain explicit in the carbon bond mechanism, organic compounds are divided into the different bond types composing their structure. The approach assumes that all bond groups of similar type (such as the paraffin carbon bond C-C as represented by PAR) react at the same rate. In reality, the reactivity of the species is influenced by the size of the molecule in which the bond occurs. Also, carbon bond type mechanisms are based on the reactions of functional groups, but have not been systematically developed to take account of interactions between the different functional groups in a molecule. Their rates are based on optimisation against a limited set of smog chamber experiments which removes the potential for automation of the reduction technique.

A more automatic method of structural lumping was developed in Fish (2000) for aliphatic hydrocarbon oxidation in the troposphere. Functions are used to calculate reaction rates and chemical products based on the initial composition of VOCs in the atmosphere. One advantage of this method is that different reduced mechanisms are generated for different emissions scenarios. Thus the method can be used to assess the success at limiting ozone formation of reactions control strategies that change VOC composition. In order to investigate a new set of emissions conditions a new mechanism must be generated however.

The disadvantages of several of the methods discussed above are due to the fact that they deal with the lumping of primary VOCs. This means that in the event of wanting to investigate different emissions scenarios, the entire reduced mechanism has to be recalculated. A more general approach

would require that the same reduced mechanism could be used to generate data for a wide range of emissions scenarios. In order to achieve this, and to provide a more general mechanism, lumping based on intermediate compounds should be considered. Jenkin et al. (2002) developed a chemical approach to lumping where the key assumption is that the potential for ozone formation from a given VOC is related to the number of reactive, that is C-C and C-H, bonds it contains. This quantity is then used to identify a series of generic intermediate radicals and products which represent species generated from the degradation of a variety of VOCs. The resultant mechanism contains approximately 570 reactions and 250 species giving a good reproduction of the selected trajectories produced by the full MCM. This type of lumping however, requires a high level of familiarity with the chemical details of the mechanism which is not straightforward when dealing with a mechanism of almost 11 000 reactions.

A more mathematically based method is therefore desirable which can be systematically and automatically applied to large and complex mechanisms. In order to carry out species lumping for the MCM, an automatic mathematically based technique is developed here, based on combining information related to the chemical structure of the mechanism through species chemical lifetimes, with techniques based on a more formal mathematical lumping as described in Tomlin et al. (1998); Li and Rabitz (1989, 1990); Wang et al. (1998). In the following sections criteria for the selection of lumped groups will be established and a function for the mapping of the lumped species back to the un-lumped species developed.

3 Formal mathematical approaches to lumping

The mathematical approach to linear lumping taken in Li and Rabitz (1989, 1990); Wang et al. (1998) shows the reduction of a $n-$dimensional system of equations describing the rate of change of chemical species $vecc$

$$\frac{d\boldsymbol{c}}{dt} = \boldsymbol{f}(\boldsymbol{c}, \boldsymbol{k}), \boldsymbol{c}(0) = \boldsymbol{c}^0 \tag{2}$$

to an $\hat{n}-$ dimensional lumped set,

$$\frac{d\hat{\boldsymbol{c}}}{dt} = \hat{\boldsymbol{f}}(\hat{\boldsymbol{c}}), \tag{3}$$

where $\hat{n} \leq n$ and $\boldsymbol{k}$ is the vector of reaction rate coefficients.

This is achieved through the transformation:

$$\hat{\boldsymbol{c}} = \mathbf{h}(\boldsymbol{c}), \tag{4}$$

where $\mathbf{h}$ is some linear or non-linear function of the original variables $\boldsymbol{c}$.

For linear lumping the new lumped variables are a linear combination of the original ones;

$$\hat{c}_i = m_{i,1} c_1 + m_{i,2} c_2 + \ldots + m_{i,n} c_n.$$

Table 1. Range of time-scales left at each reduction stage for trajectory 7 as defined in Whitehouse et al. (2004), after 36 h of simulation.

range of time-scales $\frac{1}{|\lambda|}$ (s)	number remaining			
	full	stage 1	stage 3	stage 5
fast $<1\times10^{-4}$	711	710	434	6
intermediate	2359	2175	1915	1863
slow $>2\times10^6 (\approx 24 \text{days})$	416	205	100	100
total	3487	3091	2454	1969

Therefore Eq. (4) can be simplified to

$$\hat{c} = \mathbf{M}c. \tag{5}$$

where $\mathbf{M}$ is a $\hat{n}\times n$ real constant matrix called the lumping matrix, and the new $\hat{n}$ set of odes for the lumped system is given by

$$\frac{d\hat{c}}{dt} = \mathbf{M}f(c). \tag{6}$$

For exact lumping $\mathbf{M}f(c)$ must be a function of $\hat{c}$ so that the reduced system can be expressed in terms of the new variables. Therefore we need to know the generalised inverse (Campbell and Meyer, 1979) of $\mathbf{M}$ since

$$c = \mathbf{M}^{-1}\hat{c}, \tag{7}$$

so that

$$\frac{d\hat{c}}{dt} = \mathbf{M}f(\mathbf{M}^{-1}\hat{c}). \tag{8}$$

The inverse mapping from the $\hat{c}$ space to the c space is equally important as the forward mapping, not only because it provides a link between the lumped species and the original species, but because its existence is a necessary and sufficient condition for exact lumping.

Wei and Kuo (1969) and more recently Li and Rabitz (1989, 1990) and also Wang et al. (1998) have set out conditions for the exact and approximate linear lumping of ordinary differential equations, and give examples of techniques which can be used to find the lumped schemes. In the linear case the method involves finding suitable lumping matrices of a chosen dimension and their inverses, i.e. an invertible mapping. The inverse is not unique and any generalised M^{-1} can be used to generate a lumped system. The following section describes how this generalised inverse can be approximately found using information related to the system time-scales.

3.1 Species lumping and system time-scales

Following the earlier stages of reducing the MCMv2.0 using sensitivity analysis and QSSA based techniques, 1969

species and 6168 reactions remain. Most of the very fast and very slow time-scales have been removed using these methods, but a large number of time-scales in the intermediate range are still present. Many groups of species exhibit the same or similar time-scales. Table 1 shows an example the numbers of species remaining in the fast, intermediate and slow categories as determined in Whitehouse et al. (2004). Here, λ represents the eigenvalue of the system Jacobian.

The similarity of many of the system time-scales may in part be due to the fact that many of the rate coefficients for this section of the mechanism have been approximated using structural addivity relationships, Jenkin et al. (1997). This means that there are groups of reactions which have the same or similar rate coefficients. If a group of species all react through the same paths, in reactions with the same rate coefficients then their time-scales and therefore their chemical dynamics will be identical. This indicates useful criteria that can be exploited when developing an automatic method for choosing groups in which species can be lumped without loss of accuracy to the mechanism.

4 Reduction of the MCM using time-scale based lumping techniques

4.1 Selection of lumped groups

The chemical lifetime τ_i for each of the remaining species, is given by $\tau_i = -\frac{1}{J_{i,i}}$ where J_{ii} is the ith diagonal entry of the Jacobian of the system. Analysis reveals that there are large groups of species with identical or very similar lifetimes across each simulated trajectory. In the case of reactions with some species, for example OH, groups of species have rate coefficients that are either identical or sufficiently similar to each other. As the similarity of lifetimes coincides with species of a similar type taking part in reactions of the same type e.g. with the same other species, such as OH or NO, or decomposition, these characteristics form a good basis from which to devise a lumping strategy. If enough suitable lumping groups can be identified this will again lead to considerable reduction in the computational time required to solve the system of chemical rate equations. This level of reduction will be referred to as Stage 6 in future discussion.

The admission of a species to a lumped group with primary member S_1 can be carried out according to the procedure described in Fig. 2 which is summarised below:

- Select a group of species with similar lifetimes at chosen timepoints.

- Taking the first species, see how many reactions it is present in as a reactant. Does the next possible member of the group react in the same reaction types? This feature can be determined automatically by the code from the input file describing the reaction mechanism. If yes move on to the next criteria, if no discard.

- Do each of the reactions in the new set have rate coefficients within a certain percentage of the reactions of the first species? If yes, continue, if no, discard.

- For each matched reaction, does the species under consideration react with the same species as the primary reaction? If yes, add the species to the lumping group. If no discard.

Once the group of species which is to form the lumped group has been selected, the new lumped species is formed by summing the species in the group. The lumping is therefore linear where the lumping matrix $\mathbf{M}$ consists of entries which are either 1 or zero. This lumped species will then replace the separate species in the production reactions. Only one of each reaction type will need to be retained for those reactions where the original species appeared as reactants. The product ratios of these new equations will be determined from the products of the unlumped reactions.

4.2 Formation of lumped equations

4.2.1 Example 1

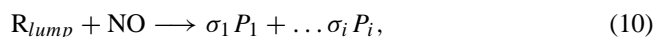In the following example species of the type R_j are formed in only one reaction as follows,

$$S_1 + \ldots + S_l \longrightarrow R_j + \text{products} \tag{9}$$
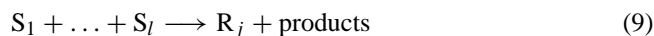
where each $S_1, \ldots, S_l$ are the reactants and k_m is the rate coefficient.

If the selected group of i species $R_1 \ldots R_i$ satisfy the lumping criteria described above and all react with NO at a rate of k_1 then i equations

$$R_j + \text{NO} \longrightarrow P_j, \ j = 1, i, \text{ where } P_j \text{ are products,}$$
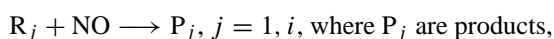
can replaced by a single equation,

$$R_{lump} + \text{NO} \longrightarrow \sigma_1 P_1 + \ldots \sigma_i P_i, \tag{10}$$

where $R_{lump} = R_1 + R_2 + \ldots + R_i$.

It is necessary when implementing this lumping method to devise some way in which the relative concentrations of the products can then be calculated, as the original reactions in which the products were formed, have been removed. As the new lumped species R_{lump} will still be formed through i production channels, the ratio between the rate at which the lumped species is formed through each channel can be used to calculate a variable coefficient for each product species in the lumped equation. This is essentially equivalent to specifying a generalised inverse of the mapping $\mathbf{M}$ as discussed above. Here $\mathbf{M}$ is the forward mapping $\mathbf{M} = \begin{pmatrix} 1 & 1 & \ldots & \ldots & 1 \end{pmatrix}$ such that $\hat{c}_1 = \mathbf{M}c = c_1 + c_2 + \ldots + c_n$, where $\hat{c}_1$ is the lumped variable and $c_1 \ldots c_n$ are the original variables. So in this
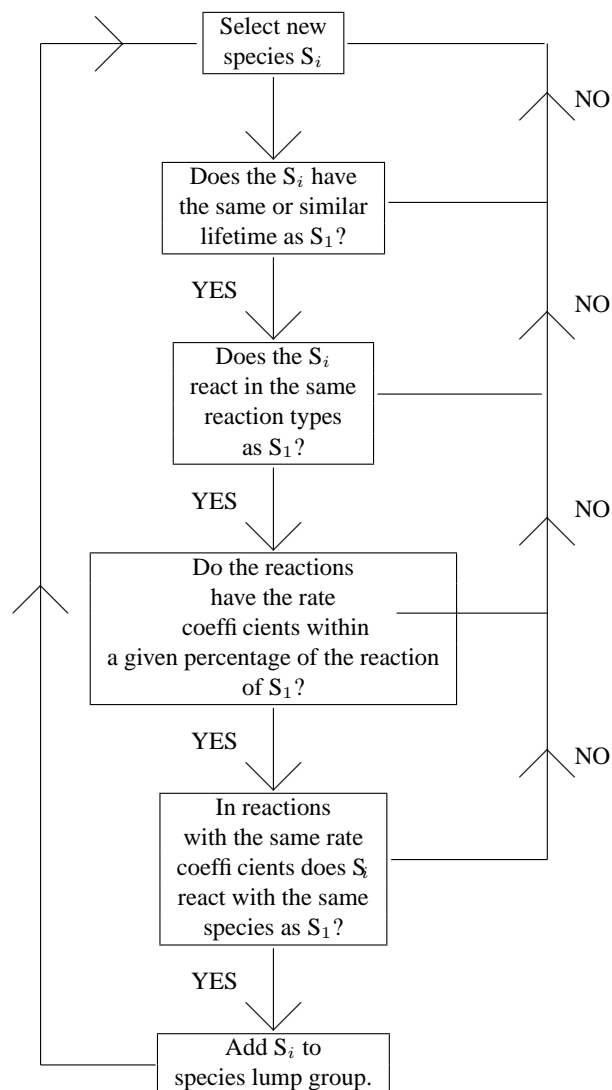

Fig. 2. A flow chart detailing the manner in which species are selected to join lumps.

instance the inverse of $\mathbf{M}$ will have the form $\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ \vdots \\ x_n \end{pmatrix}$ where $x_1 + x_2 + \ldots + x_n = 1$. Bearing this in mind, it is possible to define σ_1 as referred to in Eq. (10) as,

$$\sigma_1 = \frac{\phi_1}{\sum_{j=1}^{i} \phi_j}, \tag{11}$$

where

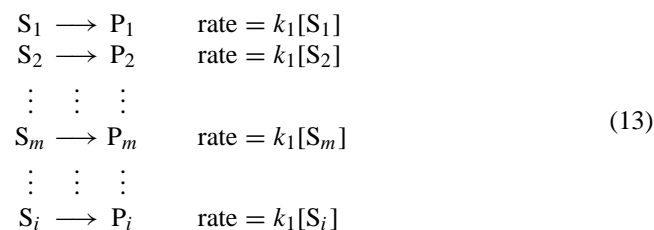$$\phi_1 = k_m([S_1] \times \ldots \times [S_l]). \tag{12}$$

The remaining $(i - 1)$ ϕ_k are defined in the same way using the production reactions for R_2 to R_i. These values must be

recalculated at each time-point as the Jacobian of the system is time-dependent.

4.2.2 Example 2

Alternatively, the use of the ratios defined above can be justified in the following manner.

If there is a set of i first order reactions as follows,

$$
\begin{array}{lll}
S_1 \longrightarrow P_1 & \text{rate} = k_1[S_1] \\
S_2 \longrightarrow P_2 & \text{rate} = k_1[S_2] \\
\vdots \quad \vdots \quad \vdots \\
S_m \longrightarrow P_m & \text{rate} = k_1[S_m] \\
\vdots \quad \vdots \quad \vdots \\
S_i \longrightarrow P_i & \text{rate} = k_1[S_i]
\end{array}
\tag{13}
$$

which all have rate coefficient k_1, then $\frac{\partial P_m}{\partial t} = k_1[S_m] +$ other terms. Here the "other terms" come from any other reactions in which P_m is present as a reactant or product. If the species $S_1 \ldots S_i$ are lumped to form $[S_{lump}] = [S_1] + \ldots + [S_m] + \ldots + [S_i]$ then Eq. (13) are lumped to give

$$
S_{lump} \longrightarrow \sigma_1 P_1 + \sigma_2 P_2 + , \ldots , +\sigma_m P_m + \ldots + \sigma_i P_i. \tag{14}
$$
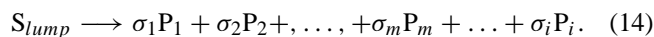
So that

$$
\begin{aligned}
\frac{\partial P_m}{\partial t} &= \sigma_m k_1[S_{lump}] + \text{other terms} \\
&= \sigma_m k_1([S_1] +, \ldots , +[S_m] +, \ldots , +[S_i]) + \text{other terms}
\end{aligned}
\tag{15}
$$

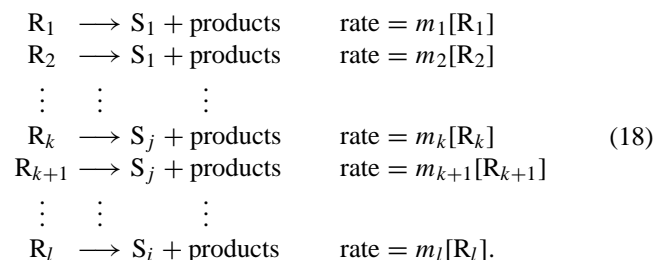Since $\frac{\partial P_m}{\partial t} = k_1[S_m] +$ other terms, it can then be said that,

$$
k_1[S_m] = \sigma_m k_1 ([S_1] +, \ldots , +[S_m] + \ldots + [S_i]). \tag{16}
$$

From Eq. (16) it can then be deduced that

$$
\sigma_m = \frac{k_1[S_m]}{k_1([S_1] +, \ldots , +[S_m] +, \ldots , +[S_i])}. \tag{17}
$$

As the individual quantities of $S_1, \ldots, S_m, \ldots, S_i$ are no longer calculated, the ratio needs to be expressed in terms of different variables.

Assume that $S_1, \ldots, S_m, \ldots, S_i$ are produced in a group of first order reactions such that

$$
\begin{array}{lll}
R_1 \longrightarrow S_1 + \text{products} & \text{rate} = m_1[R_1] \\
R_2 \longrightarrow S_1 + \text{products} & \text{rate} = m_2[R_2] \\
\vdots \quad \vdots \qquad \vdots \\
R_k \longrightarrow S_j + \text{products} & \text{rate} = m_k[R_k] \\
R_{k+1} \longrightarrow S_j + \text{products} & \text{rate} = m_{k+1}[R_{k+1}] \\
\vdots \quad \vdots \qquad \vdots \\
R_l \longrightarrow S_i + \text{products} & \text{rate} = m_l[R_l].
\end{array}
\tag{18}
$$

As the amount of S_j that was produced in the original system is proportional to $(m_k[R_k]) + (m_{k+1}[R_{k+1}])$, the ratio 17 can be accurately expressed in terms of $(m_k[R_k]) + (m_{k+1}[R_{k+1}])$. So therefore

$$
\sigma_k = \frac{(m_k[R_k]) + (m_{k+1}[R_{k+1}])}{m_1[R_1] + \ldots + (m_k[R_k]) + (m_{k+1}[R_{k+1}]) + \ldots + m_l[R_l]}. \tag{19}
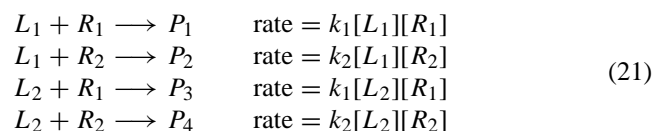$$

Although in this example only first order reactions have been used, this technique can easily be generalised in order to encompass reactions of any order. The σ values can be calculated in the following manner,

$$
\sigma_m = \frac{\text{rate of formation of } S_j}{\text{sum of rate of formation of all species in lumped group}}. \tag{20}
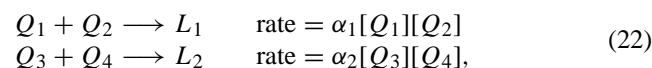$$

The accuracy of this expression does depend on the species S_j being intermediate compounds as it gives rise to the assumption that the initial concentration of S_j is zero. However since it has been explicitly stated that only intermediate species are eligible for lumping this concern is minimised.

4.2.3 Example 3

For the example shown below, species L_1 and L_2 have identical lifetimes and both react with species R_1 and R_2 to produce various products P_1, P_2, P_3 and P_4. The reactions have rates k_1 and k_2.

$$
\begin{array}{lll}
L_1 + R_1 \longrightarrow P_1 & \text{rate} = k_1[L_1][R_1] \\
L_1 + R_2 \longrightarrow P_2 & \text{rate} = k_2[L_1][R_2] \\
L_2 + R_1 \longrightarrow P_3 & \text{rate} = k_1[L_2][R_1] \\
L_2 + R_2 \longrightarrow P_4 & \text{rate} = k_2[L_2][R_2]
\end{array}
\tag{21}
$$

Then if L_1 and L_2 are produced in the following manner,

$$
\begin{array}{lll}
Q_1 + Q_2 \longrightarrow L_1 & \text{rate} = \alpha_1[Q_1][Q_2] \\
Q_3 + Q_4 \longrightarrow L_2 & \text{rate} = \alpha_2[Q_3][Q_4],
\end{array}
\tag{22}
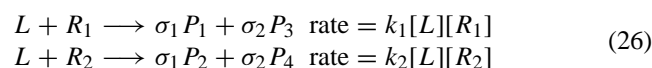$$

the rate equations of L_1 and L_2 are given by

$$
\frac{\partial[L_1]}{\partial t} = -k_1[L_1][R_1] - k_2[L_1][R_2] + \alpha_1[Q_1][Q_2] \tag{23}
$$

$$
\frac{\partial[L_2]}{\partial t} = -k_1[L_2][R_1] - k_2[L_2][R_2] + \alpha_2[Q_3][Q_4] \tag{24}
$$

and the rate equation of $[L_1] + [L_2]$ is

$$
\begin{aligned}
\frac{\partial([L_1] + [L_2])}{\partial t} =\ & -k_1[L_1][R_1] - k_2[L_1][R_2] + \alpha_1[Q_1][Q_2] \\
& -k_1[L_2][R_1] - k_2[L_2][R_2] + \alpha_2[Q_3][Q_4].
\end{aligned}
\tag{25}
$$

If $[L] = [L_1] + [L_2]$ then Reactions 21 can be reduced to:

$$
\begin{array}{lll}
L + R_1 \longrightarrow \sigma_1 P_1 + \sigma_2 P_3 & \text{rate} = k_1[L][R_1] \\
L + R_2 \longrightarrow \sigma_1 P_2 + \sigma_2 P_4 & \text{rate} = k_2[L][R_2]
\end{array}
\tag{26}
$$

where:

$$
\sigma_1 = \frac{\alpha_1[Q_1][Q_2]}{\alpha_1[Q_1][Q_2] + \alpha_2[Q_3][Q_4]}
$$

$$
\sigma_2 = \frac{\alpha_2[Q_3][Q_4]}{\alpha_1[Q_1][Q_2] + \alpha_2[Q_3][Q_4]}
\tag{27}
$$

Table 2. Time-scales for species in first group chosen for lumping, from Trajectory 7 after 36 h.

species	chemical lifetime (s)
CH_3CO_3H	8.1×10^4
$C_6H_5CO_3H$	7.0×10^4
$2-CH_3C_6H_4CO_3H$	8.1×10^4

Table 3. Rate coefficients for Reactions (28)–(36).

reaction number	rate coefficient (molecule $cm^{-3})^{1-m}s^{-1}$
(28)	$3.7 \times 10^{-12} = k_1$
(31)	$3.7 \times 10^{-12} = k_2$
(34)	$4.7 \times 10^{-12} = k_3$
(29), (32), (35)	$J(41)$
(30), (33), (36)	$KAPHO_2 \times 0.71$

4.3 Application of the lumping methodology to the MCMv2.0

Since it was not practical to base the grouping strategy on data from all of the trajectories investigated during the previous reduction stages it was decided to carry out the analysis on a single trajectory, and then to test the accuracy of this assumption on the full set of trajectories. Trajectory conditions for the application of lumping were chosen to give maximum ozone concentrations of 80 ppb and total NO_x of 300 ppb. The full set of 94 trajectories used to test the strategy are as defined in Whitehouse et al. (2004) and were designed to cover a wide range of conditions that may be typical of a UK urban area.

Information relating to the lumped groups developed is presented in Tables 7 and 8 in Appendix 1. The tables show the type of compounds that have been grouped, their chemical lifetimes after 36 h of simulation for the selected trajectory, and the rates of their reactions. Rates which are identical between all species within the lumped group are shown in the column "Equivalent Reaction Rates". Where the reaction rate of each species with OH differs, the range is shown in the right hand column. The groupings are shown to depend on each component member of the lump having a similar lifetime. The use of lifetimes therefore provides an automatic way of selecting possible groupings. Differences in lifetime, where they exist, are due to differences in the reaction rate for the species reacting with OH. Where the rate of reaction of each species with OH is identical then no entry in the right hand column is given. In this case the species have exactly equivalent lifetimes throughout the simulation and the lumping is exact. From the table it appears that many of the lifetimes are extremely similar at the chosen time-point and therefore one might expect that much larger lumped groups could be chosen. However, the difference in reaction rate with OH can be significant in determining suitable lumping groups, since using average reaction rates within the lumping procedure can lead an appreciable build up of errors over several day trajectories if the groups are not carefully chosen. Significant errors may also spread to other species that are coupled to the lumped ones. In the method therefore, the size of the lumps is controlled in order to minimize the propagation of errors. Species are added into the lumped group in order of decreasing similarity of lifetimes. When the addition of a new species to the lump leads to a signif-

icant increase in overall error of selected important species, growth of the lump is terminated and a new lump started. The lumps are therefore of differing sizes with varying ranges of reaction rates for species within the lump. In some cases the range of reaction rates with OH vary by only a few percent but in other cases, such as with the two large groups of peracids, the rate constants may vary by almost a factor of 10 without significant degradation in accuracy of the final lumped scheme. This suggests a lower sensitivity to the individual rates of reaction with OH for these compounds. For peroxy acyl nitrates, the lumps tend to be smaller with lower relative differences between the smallest and largest rates of each compound within the lump suggesting higher sensitivities. The groups detailed in Tables 7 and 8 therefore represent the largest possible lumps without leading to the build up of errors. Several examples will now be demonstrated in order to illustrate the lumping method.

4.3.1 Peracid example

The first lump shown in Table 7, L1CO3H, contains 59 species. Three species in this group are now used in order to illustrate the method, and their lifetimes are shown in Table 2. Each of these species reacts in 2 different ways, with OH and with O_2. They are formed by the reaction of HO_2 with a XCO_3, where XCO_3 is any species whose terminal group is CO_3. The set of reactions is given below in Eqs. (28) to (36). The rate coefficients for each reaction are given in Table 3, where $KAPHO_2 = 2.91 \times 10^{-13} exp(1300/T)$ molecules $cm^{-3} s^{-1}$ and T is temperature.

$$CH_3CO_3H + OH = CH_3CO_3 + H_2O \quad (28)$$
$$CH_3CO_3H + O_2 = OH + CH_3O_2 + CO_2 \quad (29)$$
$$HO_2 + CH_3CO_3 = CH_3CO_3H + O_2 \quad (30)$$

$$C_6H_5CO_3H + OH = C_6H_5CO_3 + H_2O \quad (31)$$
$$C_6H_5CO_3H + O_2 = OH + C_6H_5O_2 + CO_2 \quad (32)$$
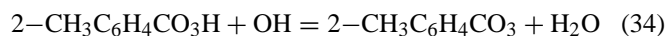$$HO_2 + C_6H_5CO_3 = C_6H_5CO_3H + O_2 \quad (33)$$
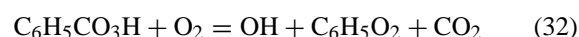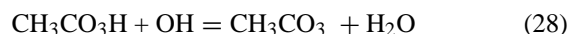
Table 4. Rate coefficients for PAN lumping example.

equation number	rate coefficient (molecule cm^{-3})$^{1-m}$ s^{-1}	m
(43)	4.30×10^{-11}	2
(46)	4.44×10^{-11}	2
(49)	4.47×10^{-11}	2
(52)	4.47×10^{-11}	2
(44), (47), (50), (53)	3.3×10^{-4}	1
(45), (48), (51), (54)	1.1×10^{-11}	2

$$2-CH_3C_6H_4CO_3H + OH = 2-CH_3C_6H_4CO_3 + H_2O \quad (34)$$

$$2-CH_3C_6H_4CO_3H + O_2 = OH + 2-CH_3C_6H_4O_2 + CO_2 \quad (35)$$

$$HO_2 + 2-CH_3C_6H_4CO_3 = 2-CH_3C_6H_4CO_3H + O_2 \quad (36)$$
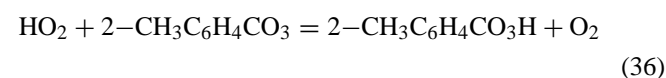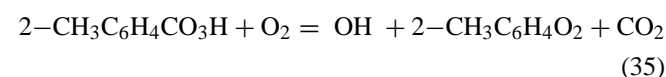
For the purposes of illustration we define a partial lump [LUMP1CO$_3$H$_{part}$]=[CH$_3$CO$_3$H] + [C$_6$H$_5$CO$_3$H] + [2−CH$_3$C$_6$H$_4$CO$_3$H] which gives equations:

$$OH + LUMP1CO_3H_{part} \longrightarrow \sigma_1 \, CH_3CO_3 + \sigma_2 C_6H_5CO_3$$
$$+ \sigma_3 (2-CH_3C_6H_4CO_3 + H_2O) \quad (37)$$

$$LUMP1CO_3H_{part} + O_2 \longrightarrow OH$$
$$+ CO_2 + \sigma_1 CH_3O_2$$
$$+ \sigma_2 C_6H_5O_2$$
$$+ \sigma_3 (2-CH_3C_6H_4O_2) \quad (38)$$

$$HO_2 + CH_3CO_3 \longrightarrow LUMP1CO_3H_{part} + O_2 \quad (39)$$

$$HO_2 + C_6H_5CO_3 \longrightarrow LUMP1CO_3H_{part} + O_2 \quad (40)$$

$$HO_2 + 2-CH_3C_6H_4CO_3 \longrightarrow LUMP1CO_3H_{part} + O_2, \quad (41)$$
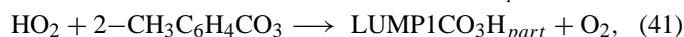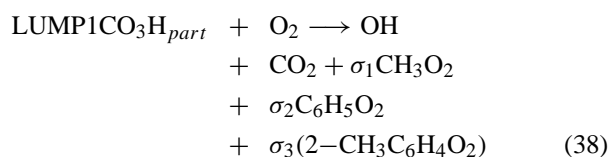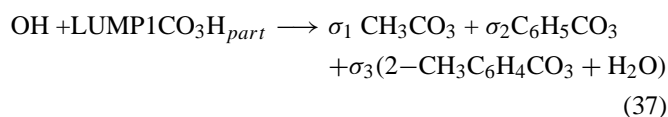
where

$$\sigma_i = \frac{\tau_i}{\sum_{j=1}^{3} \tau_j} \quad (42)$$

and $\tau_1 = k_1[HO_2][CH_3CO_3]$,
$\tau_2 = k_2[HO_2][C_6H_5CO_3]$,
$\tau_3 = k_3[HO_2][2-CH_3C_6H_4CO_3]$,
see Table 3 for details of k_1, k_2 and k_3.

The rate coefficient for Eq. (37) is obtained by taking an average of the rate coefficients of Eqs. (28), (31) and (34), that is 4.0×10^{-12}. The rate coefficient of Eq. (38) is J(41) and the rate coefficient of Eqs. (39), (40) and (41) is KAPHO$_2 \times 0.71$.

By lumping the three species, the number of species in the system is decreased by 2 and 4 reactions are removed. These three species have been lumped as part of a larger group as shown in Table 7. When complete, the creation of L1CO$_3$H leads to the removal of 58 species and 116 reactions.
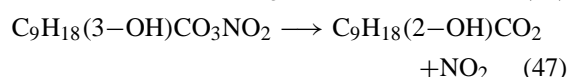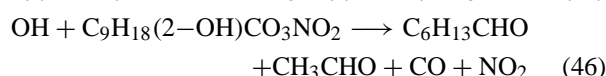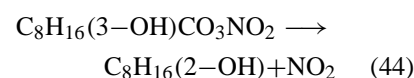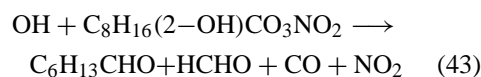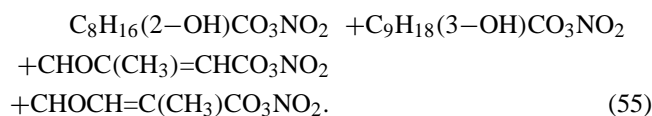
4.3.2 PAN example

The following example examines the lumping of 4 peroxy acyl nitrate (PAN) species which are grouped within PAN17. Information relating to this group can be seen in Table 8. It can be seen that although their lifetimes differ, the variation between the first and last in the group is small at the chosen time point, and is essentially driven by differences in the reaction rate of each compound with OH.

The four species react in the following manner,

$$OH + C_8H_{16}(2-OH)CO_3NO_2 \longrightarrow$$
$$C_6H_{13}CHO + HCHO + CO + NO_2 \quad (43)$$

$$C_8H_{16}(3-OH)CO_3NO_2 \longrightarrow$$
$$C_8H_{16}(2-OH) + NO_2 \quad (44)$$

$$C_8H_{16}(2-OH) + NO_2 \longrightarrow C_8H_{16}(3-OH)CO_3NO_2 \quad (45)$$

$$OH + C_9H_{18}(2-OH)CO_3NO_2 \longrightarrow C_6H_{13}CHO$$
$$+ CH_3CHO + CO + NO_2 \quad (46)$$

$$C_9H_{18}(3-OH)CO_3NO_2 \longrightarrow C_9H_{18}(2-OH)CO_2$$
$$+ NO_2 \quad (47)$$

$$C_9H_{18}(2-OH)CO_2 + NO_2 \longrightarrow$$
$$C_9H_{18}(3-OH)CO_3NO_2 \quad (48)$$

$$OH + CHOC(CH_3)=CHCO_3NO_2 \longrightarrow CH_3COCHO$$
$$+ 2CO + NO_2 \quad (49)$$

$$CHOC(CH_3)=CHCO_3NO_2 \longrightarrow$$
$$CHOC(CH_3)=CHCO_3 + NO_2 \quad (50)$$

$$CHOC(CH_3)=CHCO_3 + NO_2 \longrightarrow$$
$$CHOC(CH_3)=CHCO_3NO_2 \quad (51)$$

$$OH + CHOCH=C(CH_3)CO_3NO_2 \longrightarrow CHOCHO$$
$$+ HCHO + CO + NO_2 \quad (52)$$

$$CHOCH=C(CH_3)CO_3NO_2 \longrightarrow$$
$$CHOCH=C(CH_3)CO_3 + NO_2 \quad (53)$$

$$CHOCH=C(CH_3)CO_3 + NO_2 \longrightarrow$$
$$CHOCH=C(CH_3)CO_3 \quad (54)$$

where the rate coefficients are shown in Table 4.

The 4 individual species react along the same paths at identical or similar rates, so PAN17$_{part}$ can be replaced within the reaction by:

$$C_8H_{16}(2-OH)CO_3NO_2 + C_9H_{18}(3-OH)CO_3NO_2$$
$$+ CHOC(CH_3)=CHCO_3NO_2$$
$$+ CHOCH=C(CH_3)CO_3NO_2. \quad (55)$$

This then gives the following set of equations,

$$OH + PAN17_{part} \longrightarrow (\sigma_1 + \sigma_2)C_6H_{13}CHO +$$
$$\sigma_3 CH_3COCHO$$
$$+ \sigma_4 CHOOHO + (\sigma_1 + \sigma_4)HCHO$$

$$+\sigma_2CH_3CHO + (1+\sigma_3)CO + NO_2 \tag{56}$$

$$\begin{aligned} PAN17_{part} \longrightarrow\ & NO_2 + \sigma_1C_8H_{16}(3-OH)CO_3 \\ & +\sigma_2C_9H_{18}(3-OH)CO_3 \\ & +\sigma_3CHOC(CH_3){=}CHCO_3 \\ & +\sigma_4CHOCH{=}C(CH_3)CO_3 \end{aligned} \tag{57}$$

$$C_8H_{16}(3-OH)CO_3 + NO_2 \longrightarrow PAN17_{part} \tag{58}$$
$$C_9H_{18}(3-OH)CO_3 + NO_2 \longrightarrow PAN17_{part} \tag{59}$$
$$CHOC(CH_3){=}CHCO_3 + NO_2 \longrightarrow PAN17_{part} \tag{60}$$
$$CHOCH{=}C(CH_3)CO_3 + NO_2 \longrightarrow PAN17_{part} \tag{61}$$
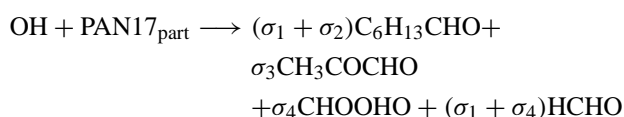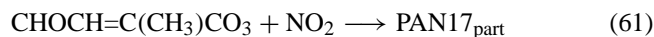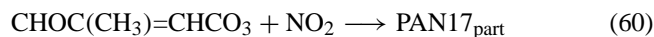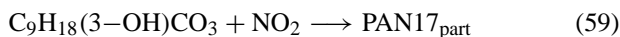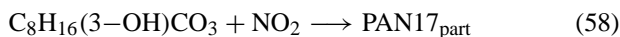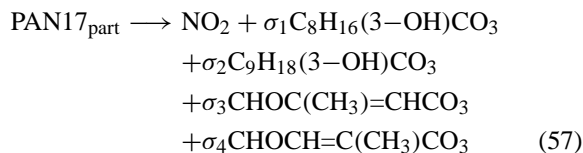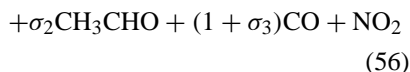
where

$$\sigma_i = \frac{\phi_i}{\sum_{j=1}^{4} \phi_j} \tag{62}$$

and

$$\begin{aligned} \phi_1 &= [C_8H_{16}(3-OH)CO_3], \\ \phi_2 &= [C_9H_{18}(3-OH)CO_3], \\ \phi_3 &= [CHOC(CH_3){=}CHCO_3], \\ \phi_4 &= [CHOCH{=}C(CH_3)CO_3]. \end{aligned} \tag{63}$$

The rate coefficient of Eq. (56) can be calculated as the mean of the rates of the four equations which have been lumped together i.e. $4.4{\times}10^{-11}$ molecule cm^{-3} s^{-1}. The lumping has therefore led to the removal of 3 species and 4 reactions.
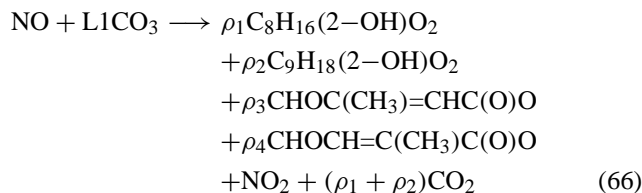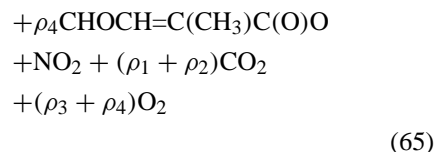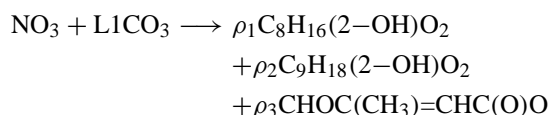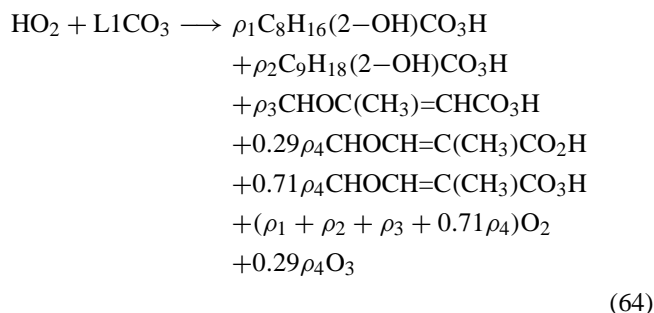
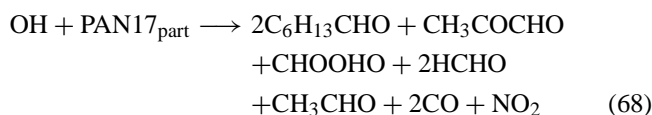On examination of Eq. (57) it can be seen that this equation would be greatly simplified if $C_8H_{16}(3-OH)CO_3$, $C_9H_{18}(3-OH)CO_3$, $CHOC(CH_3){=}CHCO_3$ and $CHOCH{=}C(CH_3)CO_3$ could be lumped to form a single variable. These four species all react with HO$_2$, NO$_3$, NO$_2$ and NO with the same set of rate coefficients. We can therefore define an exact lump L1CO$_3$ where $[L1CO_3]=[C_8H_{16}(3-OH)CO_3] + [C_9H_{18}(3-OH)CO_3] + [CHOC(CH_3){=}CHCO_3] + [CHOCH{=}C(CH_3)CO_3]$ giving the reactions,

$$\begin{aligned} HO_2 + L1CO_3 \longrightarrow\ & \rho_1C_8H_{16}(2-OH)CO_3H \\ & +\rho_2C_9H_{18}(2-OH)CO_3H \\ & +\rho_3CHOC(CH_3){=}CHCO_3H \\ & +0.29\rho_4CHOCH{=}C(CH_3)CO_2H \\ & +0.71\rho_4CHOCH{=}C(CH_3)CO_3H \\ & +(\rho_1 + \rho_2 + \rho_3 + 0.71\rho_4)O_2 \\ & +0.29\rho_4O_3 \end{aligned} \tag{64}$$

$$\begin{aligned} NO_3 + L1CO_3 \longrightarrow\ & \rho_1C_8H_{16}(2-OH)O_2 \\ & +\rho_2C_9H_{18}(2-OH)O_2 \\ & +\rho_3CHOC(CH_3){=}CHC(O)O \end{aligned}$$

$$\begin{aligned} & +\rho_4CHOCH{=}C(CH_3)C(O)O \\ & +NO_2 + (\rho_1 + \rho_2)CO_2 \\ & +(\rho_3 + \rho_4)O_2 \end{aligned} \tag{65}$$

$$\begin{aligned} NO + L1CO_3 \longrightarrow\ & \rho_1C_8H_{16}(2-OH)O_2 \\ & +\rho_2C_9H_{18}(2-OH)O_2 \\ & +\rho_3CHOC(CH_3){=}CHC(O)O \\ & +\rho_4CHOCH{=}C(CH_3)C(O)O \\ & +NO_2 + (\rho_1 + \rho_2)CO_2 \end{aligned} \tag{66}$$

The ρ values are calculated in the same manner as in Eq. (62). Equations (58) to (61) will now all have the form,

$$L1CO_3 + NO_2 \longrightarrow PAN17_{part} \tag{67}$$

Therefore as PAN17$_{part}$ is formed through 4 identical routes which have the same rate coefficient, the ratios σ_m formed in Eq. (62) will all be equal to 1. This leads to Eq. (56) having the form,

$$\begin{aligned} OH + PAN17_{part} \longrightarrow\ & 2C_6H_{13}CHO + CH_3COCHO \\ & +CHOOHO + 2HCHO \\ & +CH_3CHO + 2CO + NO_2 \end{aligned} \tag{68}$$

It can also be seen that Eq. (57) will now have the following form;

$$\begin{aligned} OH + PAN17_{part} \longrightarrow\ & (\sigma_1 + \sigma_2 + \sigma_3 + \sigma_4)L1CO_3 + NO_2 \\ & \longrightarrow L1CO_3 + NO_2 \end{aligned} \tag{69}$$
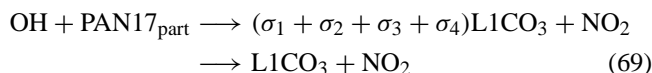
The application of the method described in the above examples results in the lumping of 802 species into 68 lumped groups leading to the removal of 734 species from the scheme. These lumps vary in size between 2 and 86 species. The details of the lumped groups can be seen in Tables 7 and 8.

4.3.3 Difficulties with further lumping

During the later stages of the lumping process problems begin to arise with the selection of further groups for lumping due to the definition of previous lumped groups. Although when looking at the updated time-scale data there are still many groups of species which have identical or similar time-scales, the interaction of these groups with previously lumped groups makes any further groups difficult to define. For example, L2CO$_3$H (see Table 7), is a lump of 42 XCO$_3$H species leading to the lumped reaction with OH having 42 different XCO$_3$ products of different variable fractional coefficients. Ideally these would be lumped together to eliminate the necessity for calculation of coefficients at each time-step. However CHOC(OH)=CHCO$_3$ which would be a member of this proposed new lump is also a product in the decomposition of PAN7. This reaction also provides an ideal group for lumping. Unfortunately some of those species present in this

Table 5. Number of time-scales in ranges after Stage 6.

range of life-times $\frac{1}{|\lambda|}$ (s)	number remaining				
	full	Stage 1	Stage 3	Stage 5	Stage 6
fast $<1\times10^{-4}$	711	710	434	6	4
intermediate	2359	2175	1915	1863	1115
slow $>1\times10^{5}$	416	205	105	100	64
total	3487	3091	2454	1969	1235

Table 6. Reactions and species remaining at each stage.

stage of reduction	number of reactions remaining	% reactions remaining	number of species remaining	% species remaining
full	10763	100	3487	100
stage 1	9539	88.6	3091	88.6
stage 2	8410	78.1	3091	88.6
stage 3	6927	64.4	2454	70.3
stage 4	6919	64.2	2454	70.3
stage 5	6168	57.3	1969	56.4
stage 6	4391	40.8	1235	35.4

alternative lump are not present as products in the reaction of L2CO$_3$H and therefore the two lumps are not mutually compatible

In order to lump the intermediate time-scale species in an optimal way (i.e. to remove as many species and reactions as possible in total), rather than beginning with the largest possible lumps, it would be necessary to allocate the groups in a global fashion. Before any lumps were firmly allocated, several protocols must be run, and the results of each compiled. The species should be grouped together first according to the similarity of their time-scale to other species. Then these groups should be divided depending on how many reactions in which the species in question appears as a reactant. The third stage of group allocation would involve looking at which species could be grouped together in the next round of groupings so that as many species as possible could be lumped. This type of global lumping strategy would be difficult to achieve using the Fortran code developed in this work, and may necessitate the use of an object oriented programming language such as C++. Having shown the success of the methodology in this work the use of such a global application of the method would be a productive area for future work.

5 Results and comparison of errors with full mechanism

Using lumping techniques a further 734 species and 1777 reactions have been removed from the mechanism. At this final stage there are 4391 reactions and 1235 species remaining. The simulation of a three day trajectory is now approximately 8 times faster than the full mechanism, with very little loss in accuracy for the calculation of the primary species.

Plots of O$_3$, NO, NO$_2$, total PAN, HONO, HNO$_3$, HCHO and OH can be seen in Figs. 3 to 11 for the full scheme and stage 6 reduced scheme for selected trajectories described in Whitehouse et al. (2004). The figures also show residual errors for the lumped stage 6 scheme as compared with reduced mechanisms from previous stages. The level of accuracy in trajectory 7 as shown in Figs. 3 to 5 is very good as would be expected given that the pre-lumping analysis was carried out on data from this trajectory. Trajectory 25 also shows little deterioration in accuracy (see Figs. 6 to 8) despite represent-

ing different conditions from those used during the lumping analysis. Trajectory 63 at Stage 6 has a lower level of accuracy than at previous stages as shown in Figs. 9 to 11 and compared to the other trajectories. It represents much lower levels of NO$_x$ and therefore the time-scale analysis applied to trajectory 7 may not have been completely relevant to trajectory 63. To increase accuracy at this stage, analysis should be carried out over a range of trajectories and time-points and lumps selected which meet life-time criteria for all of the points examined. However, the errors are still relatively small for most of the species examined.

Table 5 shows the number of species remaining in each time-scale category i.e. fast, intermediate and slow, following lumping when compared to the full and previously reduced schemes.

As can be seen there is a reduction in the number of intermediate time-scales of 748, with smaller reductions in the number of fast and slow species as a result of carrying out lumping. Table 6 shows the numbers of species and reactions remaining at each stage of the reduction, including the relative size of the reduced mechanisms compared to the full version.

The final level of reduction contains only 35% of the total number of species found in the full mechanism, and this is reflected in the time taken to solve the system. The accuracy of the final reduced mechanism with respect to ozone over all time points and all 94 trajectories can be seen in Fig. 12.

It can be seen that the ozone values for the reduced mechanism can vary from the full version by up to $\pm20\%$. However, many of the large percentage errors are present in very small total concentrations and therefore relate to small absolute errors. This often represents night-time conditions. In general the reduced model has a higher tendency to under predict ozone than to over predict. A further problem encountered when carrying out mechanism reduction on a scheme of this scale and complexity is that it is time consuming to carry out reduction analysis for all trajectories at all points as discussed in Sect. 5. The use of a limited number of trajectories can lead to some conditions being less accurately represented than others.

Fig. 3. Comparison of full and Stage 6 schemes for trajectory 7 along with errors for each stage for O_3, NO and NO_2.

Fig. 4. Comparison of full and Stage 6 schemes for trajectory 7 along with errors for each stage for OH, HCHO and HNO$_3$.

Fig. 5. Comparison of full and Stage 6 schemes for trajectory 7 along with errors for each stage for PAN and HONO.

Fig. 6. Comparison of full and Stage 6 schemes for trajectory 25 along with errors for each stage for O_3, NO and NO_2.

Fig. 7. Comparison of full and Stage 6 schemes for trajectory 25 along with errors for each stage for OH, HCHO and HNO$_3$.

Fig. 8. Comparison of full and Stage 6 schemes for trajectory 25 along with errors for each stage for PAN and HONO.

Fig. 9. Comparison of full and Stage 6 schemes for trajectory 63 along with errors for each stage for O_3, NO and NO_2.
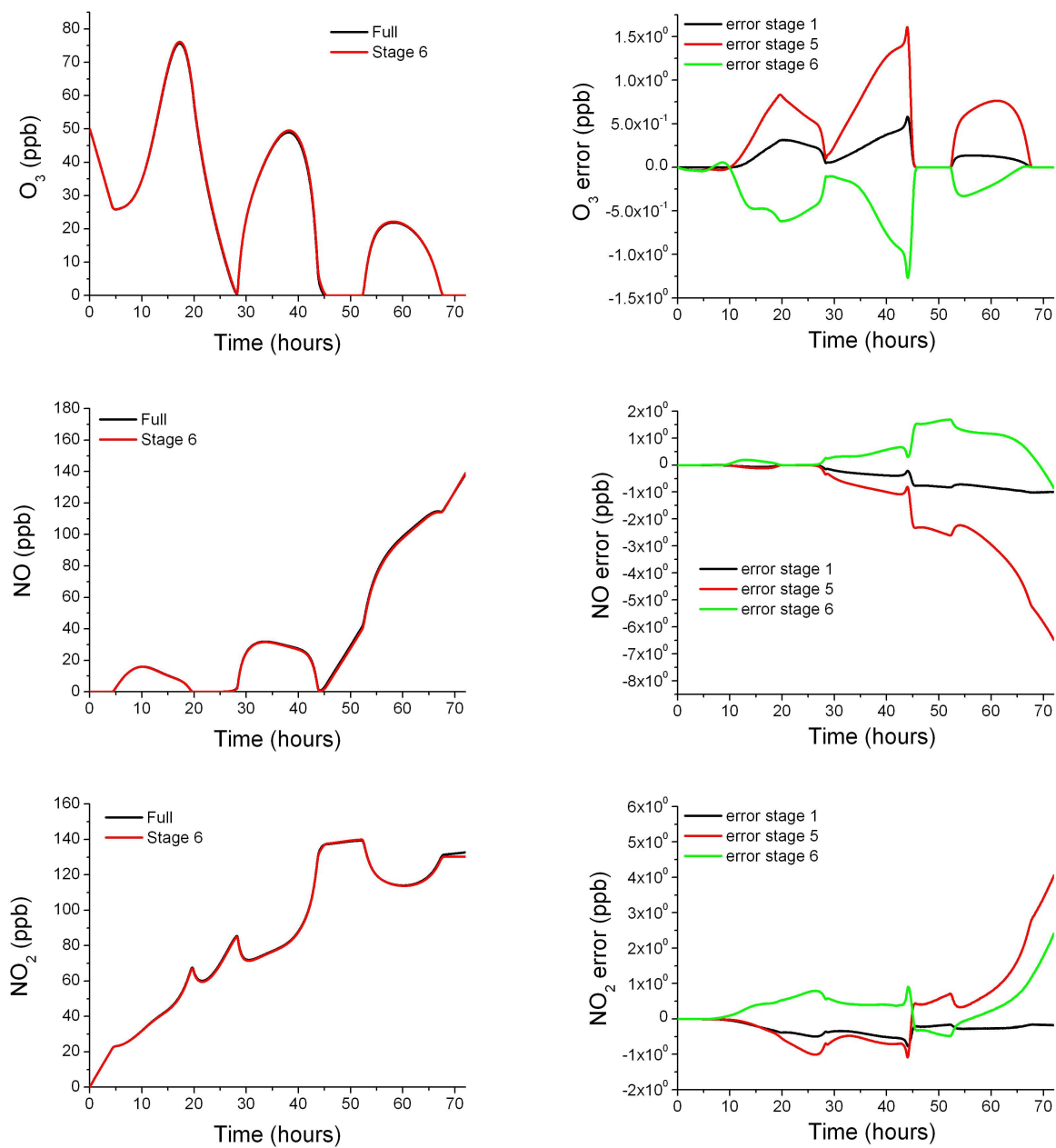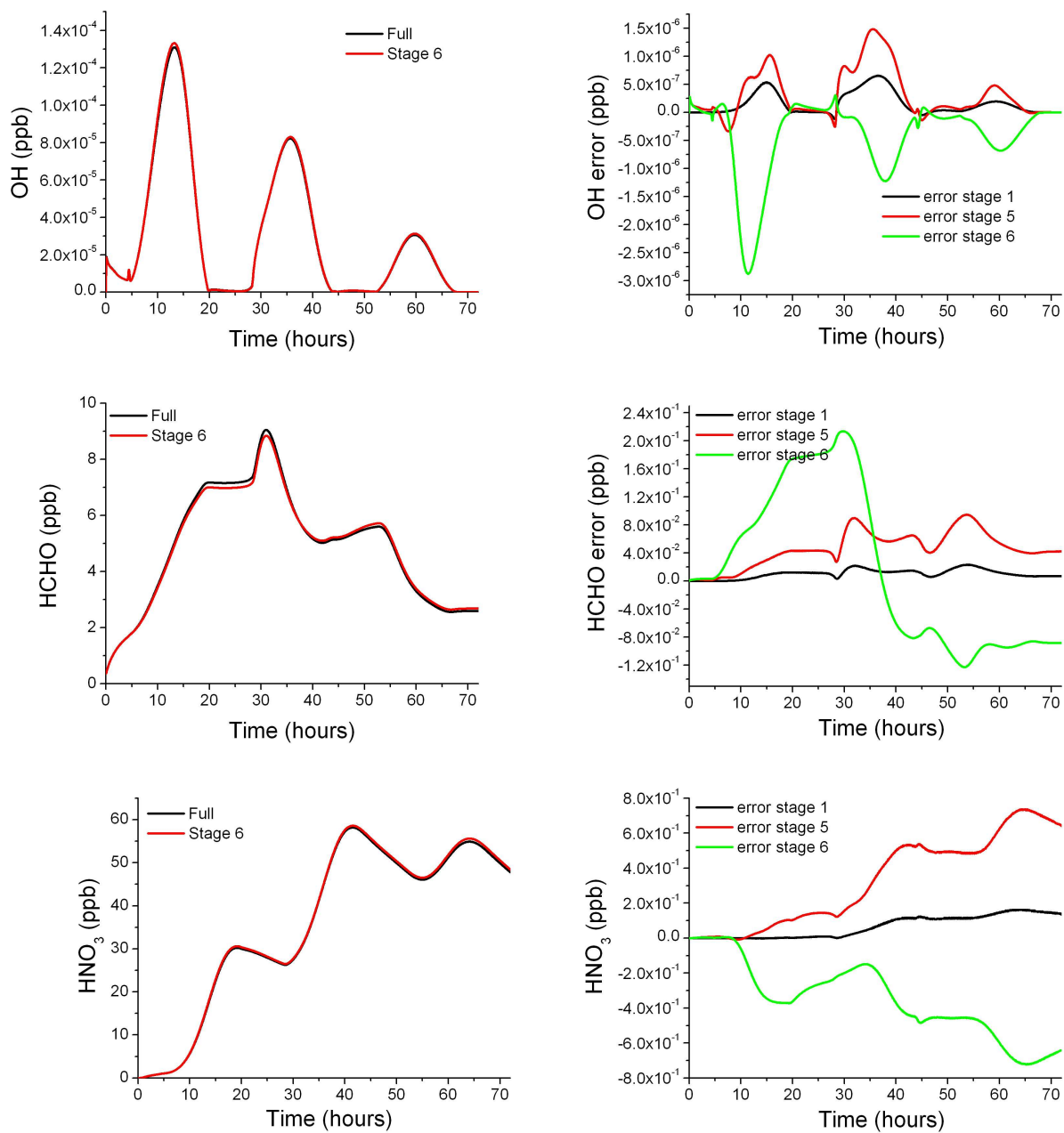
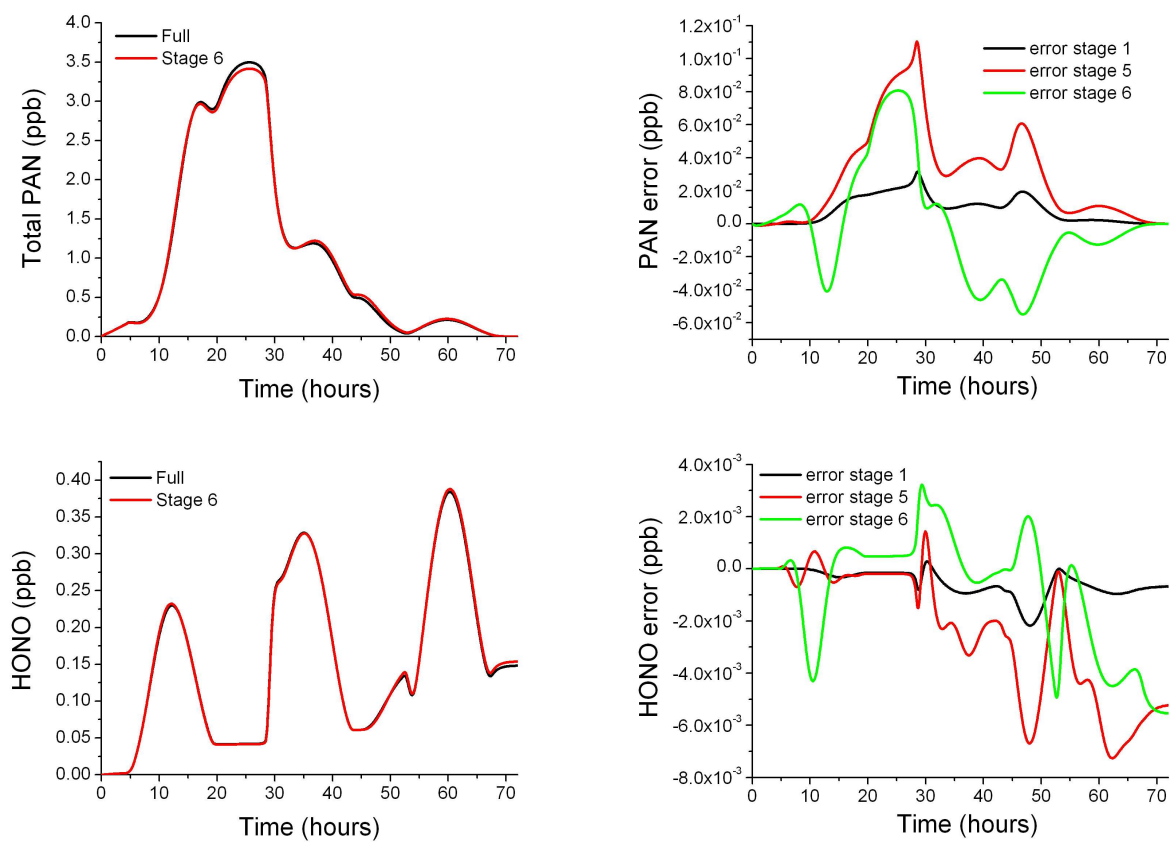**Fig. 10.** Comparison of full and Stage 6 schemes for trajectory 63 along with errors for each stage for OH, HCHO and HNO$_3$.

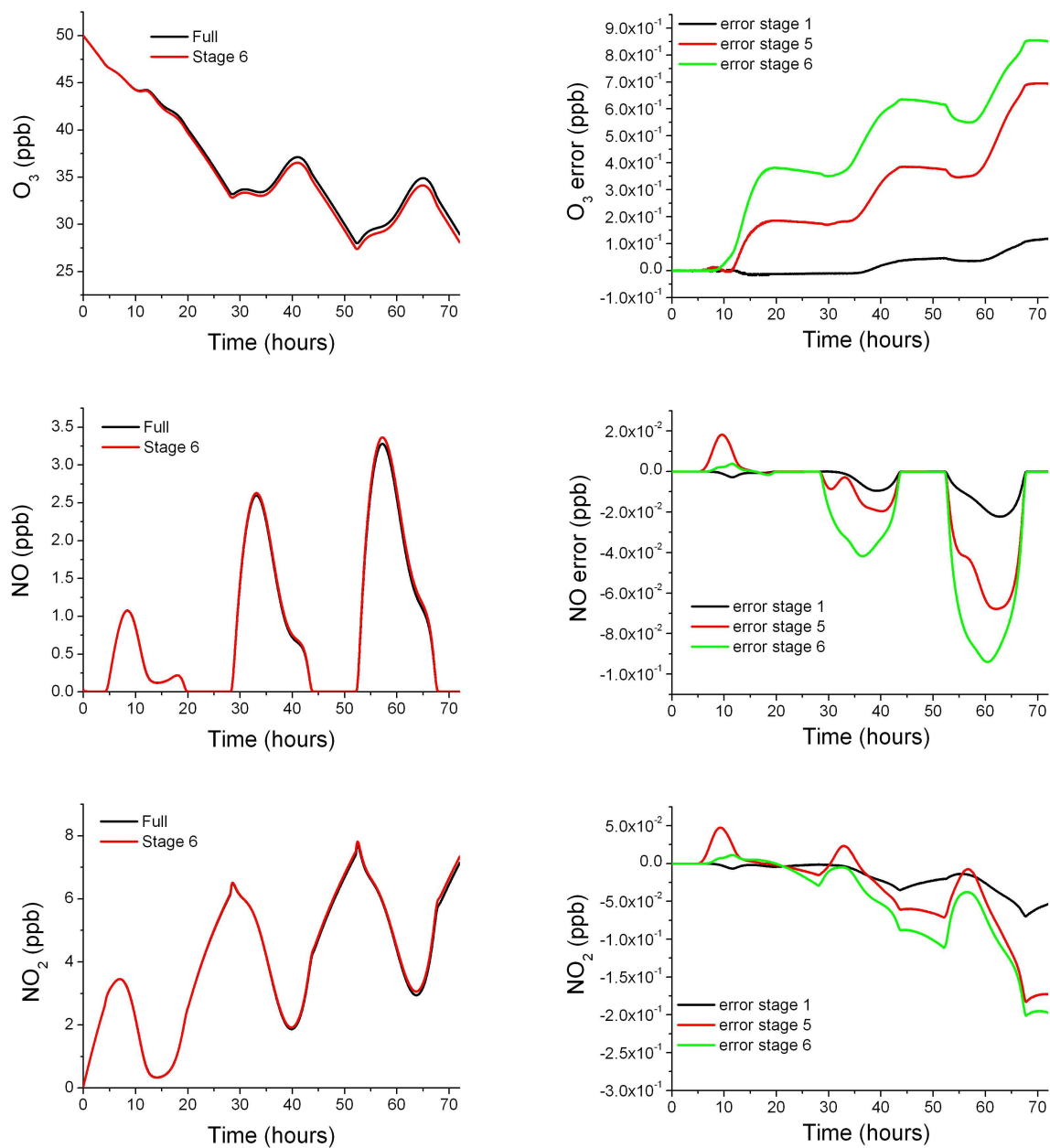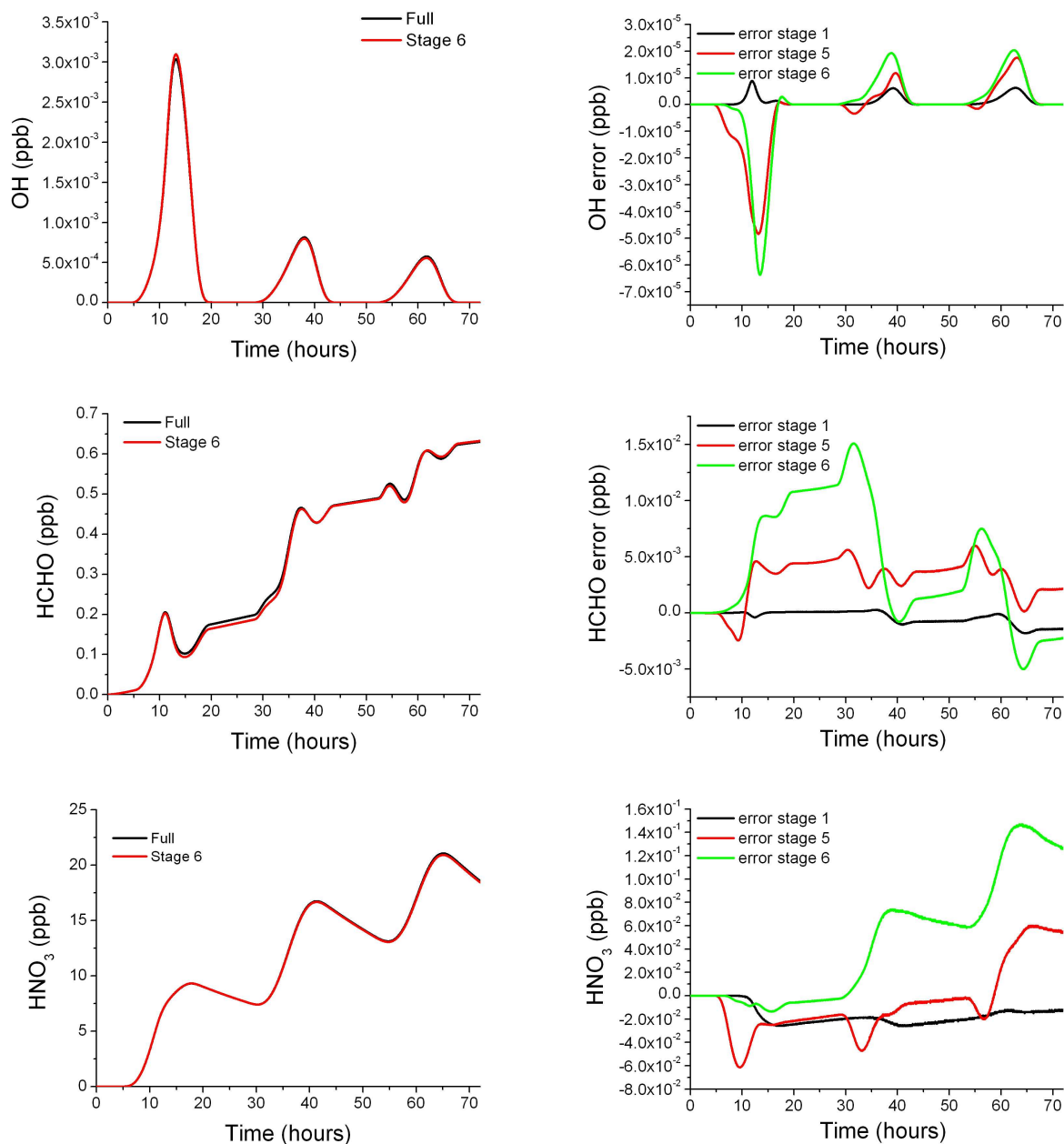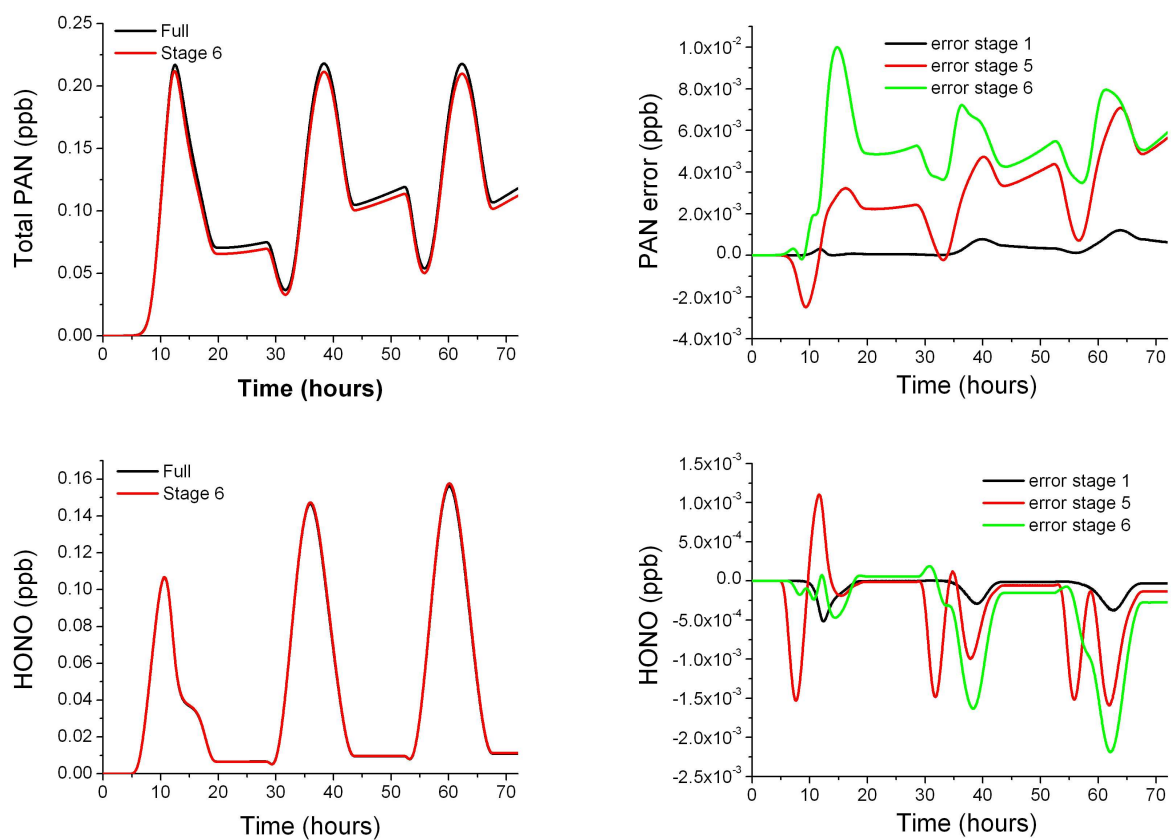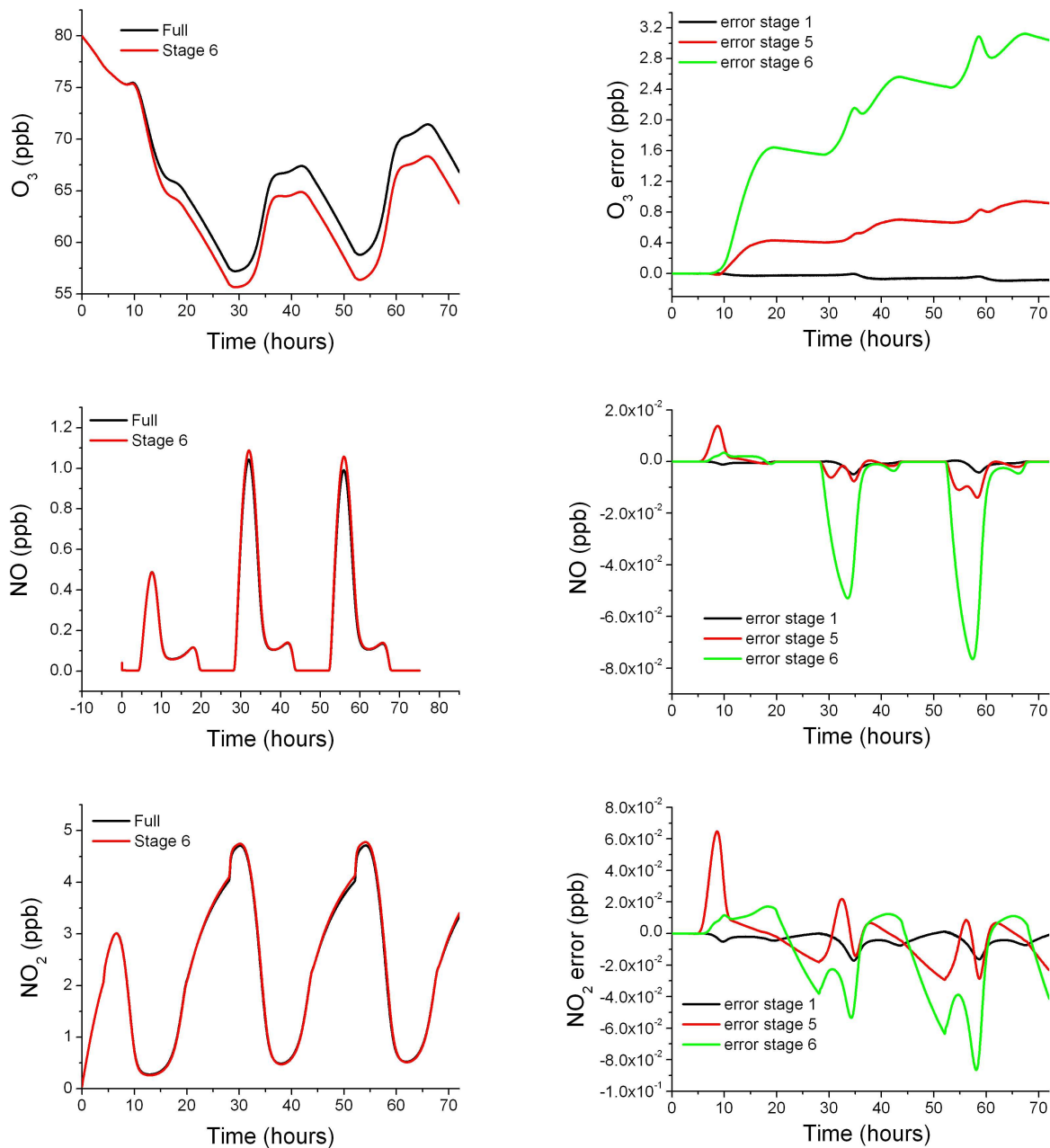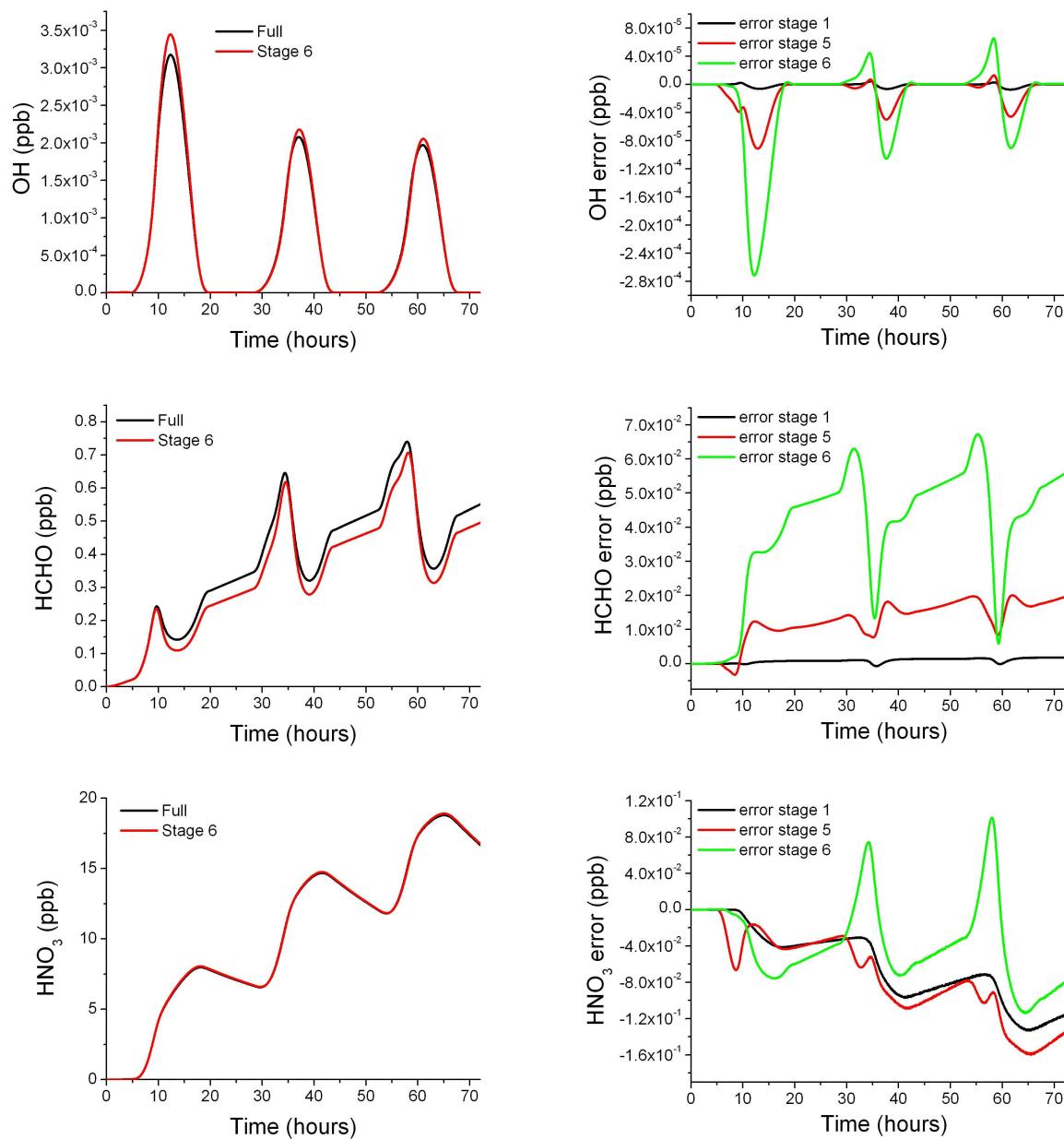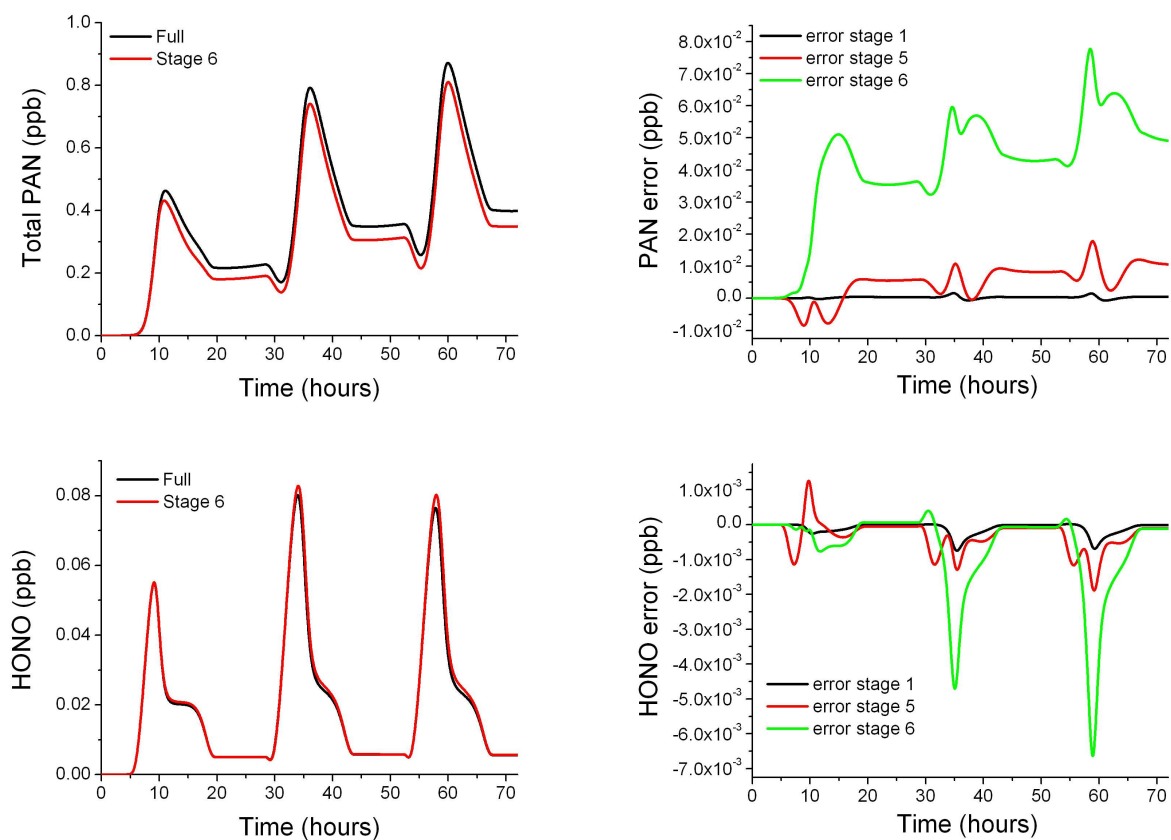**Fig. 11.** Comparison of full and Stage 6 schemes for trajectory 63 along with errors for each stage for PAN and HONO.
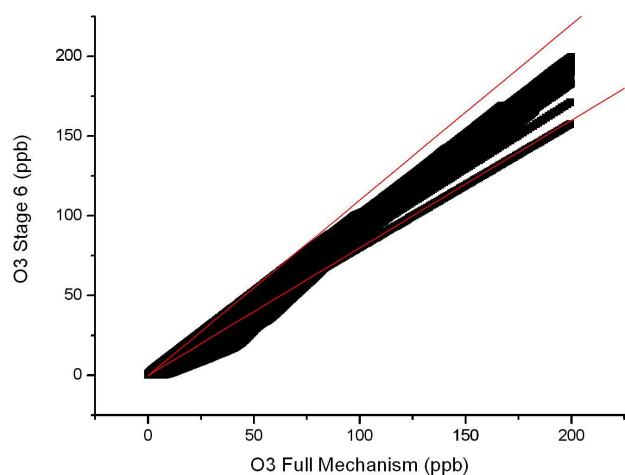
Fig. 12. Figure demonstrating the relationship between the ozone in the full mechanism and the Stage 6 lumped mechanism. The upper red lines indicates a 10% variation above the full mechanism values, and the lower red line indicates a 20% variation below.
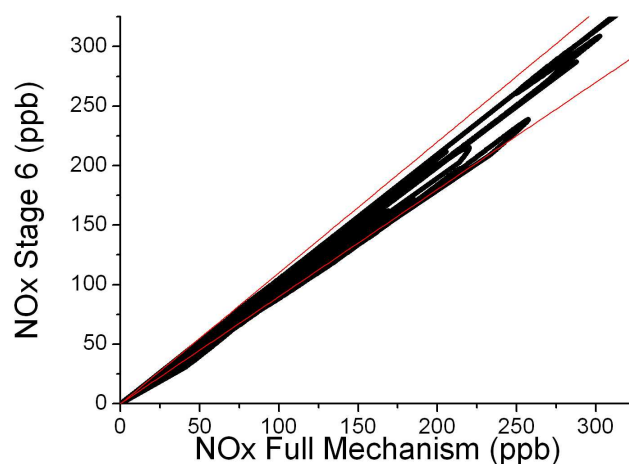
Fig. 14. Figure demonstrating the relationship between the total NOx in the full mechanism and the Stage 6 lumped mechanism. The red lines indicate a variation of 10% on either side of the full mechanism values.
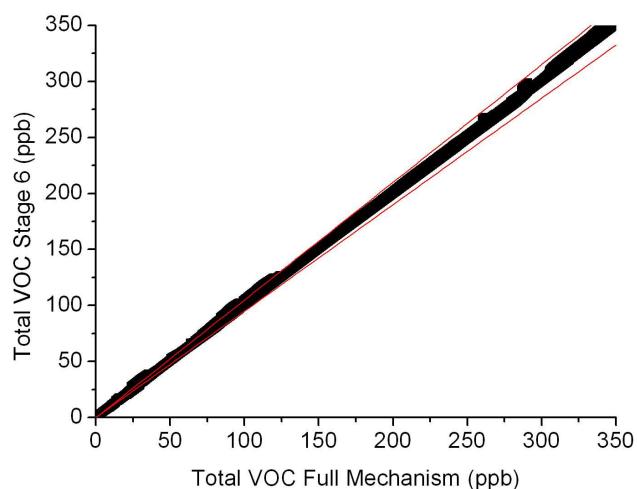
Fig. 13. Figure demonstrating the relationship between the total VOC in the full mechanism and the Stage 6 lumped mechanism. The red lines indicate a variation of 5% on either side of the full mechanism values.

Fig. 15. Figure demonstrating the relationship between OH predicted by the full mechanism and the Stage 6 lumped mechanism across all trajectories. The top red line indicates a variation of +12% and the bottom line a variation of −5%.

A plot of total VOC concentration over the three day period, as predicted by the full mechanism against those from the Stage 6 reduction for all trajectories can be seen in Fig. 13.

The total VOC values for the full mechanism and the Stage 6 reduced mechanism are within 5% of each other. The fact that none of the primary VOCs are removed from the mechanism, and are also not lumped, contributes to this high level of accuracy

The relationship between total NOx from the full mechanism and the Stage 6 reduction is shown in Fig. 14. These values can be seen to lie within 10% of each other at all times

over all trajectories. As before the highest level of absolute error is at the high NOx conditions.

The relationship between OH predicted by the full mechanism and the Stage 6 reduction is shown in Fig. 15. The OH is also well represented by the full scheme. OH and HO_2 are strongly coupled at relatively high NOx concentrations and so [OH] depends on the total rate of initiation which includes not only O^1D+H_2O, but also O_3+alkenes and a wide range of carbonyl photolyses. In addition, OH loss includes not only reaction with primary hydrocarbons, but also with a wide range of secondary carbonyl compounds. The reduced mechanism clearly captures these processes well.

6   Discussion

Lumping techniques have been widely used as a method of reducing the size of large chemical systems. Many previous techniques have concentrated on lumping primary VOCs and have used a variety of non-formal techniques for establishing lumped rate coefficients. In order to lump the MCM, a more formal approach was required in order to allow a higher level of automation than previous techniques have permitted. This work has presented such a methodology based on the lumping of species with intermediate lifetimes and the allocation of product coefficients with respect to the ratio of the forming species. Rate coefficients for the lumped reactions were taken as the mean of the rate coefficients of contributing elementary reactions in the case of inexact lumping. However, for groups of species with identical lifetimes lumping was exact. Using these techniques many lumps were identified, and, combined with other techniques such as sensitivity analysis, lumping allowed the size of the final reduced mechanism to be about 50% of the size of the full mechanism in terms of reactions and 35% in terms of species.

The lumping has been confined mainly to a few types of species. The first large group are peracids, and the second oxoacids. The third large group of species are nitrates, appearing in L21NO$_3$ to L43NO$_3$. Other groups include peroxy acyl nitrates (PANs), oxepins, carbonates and substituted phenols. Each group of species was defined according to similarities in lifetimes and reaction structures. Lumped reaction rates were formed in general by using average reaction rates for the species reacting with OH.

The computational time necessary for a single trajectory has been reduced by a factor of 8. However, in order for this lumping to be carried out in an optimum fashion an "intelligent" program would be needed in order to divide groups selected using the flowchart system into sub-groups, such that a larger number of species could be lumped in total. In this way larger continuous groups of reactions could be lumped before redivision had to occur. Although the lumping carried out here is not complete, with some development the lumping strategy developed has the potential to significantly further reduce the dimension and computational time of the MCM.

There are a number of advantages to executing lumping in this manner rather than those discussed in the introduction. Within a complex mechanism such as the MCM, identifying species with similar lifetimes throughout the simulation is a useful starting point in terms of finding groups of species that take part in the same reaction structures at similar kinetic rates. Identifying species that play a similar role within the mechanism can therefore be determined from a formal basis. Groups of species with similar lifetimes can be automatically found and then probed for similarities in terms of their reaction structures. The method is particularly successful for a mechanism like the MCM where large numbers of intermediate compounds react via many parallel but similar

reaction paths. In addition, the techniques developed here do not involve the lumping of primary compounds, making it possible to easily alter the emission profiles which the reduce mechanism represents. If primary VOCs are lumped as in many previous techniques, the whole mechanism must be reconstructed where emissions profiles change.

In a number of the lumping techniques discussed here, the rate parameters of the lumped scheme are optimised in order to fit certain experimental data or full model simulations. This removes any possibility of automating the procedures used. Also if the lumped model is to be accurate under conditions other than those for which the rate parameters are optimised it is possible that re-optimisation of the parameters should take place. In the present work, using lumped rate coefficients based on rate parameters from the full scheme means that they can be calculated simply in an automatic procedure and no further optimisation is necessary. The coefficients for the products in a lumped reaction are taken from ratios of the rates of the reactions by which the lumped species is now produced. The new data files needed in order to store this data can be generated by means of a simple Fortran routine, and with minimal adaptations to the driver code these coefficients can be calculated at each step.

A further advantage is that because both forward and reverse transformations have been defined here, it is possible to track the concentrations of the original compounds if necessary. One possible limitation of the methods is the form of the expression used for calculating the coefficients for the products in the lumped reactions where,

$$\sigma_m = \frac{\text{rate of formation of } S_j}{\text{sum of rate of formation of all species in lumped group}}, \quad (70)$$

and σ_m is the coefficient for the mth product formed from the species S_j which has now been lumped, see Sect. 4.2.2, Eq. (20). For this assumption to be valid, it was assumed that the lumped species had the same initial conditions. This was not a concern when applying it to the MCM as for the scenarios considered all intermediate species had zero initial conditions. However when considering the wider use of this method a further criterion when debating the suitability of a species for addition to a lump may be its initial value.

7   Conclusions

A formal method of species lumping which can be applied automatically to intermediate compounds within a detailed and complex tropospheric scheme has been developed. The method is based on grouping species with reference to their lifetimes. Preliminary application to the MCMv2.0 has led to the removal of 734 species and 1777 reactions from the scheme with minimal degradation of accuracy across a wide range of test trajectories. In combination with other reduction techniques such as sensitivity analysis and the application of the QSSA, a final reduced mechanism has been developed that contains 35.4% of the number of species and

40.8% the number of reactions compared to the full mechanism. This has led to a speed up of a factor of 8 in terms of computer simulation time. Analysis of the errors obtained by using the most reduced mechanism were determined to be less than 5% for many of the trajectories studied, and less than 10% for the majority.

Improvements in the lumping strategy in order to increase optimality would require that the lumps be chosen in such a way that a lumped species produces a single lumped product which is applicable for as many situations as possible. This would eliminate the need for product coefficients to be calculated from the rate coefficients at each time-point and would lead to much greater simplicity within the mechanism. In order to achieve this it would be necessary to allocate the groups in a global fashion. Before any lumps are firmly allocated, several protocols should be run and the results of all compiled. The species should be grouped according to the similarity of their lifetimes compared to the other species. It would then be necessary to establish if a secondary lump could be formed from the products of the lumped reactions of this group. Adjustments should then be made to the groups to allow this to occur in an optimum way. Future work will include the development of an automatic approach to achieving such an optimally lumped scheme.

Appendix 1

Table 7. Description of Lumped Groups for Acids, Nitrates, Oxepins and Substituted Phenols.

Lump Name	Type of Compounds	No. of Species in Lump	Range of Lifetimes (s) After 36 Hours of Simulation	Equivalent Reaction Rates (defined in Pilling et al. (1999))	Range of Reaction Rates Species + OH (molecule cm^{-3})$^{1-m}$s^{-1}
L1CO3H	peracids	59	1.08×10^4–8.10×10^4	J41	4.52×10^{-11}–3.7×10^{-12}
L2CO3H	peracids	42	9.35×10^2–1.31×10^4	J24, J41	5.44×10^{-10}–3.03×10^{-11}
L3CO3H	peracids	3	1.24×10^4–1.96×10^4	J15, J41	3.00×10^{-11}–1.47×10^{-11}
L4CO3H	peracids	7	1.48×10^4–5.29×10^4	J22, J41	3.05×10^{-11}–5.35×10^{-12}
L5CO3H	peracids	4	6.48×10^3–1.08×10^4	J18, J19, J41	6.90×10^{-11}–3.69×10^{-11}
L6OOH	oxoacids	3	7.57×10^4–8.25×10^4	J41	$1.90\times10^{-12}e^{(190/TEMP)}$–$2.20\times10^{-12}e^{(190/TEMP)}$
L7OOH	oxoacids	86	2.50×10^3–8.24×10^4	J41	3.59×10^{-12}–2.42×10^{-10}
L8OOH	oxoacids	52	3.31×10^3–5.90×10^4	J22, J41	4.33×10^{-12}–$1.42\times10^{1-0}$
L9OOH	oxoacids	79	8.36×10^3–5.33×10^4	J15, J41, $1.90\times10^{-12}e^{(190/TEMP)}$	3.44×10^2–5.56×10^{-11}
L10OOH	oxoacids	20	2.14×10^3–3.85×10^4	J22×2, J41	7.25×10^{-12}–9.61×10^{-11}
L11OOH	oxoacids	4	2.99×10^3–7.00×10^3	J15, J41	6.21×10^{-11}–1.61×10^{-10}
L12OOH	oxoacids	2	8.37×10^3–1.03×10^4	J24, J41	4.09×10^{-11}–5.26×10^{-11}
L13OOH	oxoacids	2	5.85×10^3–1.61×10^4	J24×2, J41	7.03×10^{-11}–7.26×10^{-11}
L14OOH	oxoacids	26	4.50×10^3–4.03×10^4	J22, J41, $1.90\times10^{-12}e^{(190/TEMP)}$	4.83×10^{-12}–1.07×10^{-10}
L15OOH	oxoacids	6	3.13×10^3–3.28×10^3	J22, J35, J41	1.99×10^{-11}–2.75×10^{-11}
L16OOH	oxoacids	9	6.95×10^3–8.19×10^3	J18×2, J22, J41	5.06×10^{-11}–6.19×10^{-11}
L17OOH	oxoacids	17	5.88×10^3–7.51×10^3	J15, J24, J41	5.06×10^{-11}–7.44×10^{-11}
L18OOH	oxoacids	3	3.61×10^3–4.45×10^3	J41, J56, J57	1.10×10^{-10}–1.37×10^{-10}
L19OOH	oxoacids	2	8.20×10^3–8.24×10^3	J15, J22×2, J41	4.75×10^{-11}–4.78×10^{-11}
L20OOH	oxoacids	3	4.90×10^3–5.11×10^3	J18, J19, J22, J41, $1.90\times10^{-12}e^{(190/TEMP)}$	8.51×10^{-11}–8.94×10^{-11}
L21NO3	nitrates	16	1.13×10^5–5.90×10^5		4.57×10^{-13}–8.76×10^{-13}
L22NO3	nitrates	9	9.55×10^4–1.35×10^5	J55	8.64×10^{-14}–1.68×10^{-12}
L23NO3	nitrates	44	3.98×10^4–5.14×10^5	J53	2.26×10^{-13}–1.22×10^{-11}
L25NO3	nitrates	15	1.39×10^5–2.86×10^5	J54	4.9×10^{-13}–2.7×10^{-12}
L26NO3	nitrates	15	7.72×10^4–1.19×10^5	J54	3.02×10^{-12}–6.03×10^{-12}
L27NO3	nitrates	2	6.93×10^4–7.04×10^4	J54	6.03×10^{-12}–6.15×10^{-12}
L28NO3	nitrates	5	5.71×10^4–6.81×10^4	J54	6.28×10^{-12}–7.74×10^{-12}
L29NO3	nitrates	3	4.99×10^4–5.46×10^4	J54	8.15×10^{-12}–9.04×10^{-12}
L30NO3	nitrates	8	3.42×10^4–4.41×10^4	J54	1.04×10^{-11}–1.36×10^{-11}
L31NO3	nitrates	7	2.55×10^4–3.27×10^4	J54	1.38×10^{-11}–1.97×10^{-11}
L32NO3	nitrates	11	6.61×10^4–1.56×10^5	J22	1.50×10^{-12}–6.08×10^{-12}
L33NO3	nitrates	2	5.94×10^3–8.35×10^3	J53×2	6.04×10^{-11}–8.55×10^{-11}
L34NO3	nitrates	3	2.31×10^4–1.29×10^5	J56, J57	5.23×10^{-13}–1.89×10^{-11}
L35NO3	nitrates	7	9.37×10^4–1.48×10^5	J22, J53	9.64×10^{-13}–2.99×10^{-12}
L36NO3	nitrates	2	6.25×10^3–1.48×10^4	J15, J55	2.21×10^{-12}–7.00×10^{-11}
L37NO3	nitrates	11	7.59×10^3–9.10×10^3	J18×2, J22	4.70×10^{-11}–5.83×10^{-11}
L38NO3	nitrates	8	7.00×10^3–8.26×10^3	J15, J24	4.70×10^{-11}–5.83×10^{-11}
L39NO3	nitrates	2	3.13×10^3–3.65×10^3	J35, J55	5.01×10^{-12}–2.85×10^{-11}
L40NO3	nitrates	3	3.75×10^4–5.17×10^4	J41, $1.90\times10^{-12}e^{(190/TEMP)}$	3.74×10^{-12}–7.52×10^{-12}
L41NO3	nitrates	2	1.54×10^4–1.57×10^4	J15, J54	2.25×10^{-11}–2.33×10^{-11}
L42NO3	nitrates	7	9.05×10^4–1.07×10^5	J22, J54	1.79×10^{-12}–2.65×10^{-12}
L43PIN	oxepins	15	1.32×10^3	J61, 1.00×10^{11}, 1.00×10^{-10}	
L44PIN	oxepins	12	4.39×10^3	1.00×10^{11}, 1.00×10^{-10}	
L46PIN	oxepins	4	4.45×10^3	9.20×10^{12}, 1.00×10^{-10}	
L47S	substituted phenols	8	6.31×10^3	8.20×10^{11}	

Table 8. Description of Lumped Groups for Peroxy Acyl Nitrates and Carbonates.

Lump Name	Type of Compounds	No. of Species in Lump	Range of Lifetimes (s) After 36 Hours of Simulation	Equivalent Reaction Rates (defined in Pilling et al. (1999))	Range of Reaction Rates Species + OH (molecule $cm^{-3})^{1-m} s^{-1}$
PAN1	peroxy acyl nitrates	5	3.29×10^3	KBPAN$\times 0.67$, 1.06×10^{12}	
PAN2	peroxy acyl nitrates	3	$2217.25 - 2217.27$	KBPAN	$1.15 \times 10^{-13} - 5.15 \times 10^{-13}$
PAN3	peroxy acyl nitrates	2	2217.25	KBPAN, 9.55×10^{13}	
PAN4	peroxy acyl nitrates	4	2217.24	KBPAN	$1.15 \times 10^{12} - 1.38 \times 10^{12}$
PAN5	peroxy acyl nitrates	5	$2217.18 - 2217.22$	KBPAN	$1.75 \times 10^{12} - 3.17 \times 10^{12}$
PAN6	peroxy acyl nitrates	5	$2217.12 - 2217.16$	KBPAN	$3.96 \times 10^{12} - 5.18 \times 10^{12}$
PAN7	peroxy acyl nitrates	5	$2217.10 - 2217.12$	KBPAN	$5.40 \times 10^{12} - 6.02 \times 10^{12}$
PAN8	peroxy acyl nitrates	5	$2217.02 - 2217.08$	KBPAN	$6.60 \times 10^{12} - 8.63 \times 10^{12}$
PAN9	peroxy acyl nitrates	5	$2216.99 - 2217.01$	KBPAN	$9.18 \times 10^{12} - 9.73 \times 10^{12}$
PAN10	peroxy acyl nitrates	6	$2216.92 - 2216.98$	KBPAN	$1.00 \times 10^{11} - 1.22 \times 10^{11}$
PAN11	peroxy acyl nitrates	6	$2216.87 - 2216.90$	KBPAN	$1.27 \times 10^{11} - 1.38 \times 10^{11}$
PAN12	peroxy acyl nitrates	10	$2216.76 - 2216.87$	KBPAN	$1.40 \times 10^{11} - 1.74 \times 10^{11}$
PAN13	peroxy acyl nitrates	10	$2216.43 - 2216.73$	KBPAN	$1.87 \times 10^{11} - 2.89 \times 10^{11}$
PAN14	peroxy acyl nitrates	6	$2216.29 - 2216.41$	KBPAN	$2.96 \times 10^{11} - 3.38 \times 10^{11}$
PAN15	peroxy acyl nitrates	4	$2216.23 - 2216.29$	KBPAN	$3.38 \times 10^{11} - 3.58 \times 10^{11}$
PAN16	peroxy acyl nitrates	10	$2216.06 - 2216.23$	KBPAN	$3.58 \times 10^{11} - 4.16 \times 10^{11}$
PAN17	peroxy acyl nitrates	10	$2215.96 - 2216.02$	KBPAN	$4.30 \times 10^{11} - 4.48 \times 10^{11}$
PAN18	peroxy acyl nitrates	10	$2215.73 - 2215.96$	KBPAN	$4.48 \times 10^{11} - 5.28 \times 10^{11}$
PAN19	peroxy acyl nitrates	11	$2215.36 - 2215.68$	KBPAN	$5.43 \times 10^{11} - 6.53 \times 10^{11}$
PAN20	peroxy acyl nitrates	7	$2215.22 - 2215.31$	KBPAN	$6.70 \times 10^{11} - 7.03 \times 10^{11}$
PAN21	peroxy acyl nitrates	9	$2202.48 - 2215.17$	KBPAN	$7.19 \times 10^{11} - 5.18 \times 10^{10}$
L1CO3	carbonates	4	3.38×10^4	KAPNO, KFPAN, KAPHO2, 1.6KRO2NO3	
L2CO3	carbonates	10	3.38×10^4	KAPNO, KFPAN, KAPHO2, 1.6KRO2NO3	

References

Campbell, S. L. and Meyer, C. D.: Generalized Inverses of Linear Transformations, Pitman, London, 1979.

Fish, D. J.: The automatic generation of reduced mechanisms for tropospheric chemistry modelling, Atm. Env., 34, 1563–1574, 2000.

Gery, M. W., Whitten, G. Z.and Killus, J. P., and Dodge, M. C.: A photochemical kinetics mechanism for urban and regional scale computer modelling, J. Geophys. Res., 94 D10, 12 925–12 956, 1989.

Hough, A. M.: An intercomparison of mechanisms for the production of photochemical oxidants, J. Geophys. Res., 93, 3789–3812, 1988.

Jenkin, M. E., Saunders, S. M., and Pilling, M. J.: The tropospheric degradation of volatile organic compounds :A protocol for mechanism development, Atm. Env., 31, 81–104, 1997.

Jenkin, M. E., Saunders, S. M., Derwent, R. G., and Pilling, M. J.: Development of a reduced speciated VOC degradation mechanism for use in ozone models, Atm. Env., 36, 4725–4734, 2002.

Leone, J. A. and Seinfeld, J. H.: Comparative analysis of chemical reaction mechanisms for photochemical smog, Atm. Env., 10, 437–464, 1985.

Li, G. and Rabitz, H.: A general analysis of exact lumping in chemical kinetics, Chem. Eng. Sci., 14, 1413–1430, 1989.

Li, G. and Rabitz, H.: A general analysis of approximate lumping in chemical kinetics, Chem. Eng. Sci., 45, 977–1002, 1990.

Makar, P. A. and Polavarapu, S. M.: Analytic solutions for gasphase chemical mechanism compression, Atm. Env., 31, 1025–1039, 1997.

Makar, P. A., Stockwell, W. R., and Li, S. M.: Gas-phase mechanism compression strategies: treatment of reactions, Atm. Env., 30, 831–842, 1996.

Middleton, P., Stockwell, W. R., and Carter, W. P. L.: Aggregation of volatile organic compound emissions for regional modelling, Atm. Env., 6, 1107–1133, 1990.

Pilling, M. J., Saunders, S., Carslaw, N., Pascoe, S., Jenkin, M., and Derwent, D.: Master Chemical Mechanism, http://www.chem.leeds.ac.uk/Atmospheric/MCM/mcmproj.html, 1999.

Stockwell, W. R., Middleton, P., Chang, J. S., and Tang, X. Y.: The second generation regional acid deposition model mecha-

nism for regional air-quality modelling, J. Geophys. Res., 95, 16 343–16 367, 1990.

Tomlin, A. S., Turányi, T., and Pilling, M. J.: Autoignition and low temperature combustion of hydrocarbons, chap. Mathematical tools for construction, investigation and reduction of combustion mechanisms, Elesvier, Amsterdam., 1998.

Wang, S. W., Georgopoulos, P. G., Li, G., and Rabitz, H.: Condensing complex atmospheric chemistry mechanisms – I. The direct constrained approximate lumping (DCAL) method applied to alkane photochemistry, Env. Sci. Tech., 32, 2018–2024, 1998.

Wei, J. and Kuo, J. C.: A lumping analysis in monomolecular reaction systems, Ind. Eng. Chem. Fundam., 8, 114–133, 1969.

Whitehouse, L. E., Tomlin, A. S., and Pilling, M. J.: Systematic reduction of complex tropospheric chemical mechanisms, Part I: sensitivity and time-scale analyses, Atmos. Chem. Phys., 4, 2025–2056, 2004,
SRef-ID: 1680-7324/acp/2004-4-2025.