Chempath 1.0: An open-source pathway analysis program for photochemical models

Daniel Garduno Ruiz[1], Colin Goldblatt[1], and Anne-Sofie Ahm[1]

[1]School of Earth and Ocean Sciences, University of Victoria, Victoria, British Columbia, Canada

Correspondence: Daniel Garduno Ruiz (danielgardunoruiz@uvic.ca)

Abstract. We describe the development of *Chempath*: an open-source pathway analysis program for photochemical models. This algorithm can help understand the results of complex photochemical models by identifying the most important reaction chains (pathways) for the production and destruction of a species of interest in a reaction system. The algorithm can also quantify the contribution of the pathways to the production and destruction of a species. *Chempath* is an open-source Python re-implementation of the algorithm developed by Lehmann (2004). However, *Chempath* does not include the balance of concentration changes and reaction rates that Lehmann's algorithm uses to eliminate imbalances due to numerical errors. Instead, *Chempath* quantifies the contribution of these imbalances to the production and destruction of a species.

We demonstrate how to apply *Chempath* both to a simple box model and to a one-dimensional photochemical model, using a reaction system for Earth's present-day atmosphere. *Chempath* can identify well-known chemical mechanisms for O_3 production and destruction in these models, suggesting that this algorithm can be applied to understand photochemical models of less well-known atmospheres, like past and exoplanet atmospheres.

1 Introduction

The construction of chemical pathways is essential to understand the results of photochemical models. These models typically represent hundreds of reactions producing and destroying chemical species within the atmosphere (see for example Hu et al. 2012; Tsai et al. 2017; Wogan et al. 2022). The interaction of these reactions makes it difficult to attribute the production or destruction of a species to a single reaction. Instead, to understand the mechanisms that affect the concentration of a species it is necessary to construct pathways. A pathway is a sequence of reactions that interact with each other to produce, destroy, or recycle a species. For example, stratospheric ozone destruction is explained in terms of pathways that catalyze O_3 destruction (Lary, 1997). One of these pathways involves the reaction of O_3 with chlorine species (Molina and Rowland, 1974):

$$\begin{aligned} O_3 + h\nu &\longrightarrow O_2 + O \\ Cl + O_3 &\longrightarrow ClO + O_2 \\ ClO + O &\longrightarrow Cl + O_2 \\ \text{Net: } 2\,O_3 &\longrightarrow 3\,O_2 \end{aligned} \qquad \text{(ClO pathway)}$$

This ozone-destructing pathway has three reactions, and its net effect is to convert two O_3 molecules into three O_2 molecules.

Several methods can help understand the results of photochemical models. For example, sensitivity analyses can constrain uncertainties in reaction rate constants and provide information on how variable the results of a model are when the reaction rates are perturbed (Turányi and Tomlin, 2014). Wiring diagrams can help understand the flow of molecules in a reaction system (Fishtik et al., 2006; Androulakis, 2006). However, these methods can not give quantitative information about the chains of reactions responsible for the production or destruction of a species in a chemical model. To understand these chemical mechanisms, it is necessary to construct pathways.

Chemical pathways are usually constructed manually and empirically, tracking and connecting reactions important for the production or destruction of a species of interest. However, the manual construction of pathways can not give a quantitative estimate of how important a pathway is for the production or destruction of a species relative to other pathways. Also, the manual construction of pathways has reproducibility limitations. Alternatively, pathways can be automatically constructed using algorithms (Milner, 1964; Schuster and Schuster, 1993; Clarke, 1988; Lehmann, 2002, 2004).

One of the most used algorithms to construct pathways is the "Pathway analysis program" created by Lehmann (2004). This algorithm can automatically construct all the significant pathways in a reaction system and calculate the contribution of each pathway to the production and destruction of a species of interest. The "Pathway analysis program" has been used in several studies to gain a better understanding of atmospheric chemistry models (Grenfell et al., 2006; Verronen et al., 2011; Stock et al., 2012a, b; Verronen and Lehmann, 2013; Stock et al., 2017; Gebauer et al., 2017).

In this paper, we describe the development of *Chempath*: an open-source Python re-implementation of Lehmann's (2004) algorithm for analysis of photochemical models. We aim to contribute this open-source pathway analysis program to enhance the applicability of this algorithm to photochemical models and to enhance the replicability of chemical pathway construction. Our implementation is based on the description of the algorithm in Lehmann (2004). However, there is one difference between Chempath and Lehmann's (2004) algorithm. Our implementation does not include the balance of concentration changes and reaction rates that Lehmann's algorithm uses to eliminate imbalances due to numerical errors. Instead, *Chempath* requires input information about these imbalances to quantify the contribution of numerical errors to the production and destruction of a species.

We demonstrate how to apply this algorithm to a simple box model and to the one-dimensional photochemical model *photochem* (Wogan et al. 2023, https://github.com/Nicholaswogan/photochem). *Photochem* is an updated version of *Photochempy* (Wogan, 2023), a model that has been used for exoplanet and early Earth photochemical studies (Wogan et al., 2022; Thompson et al., 2022; Garduno Ruiz et al., 2023, 2024). This model originates from *Atmos*, a photochemical model extensively used to investigate photochemistry in exoplanet and past atmospheres (see for example Kasting et al. 1979; Kasting and Donahue 1980; Segura et al. 2005; Zahnle et al. 2006; Claire et al. 2014; Arney et al. 2016).

The structure of the paper is as follows. In section 2, we review the Lehmann (2004) algorithm via a simple example. In section 3, we describe how we implemented *Chempath*. In section 4, we demonstrate how to apply *Chempath* to a simple box model and to the one-dimensional photochemical model *photochem*.

2 Algorithm review, a simple example

Here we provide a summary of the pathway analysis program presented in Lehmann (2004), using a simple example to explain each step of the algorithm (see the original paper for further details). The pathway analysis program forms pathways by the iterative connection of reactions through branching-point species (figure 1). A branching-point species is one that is used to connect reactions that produce the species with reactions that destroy it. For example, the reaction $Cl + O_3 \longrightarrow ClO + O_2$ can be connected with the reaction $ClO + O \longrightarrow Cl + O_2$ through the branching-point species ClO.

The example we use to demonstrate the algorithm consists of five reactions between six species involving hydrogen oxide radicals. The reactions are:

(R_1) $H_2O_2 + OH \longrightarrow HO_2 + H_2O$, with rate=0.1 ppb/h

(R_2) $OH + OH \longrightarrow H_2O_2$, with rate=0.6 ppb/h

(R_3) $H_2O_2 + O \longrightarrow OH + HO_2$, with rate=1 ppb/h

(R_4) $HO_2 + h\nu \longrightarrow OH + O$, with rate=0.5 ppb/h

(R_5) $HO_2 + O \longrightarrow OH + O_2$, with rate=1 ppb/h

and the species are:

(S_1) H_2O_2, with a mixing ratio of 2 ppm

(S_2) OH, with a mixing ratio of 3 ppm

(S_3) HO_2, with a mixing ratio of 2 ppm

(S_4) H_2O, with a mixing ratio of 10 ppm

(S_5) O, with a mixing ratio of 1.5 ppm

(S_6) O_2, with a mixing ratio of 100 ppm

We arbitrarily select these rates and mixing ratios for this example. Assuming this reaction system was run for 1 hour, the production and destruction by the reactions result in mixing ratio changes of -0.5, 1.2, -0.4, 0.1, -1.5, and 1 ppb for each species respectively. In the next sections, we will identify which combination of reactions is the most important to explain the OH mixing ratio change.

2.1 Assumptions and definitions

The algorithm uses the variables listed in table 1 to construct pathways. We use two square brackets to denote variables that represent matrices (for example $[[s_{ij}]]$), and one to denote vectors (for example $[c_i]$). We denote the elements of a vector or matrix with sub-indexes. For example, c_i is the i-th element of the vector $[c_i]$ and s_{ij} is the element of $[[s_{ij}]]$ located in the i-th row and j-th column. Similarly, $[s_{3j}]$ denotes the third row of the matrix $[[s_{ij}]]$. We also use the notation pos$([[x]])$ to refer to a function that makes the negative values of a vector or matrix zero. Similarly, neg$([[x]])$ is a function that makes the positive values of a vector or matrix zero.

In large reaction systems, it might not be possible to construct all the pathways of the system because the number of pathways could be computationally unmanageable. The algorithm includes the option to delete unimportant pathways to avoid the construction of an unmanageable number of pathways and reduce the computation time. However, the algorithm includes variables to keep track of the rates of these deleted pathways.

Variable	Initial value in simple example	Units in simple example	Description
$[[s_{ij}]]$	$\begin{array}{c} \begin{array}{ccccc} R_1 & R_2 & R_3 & R_4 & R_5 \end{array} \\ \begin{pmatrix} -1 & 1 & -1 & 0 & 0 \\ -1 & -2 & 1 & 1 & 1 \\ 1 & 0 & 1 & -1 & -1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & -1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{array}{c} S_1 \\ S_2 \\ S_3 \\ S_4 \\ S_5 \\ S_6 \end{array} \end{array}$	ppb	Stoichiometric matrix representing the number of molecules of species S_i produced ($s_{ij} > 0$) or destroyed ($s_{ij} < 0$) by reaction R_j. For example, in the simple example reaction R_2 destroys two molecules of species S_2, so $s_{22} = -2$.
dt	1	h	Time step.
$[dc_i]$	$\begin{array}{cccccc} S_1 & S_2 & S_3 & S_4 & S_5 & S_6 \\ \begin{bmatrix} -0.5 & 1.2 & -0.4 & 0.1 & -1.5 & 1 \end{bmatrix} \end{array}$	ppb	Vector representing the change in concentration of species S_i in the time step dt.

$[\bar{c_i}]$	$\begin{matrix} S_1 & S_2 & S_3 & S_4 & S_5 & S_6 \\ \begin{bmatrix} 1.75 & 3.6 & 1.8 & 10.05 & 0.75 & 100.5 \end{bmatrix} \end{matrix}$	ppb	Vector representing the mean concentration of species S_i in the time step dt.
$[\delta_i] = \frac{[dc_i]}{dt}$	$\begin{matrix} S_1 & S_2 & S_3 & S_4 & S_5 & S_6 \\ \begin{bmatrix} -0.5 & 1.2 & -0.4 & 0.1 & -1.5 & 1 \end{bmatrix} \end{matrix}$	ppb/h	Vector representing the rate of concentration change of species S_i.
$[f_k]$	$\begin{matrix} P_1 & P_2 & P_3 & P_4 & P_5 \\ \begin{bmatrix} 0.1 & 0.6 & 1 & 0.5 & 1 \end{bmatrix} \end{matrix}$	ppb/h	Rate of pathway P_k.
$[r_j]$	$\begin{matrix} R_1 & R_2 & R_3 & R_4 & R_5 \\ \begin{bmatrix} 0.1 & 0.6 & 1 & 0.5 & 1 \end{bmatrix} \end{matrix}$	ppb/h	Vector representing the mean rate of reaction R_j in time step dt.
$[\tilde{r_j}]$	$\begin{matrix} R_1 & R_2 & R_3 & R_4 & R_5 \\ \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \end{bmatrix} \end{matrix}$	ppb/h	Vector representing the part of the rate of reaction R_j associated with deleted pathways.
$[\tilde{p_i}]$	$\begin{matrix} S_1 & S_2 & S_3 & S_4 & S_5 & S_6 \\ \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \end{matrix}$	ppb/h	Vector representing the rate of production of species S_i by deleted pathways.
$[\tilde{d_i}]$	$\begin{matrix} S_1 & S_2 & S_3 & S_4 & S_5 & S_6 \\ \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \end{matrix}$	ppb/h	Vector representing the rate of destruction of species S_i by deleted pathways.

$[p_i] = [\tilde{p_i}] +$ pos$([[m_{ik}]]) \cdot [f_k]^T$	$\begin{matrix} S_1 & S_2 & S_3 & S_4 & S_5 & S_6 \\ \begin{bmatrix} 0.6 & 2.5 & 1.1 & 0.1 & 0.5 & 1 \end{bmatrix} \end{matrix}$	ppb/h	Vector representing the rate of production of species S_i by all pathways (including deleted pathways). The computation of this vector involves the product of pos$([[m_{ik}]])$ and the column vector $[f_k]^T$. The notation pos$([[m_{ik}]])$ represents a matrix with the positive values of $[[m_{ik}]]$ and zeros everywhere else.
$[d_i] = [\tilde{d_i}] +$ |neg$([[m_{ik}]])\| \cdot [f_k]^T$	$\begin{matrix} S_1 & S_2 & S_3 & S_4 & S_5 & S_6 \\ \begin{bmatrix} 1.1 & 1.3 & 1.5 & 0 & 2. & 0 \end{bmatrix} \end{matrix}$	ppb/h	Vector representing the rate of destruction of species S_i by all pathways (including deleted pathways). The computation of this vector involves the product of |neg$([[m_{ik}]])\|$ and the column vector $[f_k]^T$. The notation neg$([[m_{ik}]])$ represents a matrix with the negative values of $[[m_{ik}]]$ and zeros everywhere else.

$[[x_{jk}]]$	$\begin{array}{c} \begin{array}{ccccc} P_1 & P_2 & P_3 & P_4 & P_5 \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{array}{c} R_1 \\ R_2 \\ R_3 \\ R_4 \\ R_5 \end{array} \end{array}$	No units	Matrix representing the multiplicity of reaction R_j in pathway P_k. A multiplicity is the number of times a reaction occurs in a pathway. For example, in the ClO pathway presented above all reactions have multiplicities equal to 1. If a reaction does not occur in a pathway, $x_{jk} = 0$. Initially, this matrix is set equal to an identity matrix with the size of the number of reactions n_r.
$[[m_{ik}]] = [[s_{ij}]] \cdot [[x_{jk}]]$	$\begin{array}{c} \begin{array}{ccccc} P_1 & P_2 & P_3 & P_4 & P_5 \end{array} \\ \begin{pmatrix} -1 & 1 & -1 & 0 & 0 \\ -1 & -2 & 1 & 1 & 1 \\ 1 & 0 & 1 & -1 & -1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & -1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{array}{c} S_1 \\ S_2 \\ S_3 \\ S_4 \\ S_5 \\ S_6 \end{array} \end{array}$	ppb	Matrix representing the number of molecules of species S_i produced ($m_{ik} > 0$) or destroyed ($m_{ik} < 0$) by pathway P_k. This variable is equal to the matrix multiplication of $[[s_{ij}]]$ and $[[x_{jk}]]$.

$[\tau_i] = \frac{[\bar{c_i}]}{[d_i]}$	$\begin{matrix} S_1 & S_2 & S_3 & S_4 & S_5 & S_6 \\ \begin{bmatrix} 1.59 & 2.77 & 1.2 & \text{Nan} & 0.38 & \text{Nan} \end{bmatrix} \end{matrix}$	h	Vector representing the lifetime of species S_i. This vector is calculated with the element-wise division of $[c_i]$ and $[d_i]$. If the destruction by all pathways d_i is zero, the lifetime becomes undefined (Nan). This means that there are no pathways consuming S_i.

Table 1: Variables used in the pathway analysis algorithm and their initial values in the simple example used to explain the algorithm.

The algorithm assumes that the reactions are unidirectional and that the user splits the reversible reactions into their forward and backward components. It is also assumed that mass is conserved in the analyzed chemical model, and the reactions produce the exact number of molecules to explain the concentration changes:

$$[dc_i] = [[s_{ij}]] \cdot [r_j]^T dt, \tag{1}$$

This expression involves the product of $[[s_{ij}]]$ and the column vector $[r_j]^T$. For example, in our simple example we can verify that this condition is satisfied:

$$[dc_i] = \overset{[[s_{ij}]]}{\begin{pmatrix} -1 & 1 & -1 & 0 & 0 \\ -1 & -2 & 1 & 1 & 1 \\ 1 & 0 & 1 & -1 & -1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & -1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}} \overset{[r_j]^T}{\begin{pmatrix} 0.1 \\ 0.6 \\ 1 \\ 0.5 \\ 1 \end{pmatrix}} \text{ppb}/h \cdot 1h = \begin{matrix} S_1 & S_2 & S_3 & S_4 & S_5 & S_6 \\ \begin{bmatrix} -0.5 & 1.2 & -0.4 & 0.1 & -1.5 & 1 \end{bmatrix} \end{matrix} \text{ ppb} \tag{2}$$

2.2 Algorithm initialization

The algorithm requires four inputs from a chemical kinetics model: the species concentrations and the model time in two consecutive model times t and $t + dt$, the mean reaction rates in the time step dt, and the reaction system with n_r reactions

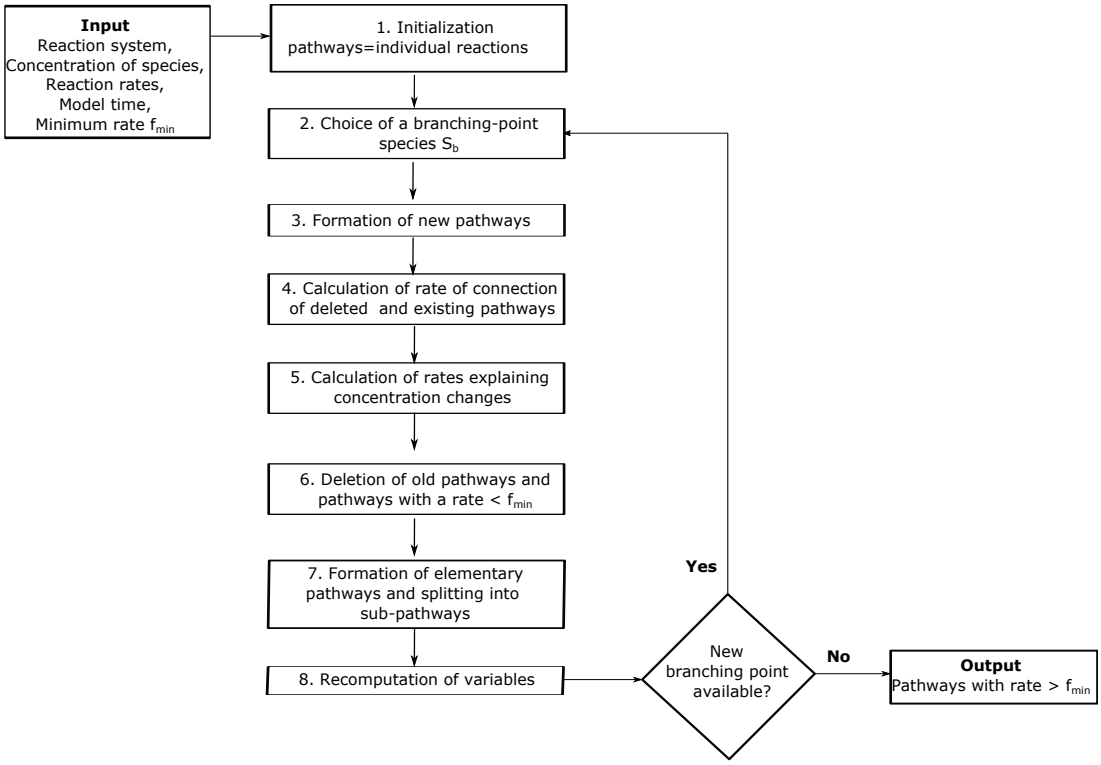

Figure 1. Pathway analysis Program algorithm. For a full description of the algorithm see Lehmann (2004).

between n_s species. We use the model times at which the solver obtains a solution for the system of equations. The user can also input a minimum rate of pathways f_{min}. With these inputs, the algorithm determines all pathways with a rate $> f_{\text{min}}$. In the simple example, we will use a minimum rate of pathways $f_{\text{min}} = 0.05$ ppb/h.

The algorithm uses the input information to initialize the variables listed in table 1. At first, each pathway is considered to have only one reaction, and the matrix $[[x_{jk}]]$ is initialized as an identity matrix with the size of the number of reactions. This means that initially pathway P_1 only contains reaction R_1, pathway P_2 only contains reaction R_2, etc. The rates of the pathways are initialized with the rates of the reactions: $[f_j] = [r_j]$. The variables $[\tilde{r_j}]$, $[\tilde{p_i}]$, and $[\tilde{d_i}]$ that store rates associated with deleted pathways are initialized as arrays of zeros.

2.3 Choice on branching-point species

Once the algorithm has been initialized, the next step is to choose a branching point species S_b to start constructing pathways. Species with shorter lifetimes are selected as branching-point species first because a small lifetime is associated with fast consumption by reactions. The algorithm forms new pathways through the iterative connection of previously formed pathways through branching-point species until there are no more branching-point species left (figure 1). A species can be a branching-point species only once. To find pathways that destroy or produce a specific species of interest, the species itself is not used as

a branching-point. Otherwise, the pathways producing and consuming this species would be connected. Sometimes it is useful to treat some species as inert or long-lived, not considering them as branching-point species. For example, in the atmosphere N_2 is long-lived and inert, so it could not be considered as a branching-point.

In the simple example, where we are investigating the change in OH, we will treat H_2O and O_2 as a long-lived inert species, not considering them as branching-points. The species with the shortest lifetime is O ($\tau_5 = 0.38$ h), so this species is the first branching-point.

2.4 Formation of new pathways and rate calculations

The algorithm forms new pathways connecting all previously formed pathways that produce a branching-point species S_b with all pathways that destroy S_b. The connections are made ensuring that the new pathways do not produce nor destroy the branching-point species S_b. If a pathway P_k produces m_{bk} molecules of the branching-point species S_b and a pathway P_l destroys m_{bl} molecules of S_b, the connection of these pathways forms a new pathway P_n with multiplicities:

$$[x_{jn}] = |m_{bl}|[x_{jk}] + m_{bk}[x_{jl}] \tag{3}$$

where k and l are the indexes of the pathways producing and destroying S_b, and b is the index of the species S_b. The rate f_n of the new pathway P_n is calculated by multiplying the rates of the producing (f_k) and destructing (f_l) pathways and dividing by the maximum of the rate of production and destruction of the branching-point species by all pathways:

$$f_n = \frac{f_k f_l}{\max(p_b, d_b)}, \tag{4}$$

where k and l are the indexes of the pathways producing and destroying the branching-point species S_b, and b is the index of the branching-point species. This equation is derived by calculating the rate at which the molecules of S_b formed by P_k are destroyed by P_l (see Lehmann 2004 for the derivation). One can think of equation 4 as distributing the rate of a pathway to new pathways using the probability that a molecule produced (or consumed) by one pathway is consumed (or produced) by another pathway. For example, if the change in concentration of the branching-point species $dc_b > 0$, then the chemical production is greater than the chemical destruction ($p_b > d_b$), and equation 4 takes the form $f_n = \frac{f_k f_l}{p_b}$. The ratio $\pi = \frac{f_k}{p_b}$ can be interpreted as the probability that a molecule of the branching-point species is produced by the pathway P_k. Since the pathway P_l is going to be connected with all pathways P_k producing S_b, and the sum of the probabilities π over all pathways producing the branching-point species S_b is equal to one, the rate f_l is going to be completely distributed to the new pathways.

The multiplicities $[x_{jn}]$ of the new pathway are divided by their greatest common divisor g to keep them as simple as possible. The rate of the new pathway is multiplied by g to avoid altering the number of molecules that the new pathway produces or destroys. The new pathway and its rate are appended to $[[x_{jk}]]$ and $[f_k]$ respectively.

In the simple example, there are two pathways destroying the branching-point species O (P_3 and P_5), and one pathway producing it (P_4). The connection of these pathways will result in two new pathways. For example, the pathway P_3 destroys

one O molecule ($m_{53} = -1$), and the pathway P_4 produces one O molecule ($m_{54} = 1$), so the connection of these pathways results in a new pathway with multiplicities:

$$[x_{jn}] = |m_{53}|[x_{j4}] + m_{54}[x_{j3}] = [0,0,0,1,0] + [0,0,1,0,0] = [0,0,1,1,0]. \tag{5}$$

The rate of this new pathway is:

$$f_n = \frac{f_4 f_3}{\max(p_5, d_5)} = \frac{0.5 \text{ ppb/h} \cdot 1 \text{ ppb/h}}{2 \text{ ppb/h}} = 0.25 \text{ ppb/h} \tag{6}$$

After forming all new pathways at the branching-point species O, the multiplicities $[[x_{jk}]]$ and the pathway rates $[f_k]$ will have the following form:

$$[[x_{jk}]] = \begin{array}{c} \begin{array}{ccccccc} P_1 & P_2 & P_3 & P_4 & P_5 & P_6 & P_7 \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 \end{pmatrix} \begin{array}{c} R_1 \\ R_2 \\ R_3 \\ R_4 \\ R_5 \end{array} \end{array} \tag{7}$$

$$[f_k] = \begin{array}{c} \begin{array}{ccccccc} P_1 & P_2 & P_3 & P_4 & P_5 & P_6 & P_7 \end{array} \\ \begin{bmatrix} 0.1 & 0.6 & 1 & 0.5 & 1 & 0.25 & 0.25 \end{bmatrix} \end{array} \text{ ppb/h}$$

We can think of a column in $[[x_{jk}]]$ as a pathway. For example, the pathway that we formed before (P_6) is located in the 6th column of the matrix $[[x_{jk}]]$ in equation 7, and contains one times reaction R_3 and one times reaction R_4:

$$\begin{aligned} &\mathrm{H_2O_2 + O \longrightarrow OH + HO_2} \\ &\mathrm{HO_2 + h\nu \longrightarrow OH + O} \\ &\text{Net: } \mathrm{H_2O_2 \longrightarrow 2\,OH} \end{aligned} \tag{P_6}$$

2.5 Calculation of rate of connection of deleted and existing pathways

In this step, the algorithm calculates the rates of the connection of deleted pathways with existing pathways. Pathways with a rate lower than f_{min} will be deleted in a subsequent step (see section 2.7). In the first iteration of the algorithm there are no deleted pathways, but in other iterations some pathways could be deleted by this point. In that case, the deleted pathways would not be connected with existing pathways. However, the algorithm keeps track of the rates of production and destruction

of the branching-point species S_b that would have been computed in these connections. The deleted pathways produce the branching-point species S_b at a rate $\tilde{p_b}$, so according to equation 4, the rate of the connection of deleted pathways that produce S_b with an existing pathway P_e that destroys S_b is:

$$\tilde{f_e} = \frac{f_e \tilde{p_b}}{\max(p_b, d_b)}, \tag{8}$$

where e represents the index of the pathway that destroys the branching-point species S_b. Similarly, deleted pathways consuming S_b at a rate $\tilde{d_b}$ would have been connected with an existing pathway P_e producing S_b at a rate:

$$\tilde{f_e} = \frac{f_e \tilde{d_b}}{\max(p_b, d_b)}, \tag{9}$$

where e represents the index of the pathway that produces S_b. This rate is added to the the variables $[\tilde{r_j}]$, $[\tilde{p_i}]$ and $[\tilde{d_i}]$ that store rates associated with deleted pathways:

$$\begin{aligned} [\tilde{r_j}] &= [\tilde{r_j}] + [x_{je}] \cdot \tilde{f_e} \\ [\tilde{p_i}] &= [\tilde{p_i}] + \text{pos}([m_{ie}]) \cdot \tilde{f_e} \\ [\tilde{d_i}] &= [\tilde{d_i}] + |\text{neg}([m_{ie}])| \cdot \tilde{f_e} \end{aligned} \tag{10}$$

where $e =$ index of pathway producing or destroying the branching-point species S_b. These operations are repeated for all existing pathways producing and destroying the branching-point species. In the simple example, there are no deleted pathways yet, so $\tilde{f_e} = 0$.

2.6 Calculation of rates explaining concentration changes

In this step, the algorithm redefines the rates of the pathways that contribute to the concentration change of the branching-point species S_b. If $dc_b > 0$, the pathways that produce molecules of the branching point species S_b contribute to its concentration change dc_b. Similarly, if $dc_b < 0$, the pathways that destroy molecules of the branching point species S_b contribute to its concentration change dc_b. The algorithm redefines the rate f_k of pathways contributing to the concentration change of the branching-point-species, keeping the fraction $\hat{f_k}$ of f_k that contributes to the concentration change dc_b. This fraction is calculated by multiplying the rate f_k by the absolute value of the rate of concentration change of the branching point species δ_b, and dividing by the maximum of the production and destruction of the branching-point species by all pathways:

$$\hat{f_k} = \frac{f_k |\delta_b|}{\max(p_b, d_b)} \tag{11}$$

where $k =$ index of the pathway producing or destroying S_b, and b is the index of the branching-point species S_b. This rate is derived by calculating the probability that a molecule of S_b produced or destroyed by a pathway contributes to the concentration change of the branching point species dc_b (see Lehmann 2004 for the derivation). If $dc_b = 0$ there is no redefinition of rates.

In the simple example, the branching-point species O has a concentration change $dc_5 = -1.5$ppb. Since $dc_5 < 0$, we will redefine the rates of the pathways that destroy O because they contribute to the concentration change dc_5. For example, the pathway P_3 has a single reaction destroying O at a rate of 1 ppb/h. The part of this rate that contributes to the O concentration change is:

$$\hat{f}_3 = \frac{f_3|\delta_5|}{\max(p_5, d_5)} = \frac{1 \text{ ppb/h} \cdot 1.5 \text{ ppb/h}}{2 \text{ ppb/h}} = 0.75 \text{ ppb/h}. \tag{12}$$

After all the rates of the pathways destroying S_b are redefined, $[f_k]$ has the form:

$$[f_k] = \begin{matrix} P_1 & P_2 & P_3 & P_4 & P_5 & P_6 & P_7 \\ \end{matrix} \begin{bmatrix} 0.1 & 0.6 & 0.75 & 0.5 & 0.75 & 0.25 & 0.25 \end{bmatrix} \quad \text{ppb/h} \tag{13}$$

2.7 Deletion of old pathways and pathways with a rate < f_{min}

After a pathway producing the branching-point species S_b has been connected with all pathways destroying S_b, it is eliminated from the matrix $[[x_{jk}]]$ if it does not contribute to the concentration change dc_b. If $dc_b > 0$, the pathways that destroy the branching-point species are deleted. If $dc_b < 0$, the pathways that produce the branching-point species are deleted, and if $dc_b = 0$, both the pathways that produce and destroy the branching-point species are deleted.

In the simple example $dc_5 < 0$, so the pathways P_3 and P_5 that were used to form new pathways will not be deleted because they destroy molecules of the branching-point species O, contributing to its concentration change. However, the pathway P_4 does not contribute to dc_5, so it is deleted.

The algorithm also deletes pathways with a rate $< f_{\text{min}}$ to avoid constructing an unmanageable number of pathways and to reduce the computing time. If a pathway has a rate $f_q < f_{\text{min}}$, it is deleted from the matrix $[[x_{jk}]]$ and the the variables $[\tilde{r_j}]$, $[\tilde{p_i}]$ and $[\tilde{d_i}]$ are updated according to:

$$\begin{aligned} [\tilde{r_j}] &= [\tilde{r_j}] + [x_{jq}] \cdot f_q \\ [\tilde{p_i}] &= [\tilde{p_i}] + \text{pos}([m_{iq}]) \cdot f_q \\ [\tilde{d_i}] &= [\tilde{d_i}] + |\text{neg}([m_{iq}])| \cdot f_q \end{aligned} \tag{14}$$

where $q =$ index of the pathway with rate $< f_{\text{min}}$. These operations are repeated for all pathways with rate $< f_{min}$. The rates of these pathways are also deleted from $[f_k]$.

In the simple example, none of the pathways have a rate < f_{min}, so there is no deletion of pathways in this first iteration. See section 2.10 to see an example of how pathways with a rate < f_{min} are deleted. After this step, the multiplicities $[[x_{jk}]]$ and the pathway rates $[f_k]$ will have the following form:

$$[[x_{jk}]] = \begin{array}{c} \begin{matrix} P_1 & P_2 & P_3 & P_4 & P_5 & P_6 \end{matrix} \\ \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 \end{pmatrix} \end{array} \begin{matrix} \\ R_1 \\ R_2 \\ R_3 \\ R_4 \\ R_5 \end{matrix} \tag{15}$$

$$[f_k] = \begin{matrix} P_1 & P_2 & P_3 & P_4 & P_5 & P_6 \end{matrix} \\ \begin{bmatrix} 0.1 & 0.6 & 0.75 & 0.75 & 0.25 & 0.25 \end{bmatrix} \quad \text{ppb/h}$$

Note that when pathways are deleted from the matrix $[[x_{jk}]]$ there is a redefinition of pathways. For example, since pathway P_4 was deleted, now there is a new pathway P_4 that corresponds to the fourth column of $[[x_{jk}]]$ in equation 15.

2.8 Formation of elementary pathways and splitting into sub-pathways

The steps above can produce pathways with a large and unnecessary number of reactions. The algorithm splits these complex pathways into shorter, simpler pathways. The first step in this process is to find the elementary sub-pathways of a complex pathway. A pathway P_s is a sub-pathway of a pathway P_c if all the reactions in P_s are in P_c. Elementary pathways do not have sub-pathways. The algorithm finds the elementary sub-pathways of a pathway P_c by forming new pathways with the reactions contained in P_c and keeping only the elementary pathways (see Lehmann 2004 for a full description of this process). If a pathway P_c with multiplicities $[x_{jc}]$ has n_e elementary sub-pathways with multiplicities $[[x'_{je}]]$, the algorithm splits $[x_{jc}]$ into the sub-pathways $[[x'_{je}]]$ finding weights $[w_e]$ that fulfill the equation:

$$[x_{jc}] = [[x'_{je}]]^T \cdot [w_e], \text{ where } c = \text{ index of pathway to be split.} \tag{16}$$

Equation 16 involves the matrix multiplication of $[[x'_{je}]]^T$ and the row vector $[w_e]$. In other words: the sum of the multiplicities of the sub-pathways multiplied by the weights must be equal to the multiplicities of the split pathway. The rate f_c of the pathway P_c is distributed to the sub-pathways using the weights $[w_e]$:

$$[f_e] = f_c[w_e], \text{ where } e = 1 \ldots n_e, \text{ and } c = \text{ index of pathway to be split.} \tag{17}$$

After finding the sub-pathways, the algorithm deletes $[x_{jc}]$ from $[[x_{jk}]]$ and appends the multiplicities of the new sub-pathways $[[x'_{je}]]$ into $[[x_{jk}]]$. Similarly, the rate f_c is deleted, and the rates $[f_e]$ are appended to $[f_k]$. If a sub-pathway already exists in the matrix $[[x_{jk}]]$, its rate is added to the already existing pathway.

As noted by Lehmann (2004), equation 16 can have multiple solutions, leading to slightly different results according to the solution that one chooses. However, these differences tend to be small and the overall result of the algorithm is similar even when equation 16 has multiple solutions (Lehmann, 2004). Our implementation includes two options to solve equation 16. The first option uses Scipy's "lsq_linear" function, minimizing the expression:

$$0.5||A \cdot x - b||^2 \text{ with constraints } 0 \leq x < \infty, \tag{18}$$

where $||x||$ is the norm of x, $A = [[x'_{je}]]^T$, $x = [w_e]$ and $b = [x_{jc}]$. When there are multiple solutions to equation 16, we choose the first solution that minimizes equation 18 found by the "lsq_linear" algorithm. The second option to solve equation 16 implements the method proposed in section 5.5.2 of Lehmann (2004). This method chooses the solution that produces more probable pathways in the sense that this solution produces simpler pathways with higher rates compared to other solutions. Both methods of solving equation 16 produce similar results.

In the first iteration of the algorithm in the simple example there are no pathways to split. See section 2.10 for an example of how to split pathways into sub-pathways.

2.9 Re-computation of variables

The variables $[[m_{ik}]]$, $[p_i]$, and $[d_i]$ are recomputed using the definitions presented in table 1 to match the information from the new pathways formed in the above steps. This re-computation is done after deleting pathways and after splitting pathways into sub-pathways.

In the simple example, this re-computation results in the following values for $[[m_{ik}]]$, $[p_i]$, and $[d_i]$:

$$[[m_{ik}]] = \overset{[[s_{ij}]]}{\begin{pmatrix} -1 & 1 & -1 & 0 & 0 \\ -1 & -2 & 1 & 1 & 1 \\ 1 & 0 & 1 & -1 & -1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & -1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}} \overset{[[x_{jk}]]}{\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 \end{pmatrix}} = \begin{matrix} P_1 & P_2 & P_3 & P_4 & P_5 & P_6 \\ \begin{pmatrix} -1 & 1 & -1 & 0 & -1 & 0 \\ -1 & -2 & 1 & 1 & 2 & 2 \\ 1 & 0 & 1 & -1 & 0 & -2 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \end{pmatrix} \end{matrix} \begin{matrix} S_1 \\ S_2 \\ S_3 \\ S_4 \\ S_5 \\ S_6 \end{matrix} \tag{19}$$

$$[p_i] = \begin{bmatrix} 0 & 0 & [\tilde{p_i}] & 0 & 0 \end{bmatrix} + \overset{\text{pos}([[m_{ik}]])}{\begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 2 & 2 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \end{pmatrix}} \overset{[f_k]}{\begin{pmatrix} 0.1 \\ 0.6 \\ 0.75 \\ 0.75 \\ 0.25 \\ 0.25 \end{pmatrix}} = \overset{\begin{matrix} S_1 & S_2 & S_3 & S_4 & S_5 & S_6 \end{matrix}}{\begin{bmatrix} 0.6 & 2.5 & 0.85 & 0.1 & 0 & 1 \end{bmatrix}} \text{ ppb/h} \tag{20}$$

$$[d_i] = \begin{bmatrix} 0 & 0 & [\tilde{d_i}] & 0 & 0 \end{bmatrix} + \overset{|\text{neg}([[m_{ik}]])|}{\begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}} \overset{[f_k]}{\begin{pmatrix} 0.1 \\ 0.6 \\ 0.75 \\ 0.75 \\ 0.25 \\ 0.25 \end{pmatrix}} = \overset{\begin{matrix} S_1 & S_2 & S_3 & S_4 & S_5 & S_6 \end{matrix}}{\begin{bmatrix} 1.1 & 1.3 & 1.25 & 0 & 1.5 & 0 \end{bmatrix}} \text{ ppb/h} \tag{21}$$

2.10 Second iteration in simple example

After the first iteration at the branching point species O in the simple example, we ended up with six pathways (equation 15). The species with the smallest lifetime with respect to these new pathways and the next branching-point species is HO_2 (S_3). Looking at the third row of $[[m_{ik}]]$ in equation 19 ($[m_{3k}]$), we can see that pathways P_4 and P_6 destroy HO_2 and pathways P_1 and P_3 produce HO_2. The connection of these pathways will lead to the formation of 4 new pathways (section 2.4). For example, connecting pathways P_6 and P_3 we obtain a new pathway P_n with the following multiplicities and rate:

$$\begin{aligned} [x_{jn}] &= |m_{36}|[x_{j3}] + m_{33}[x_{j6}] = 2 \cdot [0,0,1,0,0] + 1 \cdot [0,0,0,1,1] = [0,0,2,1,1] \\ f_n &= \frac{f_3 f_6}{\max(p_3, d_3)} = \frac{0.75 \text{ ppb/h} \cdot 0.25 \text{ ppb/h}}{1.25 \text{ ppb/h}} = 0.15 \text{ ppb/h} \end{aligned} \tag{22}$$

Similarly, the combination of pathways P_1 and P_6 will result in a pathway P_m with multiplicities $[x_{jm}] = [2,0,0,1,1]$, and a rate $f_m = 0.02$ ppb/h. This pathway produces and destroys the number of molecules $[m_{im}] = [-2,0,0,2,0,1]$.

At this point there is no production or destruction of HO_2 by deleted pathways, so the rates of connection of existing pathways with deleted pathways $\tilde{f_e} = 0$ (section 2.5). Since the concentration change of the branching-point species HO_2 is $dc_3 = -0.4$ ppb < 0, the rates of the pathways destroying HO_2 are redefined to keep the fraction that contributes to dc_3 (section 2.6), and the pathways that produce HO_2 are deleted because they do not contribute to dc_3 (section 2.7).

The rate of the pathway P_m is lower than f_{min}, so this pathway will be deleted and the variables that store the rates of deleted pathways are updated using equations 14 (section 2.7):

$$\begin{aligned}
[\tilde{r_j}] &= [\tilde{r_j}] + [x_{jm}] \cdot f_m = [0,0,0,0,0] + [2,0,0,1,1] \cdot 0.02 \text{ ppb/h} = [0.04,0,0,0.02,0.02] \text{ ppb/h} \\
[\tilde{p_i}] &= [\tilde{p_i}] + \text{pos}([m_{im}]) \cdot f_m = [0,0,0,0,0,0] + [0,0,0,2,0,1] \cdot 0.02 \text{ ppb/h} = [0,0,0,0.04,0,0.02] \text{ ppb/h} \\
[\tilde{d_i}] &= [\tilde{d_i}] + \text{neg}([m_{im}]) \cdot f_m = [0,0,0,0,0,0] + [2,0,0,0,0,0] \cdot 0.02 \text{ ppb/h} = [0.04,0,0,0,0,0] \text{ ppb/h}
\end{aligned} \tag{23}$$

The next step is to split pathways into sub-pathways (section 2.8). The pathway P_n formed above (equation 22) contains two times reaction R_3, and one times reactions R_4 and R_5:

$$\begin{aligned}
2\,(\mathrm{H_2O_2} + \mathrm{O} &\longrightarrow \mathrm{OH} + \mathrm{HO_2}) \\
\mathrm{HO_2} + \mathrm{h}\nu &\longrightarrow \mathrm{OH} + \mathrm{O} \\
\mathrm{HO_2} + \mathrm{O} &\longrightarrow \mathrm{OH} + \mathrm{O_2} \\
\text{Net: } 2\,\mathrm{H_2O_2} + 2\,\mathrm{O} &\longrightarrow 4\,\mathrm{OH} + \mathrm{O_2}
\end{aligned} \tag{P_n}$$

This pathway can be split into two simpler pathways:

$$\begin{aligned}
\mathrm{H_2O_2} + \mathrm{O} &\longrightarrow \mathrm{OH} + \mathrm{HO_2} \\
\mathrm{HO_2} + \mathrm{h}\nu &\longrightarrow \mathrm{OH} + \mathrm{O} \\
\text{Net: } \mathrm{H_2O_2} &\longrightarrow 2\,\mathrm{OH}
\end{aligned} \tag{P_{e1}}$$

$$\begin{aligned}
\mathrm{H_2O_2} + \mathrm{O} &\longrightarrow \mathrm{OH} + \mathrm{HO_2} \\
\mathrm{HO_2} + \mathrm{O} &\longrightarrow \mathrm{OH} + \mathrm{O_2} \\
\text{Net: } \mathrm{H_2O_2} + 2\,\mathrm{O} &\longrightarrow 2\,\mathrm{OH} + \mathrm{O_2}
\end{aligned} \tag{P_{e2}}$$

In this case, the solution to equation 16 is $[w_e] = [1,1]$ and we can verify that equation 16 is satisfied:

$$\begin{bmatrix} 0 & 0 & 2 & 1 & 1 \end{bmatrix}_{[x_{jc}]} = \begin{pmatrix} 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \end{pmatrix}_{[[x'_{je}]]^T} \begin{pmatrix} 1 & 1 \end{pmatrix}_{[w_e]} \tag{24}$$

The sub-pathways have the same rate as the initial pathway because the weights w_e are equal to 1. The sub-pathways P_{e1} and P_{e2} were formed before, when the connection of pathways that produce $\mathrm{HO_2}$ with pathways that destroy $\mathrm{HO_2}$ was made, so their rate is going to be added to the rate of the already existing pathways, and the initial pathway will be deleted from $[[x_{jk}]]$.

At the end of this iteration, the variables $[[x_{jk}]]$, $[f_k]$, $[[m_{ik}]]$, $[p_i]$, $[d_i]$, $[\tilde{r_j}]$, $[\tilde{p_i}]$ and $[\tilde{d_i}]$ will have the following form:

$$[[x_{jk}]] = \begin{array}{c} \begin{matrix} P_1 & P_2 & P_3 & P_4 & P_5 & P_6 \end{matrix} \\ \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 1 \end{pmatrix} \end{array} \begin{matrix} R_1 \\ R_2 \\ R_3 \\ R_4 \\ R_5 \end{matrix} \tag{25}$$

$$[f_k] = \begin{matrix} P_1 & P_2 & P_3 & P_4 & P_5 & P_6 \end{matrix} \\ \begin{bmatrix} 0.6 & 0.24 & 0.4 & 0.08 & 0.06 & 0.6 \end{bmatrix} \quad \text{ppb/h} \tag{26}$$

$$[[m_{ik}]] = \begin{array}{c} \begin{matrix} P_1 & P_2 & P_3 & P_4 & P_5 & P_6 \end{matrix} \\ \begin{pmatrix} 1 & 0 & -1 & 0 & -1 & -1 \\ -2 & 1 & 2 & 2 & 0 & 2 \\ 0 & -1 & 0 & -2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 & -1 & -2 \\ 0 & 1 & 0 & 1 & 1 & 1 \end{pmatrix} \end{array} \begin{matrix} S_1 \\ S_2 \\ S_3 \\ S_4 \\ S_5 \\ S_6 \end{matrix} \tag{27}$$

$$[p_i] = \begin{matrix} S_1 & S_2 & S_3 & S_4 & S_5 & S_6 \end{matrix} \\ \begin{bmatrix} 0.6 & 2.4 & 0 & 0.1 & 0 & 1 \end{bmatrix} \quad \text{ppb/h} \tag{28}$$

$$[d_i] = \begin{matrix} S_1 & S_2 & S_3 & S_4 & S_5 & S_6 \end{matrix} \\ \begin{bmatrix} 1.1 & 1.2 & 0.4 & 0. & 1.5 & 0 \end{bmatrix} \quad \text{ppb/h} \tag{29}$$

$$[\tilde{p_i}] = \begin{matrix} S_1 & S_2 & S_3 & S_4 & S_5 & S_6 \end{matrix} \\ \begin{bmatrix} 0 & 0 & 0 & 0.04 & 0 & 0.02 \end{bmatrix} \quad \text{ppb/h} \tag{30}$$

$$[\tilde{d_i}] = \begin{matrix} S_1 & S_2 & S_3 & S_4 & S_5 & S_6 \end{matrix} \\ \begin{bmatrix} 0.04 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad \text{ppb/h} \tag{31}$$

$$[\tilde{r_j}] = \begin{matrix} R_1 & R_2 & R_3 & R_4 & R_5 \end{matrix} \\ \begin{bmatrix} 0.04 & 0 & 0 & 0.02 & 0.02 \end{bmatrix} \text{ ppb/h} \tag{32}$$

2.11 Final iteration in simple example

The final branching-point species in the simple example is H_2O_2 (S_1). In this final iteration of the algorithm there are six pathways in the matrix $[[x_{jk}]]$ (equation 25). Looking at the first row of $[[m_{ik}]]$ we can see that the pathways P_3, P_5 and P_6 consume H_2O_2 and the pathway P_1 produces H_2O_2. The connection of these pathways will lead to the formation of three new pathways.

At this point, the deleted pathways destroy 0.04 H_2O_2 ppb/h (equation 31). This means that we will need to account for the connection of deleted pathways with existing pathways (section 2.5). In this case, the pathway P_1 with rate $f_1 = 0.6$ ppb/h produces one molecule of H_2O_2. This pathway would have been connected with the deleted pathways at a rate (equation 9):

$$\tilde{f}_1 = \frac{f_1 \tilde{d}_1}{\max(p_1, d_1)} = \frac{0.6 \text{ ppb/h} \cdot 0.04 \text{ ppb/h}}{1.1 \text{ ppb/h}} \approx 0.022 \text{ ppb/h} \tag{33}$$

This rate is used to update the variables that store the deleted rates (equation 10):

$$\begin{aligned}
[\tilde{r_j}] &= [\tilde{r_j}] + [x_{j1}] \cdot \tilde{f}_1 = [0.04, 0, 0, 0.02, 0.02] + [0, 1, 0, 0, 0] \cdot 0.022 \text{ ppb/h} \\
&= \begin{matrix} R_1 & R_2 & R_3 & R_4 & R_5 \end{matrix} \\ &\quad \begin{bmatrix} 0.04 & 0.022 & 0 & 0.02 & 0.02 \end{bmatrix} \text{ ppb/h} \\
[\tilde{p_i}] &= [\tilde{p_i}] + \text{pos}([m_{i1}])\tilde{f}_1 = [0, 0, 0, 0.04, 0, 0.02] + [1, 0, 0, 0, 0, 0] \cdot 0.022 \text{ ppb/h} \\
&= \begin{matrix} S_1 & S_2 & S_3 & S_4 & S_5 & S_6 \end{matrix} \\ &\quad \begin{bmatrix} 0.022 & 0 & 0 & 0.04 & 0 & 0.02 \end{bmatrix} \text{ ppb/h} \\
[\tilde{d_i}] &= [\tilde{d_i}] + |\text{neg}([m_{i1}])|\tilde{f}_1 = [0.04, 0, 0, 0, 0, 0] + [0, 2, 0, 0, 0, 0] \cdot 0.022 \text{ ppb/h} \\
&= \begin{matrix} S_1 & S_2 & S_3 & S_4 & S_5 & S_6 \end{matrix} \\ &\quad \begin{bmatrix} 0.04 & 0.043 & 0 & 0 & 0 & 0 \end{bmatrix} \text{ ppb/h}
\end{aligned} \tag{34}$$

After accounting for the connection of deleted and existing pathways, the rates of the pathways P_3, P_5 and P_6 are redefined because they contribute to the H_2O_2 concentration change $dc_5 = -0.5$ ppb (section 2.6). The pathway P_1 is deleted because it does not contribute to the H_2O_2 concentration change, one pathway with rate $< f_{\text{min}}$ is deleted (section 2.7), and no pathways are split into sub-pathways (section 2.8). At the end of this iteration, the variables $[[x_{jk}]]$, $[f_k]$, $[[m_{ik}]]$, $[p_i]$, $[d_i]$, $[\tilde{p_i}]$ and $[\tilde{d_i}]$ will have the following form:

$$[[x_{jk}]] = \begin{array}{c} \begin{matrix} P_1 & P_2 & P_3 & P_4 & P_5 & P_6 \end{matrix} \\ \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 \end{pmatrix} \end{array} \begin{matrix} R_1 \\ R_2 \\ R_3 \\ R_4 \\ R_5 \end{matrix} \tag{35}$$

$$[f_k] = \begin{array}{c} \begin{matrix} P_1 & P_2 & P_3 & P_4 & P_5 & P_6 \end{matrix} \\ \begin{bmatrix} 0.24 & 0.182 & 0.08 & 0.273 & 0.218 & 0.327 \end{bmatrix} \end{array} \text{ ppb/h} \tag{36}$$

$$[[m_{ik}]] = \begin{array}{c} \begin{matrix} P_1 & P_2 & P_3 & P_4 & P_5 & P_6 \end{matrix} \\ \begin{pmatrix} 0 & -1 & 0 & -1 & 0 & 0 \\ 1 & 2 & 2 & 2 & 0 & 0 \\ -1 & 0 & -2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & -2 & 0 & -2 \\ 1 & 0 & 1 & 1 & 0 & 1 \end{pmatrix} \end{array} \begin{matrix} S_1 \\ S_2 \\ S_3 \\ S_4 \\ S_5 \\ S_6 \end{matrix} \tag{37}$$

$$[p_i] = \begin{array}{c} \begin{matrix} S_1 & S_2 & S_3 & S_4 & S_5 & S_6 \end{matrix} \\ \begin{bmatrix} 0.022 & 1.309 & 0 & 0.1 & 0 & 1.0 \end{bmatrix} \end{array} \text{ ppb/h} \tag{38}$$

$$[d_i] = \begin{array}{c} \begin{matrix} S_1 & S_2 & S_3 & S_4 & S_5 & S_6 \end{matrix} \\ \begin{bmatrix} 0.522 & 0.109 & 0.4 & 0 & 1.5 & 0.0 \end{bmatrix} \end{array} \text{ ppb/h} \tag{39}$$

$$[\tilde{p_i}] = \begin{array}{c} \begin{matrix} S_1 & S_2 & S_3 & S_4 & S_5 & S_6 \end{matrix} \\ \begin{bmatrix} 0.022 & 0 & 0 & 0.1 & 0 & 0.08 \end{bmatrix} \end{array} \text{ ppb/h} \tag{40}$$

$$[\tilde{d_i}] = \begin{array}{c} \begin{matrix} S_1 & S_2 & S_3 & S_4 & S_5 & S_6 \end{matrix} \\ \begin{bmatrix} 0.067 & 0.109 & 0 & 0 & 0.06 & 0.0 \end{bmatrix} \end{array} \text{ ppb/h} \tag{41}$$

$$[\tilde{r_j}] = \begin{matrix} R_1 & R_2 & R_3 & R_4 & R_5 \end{matrix} \\ \begin{bmatrix} 0.1 & 0.055 & 0 & 0.02 & 0.08 \end{bmatrix} \text{ ppb/h} \tag{42}$$

2.12 Calculation of contributions

We calculate the contributions $[C_k]$ of the pathways P_k to the production or destruction of a species S_i as the number of molecules of S_i produced or destroyed by P_k over the number of molecules of S_i produced or destroyed by all pathways:

$$\begin{aligned} [C_k] &= \frac{[m_{ik}][f_k]}{p_i} \quad \text{if } P_k \text{ produces } S_i, \\ [C_k] &= \frac{[m_{ik}][f_k]}{d_i} \quad \text{if } P_k \text{ destroys } S_i \end{aligned} \tag{43}$$

This expression involves the element-wise multiplication of two vectors. For example, for the species OH in our simple example:

$$\begin{aligned} C_k = \frac{[m_{3k}][f_k]}{p_3} &= \left[\frac{1 \cdot 0.24}{1.309}, \frac{2 \cdot 0.182}{1.309}, \frac{2 \cdot 0.08}{1.309}, \frac{2 \cdot 0.273}{1.309}, \frac{0 \cdot 0.218}{1.309}, \frac{0 \cdot 0.327}{1.309} \right] \frac{\text{ppb/h}}{\text{ppb/h}} \\ &= \begin{matrix} P_1 & P_2 & P_3 & P_4 & P_5 & P_6 \end{matrix} \\ &\quad \begin{bmatrix} 0.183 & 0.278 & 0.122 & 0.417 & 0 & 0 \end{bmatrix} \end{aligned} \tag{44}$$

We use a similar expression to calculate the contribution of deleted pathways to the production or destruction of a species S_i:

$$\begin{aligned} \tilde{C_p} &= \frac{\tilde{p_i}}{p_i} \quad \text{for deleted pathways producing } S_i, \\ \tilde{C_d} &= \frac{\tilde{d_i}}{d_i} \quad \text{for deleted pathways destroying } S_i, \end{aligned} \tag{45}$$

In this simple example the pathway P_4 involving the interaction between reactions R_3 and R_5 is the most important chain of reactions for the production of OH, contributing 41.7% of the OH production (table 2). The interaction between reactions R_3 and R_4 (pathway P_2) is also important, contributing 27.8% of the OH production. Deleted pathways do not contribute to the OH production, but they contribute 100% of the OH destruction ($\tilde{C_d} = \tilde{d_i}/d_i = 0.109/0.109 = 1$). This means that if we were interested in understanding the chemical reaction chains that destroy OH in this example, we would need to repeat the analysis with a smaller f_{min}.

In this simple example, it is easy to see that pathways P_4 and P_2 are important for OH production without using the algorithm, but when there are hundreds of reactions interacting the pathway analysis program is a valuable tool to understand

the chemical mechanisms that produce the concentration change of a species in an atmospheric chemistry model. Also, even in this simple example we can see how this algorithm provides valuable quantitative information about the contribution of each pathway to the production of a species.

ID	Pathway	Contribution (%)	Rate (ppb/h)
P_4	$H_2O_2 + O \longrightarrow OH + HO_2$ $HO_2 + O \longrightarrow OH + O_2$ Net: $H_2O_2 + 2\,O \longrightarrow 2\,OH + O_2$	41.7	0.273
P_2	$H_2O_2 + O \longrightarrow OH + HO_2$ $HO_2 + h\nu \longrightarrow OH + O$ Net: $H_2O_2 \longrightarrow 2\,OH$	27.8	0.182
P_1	$HO_2 + O \longrightarrow OH + O_2$ Net: $HO_2 + O \longrightarrow OH + O_2$	18.3	0.24
P_3	$HO_2 + h\nu \longrightarrow OH + O$ $HO_2 + O \longrightarrow OH + O_2$ Net: $2\,HO_2 \longrightarrow 2\,OH + O_2$	12.2	0.08

Table 2: Contribution of pathways to the production of OH in the simple example used to explain the algorithm.

3 Implementation

We implement Lehmann's (2004) algorithm in Python. We designed an object-oriented code defining a class to store the variables listed in table 1 and separating the steps described in section 2 into different class methods. We run these methods in a main method that performs the loop shown in figure 1. We represent the multiplicities $[[x_{jk}]]$ of the pathways as a sparse matrix to optimize memory usage. Our implementation includes the option to find pathways using multiprocessing to speed up the computation time.

The code includes several functions that are useful for analyzing the pathways. After the main algorithm loop ends, the variables in table 1 have all the information of the pathways that have been found. The code includes functions to save these variables to binary files and to read them for future analysis. The code also has functions to transform the representation of a pathway from an array of multiplicities to a string, to get the net reaction of a pathway, to create a latex table with the pathways that contribute to the production or destruction of a species of interest, and to assign a unique identifier to each pathway. This identifier is a string containing the multiplicities and the indexes of the reactions in a pathway. The same pathway can be formed multiple times during the algorithm, so we use this identifier to avoid repeating pathways. We also include a function to calculate the contribution of all pathways to the production or destruction of a species.

Our implementation includes the option to specify species that will be ignored as branching point species. If we are interested in finding pathways at a specific timescale, it is convenient not to consider species with lifetimes higher than the timescale of

interest as branching-points. Our implementation also includes the option to ignore species as branching-point species by specifying a maximum lifetime of branching-point species.

3.1 Tests

The code includes run-time tests to ensure that the code works well. If the construction of pathways is correct, the rates of the reactions must be completely distributed to the pathways:

$$[r_j] = [\tilde{r_j}] + [[x_{jk}]] \cdot [f_k]^T, \quad j = 1 \dots n_r \tag{46}$$

This condition ensures that the number of molecules of a species produced or destroyed by the initial reactions is the same as the number of molecules produced or destroyed by the pathways. The code checks that equation 46 is fulfilled in each iteration of the algorithm, and displays a warning if it is not satisfied.

The code also includes unit tests to ensure that the code works as expected in a simple scenario with four reactions representing Chapman's O_3 destruction mechanism (Chapman, 1930). This scenario is used by Lehmann (2004) to explain how the algorithm works. We include tests to ensure that our implementation finds the same pathways and rates as those found by Lehmann (2004) in this very simple example.

3.2 Tracking imbalances due to numerical errors

Before the construction of pathways, it is essential to ensure that the concentration changes of all species are balanced by the reaction's production and destruction (equation 1). The balance might not be fulfilled due to numerical errors. To quantify this problem, we assume that the difference between concentration changes and the total production by all reactions is due to the solver's numerical error, and we include error pseudo-reactions in the chemical system that produce or destroy a species at the rate required to fulfill the balance. We include the error pseudo-reactions in the construction of pathways. After the pathway construction finishes, we delete the pathways containing error pseudo-reactions, updating $\tilde{r_j}$, $\tilde{p_i}$ and $\tilde{d_i}$. We also include variables similar to $\tilde{r_j}$, $\tilde{p_i}$ and $\tilde{d_i}$ to track the rates of the pathways containing error pseudo-reactions. This approach gives the user information on how important the numerical error is in explaining the concentration changes. Ideally, the pathways containing error pseudo-reactions will not contribute significantly to the concentration change one is interested in understanding.

3.3 Using Chempath

Chempath is available at https://github.com/DanyIvan/chempath. This code repository includes a tutorial on how to use Chempath, as well as some examples of how to apply Chempath to a box model and to a one-dimensional photochemical model (section 4).

There are two important steps to use Chempath. First, the user needs to transform the output of a photochemical model into files readable by Chempath, and balance the concentration changes if necessary (section 3.2). The code repository includes some examples of how to create these input files. Second, the user needs to choose a minimum rate of pathways f_{min}. This can

be done by trial and error, or setting it as a fraction of the rate of total production or destruction by the reactions of the species the user is interested in finding pathways for. However, this way of setting f_{min} might still require further trial and error to find an appropriate fraction of the total production or destruction by all reactions. Ideally, f_{min} will be small enough so that the deleted pathways do not contribute significantly to the production or destruction of a species of interest. However, if f_{min} is too small, the code might take a long time to run.

4 Application examples

Here we provide two examples of how to apply *Chempath* to a photochemical box model (section 4.1) and to a 1D photochemical model (section 4.2)

4.1 Pathways in a photochemical box model

To give an example of the application of *Chempath* in a simple scenario we developed a photochemical box model that solves the equation:

$$\frac{d\rho_i}{dt} = \Pi_i - L_i \tag{47}$$

where ρ_i is the number density of a species i, Π_i is its chemical production and L_i is its chemical destruction. We use a reaction system with 15 reactions between 8 species involving O_3 and hydrogen oxide species. All the reactions and reaction rate parameters that we use can be consulted here: https://github.com/DanyIvan/chempath/tree/main/examples/box_model_pathways. To keep the model as simple as possible, we assume that photochemical reactions have constant rate constants. We obtain these rate constants by running a more complex photochemical model that solves radiative transfer (Wogan, 2023). We use similar rates to the ones obtained by this model at 20km of altitude. The purpose of this simple model is to provide an example of the use of the pathway analysis algorithm, not to be an accurate representation of the stratospheric O_3 chemistry.

We run the box model for 80 days. All species in the model reach a steady state during this time, and there is an increase in the O_3 number density (figure 2a). The O_3 concentration change is the result of an imbalance between O_3 production and loss (figure 2d). This imbalance causes an increase in O_3 until a steady state is reached. We use Chempath to see what are the most important pathways for the production and destruction of O_3 in this model run.

To balance the concentration changes we include error pseudo-reactions in the reaction system that produce or destroy a species at the rate required to fulfill the balance. We calculate the rate of these error pseudo-reactions as the difference between the left and right hand side of equation 47. After including the error pseudo-reactions, the concentration changes of all the species are exactly explained by the production and destruction of the reactions (figure 2b shows an example of this for the O_3 concentration change). The pathways containing error pseudo-reactions contribute less than 0.001% to the O_3 concentration change across the model run (figure 2f).

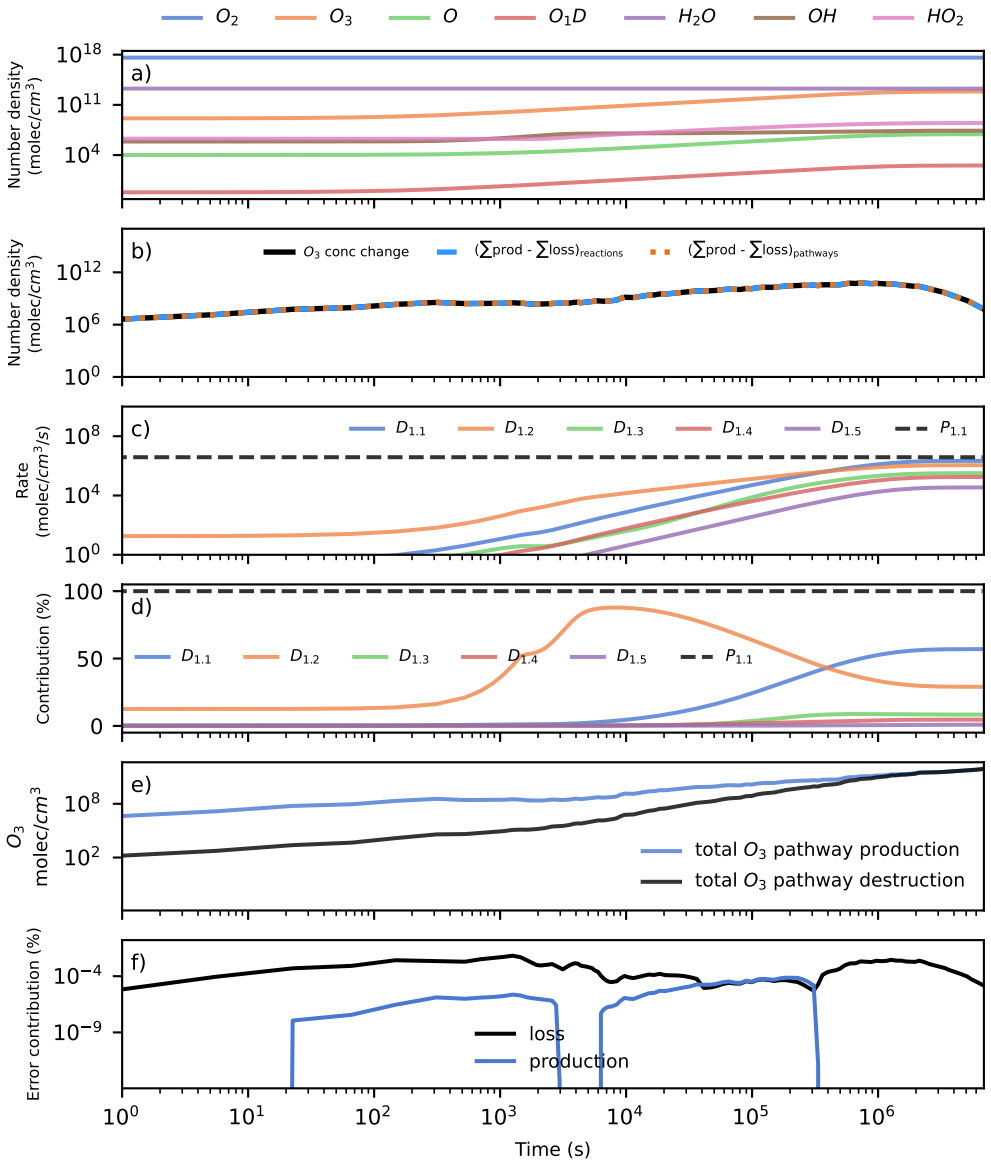

Figure 2. Application of *Chempath* to a simple photochemical box model. Panel *a* presents the evolution of the model over time. Panel *b* shows the O_3 concentration change over time (black line), the total production minus the total loss by all reactions (blue dashed line), and by all pathways (orange dotted line). This panel shows that the reactions and pathways completely explain the O_3 concentration change, and equations 1 and 46 are satisfied for O_3. Panels *c* and *d* show the rate and contribution of the 5 most important pathways for O_3 destruction (colored continuous lines), and the only pathway for O_3 production (black dotted line). These pathways can be consulted in table 3. Panel *e* shows the total O_3 production and destruction by all pathways. Panel *f* shows the contribution of pathways containing error pseudo-reactions to O_3 production or loss.

We run Chempath to find pathways in all the model time intervals at which the solver obtained a solution for the system, setting $f_{\text{min}} = 0$. The reaction rates are completely distributed to the pathways (equation 46 is fulfilled), so the concentration changes are exactly explained by the production and destruction by the pathways (figure 2b shows an example of this for the O_3 concentration change). There is only one O_3 producing pathway in this reaction system (pathway P_1 in table 3 and figure 2c). The algorithm identified 49 O_3 destruction pathways (figure 2c and table 3 present the five most important pathways). These pathways destroy O_3 through hydrogen oxide (HO_x) catalytic cycles.

ID	Pathway	ID	Pathway
$P_{1.1}$	$O_2 + h\nu \longrightarrow O + O$ $2(O_2 + O \longrightarrow O_3)$ Net: $3\,O_2 \longrightarrow 2\,O_3$	$D_{1.1}$	$OH + O_3 \longrightarrow HO_2 + O_2$ $HO_2 + O_3 \longrightarrow OH + O_2 + O_2$ Net: $2\,O_3 \longrightarrow 3\,O_2$
$D_{1.2}$	$O_3 + h\nu \longrightarrow O_2 + O(^1D)$ $H_2O + O(^1D) \longrightarrow OH + OH$ $OH + O_3 \longrightarrow HO_2 + O_2$ $OH + HO_2 \longrightarrow H_2O + O_2$ Net: $2\,O_3 \longrightarrow 3\,O_2$	$D_{1.3}$	$O_3 + h\nu \longrightarrow O_2 + O(^1D)$ $H_2O + O(^1D) \longrightarrow OH + OH$ $OH + O_3 \longrightarrow HO_2 + O_2$ $HO_2 + HO_2 \longrightarrow H_2O_2 + O_2$ $H_2O_2 + OH \longrightarrow HO_2 + H_2O$ Net: $2\,O_3 \longrightarrow 3\,O_2$
$D_{1.4}$	$O_3 + h\nu \longrightarrow O_2 + O$ $HO_2 + O \longrightarrow OH + O_2$ $OH + O_3 \longrightarrow HO_2 + O_2$ Net: $2\,O_3 \longrightarrow 3\,O_2$	$D_{1.5}$	$O_3 + h\nu \longrightarrow O_2 + O$ $O_3 + O \longrightarrow O_2 + O_2$ Net: $2\,O_3 \longrightarrow 3\,O_2$

Table 3: Pathways for O_3 production and destruction in the simple photochemical box model shown in figure 2d.

4.2 Pathways in a one-dimensional photochemical model

In this section, we show how to apply *Chempath* to the one-dimensional photochemical model *photochem* (Wogan et al. 2023, https://github.com/Nicholaswogan/photochem). We run the *photochem* model with the "Modern Earth" reaction scheme that includes 1281 reactions between 113 species. We run the model to photochemical equilibrium using surface flux boundary conditions for O_2, CH_4, CO, and H_2. We choose fluxes of $3.3 \times 10^{11}, 5 \times 10^{10}, 1.2 \times 10^8$ and 3×10^9 molecules$/cm^2/s$ for each of these species respectively. The choice of these fluxes is arbitrary and motivated to get similar conditions to the present atmosphere. For all other species, we use the default boundary conditions of the "Modern Earth" reaction scheme. After the model reaches equilibrium, we decrease the O_2 surface flux to 2.5×10^{11} molecules$/cm^2/s$ and run the model for 1.2 million years. The idea behind the reduction of the O_2 surface flux is to create a perturbation that causes concentration changes to explore with *Chempath*. The concentration of O_2 in the model is controlled mainly by the surface flux and by oxidation of reduced species like CH_4, CO, and H_2 in a timescale of millions of years (the estimated lifetime for O_2 in the modern

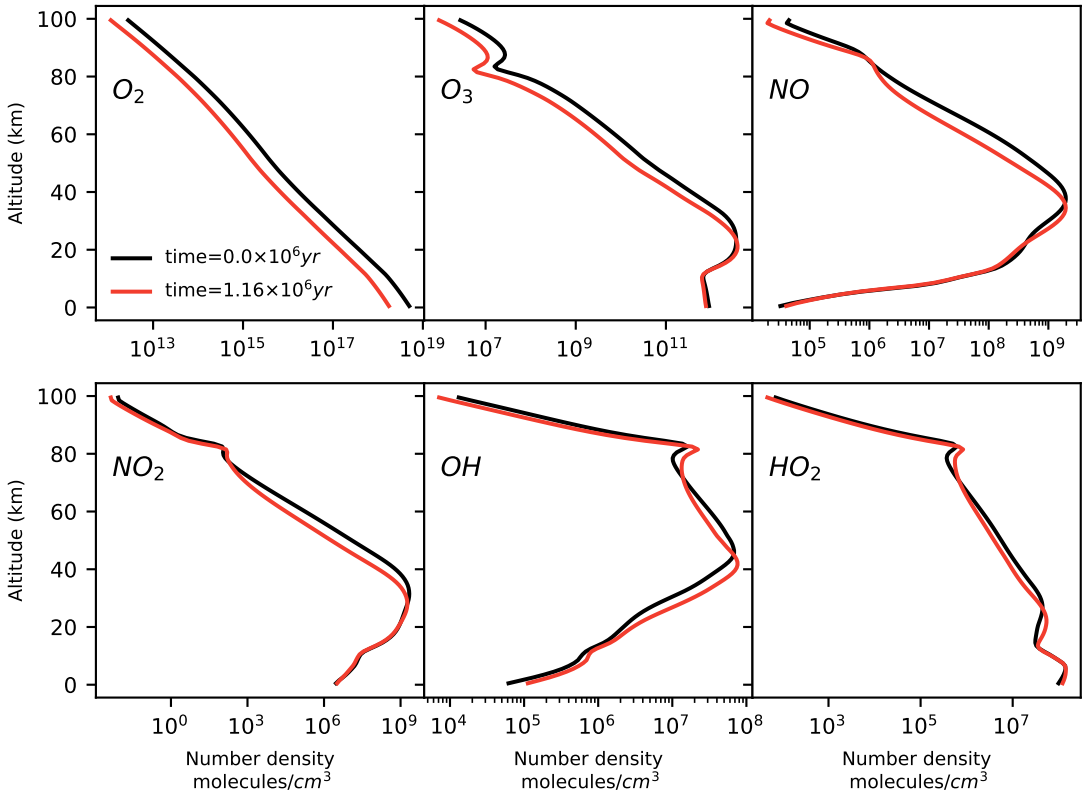

Figure 3. Number density profiles of O_2, O_3, NO, NO_2, OH and HO_2 calculated by the *photochem* model at time$=0.15$ million years (black line) and at time$=1.16$ million years (red line). In this model run we decrease the O_2 input surface flux. As a result, the O_2 and O_3 number densities tend to decrease at all altitudes. The concentration of NO, NO_2, OH and HO_2 decrease and increase at different altitudes.

atmosphere is $\sim$2 million years (Kasting, 2013)). The *photochem* model uses a solver with an adaptive timestep (CVODE BDF method created by Sundials Computing). In our simulation the timestep varies from 10^{-5}s to 10^{12}s.

The model output shows that O_2 and O_3 concentrations tend to decrease at all altitudes as a consequence of the decrease in the O_2 surface input flux (figure 3). We apply *Chempath* to the *photochem* model output to gain insight into the chemical reaction chains that produce and destroy O_3 in this model run.

4.3 Methods: How to find pathways in the *photochem* model

The application of *Chempath* to the *photochem* model output requires the inclusion of pseudo-reactions for processes that affect the concentration of species in a 1D photochemical model. The *photochem* model calculates the concentration changes of long-lived species solving the equation:

$$\frac{\partial \rho_i}{\partial t} = \frac{\partial}{\partial z}\Phi_i + \Pi_i - L_i - \Omega_i + F_i, \tag{48}$$

where ρ_i is the number density of species i, z is altitude, Π_i and L_i are the chemical production and destruction of species i, Φ_i is the vertical transport flux of species i, Ω_i is the destruction of species i from rainout, and F_i is a vertically distributed input flux (see Wogan et al. 2022 for more details).

We include pseudo-reactions for transport, rainout, and vertically distributed input fluxes in the reaction system at each altitude and model time in which we perform the pathway analysis:

$$\begin{aligned} \mathrm{S}_i &\longrightarrow \mathrm{S}_{i,\mathrm{trpt}} \text{ if transport removes } S_i \\ \mathrm{S}_{i,\mathrm{trpt}} &\longrightarrow \mathrm{S}_i \text{ if transport supplies } S_i \\ \mathrm{S}_i &\longrightarrow \mathrm{S}_{i\ \mathrm{rainout}}, \\ \mathrm{S}_{i,\mathrm{dist}} &\longrightarrow \mathrm{S}_i \text{ if there is a vertically distributed flux for } S_i \end{aligned} \tag{49}$$

We obtain the rate of these pseudo-reactions from the vertical transport fluxes, rainout rates, and vertically distributed fluxes calculated in the model.

To balance the concentration changes we include error pseudo-reactions that produce or destroy a species at the rate required to achieve mass balance. We calculate the rate of these pseudo-reactions as the difference between the left and right-hand side of equation 48. The pathways containing error pseudo-reactions contribute less than 1% of the O_3 concentration change at all times and altitudes in our analysis.

We run *Chempath* with the augmented reaction system at 32 time points distributed across the model run, ranging from 1 second to 1.2 million years. We find pathways at all the model altitudes, except at the lower and upper boundary. We prescribe a variable minimum pathway rate f_{min} that we calculate as the minimum of the chemical production by reactions (including transport pseudo-reactions) of O_2, O_3, CO, H_2 and CH_4 divided by 1000. We use these species to calculate f_{min} because we are interested in understanding their concentration changes. We do not consider these species as branching-points. We also ignore N_2, CO_2, and H_2O as branching-point species, treating them as long-lived inert species. Our f_{min} choice keeps the contribution of deleted pathways to O_3 production and destruction below 5% at all altitudes and times in our analysis. The number of reactions with rate $> f_{min}$ varies with altitude, and ranges from 94 to 160.

4.4 Results: Ozone destruction and production pathways in the *photochem* model

Chempath allows us to identify the most important pathways for O_3 production and destruction at a given altitude and time in our *photochem* model run (figures 4 and table 4). These pathways are similar to those found in a previous study of chemical pathways affecting O_3 in the atmosphere (Grenfell et al., 2006).

In the troposphere (below 11 km in our model run), O_3 production in the *photochem* model is dominated by transport (pathway $P_{2.1}$) and CH_4 and CO oxidation (pathways $P_{2.2}$ and $P_{2.3}$) under the presence of nitrogen oxide radicals (NO_x). These

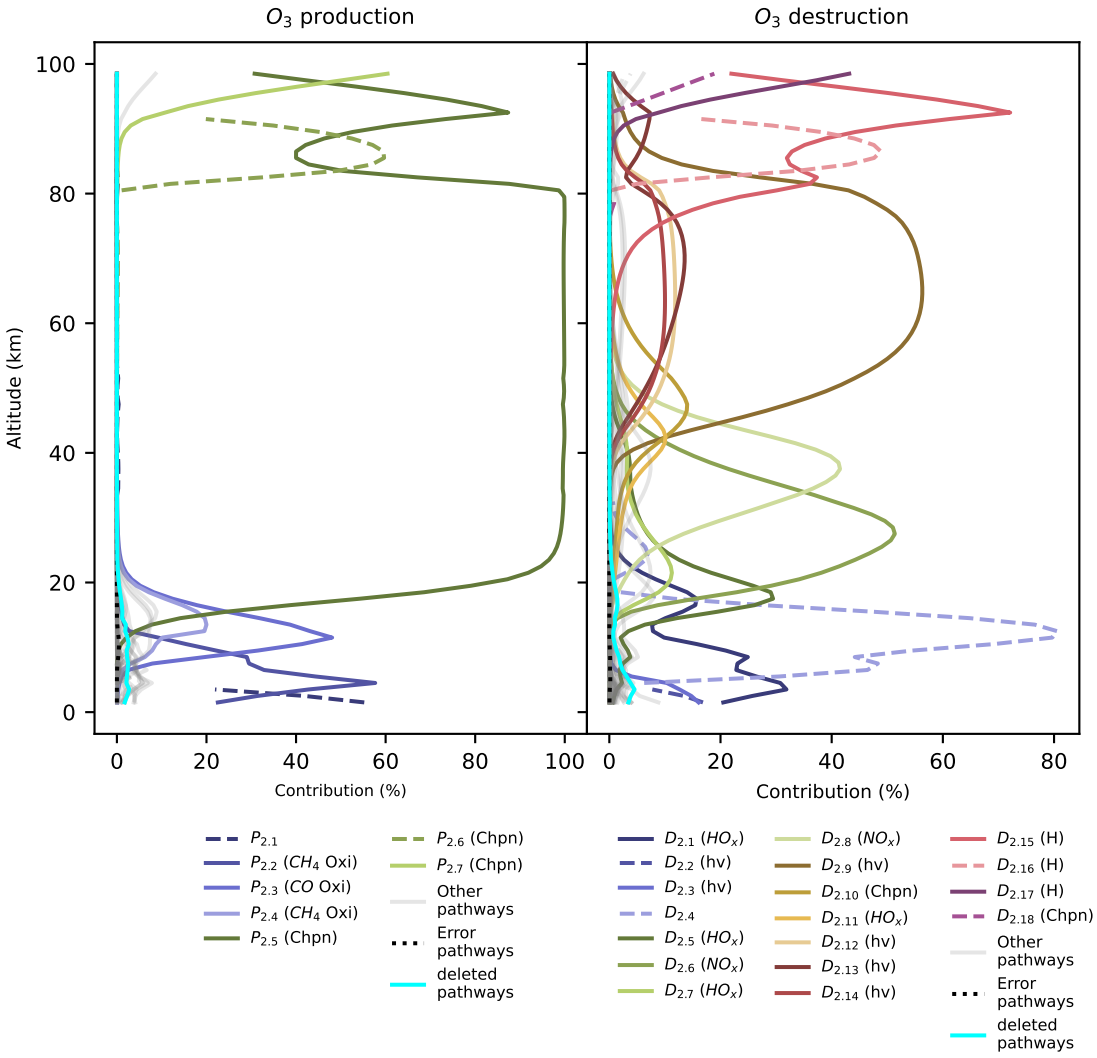

Figure 4. Contribution of pathways to O_3 production and destruction as a function of altitude at the model time= 1.16 million years. The main pathways are plotted in color and the less important pathways are plotted in gray. Pathways that include transport pseudo-reactions are plotted in dashed lines. These pathways show discontinuities because transport can either supply or remove O_3 at different altitudes. The symbols in the legend group the pathways into six categories: Oxidation (Oxi), Chapman-like (Chpn), photolysis (hν), hydrogen oxide (HO_x) and nitrogen oxide (NO_x) pathways, and destruction by the H atom (H). The pathways are listed in table 4. The cyan line shows the contribution of deleted pathways and the black dotted line shows the contribution of pathways containing error pseudo-reactions.

oxidation pathways are similar to the "smog mechanism" that produces tropospheric O_3 through oxidation of hydrocarbons (Haagen-Smit, 1952; Volz-Thomas A., 1994). The tropospheric O_3 destruction is dominated by destruction by HO_x radicals (pathway $D_{2.1}$), O_3 photolysis and subsequent CH_4 oxidation (pathways $D_{2.2}$ and $D_{2.3}$), and transport (pathway $D_{2.4}$).

In the stratosphere (11-50km), O_3 production is mainly occurring via CO and CH_4 oxidation under the presence of NO_x radicals below 25km (pathways $P_{2.3}$ and $P_{2.4}$), and by the Chapman production pathway $P_{2.5}$ above 25km. The main stratospheric O_3 destruction mechanisms involve transport (pathway $D_{2.4}$), photolysis (pathway $D_{2.10}$), and destruction by HO_x radicals (pathways $D_{2.1}$, $D_{2.5}$, $D_{2.7}$ and $D_{2.11}$) and NO_x radicals (pathways $D_{2.6}$ and $D_{2.8}$). Catalytic cycles involving NO_x and HO_x radicals are important for stratospheric O_3 destruction (Lary, 1997; Jacob, 1999). *Chempath* can identify these well-known catalytic cycles in the photochem model.

Above 50km of altitude, the main O_3 production mechanisms are Chapman-like production pathways ($P_{2.5}$ to $P_{2.7}$), and the main O_3 destruction mechanisms involve O_3 photolysis coupled to HO_x radicals cycles, destruction by HO_x radicals, Chapman-like destruction pathways, and destruction by the H atom ($D_{2.9}$ to $D_{2.18}$).

The decrease in O_3 concentration in our model run is likely the result of a decrease in O_3 production and destruction caused by the decrease in the O_2 input flux. For example, the rate of the main stratospheric O_3 producing and destroying pathways ($P_{2.5}$ and $D_{2.6}$) decreases over time (figure 5). This is likely the result of the decrease in the O_2 concentration leading to a decrease in O_3 production trough pathway $P_{2.5}$, causing a decrease in O_3 concentration, and a subsequent decrease in the O_3 loss rate. However, the contribution profiles shown in figure 4 have a similar structure in all the time steps we analyzed. Consequently, the pathways shown in figure 4 and listed in table 4 are a good representation of the pathways that produce and destroy O_3 across all times we analyzed in our model run.

The presence of the supply by transport of the methylperoxy radical (CH_3O_2) in the pathway in $D_{2.2}$ is surprising because CH_3O_2 has a lifetime < 1 minute in the troposphere (Wolfe et al., 2014), and its concentration should be controlled by reactions, not by transport. This is the result of an incomplete representation of the chemistry of this species in the reaction system we used. This reaction system comes from a legacy of models designed to study Early Earth and exoplanet anoxic atmospheres. Thus, the reaction system lacks some reactions for the oxidized modern Earth atmosphere. This example illustrates how Chempath can be a great tool to understand and validate the results of photochemical models, allowing the modelers to detect potential problems with their reaction systems.

ID	Pathway	Contrib %	Rate $\frac{molec}{cm^3 s}$	Alt km
$P_{2.1}$	$O_{3trpt} \longrightarrow O_3$ Net: $O_{3trpt} \longrightarrow O_3$	55.233	1.376e+04	1.5

$P_{2.2}$	$2(CH_4 + OH \longrightarrow CH_3 + H_2O)$ $2(CH_3 + O_2 + M \longrightarrow CH_3O_2 + M)$ $2(CH_3O_2 + NO \longrightarrow CH_3O + NO_2)$ $2(CH_3O + O_2 \longrightarrow H_2CO + HO_2)$ $2(NO_2 + h\nu \longrightarrow NO + O)$ $2(H_2CO + h\nu \longrightarrow CO + H_2)$ $HO_2 + HO_2 \longrightarrow H_2O_2 + O_2$ $H_2O_2 + h\nu \longrightarrow OH + OH$ $2(O + O_2 + M \longrightarrow O_3 + M)$ Net: $5\,O_2 + 2\,CH_4 \longrightarrow 2\,H_2 + 2\,H_2O + 2\,CO + 2\,O_3$	57.683	8.074e+03	4.5
$P_{2.3}$	$CO + OH \longrightarrow CO_2 + H$ $H + O_2 + M \longrightarrow HO_2 + M$ $NO + HO_2 \longrightarrow NO_2 + OH$ $NO_2 + h\nu \longrightarrow NO + O$ $O + O_2 + M \longrightarrow O_3 + M$ Net: $2\,O_2 + CO \longrightarrow CO_2 + O_3$	48.059	1.639e+04	11.5
$P_{2.4}$	$CH_4 + OH \longrightarrow CH_3 + H_2O$ $CH_3 + O_2 + M \longrightarrow CH_3O_2 + M$ $CH_3O_2 + NO \longrightarrow CH_3O + NO_2$ $CH_3O + O_2 \longrightarrow H_2CO + HO_2$ $H_2CO + h\nu \longrightarrow CO + H_2$ $NO + HO_2 \longrightarrow NO_2 + OH$ $2(NO_2 + h\nu \longrightarrow NO + O)$ $2(O + O_2 + M \longrightarrow O_3 + M)$ Net: $4\,O_2 + CH_4 \longrightarrow H_2 + H_2O + CO + 2\,O_3$	20.0	4.352e+03	13.5
$P_{2.5}$	$O_2 + h\nu \longrightarrow O + O$ $2(O + O_2 + M \longrightarrow O_3 + M)$ Net: $3\,O_2 \longrightarrow 2\,O_3$	99.953	1.393e+05	76.5
$P_{2.6}$	$O_{trpt} \longrightarrow O$ $O + O_2 + M \longrightarrow O_3 + M$ Net: $O_2 + O_{trpt} \longrightarrow O_3$	59.658	1.675e+05	85.5

$P_{2.7}$	$O_2 + h\nu \longrightarrow O + O(^1D)$ $O(^1D) + N_2 \longrightarrow O + N_2$ $2(O + O_2 + M \longrightarrow O_3 + M)$ Net: $3\,O_2 \longrightarrow 2\,O_3$	60.429	8.510e+03	98.5
$D_{2.1}$	$2(OH + O_3 \longrightarrow HO_2 + O_2)$ $HO_2 + HO_2 \longrightarrow H_2O_2 + O_2$ $H_2O_2 + h\nu \longrightarrow OH + OH$ Net: $2\,O_3 \longrightarrow 3\,O_2$	31.868	4.030e+03	3.5
$D_{2.2}$	$O_3 + h\nu \longrightarrow O(^1D) + O_2$ $O(^1D) + H_2O \longrightarrow OH + OH$ $2(CH_4 + OH \longrightarrow CH_3 + H_2O)$ $2(CH_3 + O_2 + M \longrightarrow CH_3O_2 + M)$ $2(CH_3O_2 \longrightarrow CH_3O_{2\text{trpt}})$ Net: $O_2 + 2\,CH_4 + O_3 \longrightarrow H_2O + 2\,CH_3O_{2\text{trpt}}$	16.846	4.197e+03	1.5
$D_{2.3}$	$2(O_3 + h\nu \longrightarrow O + O_2)$ $2(CH_3O_2 + O \longrightarrow CH_3O + O_2)$ $2(CH_3O + O_2 \longrightarrow H_2CO + HO_2)$ $2(H_2CO + h\nu \longrightarrow CO + H_2)$ $HO_2 + HO_2 \longrightarrow H_2O_2 + O_2$ $H_2O_2 + h\nu \longrightarrow OH + OH$ $2(CH_4 + OH \longrightarrow CH_3 + H_2O)$ $2(CH_3 + O_2 + M \longrightarrow CH_3O_2 + M)$ Net: $2\,CH_4 + 2\,O_3 \longrightarrow 2\,H_2 + 2\,H_2O + O_2 + 2\,CO$	16.06	2.000e+03	1.5
$D_{2.4}$	$O_3 \longrightarrow O_{3\text{trpt}}$ Net: $O_3 \longrightarrow O_{3\text{trpt}}$	80.525	3.215e+04	12.5
$D_{2.5}$	$O_3 + h\nu \longrightarrow O(^1D) + O_2$ $O(^1D) + H_2O \longrightarrow OH + OH$ $OH + O_3 \longrightarrow HO_2 + O_2$ $OH + HO_2 \longrightarrow H_2O + O_2$ Net: $2\,O_3 \longrightarrow 3\,O_2$	29.455	9.010e+03	17.5
$D_{2.6}$	$O_3 + h\nu \longrightarrow O + O_2$ $NO_2 + O \longrightarrow NO + O_2$ $NO + O_3 \longrightarrow NO_2 + O_2$ Net: $2\,O_3 \longrightarrow 3\,O_2$	51.365	4.124e+05	27.5

$D_{2.7}$	$O_3 + h\nu \longrightarrow O + O_2$ $HO_2 + O \longrightarrow OH + O_2$ $OH + O_3 \longrightarrow HO_2 + O_2$ Net: $2O_3 \longrightarrow 3O_2$	11.255	1.307e+04	21.5
$D_{2.8}$	$O_3 + h\nu \longrightarrow O(^1D) + O_2$ $O(^1D) + N_2 \longrightarrow O + N_2$ $NO_2 + O \longrightarrow NO + O_2$ $NO + O_3 \longrightarrow NO_2 + O_2$ Net: $2O_3 \longrightarrow 3O_2$	41.503	8.844e+05	37.5
$D_{2.9}$	$2(O_3 + h\nu \longrightarrow O(^1D) + O_2)$ $2(O(^1D) + N_2 \longrightarrow O + N_2)$ $HO_2 + O \longrightarrow OH + O_2$ $O + OH \longrightarrow O_2 + H$ $H + O_2 + M \longrightarrow HO_2 + M$ Net: $2O_3 \longrightarrow 3O_2$	56.3	2.253e+05	64.5
$D_{2.10}$	$O_3 + h\nu \longrightarrow O(^1D) + O_2$ $O(^1D) + N_2 \longrightarrow O + N_2$ $O + O_3 \longrightarrow O_2 + O_2$ Net: $2O_3 \longrightarrow 3O_2$	14.007	1.553e+05	47.5
$D_{2.11}$	$O_3 + h\nu \longrightarrow O(^1D) + O_2$ $O(^1D) + N_2 \longrightarrow O + N_2$ $HO_2 + O \longrightarrow OH + O_2$ $OH + O_3 \longrightarrow HO_2 + O_2$ Net: $2O_3 \longrightarrow 3O_2$	10.11	1.671e+05	42.5
$D_{2.12}$	$2(O_3 + h\nu \longrightarrow O + O_2)$ $HO_2 + O \longrightarrow OH + O_2$ $O + OH \longrightarrow O_2 + H$ $H + O_2 + M \longrightarrow HO_2 + M$ Net: $2O_3 \longrightarrow 3O_2$	11.848	5.012e+04	63.5
$D_{2.13}$	$O_3 + h\nu \longrightarrow O(^1D) + O_2$ $O(^1D) + N_2 \longrightarrow O + N_2$ $O + OH \longrightarrow O_2 + H$ $H + O_3 \longrightarrow OH + O_2$ Net: $2O_3 \longrightarrow 3O_2$	13.55	3.723e+04	70.5

$D_{2.14}$	$2(O_3 + h\nu \longrightarrow O(^1D) + O_2)$ $2(O(^1D) + O_2 \longrightarrow O + O_2)$ $HO_2 + O \longrightarrow OH + O_2$ $O + OH \longrightarrow O_2 + H$ $H + O_2 + M \longrightarrow HO_2 + M$ Net: $2O_3 \longrightarrow 3O_2$	10.016	4.237e+04	63.5
$D_{2.15}$	$O_2 + h\nu \longrightarrow O + O$ $2(O + OH \longrightarrow O_2 + H)$ $2(H + O_3 \longrightarrow OH + O_2)$ Net: $2O_3 \longrightarrow 3O_2$	72.113	6.672e+04	92.5
$D_{2.16}$	$O_{trpt} \longrightarrow O$ $O + OH \longrightarrow O_2 + H$ $H + O_3 \longrightarrow OH + O_2$ Net: $O_3 + O_{trpt} \longrightarrow 2O_2$	48.848	1.664e+05	86.5
$D_{2.17}$	$O_2 + h\nu \longrightarrow O + O(^1D)$ $O(^1D) + N_2 \longrightarrow O + N_2$ $2(O + OH \longrightarrow O_2 + H)$ $2(H + O_3 \longrightarrow OH + O_2)$ Net: $2O_3 \longrightarrow 3O_2$	43.156	6.078e+03	98.5
$D_{2.18}$	$O_3 + h\nu \longrightarrow O(^1D) + O_2$ $O(^1D) + N_2 \longrightarrow O + N_2$ $O \longrightarrow O_{trpt}$ Net: $O_3 \longrightarrow O_2 + O_{trpt}$	18.893	5.321e+03	98.5

Table 4: Pathways producing and destroying O_3 at time$= 1.16$ million years of the model run. The contribution profiles of these pathways are shown in figure 4. The contributions and rates in this table correspond to the height at which the pathways contribute the most to O_3 production and destruction. Our algorithm does not yet have the functionality to automatically order the reactions in a pathway to easily follow the flow of molecules. We ordered the reactions in all the pathways in this table by hand.

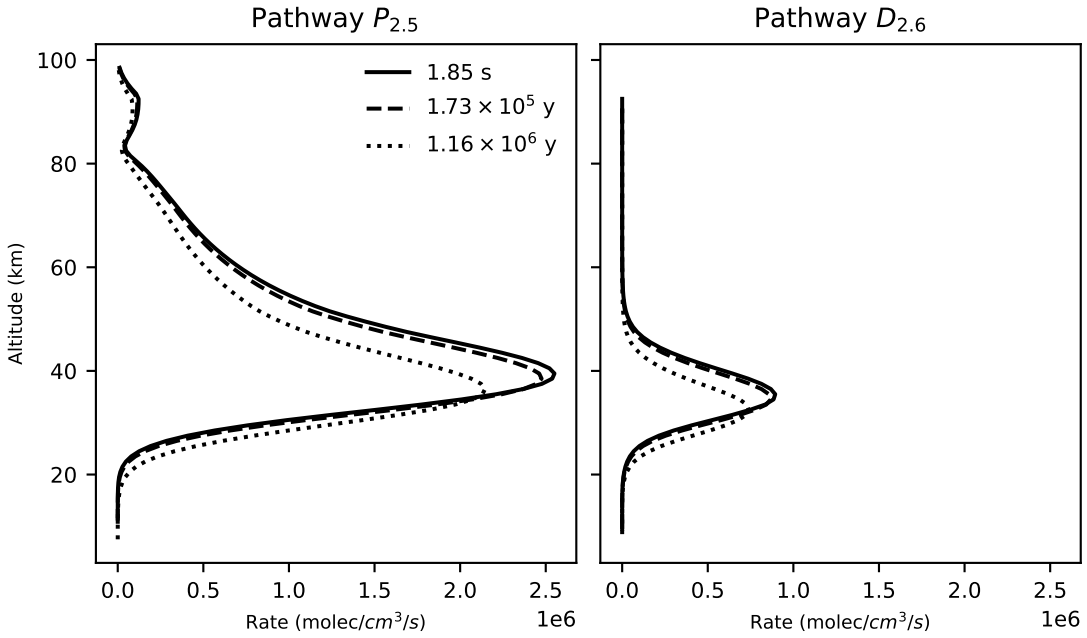

Figure 5. Rate profiles of pathway $P_{2.5}$ (right) and $D_{2.6}$ at three different points in time across our *photochem* model run.

5 Conclusions

In this paper, we described the development of *Chempath*: an open-source pathway analysis program for photochemical models that can automatically construct the most relevant pathways of a reaction system and identify the most important pathways for the production and destruction of a species of interest. We showed how to use *Chempath* in a simple box model and in a one-dimensional photochemical model. *Chempath* identified well-known pathways for O_3 destruction and production in Earth's atmosphere, suggesting that this algorithm can be used to understand chemical mechanisms in photochemical models of less well-known atmospheres, like exoplanet or past atmospheres.

Code availability. A frozen version of the code used in this paper is available at https://doi.org/10.5281/zenodo.13715328. For up-to-date developments see the *Chempath* GitHub repository: https://github.com/DanyIvan/chempath. This repository includes Jupyter notebooks that describe how to run and reproduce the examples presented in this paper.

Author contributions. DGR: Conceptualization, Software, Investigation, Writing - original draft preparation. CG: Supervision, Funding acquisition, Writing - review & editing. ASA: Supervision, Funding acquisition, Writing - review & editing.

Competing interests. We declare that none of the authors has any competing interests.

Acknowledgements. We acknowledge and respect the lək$^{\text{'w}}$əŋən peoples on whose traditional territory the university of Victoria stands and the Songhees, Esquimalt and W̱SÁNEĆ peoples whose historical relationships with the land continue to this day. We thank Ralph Lehmann for answering questions about the pathway analysis program. We thank Aurélien Stolzenbach for pointing to an incorrect reaction in the simple example we used in a previous version of the manuscript. Primary financial support came from Natural Science and Engineering Research Council of Canada (NSERC) Discovery Grants to Colin Goldblatt (RGPIN-2018-05929) and Anne-Sofie Ahm (RGPIN-2022-03912). High-performance computing facilities were provided via a NSERC Research Tools and Infrastructure Grant (RTI-2020-00277).

References

Androulakis, I. P.: New approaches for representing, analyzing and visualizing complex kinetic transformations, Computers & Chemical Engineering, 31, 41–50, https://doi.org/10.1016/j.compchemeng.2006.05.027, 2006.

Arney, G., Domagal-Goldman, S. D., Meadows, V. S., Wolf, E. T., Schwieterman, E., Charnay, B., Claire, M., Hébrard, E., and Trainer, M. G.: The Pale Orange Dot: The Spectrum and Habitability of Hazy Archean Earth, Astrobiology, 16, 873–899, https://doi.org/10.1089/ast.2015.1422, 2016.

Chapman, S.: A Theory of Upper-atmospheric Ozone, Memoirs of the Royal Meteorological Society, Edward Stanford, https://books.google.ca/books?id=Dd0VGwAACAAJ, 1930.

Claire, M. W., Kasting, J. F., Domagal-Goldman, S. D., Stüeken, E. E., Buick, R., and Meadows, V. S.: Modeling the signature of sulfur mass-independent fractionation produced in the Archean atmosphere, Geochimica et Cosmochimica Acta, 141, 365–380, https://doi.org/10.1016/j.gca.2014.06.032, 2014.

Clarke, B. L.: Stoichiometric network analysis, Cell Biophysics, 12, 237–253, https://doi.org/10.1007/bf02918360, 1988.

Fishtik, I., Callaghan, C. A., and Datta, R.: Wiring Diagrams for Complex Reaction Networks, Industrial & Engineering Chemistry Research, 45, 6468–6476, https://doi.org/10.1021/ie050814u, 2006.

Garduno Ruiz, D., Goldblatt, C., and Ahm, A.-S.: Climate shapes the oxygenation of Earth's atmosphere across the Great Oxidation Event, Earth and Planetary Science Letters, 607, 118 071, https://doi.org/10.1016/j.epsl.2023.118071, 2023.

Garduno Ruiz, D., Goldblatt, C., and Ahm, A.: Climate Variability Leads to Multiple Oxygenation Episodes Across the Great Oxidation Event, Geophysical Research Letters, 51, https://doi.org/10.1029/2023gl106694, 2024.

Gebauer, S., Grenfell, J., Stock, J., Lehmann, R., Godolt, M., Paris, P. v., and Rauer, H.: Evolution of Earth-like Extrasolar Planetary Atmospheres: Assessing the Atmospheres and Biospheres of Early Earth Analog Planets with a Coupled Atmosphere Biogeochemical Model, Astrobiology, 17, 27–54, https://doi.org/10.1089/ast.2015.1384, 2017.

Grenfell, J. L., Lehmann, R., Mieth, P., Langematz, U., and Steil, B.: Chemical reaction pathways affecting stratospheric and mesospheric ozone, Journal of Geophysical Research: Atmospheres, 111, https://doi.org/10.1029/2004jd005713, 2006.

Haagen-Smit, A. J.: Chemistry and Physiology of Los Angeles Smog, Industrial & Engineering Chemistry, 44, 1342–1346, https://doi.org/10.1021/ie50510a045, 1952.

Hu, R., Seager, S., and Bains, W.: Photochemistry In Terrestrial Exoplanet Atmospheres. I. Photochemistry Model And Benchmark Cases, The Astrophysical Journal, 761, 166, https://doi.org/10.1088/0004-637x/761/2/166, 2012.

Jacob, D. J.: Introduction to Atmospheric Chemistry, Princeton University Press, ISBN 9780691001852, http://www.jstor.org/stable/j.ctt7t8hg, 1999.

Kasting, J. F.: What caused the rise of atmospheric O_2?, Chemical Geology, 362, 13–25, https://doi.org/10.1016/j.chemgeo.2013.05.039, 2013.

Kasting, J. F. and Donahue, T. M.: The evolution of atmospheric ozone, Journal of Geophysical Research: Oceans, 85, 3255–3263, https://doi.org/10.1029/jc085ic06p03255, 1980.

Kasting, J. F., Liu, S. C., and Donahue, T. M.: Oxygen levels in the prebiological atmosphere, Journal of Geophysical Research: Oceans, 84, 3097–3107, https://doi.org/10.1029/jc084ic06p03097, 1979.

Lary, D. J.: Catalytic destruction of stratospheric ozone, Journal of Geophysical Research: Atmospheres, 102, 21 515–21 526, https://doi.org/10.1029/97jd00912, 1997.

Lehmann, R.: Determination of Dominant Pathways in Chemical Reaction Systems: An Algorithm and Its Application to Stratospheric Chemistry, Journal of Atmospheric Chemistry, 41, 297–314, https://doi.org/10.1023/a:1014927730854, 2002.

Lehmann, R.: An Algorithm for the Determination of All Significant Pathways in Chemical Reaction Systems, Journal of Atmospheric Chemistry, 47, 45–78, https://doi.org/10.1023/b:joch.0000012284.28801.b1, 2004.

Milner, P. C.: The Possible Mechanisms of Complex Reactions Involving Consecutive Steps, Journal of The Electrochemical Society, 111, 228–232, https://doi.org/10.1149/1.2426089, 1964.

Molina, M. J. and Rowland, F. S.: Stratospheric sink for chlorofluoromethanes: chlorine atom-catalysed destruction of ozone, Nature, 249, 810–812, https://doi.org/10.1038/249810a0, 1974.

Schuster, R. and Schuster, S.: Refined algorithm and computer program for calculating all non–negative fluxes admissible in steady states of biochemical reaction systems with or without some flux rates fixed, Bioinformatics, 9, 79–85, https://doi.org/10.1093/bioinformatics/9.1.79, 1993.

Segura, A., Kasting, J. F., Meadows, V., Cohen, M., Scalo, J., Crisp, D., Butler, R. A., and Tinetti, G.: Biosignatures from Earth-Like Planets Around M Dwarfs, Astrobiology, 5, 706–725, https://doi.org/10.1089/ast.2005.5.706, 2005.

Stock, J., Grenfell, J., Lehmann, R., Patzer, A., and Rauer, H.: Chemical pathway analysis of the lower Martian atmosphere: The CO_2 stability problem, Planetary and Space Science, 68, 18–24, https://doi.org/10.1016/j.pss.2011.03.002, 2012a.

Stock, J. W., Boxe, C. S., Lehmann, R., Grenfell, J. L., Patzer, A. B. C., Rauer, H., and Yung, Y. L.: Chemical pathway analysis of the Martian atmosphere: CO_2-formation pathways, Icarus, 219, 13–24, https://doi.org/10.1016/j.icarus.2012.02.010, 2012b.

Stock, J. W., Blaszczak-Boxe, C. S., Lehmann, R., Grenfell, J. L., Patzer, A. B. C., Rauer, H., and Yung, Y. L.: A detailed pathway analysis of the chemical reaction system generating the Martian vertical ozone profile, Icarus, 291, 192–202, https://doi.org/10.1016/j.icarus.2016.12.012, 2017.

Thompson, M. A., Krissansen-Totton, J., Wogan, N., Telus, M., and Fortney, J. J.: The case and context for atmospheric methane as an exoplanet biosignature, Proceedings of the National Academy of Sciences, 119, e2117933 119, https://doi.org/10.1073/pnas.2117933119, 2022.

Tsai, S.-M., Lyons, J. R., Grosheintz, L., Rimmer, P. B., Kitzmann, D., and Heng, K.: VULCAN: An Open-source, Validated Chemical Kinetics Python Code for Exoplanetary Atmospheres, The Astrophysical Journal Supplement Series, 228, 20, https://doi.org/10.3847/1538-4365/228/2/20, 2017.

Turányi, T. and Tomlin, A. S.: Sensitivity and Uncertainty Analyses, pp. 61–144, Springer Berlin Heidelberg, ISBN 978-3-662-44562-4, https://doi.org/10.1007/978-3-662-44562-4_5, 2014.

Verronen, P. T. and Lehmann, R.: Analysis and parameterisation of ionic reactions affecting middle atmospheric HO_x and NO_y during solar proton events, Annales Geophysicae, 31, 909–956, https://doi.org/10.5194/angeo-31-909-2013, 2013.

Verronen, P. T., Santee, M. L., Manney, G. L., Lehmann, R., Salmi, S., and Seppälä, A.: Nitric acid enhancements in the mesosphere during the January 2005 and December 2006 solar proton events, Journal of Geophysical Research: Atmospheres, 116, https://doi.org/10.1029/2011jd016075, 2011.

Volz-Thomas A., B. R.: Scientific Assesment of ozone depletion:1994, chap. Tropospheric Ozone, World Meteorological Organization,, https://csl.noaa.gov/assessments/ozone/1994/, 1994.

Wogan, N.: PhotochemPy: 1-D photochemical model of rocky planet atmospheres, Astrophysics Source Code Library, record ascl:2312.011, 2023.

Wogan, N. F., Catling, D. C., Zahnle, K. J., and Claire, M. W.: Rapid timescale for an oxic transition during the Great Oxidation Event and the instability of low atmospheric O_2, Proceedings of the National Academy of Sciences, 119, e2205618 119, https://doi.org/10.1073/pnas.2205618119, 2022.

Wogan, N. F., Catling, D. C., Zahnle, K. J., and Lupu, R.: Origin-of-life Molecules in the Atmosphere after Big Impacts on the Early Earth, The Planetary Science Journal, 4, 169, https://doi.org/10.3847/psj/aced83, 2023.

Wolfe, G. M., Cantrell, C., Kim, S., III, R. L. M., Karl, T., Harley, P., Turnipseed, A., Zheng, W., Flocke, F., Apel, E. C., Hornbrook, R. S., Hall, S. R., Ullmann, K., Henry, S. B., DiGangi, J. P., Boyle, E. S., Kaser, L., Schnitzhofer, R., Hansel, A., Graus, M., Nakashima, Y., Kajii, Y., Guenther, A., and Keutsch, F. N.: Missing peroxy radical sources within a summertime ponderosa pine forest, Atmospheric Chemistry and Physics, 14, 4715–4732, https://doi.org/10.5194/acp-14-4715-2014, 2014.

Zahnle, K., Claire, M., and Catling, D.: The loss of mass-independent fractionation in sulfur due to a Palaeoproterozoic collapse of atmospheric methane, Geobiology, 4, 271–283, https://doi.org/10.1111/j.1472-4669.2006.00085.x, 2006.