Final author response

Daniel Garduno Ruiz, Colin Goldblatt, Anne-Sofie Ahm

December 2, 2024

Dear Editor:

Here we provide responses to all of the comments made by the reviewer of the manuscript and specify the changes made to the manuscript.

In summary, there are two main changes to the manuscript. First, we updated the simple example we used to explain how the algorithm works. A person pointed us to a wrong reaction in our simple example ($H_2O + hv \longrightarrow OH + O$ which should be $H_2O + hv \longrightarrow OH + H$). We updated the simple example with new reactions to correct this error and to make it simpler. Second, we included a discussion sub-section in section 4 of the manuscript to discuss the anomalous CH_3O_2 pathways in our results and the validation of Chempath.

Responses to comments

- Before the actual determination of pathways, the original algorithm checks the balance of the input data (= output of a chemical model) (Lehmann, 2002, Section 3.1): Are the reaction rates consistent with the concentration changes calculated by the model? This is an essential step to detect (and possibly correct) imbalances, which may arise from numerical inaccuracies (e.g., because of a large time step in the chemical model). Balanced input data are indispensable for the actual pathway formation. The authors do not mention how they solve this problem in their implementation.

  Response: As mentioned in the open discussion, we assume that the reader ensures that mass balance is achieved in their model output. In response to this comment, we updated the code to include a function that displays a warning if any of the species are not balanced by the reactions. We added the following text to section 3 of the manuscript to describe this new function:

  "Before the construction of pathways it is essential to ensure the concentration changes of all species are balanced by the reaction's production and destruction (equation 1). Our implementation does not try to correct for imbalances, but we include a function that displays a warning if mass balance is not fulfilled. The warning is displayed if the unbalance is greater than a relative tolerance of 1×10^{-3} if the concentration change is greater or equal to 1 molecule/cm^3, and to an absolute tolerance of 1×10^{-3} if the concentration change is lower than 1 molecule/cm^3. We use an absolute tolerance for concentration changes lower than 1 molecule/cm^3 because we consider that concentration changes lower than 1×10^{-3} molecules/cm^3 are unimportant. If this warning is displayed, the model output should be checked for potential problems or corrected (see (Lehmann, 2002) for an example of how to do this)."

- How does the algorithm choose a splitting of a pathway into sub-pathways if this splitting is not unique? Does the formulation"we choose the solution that minimizes the most equation 19" mean that a solution of equation 19 is chosen? Although it is true that the optimization problem (19) has a global minimum (line 221), it is not guaranteed that there is a unique solution. In fact, if Equation (17) has multiple solutions, this will also be the case for the related optimization problem (19).

  Response: As mentioned in the open discussion, we chose the first solution that minimizes equation 19 found by Scipy's lsq_linear algorithm. As a response to this comment, we updated the code to include the option to solve equation 17 of the manuscript with the procedure described in section 5.5.1 of Lehmann (2004). We find similar results with both methods. We updated line 217 of the manuscript to describe this new potion:

  "Our implementation includes two options to solve equation 16. The first option uses Scipy's "lsq_linear" function, minimizing the expression:

$$0.5||Ax - b||^2 \text{ with constraints } 0 \leq x < \infty, \tag{1}$$

  where $||x||$ is the norm of x, $A = [[x'_{je}]]$, $x = [w_e]$ 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 5.2.2 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."

- Table 3: Pathways D2,1, P2,3, P2,4, P2,5: In the troposphere CH3O2 has a short chemical lifetime (usually < 1 min). Therefore we would expect that its fate is controlled by local chemistry, not by transport. Nevertheless, the pathways mentioned above involve transport of CH3O2. Is it possible that this transport of CH3O2 is a numerical artifact, resulting from its calculation according to Eq. (46)? There the (possibly small) contribution of transport Ti is calculated as the difference of (possibly larger) chemical terms. Pathway D2,5: The same type of question applies to D2,5: This pathway involves transport of atomic oxygen, which has a short chemical lifetime in the altitude region indicated (around 15.5 km).

  Response: As mentioned in the open discussion, CH_3O_2 and O are present in these pathways because of an incomplete representation of their chemistry in the reaction system we used. We updated the manuscript to include a new section to discuss the presence of these species in these pathways.

- 1 As mentioned above, the authors describe a re-implementation of the (existing) pathway analysis program of Lehmann (2004) in Python. This is said correctly in the body of the text, but it would be good to state this more clearly in the abstract (and, maybe, conclusions) for the hasty reader.

  Response: We modified the abstract to make this clear, including the following sentence

  "*Chempath* is a Python re-implementation of the algorithm developed by Lehmann (2004)."

- 36-37 This leads to a difficult discussion. Available on request may also be considered as a form of open - with the additional advantage that through the personal contact the user can obtain all support needed. Anonymous download of a program is risky if the user does not understand perfectly the functionality and limitations of the program, which may be hard to achieve even if there is a good documentation. I understand that the authors want to provide a reason for their re-programming effort, but I think that the formulation in line 38 is sufficient for that.

  Response: We deleted these lines.

- 64-72 Strictly speaking, ppb is the unit for mixing ratio, not concentration.

  Table 1: Row sij : For readers familiar with chemical systems you might add stoichiometric matrix.

  Table 1: Rows $\hat{r_j}, \hat{p_i}, \hat{d_i}$: The reader might be surprised by these definitions, because deleted pathways have not been mentioned before. An earlier mentioning of deleted pathways would also be beneficial for Section 2.5 (describing the calculation of rates of deleted pathways), which is placed before Section 2.7, where deleted pathways are introduced.

  Response: We changed concentration to mixing ratio. We added stoichiometric matrix to table 1. We also added the following paragraph to mention deleted pathways before table 1:

  "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 enhance the computation time. However, the algorithm includes variables to keep track of the rates of these deleted pathways."

- 81 (and several other lines) The official symbol four hour is h. Please insert blanks between numbers and units

  Response: We changed hr to h and inserted blanks between numbers and units.

- 81 The notation $\sum_{j=1}^{5}[1,0,1,0,0] \cdot [1,05,1.5,5,0.1]$ does not make sense, since there is no j in the terms after the sum. A workaround might consist in writing the sum explicitly (1 ů 1 + 0 ů 0.5 + ...) or as scalar product.

  Response: We updated the text to include the recommended notation.

- 83 I would not mention the reaction system together with in two consecutive time steps (it the same for both time steps). 84 mean reaction rates: in two consecutive time steps or between two consecutive time steps or within one time step?

  Response: We changed the sentence to:

  "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 between n_s species."

- 84 It may happen that the concentration of a species at two points in time t1 and t1 + t is not sufficient as input, but information on the concentration between t1 and t1 + t, e.g. its mean, is also needed: For instance, if you analyse tropospheric chemistry over a full day

(from midnight t1 to the next midnight t1 + t, t = 24 h), you will obtain [OH] zero at t1 and t1 + t (although OH is present during the day), which leads to a wrong estimate of the lifetime of OH in [t1, t1 + t] (needed in line 94).

Response: We included the following sentence to clarify that the two points in time must be chosen in a way that the processes one is interested in understanding are resolved by the time step:

"The consecutive model time steps must be the time steps in which the solver obtains a solution for the system of equations. The algorithm could also be applied in two model times that are not consecutive, but the time step must be small enough to resolve the processes one is interested in understanding."

- 90 $f_j = r_j$ instead of $f_{kinit} = r_j$?

  Response: We included the suggested change.

- Fig. 1: Box 8: Recomputation of variables: Which variables? How?

  Response: This is explained in section 2.9

- 117 ... produced (or consumed) by one pathway is consumed (or produced) by another pathway might be a bit clearer than the present formulation.

  Response: We included the suggested change.

- 123-124 As these operations are carried out for each pathway separately (i.e. g may differ from pathway to pathway), the sentence should be formulated in singular: The multiplicities xij of a new pathway ...The rate of the new pathway is multiplied...

  Response: We included the suggested change

- 125, 130 (and several other lines) Throughout the manuscript the authors use identical denotations for elements of a matrix (or vector) and the whole matrix (or vector). Although the reader may guess what is meant, I recommend a stricter notation, especially since the number of indices does not always indicate the dimension of the object, e.g. xjn in line 130 denotes a vector.

  Response: We updated our notation, using two square brackets around variables that denote matrices (for example $[[s_{ij}]]$) and one around variables that denote vectors (for example $[r_j]$).

- 134, 197, 256, 287: Why similar to instead of has the following form?

  Response: We replaced similar with has the following form.

- 143 in the previous step from the previous step (or omit completely)?

  Response: We omitted these words.

- 153 Eq. (10): $\sum_k \rightarrow \sum_e$ (3 times)

  Response: We included the correction

- 179-181 Lines 176-181 describe the deletion of pathways that have been used (i.e. connected to other pathways). This step does not involve any consideration of deletion due to rates < fmin. Why are $\hat{r_j}$, $\hat{p_i}$, $\hat{d_i}$ mentioned nevertheless? (And what exactly does this case refer to?)

Response: We omitted the sentence

- 218 minimizing the equation → minimizing the expression

  Response: We included the suggested change

- 219 $x \leq \infty \rightarrow x < \infty$

  Response: We made the correction

- 220 A comparison with the variables in (17) indicates that: $A = ((x'je))j = 1, ..., n_r, e = 1, ..., n_e$ $b = (xjc)j = 1, ..., n$

  Response: Thanks for noticing this error. We included the correction.

- 277 0.007 → 0.0073 (in order to be consistent with Eq. (30))

  Response: We used the same number of decimals in the corrected simple example.

- 304 this contribution → the contribution

  Response: We included the correction

- 306 Eq. (42): Sb → Si (2 times)

  Response: Thanks for noticing this error. We included the correction.

- 308-309 It seems that For example, ... (43) should be placed directly after Eq. (42), not after text about deleted pathways.

  Response: We moved the text about deleted pathways to the end of the section.

- 309 Eq. (43): This equation involves element-wise multiplication of two vectors (not scalar product). Perhaps this should be said explicitly.

  Response: We included the following text after equation 42 to clarify this: Expression 42 involves the element-wise multiplication of two vectors.

- Table 2: HV → $h\nu$ (several times)

  Response: We included the correction

- 353-354 setting it as a fraction of the rate of production of the species...: Here production refers to the total production by all reactions? If so, you might emphasize this, in order to avoid confusion with the production by pathways in Table 3. In general, this way of choosing fmin may still require further trial and error: If the species of interest (Si) is involved in zero cycles with large rates (e.g. O3 O + O2), then the rate of the (total) production of Si will be much larger than the net production or destruction, which shall be explained by pathways. This may have the consequence that an originally chosen fraction (i.e. fmin) may turn out to be too large and must be reduced.

  354 Shouldnt destruction rates also be taken into account, e.g. for species like CH4 that is only destroyed in Earths atmosphere?

  Response: We modified the text to clarify that we refer to the total production by all reactions, and that setting $f_{\min}$ as a fraction of this production might still require further

trial and error. We agree that destruction rates should be taken into account for species like CH4:

"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."

- 360 How many of the 1281 reactions have a rate > fmin, so that they actually take part in the formation of pathways?

  Response: The number of reactions with rate $> f_{min}$ varies with altitude, and ranges from 78 to 132. We included the following text:

  "The number of reactions with rate $> f_{min}$ varies with altitude, and ranges from 78 to 132. "

- 364 Please explain to the reader the idea behind the reduction of the O2 surface flux. Which processes in the model lead to a removal of O2 (eventually balancing the source by the surface flux)? What is the time scale of these processes?

  Response: We included the following text to explain this:

  "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 atmosphere is $\sim$2 million years (**?**))."

- **Response**:

- 365 every time step: How long is one time step?

  Response: We included the following text to answer this question:

  "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. We only output the model results when the simulation time is greater than $10^{11}s$."

- 368-370 ... if we want to know what are the chemical mechanisms that contribute to this O3 loss, we need to use the pathway analysis program. We apply Chempath to the photochem model output to gain insight into the chemical reaction chains that destroy O3 in this model run.: This formulation sounds as if the O3 destruction pathways directly explain the O3 decrease (probably of a few ppb / million years $\sim 10^6$ ppb/y) occurring in the model run after the reduction of the O2 surface flux. However, this is not the case. As Ox (= O3 + O + O(1D)) has a chemical lifetime of 1 year below 100 km (and much less in the middle atmosphere) (e.g., G. Brasseur and S. Solomon: Aeronomy of the Middle Atmosphere, Springer, Dordrecht, 2005: Fig. 5.3), it will be close to equilibrium in your million-year long model run, i.e. the concentration is determined by the production rate (strongly dependent on the changing [O2]) and the time scale of destruction. By the way, these arguments may serve as motivation for showing production and destruction pathways later on (Table 3).

Response: Thanks for sharing these arguments. We agree with your analysis. We modified this paragraph to clarify that both ozone production and destruction pathways are important to understand the O3 concentration change:

"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."

- 372-373 vertical transport production and destruction e.g., supply and removal by vertical transport?

  Response: Thanks for noticing this error. We included the correction.

- Figure 2 (upper left panel): m means milli = 10^3

  Response: We changed my to million years.

- 374 Eq. (45): Li ⇒ Di

  Response: Thanks for noticing this error. We included the correction.

- 376 Does production by rainout mean that evaporation of rain and release of trace gases from the liquid phase to the gas phase is included in the model? According to Eq. (47) this seems not to be the case.

  Response: Thanks for noticing this error. We updated the text to only say destruction by rainout.

- 381 You might include di/dt in the list of values obtained from the model.

  Response: We already mentioned that we obtained the number density and the model time at two consecutive model times to calculate this expression.

- 384 Eq. (47): It seems that supplies and removes should be interchanged.

  Response: Thanks for noticing this error. We included the correction.

- 388-389 CH4 is not photochemically produced in Earths atmosphere (G. Brasseur and S. Solomon: Aeronomy of the Middle Atmosphere, Springer, Dordrecht, 2005: p. 296). If this is true also in your model, then fmin as defined in lines 388-389 will be zero. Please clarify.

  Response: Our reaction system also includes a pseudo-reaction for CH_4 supply from transport. For this reason, f_{min} is different from zero. We modified this sentence to make this clear:

  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

- Figure 3: It would be nice to use similar colours (or additional symbols) to indicate pathways of the same family (HOx, NOx etc.).

  Response: We included additional symbols in the legend of the figure to group the pathways into five categories: Oxidation, Chapman-like, photolysis, HOx, and NOx pathways. We also updated the figure caption to explain the additional symbols.

- Figure 3, Table 3: It might be more logical to present production pathways before destruction pathways.

  Response: We changed the order of the pathways

- Figure 3, Table 3 and related text: All pathways have the same first index 2. Therefore it might be omitted in the manuscript (probably it results from the fact that O3 is species no. 2 in the model).

  Response: We omitted the index

- Table 3: O^1D *Rightarrow* $O(^1D)$ (several times)

  Response: We made the correction

- Table 3: It seems that the algorithm does not attempt to order the reactions in a way that the flux of molecules can be easily followed by the user. I suggest to do that by hand" (and clearly state that it was done in this way); e.g. pathway D(2)2 might be reordered in the following way:

  Response: We ordered the reactions by hand and included the following sentence in the caption of table 3: Our algorithm does not yet have the functionality to automatically order the reactions to easily follow the flow of molecules. We ordered the reactions in all the pathways by hand.

- Figure 4: Upper panel: Delete s" in the unit of the number density

  Response: We included the correction

- Figure 4: Caption: middle panel and bottom panel does not coincide with the figure above.

  Response: We updated the figure to coincide with the caption description.

- Figure 4: Caption: tables 3 $\Rightarrow$ table 3"

  Response: We included the correction

References

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.