Review of

"Chempath 1.0: An open-source pathway analysis program for photochemical models" by D. Garduño Ruiz et al.

(Numbers refer to line numbers.)

General

The authors describe an implementation of the (existing) pathway analysis program of Lehmann (2004) in Python. This may please potential users who prefer to run a code in the younger programming language Python over Fortran.

Questions

- 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.

221-222 How does the algorithm choose a splitting of a pathway into subpathways if this splitting is not unique?
Does the formulation "we choose the solution that minimizes the most equation 19" mean that a solution of

$$\min\left\{\frac{1}{2}\cdot||Ax-b||^2 \mid x\geq 0\right\} \qquad \hat{=}\ (19)$$

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).

- Table 3: Pathways $D_{2,1}$, $P_{2,3}$, $P_{2,4}$, $P_{2,5}$: In the troposphere CH_3O_2 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 CH_3O_2. Is it possible that this transport of CH_3O_2 is a numerical artefact,

resulting from its calculation according to Eq. (46)? There the (possibly small) contribution of transport T_i is calculated as the difference of (possibly larger) chemical terms.
Pathway $D_{2,5}$: The same type of question applies to $D_{2,5}$: This pathway involves transport of atomic oxygen, which has a short chemical lifetime in the altitude region indicated (around 15.5 km).

Details

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.

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.

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

- Table 1: Row "s_{ij}": For readers familiar with chemical systems you might add "stoichiometric matrix".

- Table 1: Rows "$\tilde{r}_j$, $\tilde{p}_i$, $\tilde{d}_i$": 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.

81 (and several other lines) The official symbol four "hour" is "h".
Please insert 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 ("$1\cdot 1+0\cdot 0.5+...$") or as scalar product.

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"?

84 It may happen that the concentration of a species at two points in time t_1 and $t_1 + \Delta t$ is not sufficient as input, but information on the concentration between t_1 and $t_1 + \Delta t$, e.g. its mean, is also needed: For instance, if you analyse tropospheric chemistry over a full day (from midnight t_1 to the next midnight $t_1 + \Delta t$, $\Delta t = 24$ h), you will obtain [OH] $\approx$ zero at t_1 and $t_1 + \Delta t$ (although OH is present during the day), which leads to a wrong estimate of the lifetime of OH in $[t_1, t_1 + \Delta t]$ (needed in line 94).

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

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

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

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 x_{ij} of a new pathway ...The rate of the new pathway is multiplied..."

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. "x_{jn}" in line 130 denotes a vector.

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

143 "in the previous step" $\Rightarrow$ "from the previous step" (or omit completely)?

153 Eq. (10): "$\sum_k$" $\Rightarrow$ "$\sum_e$" (3 times)

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 $< f_{min}$. Why are $\tilde{r}_j, \tilde{p}_i, \tilde{d}_i$ mentioned nevertheless? (And what exactly does "this case" refer to?)

218 "minimizing the equation" $\Rightarrow$ "minimizing the expression"

219 "$x \leq \infty$" $\Rightarrow$ "$x < \infty$"

220 A comparison with the variables in (17) indicates that:

$$A = ((x'_{je}))_{j=1,...,n_r, e=1,...,n_e}$$
$$b = (x_{jc})_{j=1,...,n_r}$$

277 $0.007 \Rightarrow 0.0073$ (in order to be consistent with Eq. (30))

304 "this contribution" ⇒ "the contribution"

306 Eq. (42): $S_b \Rightarrow S_i$ (2 times)

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

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

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

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 f_{min} may still require further "trial and error": If the species of interest (S_i) is involved in zero cycles with large rates (e.g. $O_3 \leftrightarrow O + O_2$), then the rate of the (total) production of S_i 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. f_{min}) may turn out to be too large and must be reduced.

354 Shouldn't destruction rates also be taken into account, e.g. for species like CH_4 that is only destroyed in Earth's atmosphere?

360 How many of the 1281 reactions have a rate $> f_{min}$, so that they actually take part in the formation of pathways?

364 Please explain to the reader the idea behind the reduction of the O_2 surface flux.
Which processes in the model lead to a removal of O_2 (eventually balancing the source by the surface flux)? What is the time scale of these processes?

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

368-370 "... if we want to know what are the chemical mechanisms that contribute to this O_3 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 O_3 in this model run.":

This formulation sounds as if the O_3 destruction pathways directly explain the O_3 decrease (probably of a few ppb / million years $\sim 10^{-6}$ ppb/y) occurring in the model run after the reduction of the O_2 surface flux. However, this is not the case. As O_x (= O_3 + O + $O(^1D)$) has a chemical lifetime of $\leq$ 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 [O_2]) 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).

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

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

374 Eq. (45): "L_i ⇒ "D_i"

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.

381 You might include $\frac{d\rho_i}{dt}$ in the list of values obtained from the model.

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

388-389 CH_4 is not photochemically produced in Earth's 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 f_{min} as defined in lines 388-389 will be zero. Please clarify.

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

- Figure 3, Table 3: It might be more logical to present production pathways before destruction 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 O_3 is species no. 2 in the model).

- Table 3: "O_1D" ⇒ "$O(^1D)$" (several times)

- 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:

$$
\begin{aligned}
2\ (O_3 + \mathrm{hv} &\longrightarrow O + O_2\ ) \\
2\ (CH_4 + OH &\longrightarrow CH_3 + H_2\ O) \\
2\ (CH_3 + O_2 + M &\longrightarrow CH_3O_2 + M) \\
2\ (CH_3O_2 + O &\longrightarrow CH_3O + O_2\ ) \\
2\ (CH_3O + O_2 &\longrightarrow H_2CO + HO_2\ ) \\
2\ (H_2CO + \mathrm{hv} &\longrightarrow CO + H_2\ ) \\
HO_2 + HO_2 &\longrightarrow H_2O_2 + O_2 \\
H_2O_2 + h\nu &\longrightarrow OH + OH
\end{aligned}
$$

- Figure 4: Upper panel: Delete "s" in the unit of the number density
- Figure 4: Caption: "middle panel" and "bottom panel" does not coincide with the figure above.
- Figure 4: Caption: "tables 3" $\Rightarrow$ "table 3"