

Geosci. Model Dev. Discuss., referee comment RC2
<https://doi.org/10.5194/gmd-2021-159-RC2>, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on gmd-2021-159

James Orr (Referee)

Referee comment on "PyCO2SYS v1.8: marine carbonate system calculations in Python"
by Matthew P. Humphreys et al., Geosci. Model Dev. Discuss.,
<https://doi.org/10.5194/gmd-2021-159-RC2>, 2021

GENERAL COMMENTS

Humphreys et al. have provided an excellent manuscript on a new package (PyCO2SYS), entirely written in Python, that compute marine CO₂ system variables, buffer factors, and uncertainties. Although similar packages exist in other languages, this is the first time such a wide-based package has been available in Python itself. The package is based on CO2SYS-MATLAB, hence its name, but is not just a simple translation, containing many novelties and the most recent updates. Furthermore, the coauthors include the primary expert that developed the original version of CO2SYS (for MS-DOS), the expert that developed the subsequent Excel and MATLAB versions, and another who added major new features to the MATLAB version. This wide expertise combined with the first authors' strong lead in adding novel features in Python lends great confidence to this initiative. PyCO2SYS will be a welcome addition to the community, especially for young scientists, many of whom are adopting Python as a primary coding language

The manuscript is generally well written and requires only minor revisions. My strongest complaint is that the section on uncertainty propagation contains an equation that is incorrect and that the related Appendix G should be deleted because it too strongly resembles previous work, both rather easy fixes. I would also like to see an evaluation of the computational efficiency of PyCO2SYS along the lines suggested by the other reviewer, Guy Munhoven. My minor comments are listed below.

ABSTRACT

line 8: The authors state "but no fully functional and rigorously validated tool was previously available for Python". They might want to say "written in Python" since seacarb is available for use in Python (through Rmagic or rpy2).

line 12: "every solute is included"? "every" is a dangerous word.

INTRODUCTION

line 32: what the authors refer to as "metadata" is not. Metadata is data that provides information about other data.

line 64: It would be useful for readers if the authors could mention that "molality" (number of moles of solute per kilogram of the mixture). I think molality is more specific than the definition of "substance content" that is given in the IUPAC (1997) Goldbook, where the latter is defined as "Amount-of-substance of a component divided by the mass of the system." Furthermore, the authors' use of "content" as shorthand for "substance content" may escape all readers that have not read the Introduction.

line 66: "in-water pressure"? Do the authors mean the "hydrostatic pressure"?

line 83: The authors should clarify what is meant by "total salt contents". It would be useful to see the full list in parentheses.

line 131-132: The phrase "because these are all directly proportional to each other" may mislead readers because the relationships between the terms mentioned depend on temperature, salinity, and atmospheric pressure. Thus it would be useful to add at the end of the sentence, "at a given temperature, salinity, and atmospheric pressure".

line 159, 162: Use of "every" may be misunderstood by readers. PyCO2SYS does not account for alkalinity from organic acids except for advanced users that have the ability to specify the appropriate parameters for their regional conditions. And even they will not be able to precisely add in all the missing components.

3.1.3 pH SCALE CONVERSIONS

line 197: "negligible"? Please provide the readers with approximate quantitative estimates of what the differences in pH from this error actually are at typical seawater pH when the user selects the free and seawater pH scales.

3.3.2 GAS CONSTANT

line 276: Please be more quantitative. Some might take the qualitative "minor" to be on the order of a percent or more, but the actual effect on calculated variables from changing the 4th decimal point of R must be very much smaller.

3.3.4 BUFFER FACTORS

lines 296-298: The typographical errors in the equations by Egleston et al. were corrected well before 2018. They were first corrected by Orr (2011), who provided a function in seacarb `\textit{buffesm}`, released also in 2011, that calculated the Egleston et al. buffer factors but with the corrected equations. Alvarez et al. (2014) later provided equations the same corrections. Those earlier fixes were mentioned in section 2.2 of Orr et al (2018).

lines 298-299: The authors are right to mention that the Egleston et al. equations "do not include the effect of species beyond the carbonate, borate, and water contributions to total alkalinity." However, these equations were extended by Orr et al. (2018) to also include contributions to total alkalinity from phosphoric and silicic acid systems. See their Appendix B. They also updated their `\textit{buffesm}` routine, available in both seacarb and mocsy, to include these extensions.

lines 304-306: Another reviewer, Guy Munhoven, seems to suggest adding an option for computing the analytical solution for the Revelle factor. This solution, with the corrected equations from Egleston et al. (2010) and the Orr et al (2018) extensions to include the phosphoric and silicic acid contributions to total alkalinity is available in the `\textit{buffesm}` function of both seacarb and mocsy.

3.4 NO-SOLVE MODES

line 318: It would be useful to mention what assumption is made in PyCO2SYS about atmospheric pressure when converting to and from xCO2 ($P_{atm} = 1 \text{ atm}$). This may be a large correction, particularly in the very high latitudes, and CO2SYS does not account for it.

3.6 UNCERTAINTY PROPAGATION

line 332: "This feature" seems ambiguous. Do the authors mean "Propagating the uncertainty" or "Calculating the derivative"?

line 335: what is the difference between "totally independent" and just "independent"? Or

perhaps one could just say "differs".

line 336: I did not realize that "automatic differentiation, in general, was so computationally inefficient. Over what parts of PyCO2SYS is automatic differentiation inefficient? It would be very nice if the authors could provide quantitative statements regarding this statement.

Equation (23): I am a little surprised by this scaling, which would seem to be applied almost always since the median of "a" would almost never be exactly zero. What about very small numbers, where the median was close to but not quite zero? This would make the step size much too small. The choice of Δa seems arbitrary. Have the authors tried to seek out optimal values by comparing their partial derivatives calculated by finite differences with those that they could calculate from the analytical solutions based on the buffer factors from Egleston et al.?

This scaling by median(a) also means that the uncertainties for a particular line of input data will depend on the values of the other lines in the data set. This makes calculations less reproducible.

line 355: Equation 24 is incorrect. It is missing the partial derivatives, something which might be suspected because the units of the left-side term with differs with each right-hand side term. See equation (1) in Orr et al. (2018).

line 358: The authors state, "The user can therefore build". This phrase is not clear. Does PyCO2SYS actually offer an easy way for the user to compute a set of covarying argument uncertainties? This is a tricky business because correlations between the *uncertainties* of various input variables are hard to ascertain. Does PyCO2SYS offer something original in this regard, beyond what is now available in other packages?

4.1.2 BUFFER FACTORS

lines 380-390: Just after this subsection it would be useful to insert another subsection where the partial derivatives calculated by finite differences would be compared to those that could be calculated from the buffer factors of Egleston et al. (either the automatic derivatives or the analytical solutions). See Orr et al. (2018, equations B9 and B13), which details the conversion between buffer factors and partial derivatives.

4.1.3 UNCERTAINTY PROPAGATION SIMULATIONS

The agreement of 3% is not bad, but I think it could be much better if the authors would use more iterations. Based on previous such comparison (Orr et al., 2018, Fig. 3), I would suggest that the authors try to repeat this comparison but use $n=10^6$ instead of $n=10^4$.

line 441: I struggled to understand "from one of these, [H+], and K1 and K2 equilibrium constants using the equations in Appendix C". Please try to clarify. I would suggest (if it is what is actually meant by the authors) to change "these" to "these three variables". Also, I would recommend to be more specific by referring to the 3 relevant subsections of Appendix C, rather than the whole Appendix, which is quite long.

lines 451-453: Differences of 20% between this study's estimates and those from Orr et al. (2018) for partial derivatives of [H+] do not seem negligible, unlike what is suggested by the authors. I do not believe that these differences are due to the "pH scale conversion simplification (Sect. 3.1.3)" as suggested by the authors. Indeed, those derivatives from Orr et al. (2018) were calculated for the total [H+] scale, for which the authors say this issue presents no problem. The authors also suggest that the 20% difference could come from "other differences in pH-solving from AT and TC (Sect. 3.1)", but I still don't understand how. In Orr et al. (2018), we got the same results for the partial derivatives in multiple packages (seacarb, CO2SYS, mocsy) as well as with different approaches in the same package, including finite differences, automatic derivation and analytical (symbolic) solutions. Furthermore, because [H+] and pCO₂ are nearly linearly related, it does not make sense to me that the authors find a 20% difference for partial derivatives of [H+] but not for partial derivatives of pCO₂. Did the authors use the free pH scale for the calculations of the partial derivatives of [H+]?

5.3 PRESSURE CORRECTIONS FOR pCO₂

lines 529-531: The authors might mention that these pressure corrections have already been implemented in two other packages, seacarb and mocsy, see Orr et al. (2015, Tables 9 and 10), as well as Orr and Epitalon (2015).

APPENDIX B:

line 627: The authors state, "Undissociated H₂CO₃ is considered negligible and thus not modelled (Zeebe and Wolf-Gladrow, 2001)." Actually, H₂CO₃ is modelled, but not separately. The undissociated carbonic acid [H₂CO₃] is indeed a small fraction of the total, but in the CO₂ system equations, it is actually taken into account together with the dissolved CO₂ (molecular or aqueous CO₂). Please refer to the best practices guides (Dickson et al., 2007; Dickson, 2010). For instance, in Dickson et al. (2007) Chapter 2 states, "It is usual to combine the concentrations of CO₂(aq) and H₂CO₃(aq) and to express this sum as the concentration of a hypothetical species, CO₂(aq)." Thus in all equations where the authors have used CO₂(aq), that should be replaced by CO₂*(aq) or simply CO₂*.

APPENDIX C:

line 689: "in some cases, simpler alternatives are used instead". This is vague. Can the authors be more specific as to exactly where simplifications are used (and where they come from). These details could be given when the simplifications themselves are presented.

It might also be appropriate to mention the best-practice guides from Dickson et al. (2007) and Dickson (2010) in addition to Zeebe and Wolf-Gladrow (2001).

C1.2 KNOWN pCO₂, xCO₂ or [CO₂(aq)]

line 704: I have not seen the name "vapour pressure factor" before. Does that even have to be named? If so, why not a simpler term, and one that is more relatable, such as "humidity correction" or "wet-air correction".

line 706: Can the authors please clarify what is meant by "(assumed to be 1 atm)"? Do users have a choice for P (an input argument) or is the value of P=1 atm hard wired into the code. As mentioned by Orr et al. (2017), "the average surface atmospheric pressure between 60 and 30S is 3% lower than the global mean, thus reducing surface-ocean pCO₂ by 10 uatm". This seems like a big correction, and it would be nice for users to know if PyCO₂SYs provides for it.

APPENDIX F:

Please be more quantitative? What does "mostly negligible" mean? For example, to what decimal place is the Revelle factor typically affected by these errors?

APPENDIX G: PROPAGATION OF CO-VARYING UNCERTAINTIES

lines 825-837: This appendix too greatly resembles Appendix A from Orr et al. (2018). It has the same equations and similar text. It is not original and should be deleted. The appropriate citation could be given in the text. Overall, the authors do not do justice to the complicated subject of propagating uncertainties when uncertainties of the different input parameters are correlated, and just adding this appendix is not a good solution. Referring to previous work and mentioning what new features or facility that PyCO₂SYs offers would seem a better option.

GLOBAL CHANGE

For consistency with the best-practice guides and with the A_T , I would recommend to change T_C symbol to C_T .

PUNCTUATION/LANGUAGE (minor and few)

* The colon is used incorrectly in lines 29, 90, 146, 215, 223, 233, 248, 769, 774, 792, 800, 817. This misuse has been referred to as the most common colon error ever. In all these cases, the colon should just be deleted.

* line 167: "too" could be deleted.

REFERENCES

Álvarez, M., Sanleón-Bartolomé, H., Tanhua, T., Mintrop, L., Luchetta, A., Cantoni, C., Schroeder, K., and Civitarese, G.: The CO₂ system in the Mediterranean Sea: a basin wide perspective, *Ocean Sci.*, 10, 69--92, <https://doi.org/10.5194/os-10-69-2014>, 2014.

Dickson, A. G.: The carbon dioxide system in seawater: equilibrium chemistry and measurements, in: Guide to best practices for ocean acidification research and data reporting, edited by: Riebesell, U., Fabry, V. J., Hansson, L., and Gattuso, J.-P., Publications Office of the European Union, 17–40, 2010.

Dickson, A. G., Sabine, C. L., and Christian, J. R.: Guide to best practices for ocean CO₂ measurements, PICES Special Publication 3, 191 pp., <http://aquaticcommons.org/1443/>, 2007.

Munhoven, G.: Mathematics of the total alkalinity-pH equation - pathway to robust and universal solution algorithms: the SolveSAPHE package v1.0.1, *Geoscientific Model Development*, 6, 1367--1388, <https://doi.org/10.5194/gmd-6-1367-2013>, 2013.

Orr, J. C., Recent and future changes in ocean carbonate chemistry, In: Gattuso, J.-P. and Hansson, L. (eds.), *Ocean Acidification*, Oxford Univ. Press, 41--66, 2011.

Orr, J. C. and Epitalon, J.-M.: Improved routines to model the ocean carbonate system:

mocsy 2.0, *Geoscientific Model Development*, 8, 485-499, <https://doi.org/10.5194/gmd-8-485-2015>, 2015.

Orr, J. C., Epitalon, J.-M., and Gattuso, J.-P.: Comparison of ten packages that compute ocean carbonate chemistry, *Biogeosciences*, 12, 1483--1510, <https://doi.org/10.5194/bg-12-1483-2015>, 2015.

Orr, J. C., Najjar, R. G., Aumont, O., Bopp, L., Bullister, J. L., Danabasoglu, G., Doney, S. C., Dunne, J. P., Dutay, J.-C., Graven, H., Griffies, S. M., John, J. G., Joos, F., Levin, I., Lindsay, K., Matear, R. J., McKinley, G. A., Mouchet, A., Oschlies, A., Romanou, A., Schlitzer, R., Tagliabue, A., Tanhua, T., and Yool, A.: Biogeochemical protocols and diagnostics for the CMIP6 Ocean Model Intercomparison Project (OMIP), *Geosci. Model Dev.*, 10, 2169–2199, <https://doi.org/10.5194/gmd-10-2169-2017>, 2017.

Soli, A.L., Byrne, R.H., 2002. CO₂ system hydration and dehydration kinetics and the equilibrium CO₂ / H₂CO₃ ratio in aqueous NaCl solution. *Mar. Chem.* 78, 65 – 73, [https://doi.org/10.1016/S0304-4203\(02\)00010-5](https://doi.org/10.1016/S0304-4203(02)00010-5), 2002.