Comment on cp-2021-3
Anonymous Referee #1

Referee comment on "Stalagmite carbon isotopes suggest deglacial increase in soil respiration in western Europe driven by temperature change" by Franziska A. Lechleitner et al., Clim. Past Discuss., https://doi.org/10.5194/cp-2021-3-RC1, 2021

Review for:

Stalagmite carbon isotopes suggest deglacial increase in soil respiration in Western Europe driven by temperature change

This paper presents a useful modeling approach for reconstructing soil respiration using stalagmite carbon isotopes and proxy constraints on prior calcite precipitation and bedrock dissolution effects. Carbon isotopes have been a particularly messy avenue in speleothem science due to the complex interplay of these effects (and others) and the paper represents an exciting effort toward rigorously disentangling this mess. The model is an important step toward understanding how soil respiration changes with climate and, given the breadth of data available in the SISAL database, could be quickly applied on a large scale (assuming it can be appropriately constrained). But the modeling approach carries some critical (and likely invalid) assumptions that need to be addressed. This paper can be of sufficient interest for Climate of the Past and I think my points can be addressed with major revisions. I commend the authors for their coupled proxy-model approach and hope my feedback is useful as they refine their work.

My expertise most closely aligns with the modeling work, so I focus my feedback on this part of the manuscript. I can’t speak much to the analytical methods. Below, I’ve divided my feedback into points about the main modeling approach (outlining my own confusion) and line-by-line items. My biggest concerns are that the modern calibration of the soil respired end member seems invalid (or at least should use modern CO2 levels), and that it is not a safe assumption that the mixing end members are time-invariant.

Main modeling points

My understanding of the modeling approach involves three steps: (1) use modern data to derive a relationship between atmospheric CO2 and soil-respired end members; (2) assuming this mixing relationship holds through time prescribe the full range of soil CO2 and d13C possibilities to solve for proxy data with CaveCalc; (3) using the model output, select and analyze the combination of input parameters that yield results closely in line with the measured data. Below, I dive into my concerns on steps 1 and 2 in more detail...
and include one note on step 3.

**Step 1: calibrating a soil respiration end member**

This calibration exercise (outlined in Figure 2) carries four big assumptions that I think need to be addressed. The biggest has to do with using modern data to calibrate a pre-industrial mixing curve (see point 2).

- First, cave-monitored CO2 and d13C are used to calibrate a soil CO2 mixing line. This assumes that the mixing of atmosphere and respired CO2 in the cave falls the exact functional form of mixing in the soil (put otherwise, it assumes that the bedrock contribution to cave carbon is the same as soil carbon and that there are no other carbon fluxes distinguishing cave from soil). The authors concede this does not hold true in the winter, but could more be said about this assumption in the months of April-November (when I think the monitoring data are used)? Could changes in hydrology, or cave vs soil temperatures, or other things violate this assumption? Is soil CO2 assumed to reflect the soil-column integrated conditions? Is it a problem that this assumption breaks down seasonally if calcite deposition is seasonally biased? **I think, at the least, it must be written that this assumption is made** (right now the link between cave conditions and soil conditions is a bit vague to me).

- Second, the model is calibrated to pre-industrial CO2 levels even though the data are taken in modern conditions (when CO2 is very well-constrained and much higher!) I think this is done so the same calibration end members can be applied throughout the Holocene (more on this in the “Step 2” section). The forest data are ignored because they might be influenced by “turbulence and advection effects” (Line 196), but they would probably fit really nicely on a mixing line that reaches to a modern CO2 end member (with higher CO2)! In fact, all of the data would likely fit better on such a line (given that the monitoring data residuals to mixing line 1 are mostly above the line when CO2 > 1,000 ppm). This makes sense, because the monitored data are mixing with the modern atmosphere, not the pre-industrial atmosphere. (Correcting for the Suess effect only corrects for pre-industrial d13C, it does not account for the difference in CO2 between then and now). **I highly encourage the authors to use a calibration to a modern end member.** Otherwise a much more rigorous justification for the pre-industrial end member is needed.

- Third, I don’t know how the soil-respired end member is defined as 7800ppm and -22.9‰. **I imagine that it’s an extrapolation of mixing line 1, but why not extrapolate some number other than 7800?** Is there some assumption that I’m missing?

- Fourth, **I think this modeling approach assumes that boundary layer CO2 concentration and d13C (the stuff that diffuses into the soil) also falls on the same mixing curve.** This should be stated since at least two things relevant to this study might violate this assumption. First, a shift from no canopy in the last glacial to a canopy when forests appear might lead to a “canopy effect” whereby d13C gets lower than expected for a given pCO2 due to recycling. Second, “turbulence and advection effects” (line 196) that appear to matter during the daytime (probably when photosynthesis is happening) can overprint the simple mixing relationship and propagate down to the soil respired end member. These effects might well be small, but I think the assumption should at least be recognized.
Step 2: assuming this holds through time

This analysis assumes that the mixing slope between soil respired d13C and atmospheric d13C is constant through time. This is not a good assumption because a lot of factors that matter on decadal or longer timescales (i.e. factors that are not captured by the short calibration) violate it by changing end member CO2 but not d13C (or vice versa). For example, if CO2 increases (like it did from LG to EH), then the isotopic composition of CO2 must decrease to keep the end member on the curve, but we know very well that this assumption is violated on paleoclimate timescales (e.g. Schmitt et al., 2012; Science; Figure 1). Similarly, if soil respiration decreases (thus decreasing soil-respired CO2) then the d13C of soil respiration must increase to stay on the mixing curve. I’m not sure if there’s a defensible mechanism for this, although it might occur by coincidence if water stress increases vegetation d13C (thus soil-respired d13C) while decreasing soil respiration. Either way, I am not aware of any mechanistic reason why the end members of the mixing relationship should, themselves, vary along a mixing curve. The end-members, just like the average soil CO2 values that reflect their mixing, should vary over time.

I think the modeling can still be performed if some significant changes are made (these are just suggestions and other options can be valid too!).

- Consider using the actual paleoclimate constraints on pCO2 and d13C of CO2 to parameterize the atmospheric end member.
- Instead of calibrating the soil-respired end member with modern data, be clear that constraints on this term are not great but define reasonable ranges and run sensitivity tests. Allow the soil respired d13C and CO2 to vary with time. Or consider forcing the model with different scenarios as a sensitivity test (i.e. low vegetation d13C, high vegetation d13C and variable, or decreasing, or increasing d13C). Given the strong evidence for substantial changes in vegetation, it is helpful (maybe necessary) to rule this out as the main factor affecting d13C_spel. Consider holding soil respiration constant while letting d13C-respired vary; one might find that the variability would have to be too high to be explained by changes in C3 vegetation or water stress alone (see Kohn, 2010; PNAS).

Step 3: filtering for best model results

More discussion / sensitivity analysis should be done here. Were other options for finding the “best fit” considered? How does changing the thresholds for carbon and calcium isotope data affect the results? What happens if one uses a broader DCF threshold? If these decisions affect the results (or if they don't) it would be important to know.

Smaller comments and line-by-line
- Please clarify the use of “soil carbon” vs “soil-respired carbon”. For example, line 199 states “The regression points toward a soil carbon end member...”. Is this treated as just “soil carbon” in the modeling? Because the exercise seems to imply that the constraint is a “soil-respired carbon” end member. Line 190 also refers to the “soil carbon end member” but states it was constrained with data, not the modeled regression (which I think is accurate). Since soil respired carbon is defined as a component of soil carbon (line 182) this distinction is super important. It's still not fully clear to me how soil carbon vs soil respired carbon are treated in the model.
- I imagine that the soil respired end member of the mixing curve changes seasonally. If calcite formation is seasonally biased, could this affect the results? For example, are more model solutions at higher soil CO2 conditions possible when strictly summer-time inputs are used?

Line 36: Check out Bova et al., 2021 (Nature) for updated Holocene climate constraints.

Line 49: I’m not convinced by Figure 1 that these records are “highly consistent in timing, amplitude, and absolute d13C“. I worry that the words “highly consistent“ are overstating the data. Consider focusing on the main trends that are clearly robust, like the general shift to lower d13C values from 18ka to 6ka.

Figure 1B: Consider labeling the El Pindal and La Vallina sites.

Figure 1C: Is the straight blue line from ~18ka to 15ka just due to the fact that there are no data? It might be clearer to disconnect the timeseries lines whenever there is a sufficiently long duration of no data (maybe wherever there is ~500 years of no data or something).

Line 129: I don’t know if CP allows citing papers in review, just adding it here as a note (although I assume that the authors have already confirmed that this reference is okay!)

Line 169: Is this really deriving the soil carbon “…response to temperature change”? I think the link to temperature change is solely based on interpretation, not model derivation.

Line 184: Not a paper strictly on soil CO2, but Slessarev et al., 2016 (Nature) might be useful here for linking parameters of the soil carbonate system to the water balance.

Line 194: "...by linear regression of the summer cave monitoring data". I assume these are the large-diamond points in Figure 2. But looking at figure 2 I assume that the regression data are spring, summer, and fall (since monitoring is said to be monthly and there is no indication that spring/fall data are removed). Which data are actually used in the regression?

Line 196: While I suspect the offset of the forest data may actually be due to mixing with modern pCO2 (not pre-industrial levels), if the authors wish to keep this turbulence/advection effect argument I think it is important that a reasonable hypothesis for the signature of the third, unaccounted for air mass is added. Based on atmos circulation and likely boundary layer d13C in upstream ecosystems, is this mixing trend reasonable?

Line 201: “... but they provide the best available constraints on the end-member”. Wouldn't directly measuring soil CO2 provide a better constraint? (Although, as stated above, I disagree with using a modern calibration to get a Holocene end member)

Line 217: Why was each simulation for each timeslice repeated twice? Were they varied
from one simulation to the other? (Table 1 only gives single values for each timeslice)

Line 221-226: I’m a bit confused. Is there one set of binary filtering for the three mixing line simulations, and a different filtering approach (just selecting the best 5%) for the sensitivity tests?

Figure 4: Are measurement uncertainties considered in these regressions?

Line 293-294: I don’t think that encouraging model results is confirmation that “the estimate of the soil respired end member composition is accurate”. More sensitivity tests are needed to demonstrate that other soil respiration end member compositions lead to problematic results (particularly when the end members are allowed to vary with time, as discussed above).

Line 309-311: This is another instance where I’m tripped up by terminology. I think “initial soil gas” is the same as “soil respired CO2” and not the same as just “soil gas” or “soil CO2”?

Line 310-311: I would like to know more about this. Why does the sensitivity test require such an enriched d13C end member? How is the use of this end member justified over the use of the “calibrated” one? Is it a problem that the -22 per mille value does not yield many positive results?

Line 328: What is meant by “depth” here? I don’t think the Pataki paper actually measures anything over soil depth.

Line 347: This is probably just my own problem, but I’m confused with terminology again. I thought initial soil gas might be the initial CO2 from soil respiration, but this sentence implies initial soil gas and soil respired CO2 are two distinct things.

Line 433: “...best explained by..” my version of the document just says “c.”

References


