RC3: 'Comment on acp-2022-504', Anonymous Referee #3, 22 Sep 2022

Diurnal variability of atmospheric O_2, CO_2 and their exchange ratio above a boreal forest in southern Finland

Overall, this is very nice paper presenting important results. The authors conduct challenging measurements, analyze the data intelligently and combine their own data with ancillary datasets in a clever way to extract interesting values. They focus on the O2/CO2 exchange ratio for a boreal forest (unprecedented) and then extend their work to seperately assess the exchange ratios associated with respiration and assimilation. The work is valuable, the paper is generally well organized and it definitely deserves publication.

We thank this reviewer for their assessment of our manuscript, including the detailed comments on the text. We will address remaining issues below.

That said, I do have some concerns that need to be addressed prior to full acceptance:

1. The authors use α_b, ER and OR somewhat interchangeably in the introductory part of the paper. Each of these symbols really does have a distinct and specific meaning. Although the use of these terms in the literature has been somewhat sloppy, as our field matures it becomes more important to use the right word in the right context.

   We agree with the reviewer that in our field the terms alpha_b, ER and OR are used somewhat interchangeably in the literature and that we should be careful what we mean with each of them. We tried to use the terms correctly throughout our manuscript, but we agree with the reviewer that in some parts of the text we did not succeed in making a clear distinction between the different terms. Therefore, we changed the introduction text to be more consistent. The exact changes, with respective line numbers can be found below, where we have included the list with notes from the pdf, and a reply to these specific points one by one.

2. The authors assume ERr is constant day and night. This may well be true, but it's possible it isn't true. Since this assumption is central to the subsequent analysis, there should be more discussion of this assumption and its validity.

   We agree with the reviewer that more discussion should be added to strengthen the argument that the ERr can be assumed to be constant during the day and night. Not many studies have researched the changes in ecosystem ERr and ERa.

   Hilman et al. (2022) showed that the major contributor to the changes in ERr is the bulk soil respiration (the bulk soil is the part of the soil that is not influenced by roots). The bulk soil respiration changes with soil temperature and soil moisture. These changes are not likely to affect our diurnal cycle of ERr and mainly show effect on seasonal time scales. One component that could potentially change ERr during the diurnal cycle is the respiration of plants. As the respiration of the plants involves photorespiration during daytime and only dark respiration during nighttime. Dark- and photorespiration use different pathways and therefore may have a different ERr values and could affect the ERr of the ecosystem. To our knowledge no studies so far have looked in detail at the diurnal cycle of ERr of plant respiration and it is therefore difficult to say how much effect this has on the total ecosystem ERr. However, Hilman et al. (2022) showed that the bulk soil respiration has the largest effect on the variability of ERr, so we assume that the variability of plant respiration has little effect on the total ERr of the ecosystem. Other studies that looked at the ERr also only focussed on the soil and only focussed on a longer time scale compared to the daily cycle (Angert et al., 2015; Pries et al., 2020). Therefore, more research should be performed to check if our assumption is valid.

The ERa seems to be influenced by the source of nitrogen in the soil (Bloom, 2015) and the amount of light the plant receives (Fischer et al., 2015). If the source of nitrogen is from nitrate, the ERa will go up because of nitrogen assimilation. However, as stated in line 597, the main source of nitrogen in Hyytiälä is ammonium. Ammonium does not affect the ERa. If the plants are deprived of light for a longer time, the ERa will also change. As the plants cannot produce enough small carbon compounds from respiration it has to use larger carbon compounds from its storage. However, this would mean that the plant would receive very little light for a longer time period, which did not happen during our measurement campaign. We therefore could say whether the ERa also stays constant during the day.

There are only a few studies that focus on the diurnal variability of ERr and ERa and the studies that do exist mainly focus on the effect of relative abrupt changes in the environment and their effect on the ER signals. It is therefore not yet known whether ERr and ERa stay constant throughout the day and if our assumptions are valid. However, as the existing studies suggest, only major changes in the environment could lead to large changes in ERr and ERa. This means that on diurnal time scale only small changes in ERr and ERa occur, which are too small to measure with our measurement set-up and therefore also fall outside of the scope of this research. Further research using soil, plant, branch or lab chambers is also recommended to obtain further knowledge about the variability of these process level ER signals.

We have updated the text to include further discussion on this point. The exact changes, together with their line numbers can be found in the list with notes from the pdf, which is included below.

3. The data were compromised at times by the failure of some MKS pressure/flow controllers. The authors apply a correction to the data, but there are a few points with (apparently) anomalous values where we're told that the correction simply wasn't adequate. Since we aren't told any of the details of the correction, I'd like to see evidence that the other (non-anomalous and corrected) data are valid, and not just because their values are close to what we expect.

   We agree with the reviewer that we did not discuss well enough how the pressure corrections influenced our data and why the uncertainty of this correction increased during the mid-day measurements. Therefore, we added an additional figure to the appendix of the paper and elaborated more on this topic in the text. A more detailed explanation of our correction and the changes we made to the text can be found in the response to reviewer #1, at the first major point, and changes to the text in lines 164, 320, 332.

4. The authors attribute differences between ER_{atm} and ER_{forest} to "boundary layer dynamics and entrainment" or the unique nature of boreal ecosystems. I think the first explanation misses the point and the second if very likely wrong. Whenever you see O2 and CO2

changing with time with a slope more negative than -1.2, this indicates the influence of fossil fuel combustion.

We agree with the reviewer that our phrasing of the explanation for the difference between ERatmos and ERforest and the effects of entrainment should be made clearer in the text. We understand that the ER of fossil fuel combustion has values than 1.2, however we disagree with the point that a measured ERatmos signal higher than 1.2 should automatically indicates a source of fossil fuel combustion. We calculated the footprint for the representative day (7 through 12 July), and it shows that the dominant wind direction is from the North to Northeast, where hardly any sources of fossil fuels are located.

The large value for ERatmos, especially during P2, is a point of concern for all the reviewers. We realize that we did not explain our reasoning well enough to show in a clear way how this large number of around 2 could arise. A more elaborate explanation is given in the answer to reviewer #1 in the section "Explanations for the high ERatmos values". Please see our response to reviewer #1, and our updates to the text in line 481.

We will elaborate here on the specific points that reviewer #3 made in the pdf annotations related to this major point:
Air that is entrained from the residual layer and the free troposphere is influenced by air masses with different background signals. This means that different sources, such as ocean, fossil fuel and biosphere could have contributed to this signal. Mixing different sources of air with different ER signals could create a mixture of air that has a final ER that cannot be contributed to one specific process or could even have an ER value that is higher than 2.

This happens because we cannot just average each ER signal that is mixed into the free troposphere, as an ER signal of fossil fuel has a different meaning than an ER signal of the biosphere. The ER signal of fossil fuel means that the air is depleted of O_2 and enriched of CO_2. The ER signal of net biosphere exchange means that the air is enriched with O_2 and depleted of CO_2. Averaging the ER signals would then be wrong, similarly to averaging the day and night ER signals of a forest (see our reply to point 3 of reviewer #1). Different sources contribute to the air in the free troposphere differently for O_2 and CO_2, it is therefore highly unlikely that the ER signals from different background sources can still be distinguished.

During the morning transition this air is entrained and measured by our devices. When different ratios of O_2:CO_2 are entrained in our aggregate day, we find this steep slope of around 2 during the entrainment dominant period. The 125 m measurement height also confirms that entrainment is causing these high values. At 125 m, the measurements are outside of the roughness sublayer and therefore the measurements at this height represent the mixed-layer signals that are normally more influenced by the entrainment processes, which results in a slope of 3.40. The ER_{atmos} value is therefore a useful tracer to quantify how the impact of atmospheric driven processes on the measurements. However, to link the ER_{atmos} signal to only the biosphere processes, we recommend using measurements as close as possible to the canopy.

5. There appears to be circularlity in some of the analysis. For example, the EC data are used to set a value of the free parameter K (a transport coefficient) for getting fluxes from O2 gradients. Then the O2-based fluxes are assessed by comparing them to the EC

data. Similarly, NEE is split into GPP and TER using the O2 and CO2 data. Then the O2 and CO2 data are further interpereted by taking GPP and TER as if they were known a priori.

We agree with the reviewer that there is indeed a minor degree of circularity in our analysis for the 1) flux partitioning and 2) the use of K. We will address these two points and clarify the limited impact on our analysis.

1) The circularity in the O_2 method. We agree with the reviewer that it could have been written down more carefully how we calculated the flux partitioning using the O_2 method as shown in figure 8. We used the GPP and NEE measurements (that were available from the EC measurements at Hyytiälä) to determine ERa for the representative day (7 through 12 July). Then we applied this ERa to a new set of days (13 through 15 June) using the O_2 method. For the new set of days, we used the ERa and ERr that were based on the initial representative day and the ERforest was calculated with the measurements of the vertical gradient and resulting O_2 and CO_2 fluxes for the new set of days. This means that there is a certain degree of circularity in this calculation, as the ERa is based on GPP measurements from one combination of days (7 through 12 July) and then we use this ERa again to determine GPP for a new combination of day (13 through 15 June). However, we expect only minor changes in ERa compared to the values we derived, and the good comparison between the O_2 and the EC methods for the partitioning of the fluxes shows the potential of the O_2 method. We further elaborate on this point also in our reply to reviewer #2 and updated the text in line 611.

2) The circularity in the K validation. It is indeed correct that we use the EC-based CO2 fluxes to calculate the K and then use K to derive our fluxes from the vertical gradients and compare them again to the EC-based CO_2 fluxes. This results in a circularity which is mentioned in line 517. Unfortunately, this approach cannot be tested on one day (first representative day) and then be applied to a different day (the second representative day), as we did above to check the O_2 method.

   However, for this flux calculations of O_2 and CO_2, our approach is again not completely circular, because we use the EC flux of CO_2, together with the gradient of the ICOS CO_2 mole fraction observations to determine K. Then, we use this derived K, together with the gradient of our campaign data to determine the CO_2 flux and compare it again to the EC flux. With this approach, the vertical gradient is measured by two different instruments, which makes this comparison not completely circular. This circularity will always exist even if we would test it on another day which makes it difficult to check the method of using CO_2 to determine K without being biased. The circulatory of determining K is therefore difficult to solve.

   To remove this circularity, we could choose to use another parameter measured with EC, such as potential temperature and the sensible heat flux to determine K. We have used this approach to test several options for deriving K, as we show in table 2 of the manuscript, and these different approaches can also be used to get an idea of the uncertainty of K and the impact on our results. However, we prefer to use CO_2 as it has a more directly linked to O_2 on how it is transported through the atmosphere, compared to potential temperature, which is related to the sensible heat flux and therefore an active variable in generating turbulence. Other studies (Wu et al., 2015) also use CO_2 as a measure to determine K and we therefore are confident that this approach works. This is discussed further with point 30 below.

We have included table B1 in the appendix of the paper to clarify the data used in each step of our calculations, so the level of circularity in our calculations is more clearly shown.

It's quite possible (particularly for #5) that the authors have done nothing wrong and I have simply failed to understand their work. If that's the case, then my comments should be taken as a plea for clarification and explanation in the text.

All of these concerns, along suggestions/corrections on word choice, punctuation, sentence structure and grammar, and covered in the attached "marked up" PDF. The markings are in three colors: Red - add/delete/move text, to be taken verbatim Green - questions/directives for the authors Yellow - highlighting text for which I have typed a "sticky note". Be sure to open the note and read to the end. Scrolling may be required.

We copied the sticky notes comments from the PDF file below, with the line number they were attached to, so we could provide answers with each comment. Finally, we have updated the text using the textual suggestions provided by this reviewer.

Finally, I would like to acknowledge that the writing quality is very high. Even though I have made numerous editorial markings, as a native English speaker (with a modest proficiency in German) I am in awe of the authors' ability to write so well in a second language. Well done!

Thank you for this compliment, we tried to improve the writing further using your suggestions.

1. Line 36: The semi-synonymous use of OR and ER is a problem in this field. OR is a chemical property specific to materials and investigated by elemental analysis or combustion (or similar methods) in the laboratory. ER is a behavioral “symptom” that is specific to an organism, group of organisms, or ecosystem. Then we have alpha_b, which is an effective ER for the planetary biota and has things like disturbance and wildfire built into it. You should be careful of these distinctions throughout the paper and choose your labels accordingly.

   We agree with the reviewer that there is a distinct difference between the OR, ER and alpha_b, and we should have been more careful explaining them. Therefore, we update the text accordingly and have added an extra sentence in line 57, that explains the difference between these two terms.

2. Line 58: Here again there is the problem of OR vs ER. It’s not just a temporal difference - there’s also a spatial scale. OR from elemental analysis is intrinsically tied to particular

samples. Leaves may be different from twigs or trunk wood, and you only learn about what you put in your analyzer. Your prose should be chosen with great care to reflect these distinctions.

We agree with the reviewer that the difference between OR and ER is not just based on temporal difference, but also on spatial. Therefore, we have added extra details to line 58 to make this more clear.

3. Line 60: …and here, alpha_b should be reserved for a single global number. The budget equations using alpha_b are really only meaningful on large scales, so we can't talk about local values of alpha_b. + Line 79

   We agree with the reviewer that the alpha_b that is used on large scale, e.g. to estimate the ocean sink, should be taken as a single global number that account for all the specific processes. However, when the O2 method is applied on a more local scale, for example using the APO method to determine fossil fuel emission (Pickers et al., 2022) a more local O_2:CO_2 molar ratio for the biosphere has to be used, that indicates more the processes that influence the measurement location. In this way we could talk about local values for the biosphere exchange. We therefore have removed the reference to alpha_b when talking about a more local scale.

4. Line 66: If you want the ER of the forest, any real measurements of "surface" fluxes won't work (unless "surface" includes the soil surface, leaf surface, trunk surface, petiole surface, etc.). I believe you're imagining an idealized surface, so you should probably say so explicitly.

   To be clearer with what we mean by surface, we added some text to line 64.

5. Line 78: Seibt et al's paper is a wonderful one, but they did not present continuous measurements of air above a forest- she analyzed discrete flask samples. Probably better to say "…that measure O2 and CO2 in the atmosphere above an ecosystem with sufficient frequency to derive ER values."

   We agree with the reviewer that we should have been clearer about which measurements Seibt et al. (2004) did and therefore we changed the text in line 78.

6. Line 175: I'm probably missing something simple, but it's not clear to me why a different Target Tank should yield a different std dev. Is there some simple explanation you can include in the text?

   We added a few words to line 175 to explain that the target tank of 2018 and 2019 differ because they contain different composition of air. This results in a different mean difference to the calibrated value. In principle, the std should be the same if the system did not change between the years, and when averaging over a long enough period. However, for our short campaigns we still found some differences between the two years, and therefore report these separately.

7. Line 191: Presumably you mean a negative relationship in the changes of O2 and CO2 as a function of time (i.e. dO2/dt and dCO2/dt have opposite signs). Please clarify.

We changed the line 191 to make clear that the negative relationship indeed indicates a relationship of changes over time.

8. Line 225: I think what you're trying to say is "the flux for any give entire day is the average of the fluxes for the unstable (daytime) period and the portions of the stable (nighttime) periods that lie in the chosen midnight-to-midnight 24hr window. The data from the transitional periods (excluded from the day and night periods) are not included in the full-day average." Maybe I have misunderstood what you are saying, but if so, that only strengthens my point: Please consider trying to make this clearer, rather than just saying "the flux for the entire day is the average over the entire day".

   We added some text to line 225 to make clear that we use all the points to calculate the average fluxes of O_2 and CO_2 to determine the ER_{forest} of the entire day. The transition periods are now also included.

9. Line 258: I trust that you're not actually doing anything wrong here, but this whole section ER_forest is a bit confusing since I can't clearly tell when you switch from discussing the observational approach to the theoretical one. In particular, in the observational method, you need EC-based CO2 fluxes to calculate K. Thus, "choosing the value of K that gives the best agreement with the observed EC CO2 fluxes" is circular. I don't actually think you're falling into this trap, but I hope you can see how this existing prose might be confusing to the reader. Please try to reorganize the content so that the two approaches are more clearly distinct.

   For a more elaborate explanation on the circularity, please look at major point 5, and below at point 29.

   We added some text to line 258 to make it more clear that we test the best approach to calculate K_φ based on both the observational and theoretical approach.

10. Line 272: This assumption is essential for your analysis that follows, and it's probably correct enough (given the uncertainties in your O2 measurements). However, I think you should consider and acknowledge the possibility that respiration during the day (which includes photorespiration, particularly in hot, dry conditions) may actually have a slightly different ER_r than respiration at night. I'm no plant physiologist, but I'd like to know that you have at least thought about photorespiration and its potential link to nitrogen in the tissues (e.g. see Bloom, Photosyn. Res. 123(2):117-128) and how it might make ER_r different during daylight hours.

    We agree with the reviewer that we should explain our assumption that ERr stays constant throughout the diurnal cycle. Please see our reply above for major point 2 for a detailed explanation on why we assume that ERr stays constant during the day and night and please look at line 272 for the changes made in the text.

11. Line 279: In Fig. 3 you only shade July 7-12 and it's quite hard to see anything about July 13-15. Please shade these dates too (but with a distinctive tone) and include them on the zoomed-in plots. I would like to be able to assess whether the data really do look comparable for the two intervals.

We added a shade to the days 13-15 June to make it more clear which days we choose for the analysis and calculation of the O_2 method. The zoomed-in plot of these particular days was added to the appendix and show here below in figure 1.

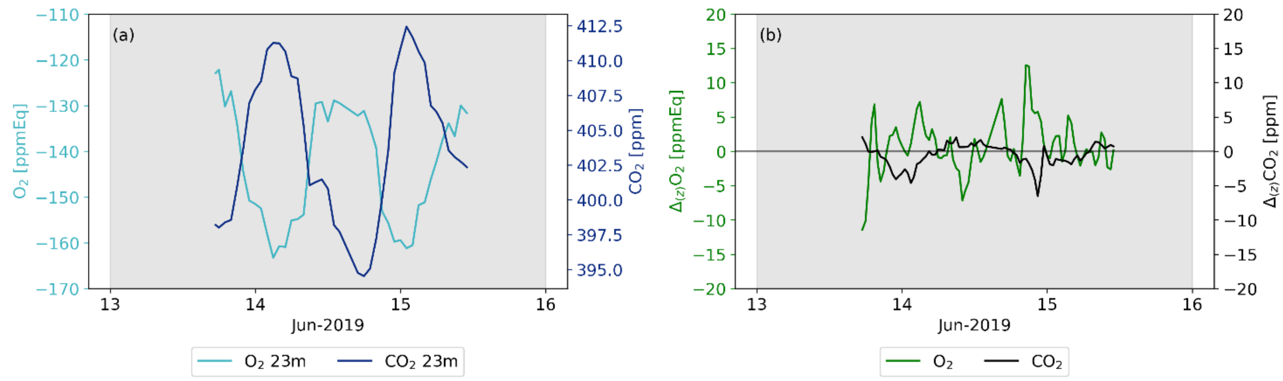

Figure 1 The O2 and CO2 measurements at 23 m, and the vertical gradients of O2 and CO2 for the second representative day (13 through 15 June 2019). This is added to figure 3 in the paper.

12. Line 288: This is a clever approach, but it only works if you have ER_a and ER_r in hand and you are confident that they are the same on the day you determine them and the day you use them to get GPP and TER. That seems like a good assumption in this case, where your two "representative days" are not very far apart (temporally and spatially), but I would like to see some prose addressing this assumption/limitation.

    We agree with the reviewer that we should elaborate more on why we justify the assumption that the ERa and ERr of the period 7 through 12 July can be used for the period 13 through 15 June and can therefore be used to calculate GPP and TER from the O_2 method. As both periods are relatively close to each other and have similar meteorological conditions, we can assume that the ERa and ERr stay constant based on the few studies that exist on this topic (see above our reply on major point 2 for further explanation). **We have added more elaborated information to the text in line 281.**

13. Line 305: This is valuable information. Could you please mark it (with vertical dashed lines, or something like that) in Figure 4?

    We included the sunrise and sunset times with vertical dashed lines in Figure 4. Be aware that Figure 4 is in local wintertime and the time that is given in the text is summertime. This means that there is 1 hour difference.

14. Figure 4 caption: "A remaining artefact" is too vague. My immediate reaction was that you were talking about the single point that jumps up. However, after reading the body of the text, I see that you're referring to a bigger problem. In this caption, you should either omit any reference to the problem or highlight the effected points on the plot and say "see text for details".

    We agree with the reviewer that we should refer to the text to highlight that this period was part of a bigger problem. Therefore, we adjusted the text in the caption of figure 4.

15. Line 319: I think you're saying that you expect the O2 concentration to remain high since you have every reason to believe that assimilation is dominating. However, O2 falls in your plot and you're saying this is due to the MKS problems that you couldn't correct for. If

I've interpreted your writing correctly, this explanation is plausible, but it does make me wonder about the robustness of the rest of your data. Why would the MKS problem be irreparable only in the late afternoon? And since this is a composite "day", are you saying that it was an irreparable problem only in the late afternoon on several consecutive days?

The reviewer indeed interpreted the text correctly and we indeed meant that O_2 should have stayed high, but due to the high uncertainty of the PMKS correction during the day the pressure correction could not account completely for the MKS problems. This happens during the day because the pressure instability issue has a high correlation with temperature.
A more elaborate explanation about the PMKS correction and why it mostly affected the mid-day data is given in the answer to the major point raised by reviewer #1. Please see for more detail the section "Further details on MKS pressure transducer instability correction". We have modified the text in lines 164, 320, 332 to explain this in more detail.

16. Line 326: I'm sure I'm just missing something obvious, but you say the CO2 gradient goes from negative to positive due to CO2 being transported downwards. Wouldn't that have the opposite effect?

    The gradient is calculated by subtracting the observations at 23 m from the observations at 125 m (125 m-23 m). This means that for example when CO_2 has a positive gradient during the day, the concentration of CO_2 at 125 m is higher compared to concentrations at 23 m, since the 23 m is closer to the sink of CO_2 driven by photosynthesis which reduces the concentration. This means that the forest is taking up more CO_2 to deplete the air of CO_2 at the height of 23 m compared to 125 m. As a result, the gradient becomes positive.

    As equation 7 also shows, the gradient and the flux are anticorrelated. Which means that with a positive gradient (CO_2 during the day), the flux is negative and points downwards into the forest.

17. Figure 5: To me, this looks rather like two different "regimes" within P2. As you say in the text, these slopes are probably not telling us about the forest, but instead changes in the degree to which entrainment and other physics is driving the concentrations. It makes me wonder about the choice of boundaries for the various time periods. I can see these points most clearly between 4am and 6am in Fig. 4. One possibility is that the stability of the layers close to the surface (i.e. around 25m) starts to break down before the plants wake up and start photosynthesizing vigorously. Is your micrometeorological data consistent with this?

    We agree with the reviewer that there indeed seems to be two periods inside P2. The steep increase is very likely caused by entrainment or other processes which are dominant and that are not related to the surface fluxes. Therefore, it is still correct to assign this period to P2, as we define this as the entrainment dominated period. We are currently working on a follow up modelling study where we aim to further investigate this pattern and try to understand how this steep increase in the morning is formed.

18. Figure 7: There is probably a simple explanation, but I am surprised that the error band for this period is so small, given that the spread in the individual points appears larger than the spread in the 0:00-4:00 period. Perhaps this is worth addressing in the text or in the figure caption.

We used the same uncertainty for the whole night period, as ERr is based on all the night measurements. Therefore, is the error band the same size for the period between 22:00-00:00 as the period 00:00-04:00. The reason why the individual points show larger variability between 22:00-0:00 is because the sun then had just set and the gradient of both O_2 and CO_2 is still relatively small compared to the period 00:00-04:00. The smaller gradient results in a larger uncertainty in determining the surface flux and therefore a more variable ERforest.

The error band therefore is representative for all nighttime measurements rather than for individual points.

19. Line 396: This is just a confusion over choice of words, but I think your "averaged fluxes" are what I would call "a flux-weighted average of the measurements". I think you are saying that you did such a flux-weighted average for each of the periods in the day, as well as the full 24-hour day. This is particularly confusing for me because all of the really big spots in Fig 7 (i.e. the times with large fluxes) have ER_forest values that are more negative than -1.0 and the biggest are more negative than -2.0, yet your full-day value is only -0.83.

We agree with the reviewer that our method could be described as a flux-weighted average. However, this applies only for the daytime and nighttime signal and not the 24 hours signal. We give an example with the figure below where we calculated the mean of all the ER data points of the daytime. The mean of the ER signals (red line) is lower compared to our daytime ER signal, which is based on the average O_2 flux divided by the average CO_2 flux (blue line) (representing the flux-weighted average). By using the averaged fluxes, we use indeed a flux-weighted average, as our daytime ER signal is higher compared to the mean of the ER signal. This means that the higher surface fluxes, with a higher ER are given more weight.

We understand that it can be confusing if we look at the 24-hour value, which is 0.83. This is because the daytime and night-time ER represent opposite fluxes of opposite sign for day and night for both O_2 and CO_2 and therefore influence the atmosphere composition differently. We cannot apply a weighted average when combining the daytime and night-time as we would then ignore that the ER represents different and contrasting processes (night respiration versus day photosynthesis and stable stratified conditions versus convective conditions) during the night compared to the day. For a more detailed explanation please see our answer to point 3 of reviewer #1. We have modified the text in lines 398 + 552 + 559.

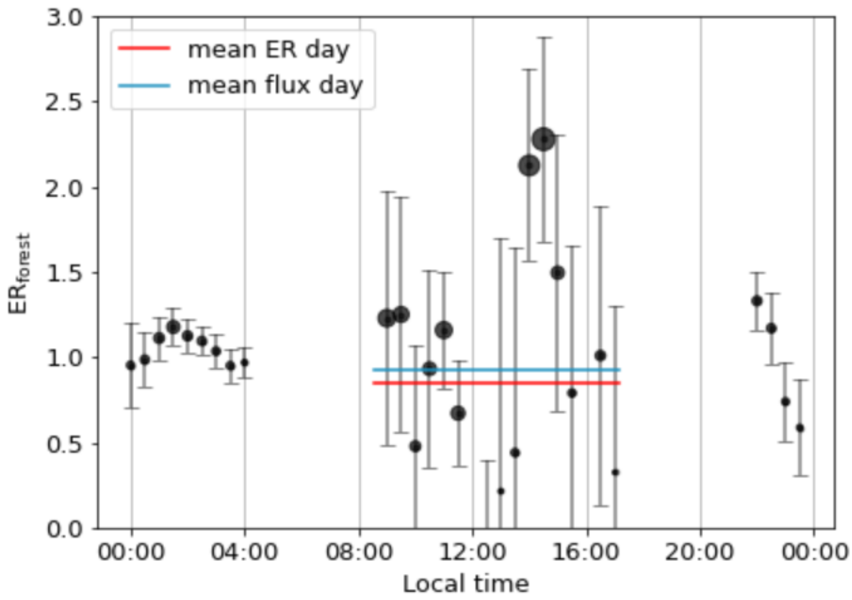

Figure 2 The same data points as Figure 7 in the paper, but now comparing two approaches to calculate the average ER for the daytime: Using the average of the ER signal (red line) and using the average O2 flux divided by the average CO2 flux (blue line).

20. Line 404: Am I correct that this calculation assumes that you have values for NÉE and GPP from the EC measurements taken by SMEAR II? If so, you should say so explicitly. Please convince me the rest of your data are properly corrected.

    The reviewer is indeed correct that the values for NEE are from the EC measurements of SMEAR II, and the GPP values are derived from these NEE EC measurements. To make it more clear how GPP is determined from NEE, we have added the reference of Kulmala et al. (2019) and more **explanation in line 275.**

21. Line 405: See comment line 272

    Please see our reply to major point 2 above for a detailed explanation and for the changes in the text in lines 405 + 585 (see also our reply to point 32 below).

22. Line 411: I think I missed something. Have you already shown (derived/proved) this, or have you just stated it? I don't think you actually show it in the text that immediately follows either.

    See next point for clarification.

23. Line 412: Actually, my confusion is more profound than just my "show" comment above. Eqns 7 & 8 have six separate terms. If you have NÉE and ER_forest and ER_a and ER_r then you can definitely solve for the two remaining ones (GPP and TER). However, as I understand it, the only way to get ER_a and ER_r is from these equations, using NEE and GPP from independent EC data (from SMEAR II, in this case). Thus, this looks like a circular (i.e. self-referencing) approach. The only way around this (which is probably what you're doing) is to calculate ER_a and ER_r for one set of data (i.e. one particular representative day) and then apply it to other data (a different day). If this is in fact what you're doing, you should be very clear about it. Otherwise, the calculation appears completely circular.

Please see our reply to major point 5 above and our reply to the second major point of reviewer #2 for a clarification on the reasons why we think there is only a minor circularity in our calculations, because we use different days. We have updated the text in line 611 to explain this further.

24. Line 418: Since you say “increasing/decreasing RE_forest”, you should put signs on the associated change in your inferred value of GPP (i.e. if RE goes up, does GPP increase or decrease?). You could do this as simply as including +/- or -/+ as appropriate.

    We agree that this was not clear in figure 8. As GPP and RE are coupled with the equation:
    NEE = -GPP + TER
    This means that when TER increases, GPP has to become more negative to keep the NEE constant in time. To make it more clear we changed the signs, as GPP is normally presented in atmospheric studies as a negative flux (meaning pointing from the atmosphere into the forest and removing CO_2 from the atmosphere), we changed Figure 8 and line 418 accordingly.

25. Line 421: How did you settle on this value? I would think that you want to be able to do roughly as well as the EC method. Or perhaps, you'd like to be able to see differences between GPP and TER comparable to those implied by the EC data. Either way, you should state how you settled on 0.05 as your threshold.

    We agree with the reviewer that we should be more clear how we determined the 0.05. We based indeed on the comparison with the EC method. With a suggested precision of 0.05 for ER_{forest}, the change in GPP and TER fluxes for the O_2 method stays in the range of the GPP and TER fluxes of the EC method. However, to create a more detailed discussion on the comparison between the EC method and the O_2 method, a higher precision is probably needed, together with an independent measurement of ERa. A detailed comparison between the EC method and the O_2 method was not the main goal of this study (see second comment reviewer #2), and we include this analysis to show the potential of the O_2 method, and therefore our suggested precision is 0.05. We added a sentence in line 421 to make it more clear why we choose 0.05, and added further discussion in line 611.

26. Line 479: It is tempting to say simply “atmospheric dynamics and entrainment can give a slope closer to -2” but I think that's a little misleading. When non-canopy air is entrained or advected, that air brings both O2 and CO2 with it. This will change the composition of the air you measure, “moving” down-right or up-left in Fig. 5, depending on the composition of the previous parcel you measured. Most importantly, the “movement” will define a slope that reflects the processes that last changed the new parcel. If you are advecting/entraining air that has only seen biological influences, the slope will always be close to -1 no matter how much O2 and CO2 are removed/introduced. The only way the slope can be close to -2 is if the air that was entrained was heavily influenced by natural gas combustion. In short, your observed values more negative than -1.3 are almost certainly indicating that you are measuring non-local air with a strong fossil-fuel signature. Furthermore, the longer the time over which you include data, the greater the footprint and the more likely you are to have non-local influences. This is consistent with your 24-hour values being much more negative than the shorter periods. When you look at it this way, you see that you have to be very close to the top of the canopy (or within it), and measuring over a short time in order to get ER_atm to be at all comparable to ER_forest. I encourage to you to rewrite this section of your discussion with these ideas in mind.

Please see our answer to major point 4 above and our answer to reviewer #1 in the section "Explanations for the high ER_{atmos} values" for a detailed explanation about this point. We have changed the text in line 481.

27. Line 481 – 487: Unless there is some reason to believe ER_forest is very, very different from -1.05 (i.e. much more negative), a much more likely explanation is non-local influences, as described in my previous note (above). As you yourselves discuss below, the OR of the organic matter in a boreal forest has values not so different from -1.0. Simple calculations show that even when there's disequilibrium between the kind of materials being synthesized and those being respired (e.g. low-nitrogen trunk wood vs. shoots and leaves) the ER will differ from the underlying OR by no more than 10-15%. For this reason, invoking this explanation as a real possibility is inappropriate.

    We agree with the reviewer that non-local effects are the reason for the high ER_{atmos} values. As described in our reply to major point 2 above and our answer to reviewer #1, we think that it is entrainment from air of the free troposphere that is causing this high ER_{atmos} signal. The air of the free troposphere is influenced by non-local sources. As already discussed, we therefore adjusted the text in line 481.

28. Line 500: Maybe I've just forgotten where you mentioned this earlier, but I'm not sure exactly what you mean by "flux-gradient" method. Please define/remind.

    We added to line 500 the reference to equation 7, which indicates what we mean with the flux-gradient method.

29. Line 517: Yes, this circularity is a problem and deserves more than an acknowledgement. I have already made comments about it on Lines 404 and 414. Your basic idea for using all the available information is very clever, but I believe the only way to make it robust is to use one dataset (the first representative day) to "tune" your method, and a different dataset (the second representative day) for drawing conclusions and deriving results. Otherwise, the results are untrustworthy and the error analysis essentially meaningless (or at least far too complicated for me to imagine/understand/trust). Perhaps you've already done this and I just missed it. If so, please make it much more explicit.

    Please look at our reply to major point 5 above for an elaborate discussion on this circularity issue and why we think it is not a major issue. We have added information in line 517.

30. Line 536: Maybe I just missed it, but I don't see any quantitative statement about the uncertainty in K and there's no sign of it in Table 2. Without that number, I'm not convinced that the error bars on the nighttime fluxes are really as small as those pictured. And of course, quantifying uncertainty in K ties back in to my comments about circular reasoning above.

    We did not include the uncertainty of K in our calculations, because the final uncertainty in our values is from the measurement uncertainty in O_2. As equation 7 shows, to determine K we used the EC flux measurements of CO_2 and the gradient of CO_2 from the heights 16 m, 67 m and 125 m from the ICOS instrument. This means that all the final uncertainties of our results where ICOS data is included, are probably somewhat higher than currently given. However, as stated in line 574, the largest factor of uncertainty in our observations is from the O_2 measurements themselves, which means that probably most of the uncertainty range is captured in the values we present.

To better quantify how the variability in K would effect the final results of the ER signals we test what happens when we use K with theta, instead of K with CO_2 , to determine our final CO2 and O2 surface fluxes (see Figure 2 below).

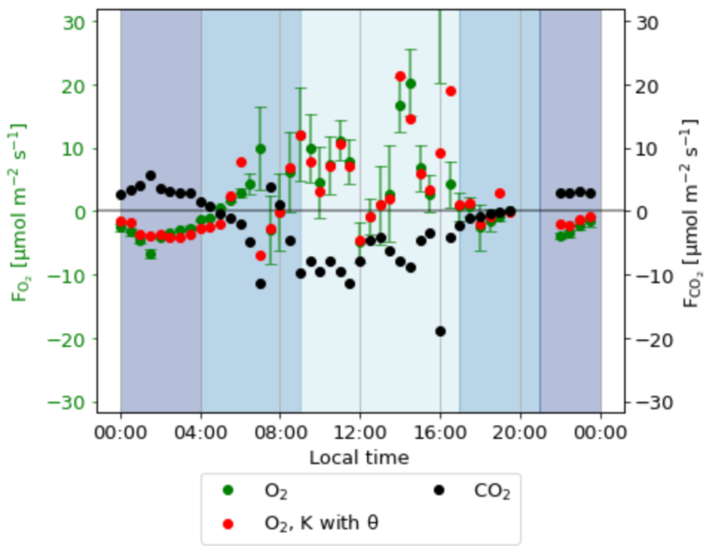

Figure 3 The same as Figure 6b in the paper, but now with the approach added where the O2 flux is determine with the K based on potential temperature.

If we use K with theta to calculate the surface fluxes of O_2 (Figure 2 above) and CO_2 (as for Figure 6a in paper), we find that the difference with the approach to use CO_2 for K is minimal (red and green data points in the figure 2 above). The largest differences arise in the transition periods. This is expected, as the gradient changes fast during this period and differences between the gradient change in theta compared to CO_2 are then more pronounced.

If we would then use the fluxes based on the K with theta, we get the following ER signals:

Table 1 The ER signals, same as Table 3 in paper, but now with the approach where K with theta is used to determine the surface fluxes. For reference the ER signals of the approach where K with CO2 is used is also added.

ER signal	**Value (based on theta)**	**Value (based on CO2, from table 3)**
ERforest_day (09:00-17:00)	0.94	0.92
ERforest_night/ERr (21:00-04:00)	1.04	1.03
ERforest_all (all data point)	0.92	0.84
ERa_day (09:00-17:00)	0.97	0.96
ERa_all (all data points)	0.99	0.96

We find by comparing the table above with the ER signals based on CO_2 from table 3 (also included in the table above), that most of the ER signals are quite similar. The biggest difference can be found between the ER signals where the transition periods are included (ERforest_all and ERa_all), which was expected as the transition periods were most difficult to determine.

By using different methods to determine K, we showed that even with a variable K the final results are still very similar. This shows that we can be quite certain about the derived K and the uncertainty of this component is low. As already stated above, the largest uncertainty is caused by our O_2 measurements and therefore it is reasonable to omit the uncertainty in the K from our uncertainty calculation.

To make it more clear that we could not include the uncertainties of the ICOS data we added some text in line 543.

31. Line 565-577: This entire section ties into my comments on Lines 478 and 485. Please revise accordingly.

As described in our reply to major point 4 above and our answer to reviewer #1, we think that entrainment is the largest contributor to the high ERatmos signals.

We added to line 466 in more detail why entrainment is the cause of this large difference between ER_{forest} and ER_{atmos} and that our measurement height is probably a first explanation why we find such different results compared to previous studies.

32. Line 583-592: I already commented extensively on this at Lines 272 and 407. To add something specific here: I am not a plant physiologist or soil scientist, but I have the distinct impression that respiration is distributed through the ecosystem, so such a close focus on soils may be myopic. I am quite willing to change my opinion about the primacy of soil composition, but I would like to see some evidence that root/trunk/foliar respiration is a second (or third) order concern, particularly in the daytime.

We agree with the reviewer that we should have been more clear in our explanation on why we think that the ERr is mainly formed by the soil. To elaborate further on this, we added more details in the text in line 585 and refer to the study of Hilman et al (2022).

33. Line 594: Perhaps I'm confused, but since ER_r is assumed to be the same day and night, then ER_a is zero at night. This means the comparison is really day-only vs. day-only + transitions. Perhaps you should present the two methods this way (rather than "day" and "all day").

We agree with the reviewer that our ERa for 24 hours is also the ERa with the day-only + transitions, as the ERa is zero during the night. We decided to present the ERa as all day/24 hours instead of day-only + transitions to be more certain of our final ERa value. The transition periods are the most uncertain as here the gradient becomes very close to zero and even switches sign. To make sure that the average NEE and TER fluxes are more robust we therefore included also the nighttime measurements. For GPP it does not matter as the GPP flux during the night is zero. For the ERa calculations, we need the NEE, TER and GPP fluxes. When these fluxes can be calculated with a lower uncertainty, the uncertainty of ERa also goes down. In principle the ERa should be the same between day + transitions and 24 hours and therefore it is better to choose the method that gives the lowest uncertainty.

Next to that, when deriving this value of the ERa for the entire day it is hopefully clear that this ERa is the ERa for this ecosystem throughout the diurnal cycle and can therefore be used for the O_2 method, presented in Figure 8. It links better with the ERforest of the entire day and hopefully makes it easier to understand how all the ER signals are linked.

34. Line 600: Unless there's a typo, the two methods give results that are identical (not just close), and you couldn't tell them apart even if they had tiny uncertainties.

    We agree with the reviewer that both the ERa calculations have the same result. After adding some references, we now expanded our explanation that we can assume that ERa stays constant over the day. As other studies shows that the ERa only changes with major environmental changes, which did not happen during our period of days used of the representative day.

35. Line 630: See comments above (lines 478 and 485)

    We think the word 'entrainment' is right here. For a more detailed explanation please see our reply to major point 4 above and our answer to reviewer #1.

Citation: https://doi.org/10.5194/acp-2022-504-RC3

References:

Angert, A., Yakir, D., Rodeghiero, M., Preisler, Y., Davidson, E. A., & Weiner, T. (2015). Using O 2 to study the relationships between soil CO 2 efflux and soil respiration. *Biogeosciences*, *12*(7), 2089–2099. https://doi.org/10.5194/bg-12-2089-2015

Bloom, A. J. (2015). Photorespiration and nitrate assimilation: A major intersection between plant carbon and nitrogen. *Photosynthesis Research*, *123*(2), 117–128. https://doi.org/10.1007/s11120-014-0056-y

Fischer, S., Hanf, S., Frosch, T., Gleixner, G., Popp, J., Trumbore, S., & Hartmann, H. (2015). Pinus sylvestris switches respiration substrates under shading but not during drought. *New Phytologist*, *207*(3), 542–550. https://doi.org/10.1111/nph.13452

Hilman, B. (2022). *The Apparent Respiratory Quotient of Soils and Tree Stems and the Processes That Control It Journal of Geophysical Research : Biogeosciences*. https://doi.org/10.1029/2021JG006676

Kulmala, L., Pumpanen, J., Kolari, P., Dengel, S., Berninger, F., Köster, K., Matkala, L., Vanhatalo, A., Vesala, T., & Bäck, J. (2019). Inter- and intra-annual dynamics of photosynthesis differ between forest floor vegetation and tree canopy in a subarctic Scots pine stand. *Agricultural and Forest Meteorology*, *271*(February), 1–11. https://doi.org/10.1016/j.agrformet.2019.02.029

Pickers, P. A., Manning, A. C., le Quéré, C., Forster, G. L., Luijkx, I. T., Gerbig, C., Fleming, L. S., & Sturges, W. T. (2022). Novel quantification of regional fossil fuel CO 2 reductions during COVID-19 lockdowns using atmospheric oxygen measurements. In *Sci. Adv* (Vol. 8). https://icos-cp.

Pries, C. H., Angert, A., Castanha, C., Hilman, B., & Torn, M. S. (2020). *Using respiration quotients to track changing sources of soil respiration seasonally and with experimental warming*. 3045–3055.

Seibt, U., Brand, W. A., Heimann, M., Lloyd, J., Severinghaus, J. P., & Wingate, L. (2004). Observations of O2: CO2 exchange ratios during ecosystem gas exchange. *Global Biogeochemical Cycles*, *18*(4), 1–18. https://doi.org/10.1029/2004GB002242

Wu, Z. Y., Zhang, L., Wang, X. M., & Munger, J. W. (2015). A modified micrometeorological gradient method for estimating O3 dry depositions over a forest canopy. *Atmospheric Chemistry and Physics*, *15*(13), 7487–7496. https://doi.org/10.5194/acp-15-7487-2015