Dear Dr. Bowling,

We appreciated the constructive comments of the reviewers. We have addressed the comments below. Reviewer/editor comments are shown in bold with our responses in blue. Line numbers refer to the tracked changes manuscript, and changes to the text are underlined.

Reviewer 1

Summary:

In this paper Byrne et al. evaluate the seasonal distribution of NEE at high latitudes using a combination of atmospheric CO2 measurements to inform model inversions of net C exchange in combination with satellite estimates of GPP to infer respiration. They note that anomalously low NEE in autumn can be attributed to greater Rh release. They also note a mismatch between their estimates and those derived from land surface models. They then provide an explanation whereby temperature lags within the soil can explain a certain fraction of this enhanced autumn respiration. This was a nice paper and could be publishable with some additional analysis and consideration of assumptions.

General Comments:

This paper profits from the high resolution XCO2 measurements which now allow us to estimate net CO2 fluxes at high resolution within specific bioregions and uncover different processes that may be affecting these seasonal fluxes. I found this analysis to be quite thorough and pretty convincing; however, I did have a few general comments. It seems as though the mismatch between observations and models may be dependent on the seasonal estimates of CUE and their assumptions. Perhaps it would be useful to look at total ecosystem respiration (GPP-NEE= TER) initially to see if the same mismatch is evident, this would suggest that the mismatch is not an artifact of the unique CUE applied over this region. Alternatively, one could use an independent estimate of CUE from an independent model (Konings et al. 2019) or use the same seasonal CUE for all regions.

We agree that the CUE is a very important parameter in decomposing NEE between NPP and Rh, and that there is considerable uncertainty in this value. We attempted to minimize the impact of this by applying the CUE estimates from the TRENDY ensemble, and thus applying the same CUE estimate to go from GPP to NEE (on average). We also propagated uncertainty from CUE (based on the TRENDY ensemble spread) into the NPP and Rh estimates (described in Sec. 2.4 and Sec. 2.5). Shortcomings of this approach are also discussed in Sec. 4.2 (L445-454). Still, we agree that it would be useful to the reader to present the more direct GPP and TER estimates. Therefore, we have added a plot of these fluxes to Sec. 3.1 (revised Fig. 2) along with the following accompanying text:

L257-271: "To further investigate the causes of differences in NEE between the TRENDY and <u>v9 OCO-2 MIP</u> ensembles, we separately examine<u> component primary productivity and respiration fluxes. As the most direct decomposition, we employ the data-driven GPP estimates to decompose NEE into GPP and terrestrial ecosystem respiration fluxes (TER) (Fig. 2). This</u>

comparison shows that the TRENDY ensemble mean GPP tends to overestimate the data-driven GPP during the autumn (Sep–Nov), largely explaining the mismatch in NEE during this season. For TER, we find good agreement for over the Warm region except for an underestimate of TER for the TRENDY ensemble mean during the summer (mirroring GPP). For the Mid regions, agreement is found between the TRENDY and data-driven TER estimates throughout the growing season. For the Cold region, we find that the TRENDY ensemble mean suggested greater TER during May–Aug, which drives the mismatch found in NEE.

We next decompose NEE into component NPP and Rh fluxes. These estimates require an additional assumption about the CUE in comparison to the GPP/TER decomposition, but also have the potential to provide more process understanding. As described in Sec. 2.4, we employ the monthly CUE estimates from the ensemble of TRENDY models, this both allows an "apples-to-apples" comparison with the TRENDY models as the CUE estimates are consistent between the data-driven and TRENDY estimates, and allows us to propagate uncertainty in CUE from the ensemble spread. The data-driven NPP and TRENDY NPP are shown in Fig. 3(d-f). The seasonality in NPP between the data-driven and TRENDY estimates show good agreement for all regions.

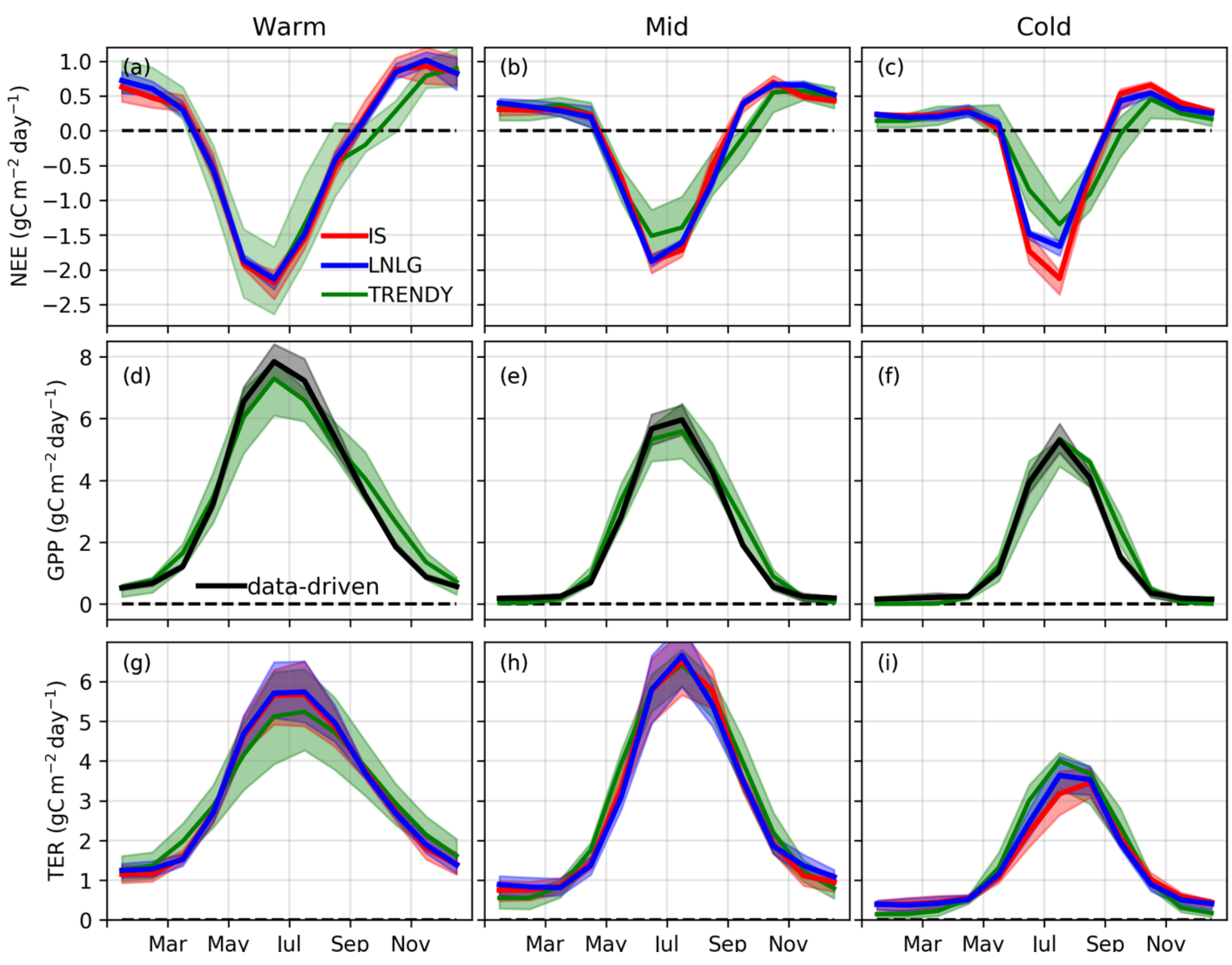

Figure 2. Monthly carbon cycle fluxes (average of 2015, 2016 and 2018; 2017 is excluded due to an OCO-2 data gap). (a-c) Mean (solid line) and interquartile range (shaded area) of NEE for the ensemble of IS (red) and LNLG (blue) v9 OCO-2 MIP and for the TRENDY ensemble (green). (d-f) GPP for the TRENDY ensemble (green) and data-driven datasets (black). (g-i) TER simulated by the TRENDY ensemble (green) and calculated from combining the data-driven GPP with the IS (red) and LNLG (blue) v9 OCO-2 MIP NEE constraints."

Furthermore, see recent analysis on Siberian warming where there is a strong relationship between spring GPP and fall TER (Kwon et al. 2021). Although the timespan for the OCO-2 inversions is too short, this citation on seasonal anomalies may help to put these results in a longer temporal context.

We have noted this and added a reference in the Discussion section (Sec. 4.1):

L422-424: "These inferred changes in Rh may in part be related to warming-induced changes in the seasonality of GPP (Liu et al., 2020; Kwon et al., 2021), but more research is needed to determine the impact of these different drivers."

I also had some comments on the soil model comparisons with the observation constrained estimates. The text (line 323) discusses regressions and statistics of those regressions, but the actual figure shows seasonal distributions from models and observations. The figure is pretty clean and easy to interpret, but the paper could benefit from a table in the main text that include your statistics, in addition to standard model performance statistics such as RMSE, MAE, and bias statistics.

We have added Table 1 to the manuscript that provides statistics on the single-layer model fits.

Table 1. Statistics on the data-model fits for the single layer models.

Region	Experiment	Slope	Intercept ($gCm^{-2}day^{-1}$)	R^2	Standard Error (SE) ($gCm^{-2}day^{-1}$)
Warm	$\alpha e^{\beta T_{surf}}$	0.94	0.11	0.89	0.067
Warm	$\alpha e^{\beta T_{1m}}$	0.91	0.14	0.93	0.051
Warm	$\alpha(t) e^{\beta T_{surf}}$	0.94	0.10	0.92	0.057
Mid	$\alpha e^{\beta T_{surf}}$	1.03	0.00	0.84	0.091
Mid	$\alpha e^{\beta T_{1m}}$	0.88	0.09	0.94	0.046
Mid	$\alpha(t) e^{\beta T_{surf}}$	1.04	-0.01	0.89	0.074
Cold	$\alpha e^{\beta T_{surf}}$	1.08	-0.01	0.66	0.16
Cold	$\alpha e^{\beta T_{1m}}$	1.04	-0.01	0.88	0.08
Cold	$\alpha(t) e^{\beta T_{surf}}$	1.11	-0.03	0.77	0.13

The model could also be tested against the eddy flux data estimates of Rh and these values could be reported in the table. This would help the reader evaluate which models are indeed superior.

We have added fits to the FLUXNET seasonal cycle of Rh. We find consistent results with the analysis for the cold region. We have added a supplementary figure and table to show these results, and have added the following text to the main manuscript:

L360-365: "To further confirm that $R_h(\alpha_c,T_{1m})$ best captures the seasonality of R_h, we fit these same models to seasonal FLUXNET R_h averaged over cold sites. This is a rather rough comparison as we drive the models with soil temperatures averaged over the Cold region rather than site specific datasets (due to absence of soil temperature data). Figure S16 shows the resulting fits and Table S2 gives the statistics of the fits. We find that $R_h(\alpha_c,T_{1m})$ performs best (R^2 = 0.96, SE = 0.08 gCm^{-2} day^{-1}), while $R_h(\alpha_t,T_{surf})$ performs second best (R^2 = 0.85, SE = 0.19 gCm^{-2} day^{-1}) and $R_h(\alpha_c,T_{surf})$ gives the poorest performance (R^2 = 0.75, SE = 0.25 gCm^{-2} day^{-1}), consistent with the regional-scale data-driven analysis."

Table S2. Statistics on the data-model fits for the single layer models against the FLUXNET inferred R_h seasonal cycle.

Experiment	Slope	Intercept ($gCm^{-2}day^{-1}$)	R^2	Standard Error ($gCm^{-2}day^{-1}$)
$\alpha e^{\beta T_{surf}}$	1.36	-0.14	0.75	0.25
$\alpha e^{\beta T_{1m}}$	1.28	-0.14	0.96	0.08
$\alpha(t)e^{\beta T_{surf}}$	1.4	-0.17	0.85	0.19

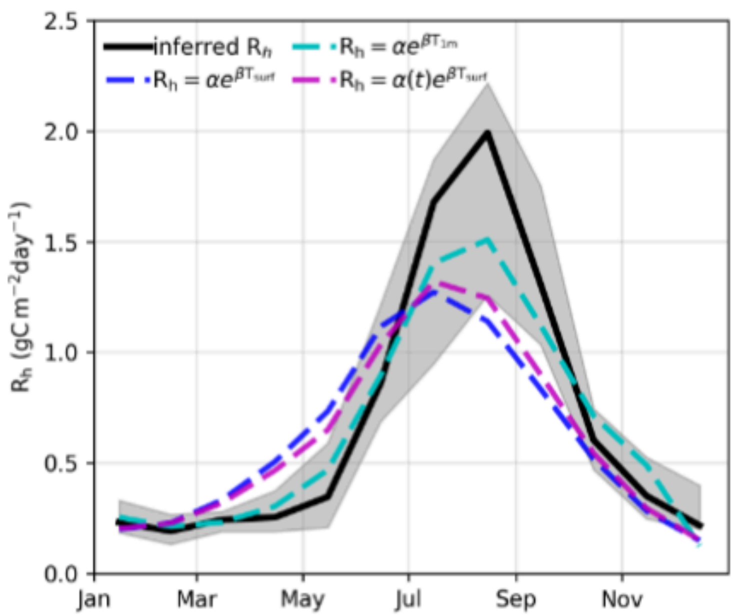

Figure S13. Mean and range in inferred monthly FLUXNET R_h with fits for single-layer R_h models that employ (navy dash) T_{surf} dependence and no seasonal variations in the carbon pool, (cyan dash) T_{1m} dependence and no seasonal variations in the carbon pool and (magenta dash) T_{surf} dependence and seasonal variations in the carbon pool.

Also litterfall estimates seem an order of magnitude too high in Fig. s15 should peak at ~2 TgC day-1 as compared to NPP estimates in Fig. 3. This may just be a units problem but check the model.

We had an error in the model and description. The annual total litterfall is defined to be equal to the annual total Rh, so that the labile carbon pool is in steady state. Thus, the monthly litterfacll fluxes are equal to "f_{npp}*sum(R_h)", where s sum(R_h) is the total Rh over a year. Please see the revised Appendix 1. The corrected figure is shown below (note that changes are very small).

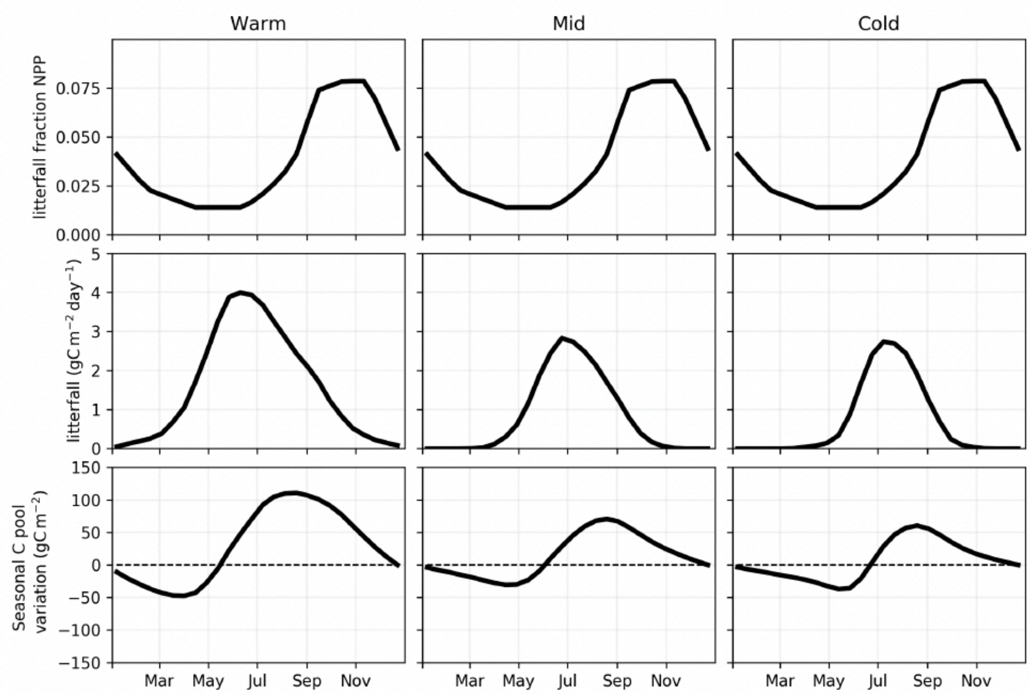

Figure S16. (top) Fraction of NPP that becomes litterfall. (middle row) Carbon flux from litterall. (bottom) Seasonal variations in the labile carbon pool due to litterfall and R_h.

Specific comments:

L26: more of a synthesis than a conventional meta-analysis

Fixed

L46: sparse

Fixed

L47: Comane et al. used a top-down approach as well, but no satellite estimates based on my recollection

Yes, that is correct. They perform a more regional analysis with measurements from Barrow tower and aircraft campaigns.

L51: seems like you need more of a justification of why Eurasia other than just the sparsity of data. See Bastos et al. on attribution of the enhanced seasonal cycle to Eurasia: https://doi.org/10.5194/acp-19-12361-2019

Added additional context:

L50-53: "We utilize these data to investigate carbon cycle dynamics over three large regions within Eurasia (Fig. 1), which are defined based on the east-west temperature gradient (see Sec. 2.1), with the coldest region in the east and warmest region in the west. We focus on Eurasia as much of this region has particularly sparse site-level observations yet is experiencing rapid change (Liu et al., 2020; Bastos et al., 2019)."

L90: what is this resolution state here.

The spatial resolution differs for each dataset, that is why we state them when introducing the specific datasets.

L145: are all these soundings weighted equally or do some higher error or bias?

Yes, clarified.

L156: why did you include land use? This seems unnecesary for your research questions and adds a potentially confounding factor.

The data-driven estimates of NEE and GPP will be impacted by land use, therefore we compare to this experiment as it relates most closely to reality.

L163-164: how much spatial variability is there among these CUE estimates?

This can be seen from comparing the regions in Fig. S5. The CUE falls to a similar value of ~0.5 during the growing season.

Eq 4: omit already given as equation 1

Done

L206-207: this of course could change considerably with surface water pooling happens above poorly drained permafrost soils

Yes, changes in the thermal regime could have these knock-on impacts.

L218-219: once again depend on the CUE estimates and assumptions

Please see the response to the general comments

L253: NPP estimates largely

fixed

L257-258: of course all of the unexplained variance is ascribed to the inferred term in the budget or Rh

This is correct. However, we propagate uncertainties in each quantity (GPP, NEE, CUE) such that systematic differences should be attributable to Rh.

L297-299: this dip in respiration could be due to the timing of snow melt as well that insulates the soil and promotes Rh: https://doi.org/10.1111/geb.12441

We have added this as a possible mechanism to the manuscript:

L321-324: "A potential mechanistic explanation for a spring pulse of Rh could be due to thawing soils that release CO2 that has been trapped within subsurface soil layers over the winter (see Sec. 4). Another plausible mechanism could be the timing of snow melt, which may insulate the soil over winter (Yu et al., 2016). However, the signal from this first peak is small relative to the uncertainties."

L318: should probably only consider the top 0.5 m considering below that is below zero year round in the cold region

We feel that using the top 1m is a reasonable average. The temperature is not below zero year-round for 50-100 cm. In fact, even the 50-200 cm interval is above zero from Jul-Nov, as shown in Fig. 4.

L322: these just seem to be the seasonal cycles with no regression or RMSE statistics reported

We have added regression statistics as described in response to the general comments.

L326: it would be helpful if you used the same terms to describe the same experiments in the figure and the text.

We have revised this section to give consistent model names in the figure and text. The experiments are now named "$R_h(\alpha_c, T_{surf})$", "$R_h(\alpha_c, T_{1m})$", and "$R_h(\alpha_t, T_{surf})$".

L339-340: do they incorporate differences in litterfall as a function of NPP?

Yes, we have clarified this in the text:

L375: "… and simulated to a depth of 300 cm using a dynamic carbon pool (D300cm, due to dynamic litterfall inputs and Rh outputs)."

L357-359: this could be tested with more years of data where litterfall may vary as a function of inter-annual variability in NPP see Kwon et al. https://doi.org/10.1088/1748-9326/ac358b

Agreed, however, we have a relatively temporally limited record from OCO-2. Further, we find that the sensitivity of these inversions to interannual variations in NEE at high latitudes was quite limited (not shown).

L422-423: this is a good idea. Have Schur et al. measured 14CO2 seasonally or just annually?

We are unaware of the density of existing ^{14}C measurements at high latitudes, but we agree that this would be something to follow-up on in future research.

References:

Konings, Alexandra G., A. Anthony Bloom, Junjie Liu, Nicholas C. Parazoo, David S. Schimel, and Kevin W. Bowman. 2019. "Global Satellite-Driven Estimates of Heterotrophic Respiration." ***Biogeosciences*** **16 (11): 2269–84.**

Kwon, Min Jung, Ashley Ballantyne, Philippe Ciais, Ana Bastos, Frédéric Chevallier, Zhihua Liu, Julia K. Green, Chunjing Qiu, and John S. Kimball. 2021. "Siberian 2020 Heatwave Increased Spring CO2 Uptake but Not Annual CO2 Uptake." ***Environmental Research Letters: ERL [Web Site]*** **16 (12): 124030.**

Reviewer 2

The manuscript by Byrne et al is an analysis of seasonal cycles of carbon fluxes over northern ecosystems, with a comparison of inversion results of NEE against DGVM estimates of NEE, and then a comparison of the inversion NEE estimates minus GPP and an estimate of Ra to infer Rh, against DGVM estimates of that as well. The main findings are (a) that NEE late-season positive fluxes are higher in the observations than the models, and (b) that the inferred Rh seasonality indicates that the DGVMs underestimate that late-season Rh.

Overall, I think the comparison between inversion results and models is really useful, and the paper should be published. But I find it an interesting but not entirely satisfying analysis. One problem is that the number of different steps from NEE to Rh seems like it introduces the potential for several errors to creep in, particularly as relate to Ra.

We agree that the CUE is a very important parameter in decomposing NEE between NPP and Rh, and that there is considerable uncertainty in this value. We attempted to minimize the impact of this by applying the CUE estimates from the TRENDY ensemble, and thus applying the same CUE estimate to go from GPP to NEE (on average). We also propagated uncertainty from CUE (based on the TRENDY ensemble spread) into the NPP and Rh estimates (described in Sec. 2.4 and Sec. 2.5). Shortcomings of this approach are also discussed in Sec. 4.2 (L445-454). Still, we agree that it would be useful to the reader to present the more direct GPP and TER estimates. Therefore, we have added a plot of these fluxes to Sec. 3.1 (revised Fig. 2) along with the following accompanying text:

L257-271: “To further investigate the causes of differences in NEE between the TRENDY and v9 OCO-2 MIP ensembles, we separately examine component primary productivity and respiration fluxes. As the most direct decomposition, we employ the data-driven GPP estimates to decompose NEE into GPP and terrestrial ecosystem respiration fluxes (TER) (Fig. 2). This comparison shows that the TRENDY ensemble mean GPP tends to overestimate the data-driven GPP during the autumn (Sep–Nov), largely explaining the mismatch in NEE during this season. For TER, we find good agreement for over the Warm region except for an underestimate of TER for the TRENDY ensemble mean during the summer (mirroring GPP). For the Mid regions, agreement is found between the TRENDY and data-driven TER estimates throughout the growing season. For the Cold region, we find that the TRENDY ensemble mean suggested greater TER during May–Aug, which drives the mismatch found in NEE.

We next decompose NEE into component NPP and Rh fluxes. These estimates require an additional assumption about the CUE in comparison to the GPP/TER decomposition, but also have the potential to provide more process understanding. As described in Sec. 2.4, we employ the monthly CUE estimates from the ensemble of TRENDY models, this both allows an “apples-to-apples” comparison with the TRENDY models as the CUE estimates are consistent between the data-driven and TRENDY estimates, and allows us to propagate uncertainty in CUE from the ensemble spread. The data-driven NPP and TRENDY NPP are shown in Fig. 3(d-f). The seasonality in NPP between the data-driven and TRENDY estimates show good agreement for all regions.

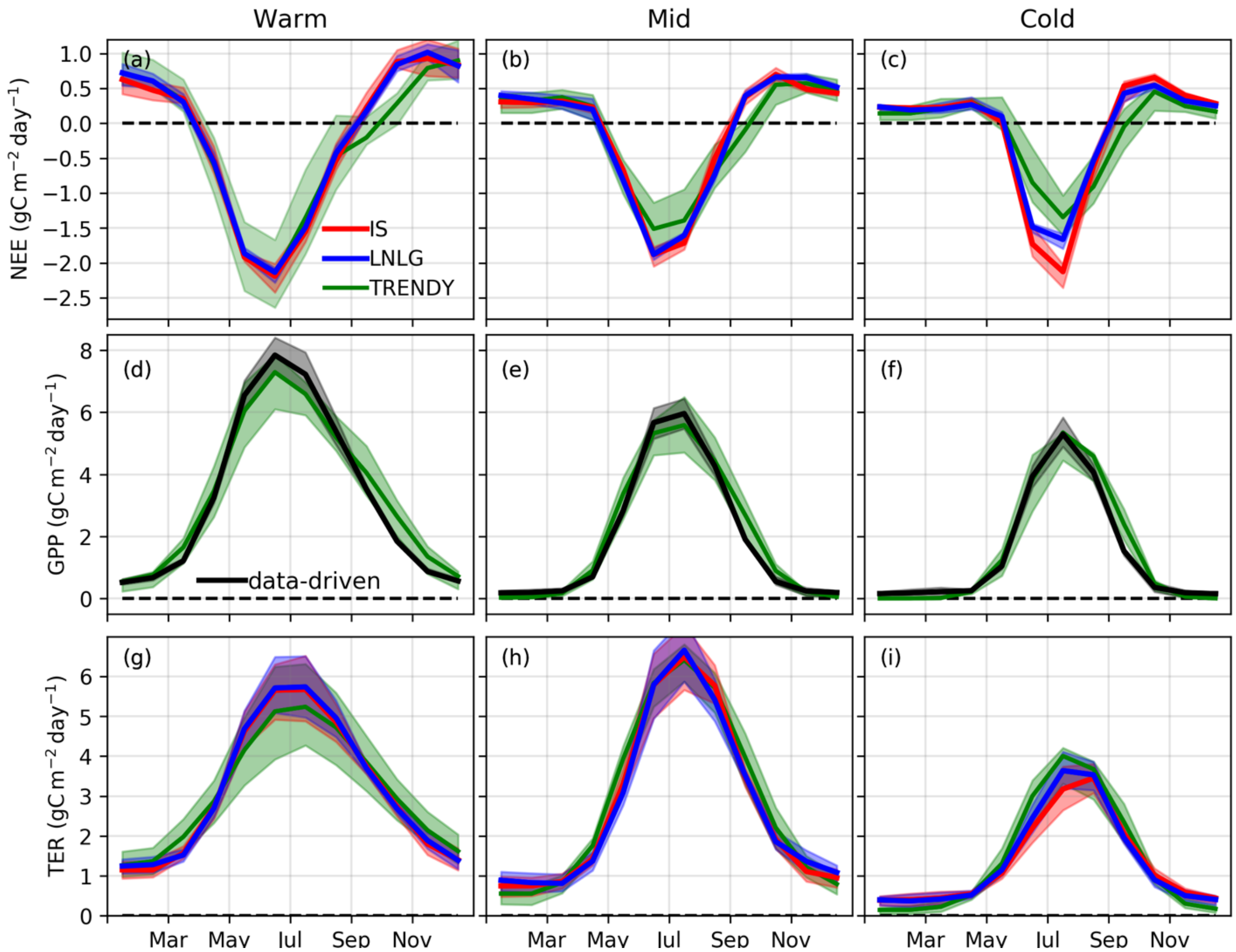

Figure 2. Monthly carbon cycle fluxes (average of 2015, 2016 and 2018; 2017 is excluded due to an OCO-2 data gap). (a-c) Mean (solid line) and interquartile range (shaded area) of NEE for the ensemble of IS (red) and LNLG (blue) v9 OCO-2 MIP and for the TRENDY ensemble (green). (d-f) GPP for the TRENDY ensemble (green) and data-driven datasets (black). (g-i) TER simulated by the TRENDY ensemble (green) and calculated from combining the data-driven GPP with the IS (red) and LNLG (blue) v9 OCO-2 MIP NEE constraints."

Second, there are any number of reasons why the DGVMs could show a bias relative to the observations, and it is certainly possible that the lack of deep-soil respiration is one reason. But the attempt to provide a mechanistic explanation here using a simple model is not very clear, and subject to somewhat arbitrary choices like how to handle substrate seasonality. I wonder if a slightly different approach of looking at the DGVMs themselves, and asking whether there are structural or parametric characteristics of the models that govern the shapes of their seasonal cycles, and which might provide some clues for identifying whether any of them do a better or worse job than others?

We agree that it would be preferable to identify model differences across the TRENDY ensemble that can explain the data-model differences. We attempted to perform such an analysis but it was unsuccessful. One of the challenges is that the models general differ in many ways, thus isolating specific factors in very challenging. This is why we have adopted the simpler approach of using an idealized one-layer diagnostic model, and one diagnostic model run with different set-ups. We address the challenges if identifying mechanisms driving these differences across the TRENDY ensemble in the discussion section:

L390-404: "Over the cold northeastern region of Eurasia, our data-driven Rh seasonal cycle allocates 64–70% of annual CO_2 emissions to outside of the summer (August - April) compared to only 52% of annual Rh emissions allocated by the TRENDY DVGMs to this period. The reason that the TRENDY models do not capture this seasonality is unclear. A plausible explanation is that the TRENDY models do not capture the contribution of subsurface layers to R_h, especially during the zero-curtain period. This is clearly the case for the subset of TRENDY models that drive R_h with air temperature. However, it is unclear if this is an important factor for models with more sophisticated soil modules. Surprisingly, a preliminary analysis did not find a relationship between model complexity and agreement with the data-driven estimate. The drivers of differences from the data-driven estimate may differ between models, and be impacted by the interplay of litterfall phenology, R_h formulation (Peylin et al., 2005), and number of soil layers, among other factors. Some potential areas of focus for improving models may be gleaned from recent studies. Seiler et al. (2021) suggest that the TRENDY models may systematically underestimate soil organic carbon at high latitudes, which could contribute to an underestimate of subsurface R_h across the models. Endsley et al. (2021) found a similarly phased bias in simulated R_h by the Terrestrial Carbon Flux (TCF) model against flux tower R_h to that reported here. They show that this bias could be largely mitigated by adding seasonally varying litterfall phenology, an O_2 diffusion limitation on R_h and a vertically resolved soil decomposition model, suggesting these may be foci for model improvements."

Line 12: please provide uncertainty range for the DGVM estimates, as you do for the data-driven estimates

Done:

Furthermore, we show that this seasonality of NEE and Rh over northeastern Eurasia is not captured by the TRENDY v8 ensemble of dynamic global vegetation models (DGVMs), which estimate that 47–57% (interquartile range) of annual Rh occurs during Aug-Apr, while the data-driven estimates suggest 59–76% of annual Rh occurs over this period.

Line 17: "is not well captured by current DGVMs." Any DGVMs, or just the ensemble mean?

We believe the current text, "… suggests that autumn Rh from subsurface soils in the northern high latitudes is not well captured by current DGVMs", to be accurate. The differences examined here are generally considered within uncertainties based on the interquartile range, thus we could consider differences to be robust at this level. We did look at individual models when trying to better understand mechanisms (not reported in the study due to inconclusive results), and all models showed deficiencies in reproducing the data-driven Rh seasonal cycle across the regions, consistent with this general statement.

Line 70-72: Could you clarify whether you are using monthly mean CUE values or annual mean values here?

Monthly, clarified.

Lines 99-100. How confident are we in the soil temperature predictions of these models? There have been a few analyses of the soil temperature dynamics and permafrost statistics of climate models at high latitudes. Does this set represent a set of best-performing soil temperature models?

We compared the ensemble to the reanalysis datasets from ERA5 and MERRA2, and they generally reproduced similar seasonality (Fig. S14). We also represent the interquartile spread among the models, shown by the shaded area.

In addition, we compared the MERRA2 soil temperature seasonality to borehole measurements (Text S2 and Fig. S13) and found good agreement overall. Given that the models generally show agreement with MERRA2 soil temperature, we can conclude that the model seasonality is largely consistent with the borehole data as-well.

Fig. 2. I think that the per-area fluxes are more meaningful here, otherwise the reader gets the suggestion that NEE is higher during the summer in the colder than the warmer regions, which is confusing. So I suggest switching fig. S7 and fig. 2, and in general reporting things per unit area.

We have changed the reported units to gC m^{-2} day^{-1}.

Figs 2 & S7: I am skeptical about the errors introduced by the GPP -> NPP conversion, I think it would be useful to include a set of GPP panels as well, since, like NEE, that is the most directly observed, with the NPP and RH much less direct.

Please see our response to the first comment.

Figs 2 and S7: A lot of the focus of the discussion is on the autumn differences, but I wonder if the more general problem is that the winter respiration in general is underestimated by the models in the cold region. This would be consistent with the findings of Natali et al., but given the larger-scale datasets used here would still be an important point to emphasize here.

We have decided against emphasizing differences during the winter as this season is less well informed by the CO_2 measurements due to small overall fluxes and sparse sampling (e.g., Byrne et al., 2017). Thus, some of the differences in the winter might be impacted by differences in the prior fluxes rather than being informed from the observational constraints. Furthermore, the absolute differences in NEE are much larger during the growing season.

Byrne, B., Jones, D. B. A., Strong, K., Zeng, Z.-C., Deng, F., and Liu, J.: Sensitivity of CO_2 Surface Flux Constraints to Observational Coverage, J. Geophys. Res.-Atmos, 112, 6672–6694, https://doi.org/10.1002/2016JD026164, 2017

Lne 264: This isn't really a shift, so much as a bias in TRENDY relative to the observations?

We were referring to a shift in the seasonality, that is, a large fraction of Rh occurs during May-Jul relative to the rest of the year.

Lines 264-270 and fig. S12. I think FLUXNET is actually telling a different story than the larger-scale datasets. The TRENDY models actually have a higher positive NEE anomaly during the shoulder season than FLUXNET, which is the opposite pattern shown in fig. 2c. If this is correct, then I think the discussion of this result needs to be revised accordingly.

It is true that there are some differences in the NEE and GPP relative to the regional estimates. However, the we are referring specifically to the R_h seasonality, which peaks later in the year than the TRENDY ensemble predicts. In the revised manuscript, we have added fits with the one-layer model to the FLUXNET-based R_h seasonal cycle, and find consistent results with the analysis for the cold region. We have added a supplementary figure and table to show these results, and have added the following text to the main manuscript:

L360-365: "To further confirm that $R_h(\alpha_c,T_{1m})$ best captures the seasonality of R_h, we fit these same models to seasonal FLUXNET R_h averaged over cold sites. This is a rather rough comparison as we drive the models with soil temperatures averaged over the Cold region rather than site specific datasets (due to absence of soil temperature data). Figure S16 shows the resulting fits and Table S2 gives the statistics of the fits. We find that $R_h(\alpha_c,T_{1m})$ performs best ($R^2 = 0.96$, SE = 0.08 $gCm^{-2}\ day^{-1}$), while $R_h(\alpha_t,T_{surf})$ performs second best ($R^2 = 0.85$, SE = 0.19 $gCm^{-2}\ day^{-1}$) and $R_h(\alpha_c,T_{surf})$ gives the poorest performance ($R^2 = 0.75$, SE = 0.25 gCm^{-2} day^{-1}), consistent with the regional-scale data-driven analysis."

Table S2. Statistics on the data-model fits for the single layer models against the FLUXNET inferred R_h seasonal cycle.

Experiment	Slope	Intercept ($gCm^{-2}day^{-1}$)	R^2	Standard Error ($gCm^{-2}day^{-1}$)
$\alpha e^{\beta T_{surf}}$	1.36	-0.14	0.75	0.25
$\alpha e^{\beta T_{1m}}$	1.28	-0.14	0.96	0.08
$\alpha(t)e^{\beta T_{surf}}$	1.4	-0.17	0.85	0.19

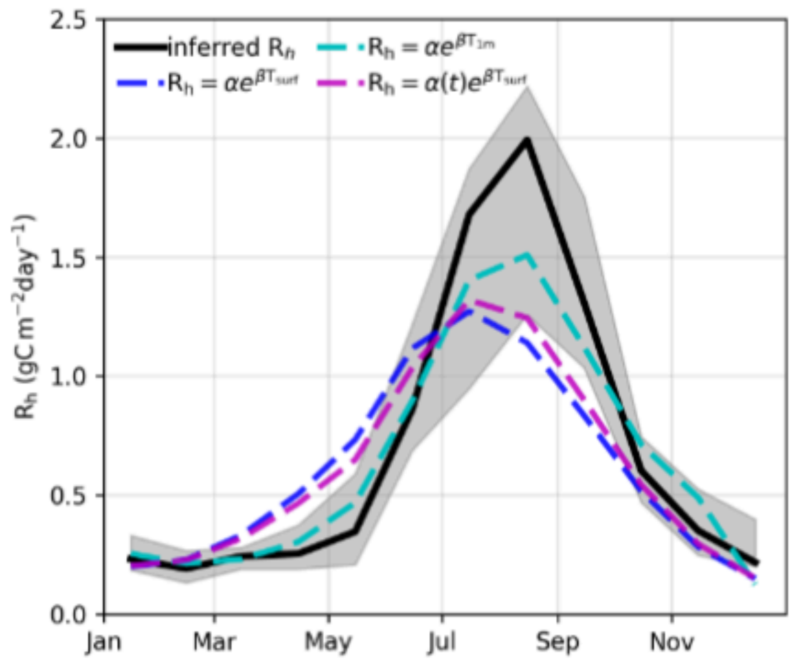

Figure S13. Mean and range in inferred monthly FLUXNET R_h with fits for single-layer R_h models that employ (navy dash) T_{surf} dependence and no seasonal variations in the carbon pool, (cyan dash) T_{1m} dependence and no seasonal variations in the carbon pool and (magenta dash) T_{surf} dependence and seasonal variations in the carbon pool.

Section 3.2, I'm not sure I understand what new information the 14-day-resolved data provides beyond what is in the monthly data. Is this analysis really necessary? If so, could the authors give a bit better motivation and explanation?

We describe the motivation for this as:

L303-304: "This higher resolution better resolves temporal changes in CO_2 fluxes throughout the growing season, particularly during the shoulder seasons, when week-to-week changes in CO_2 fluxes are large (Parazoo et al., 2018a)."

We believe that this higher resolution is worthwhile to better resolve variations in the seasonal cycle of NEE, NPP and Rh.

Fig.3. I'm very skeptical about how narrow the range of uncertainty in panels a-c are here. What is that a measure of?

The shaded regions are showing the range between three different flux inversions, and is stated in the Figure 3 caption "Median and ensemble spread…". It is a measure of the precision with which flux inversion analyses can estimate fluxes over these regions.

Lines 334-348, and figure 4. I don't understand this sensitivity analysis, or why the seasonal cycles in panels g-i are so different from the ones in panels d-f. Could you clarify a bit more what is being shown here?

We have added additional text to better motivate this experiment:

L371-373: "We further investigate these mechanisms using a soil carbon decomposition model that can simulate seasonal and vertical variations in carbon pools (Sec. 2.6). This allows for a prognostic simulation of mechanisms driving the seasonality, in contrast to the diagnostic one-layer models."

And have refined the caption to better explain the experiments:

"… (g-i, top) Normalized seasonal cycle of Rh simulated by the soil decomposition model (Sec. 2.6). The different lines show different model simulations: D300cm employs a dynamic carbon pool over 0-300 cm depth, C300cm employs a constant carbon pool over 0-300 cm depth, D10cm employs a dynamic carbon pool over 0-10 cm depth). (g-i, bottom) Differences in simulated Rh between experiments."

Further, the argument about deep soil playing a greater role should help with the autumn respiration peak, but less so with the bias in respiration in the cold region throughout the winter. What does this analysis have to say about that?

We are reticent to focus on winter differences, as these are less well informed by the atmospheric CO_2 data and the absolute differences in NEE are relatively small.

Line 441. I don't understand the line "TRENDY v8 data were downloaded from trendy-v8@trendy.ex.ac.uk.", since that is an email address, not a URL. Please provide a URL or DOI to a FAIR-aligned data archive where the data can be freely downloaded or, if the data is not available, then per this journal's data policy, a detailed explanation of why this is the case is required.

The TRENDY data were downloaded and analyzed for this study, but these data were not generated by us and we do not control their data policy. This text has been re-worded to point to the TRENDY website:

TRENDY v8 gridded data can be accessed through the website https://sites.exeter.ac.uk/trendy

Multi-year observations reveal a larger than expected autumn respiration signal across northeast Eurasia

Brendan Byrne[1], Junjie Liu[1,2], Yonghong Yi[3,4], Abhishek Chatterjee[1], Sourish Basu[5,6], Rui Cheng[2], Russell Doughty[2,7], Frédéric Chevallier[8], Kevin W. Bowman[1,3], Nicholas C. Parazoo[1], David Crisp[1], Xing Li[9], Jingfeng Xiao[10], Stephen Sitch[11], Bertrand Guenet[12], Feng Deng[13], Matthew S. Johnson[14], Sajeev Philip[15], Patrick C. McGuire[16], and Charles E. Miller[1]

[1]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA
[2]Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CA, USA
[3]Joint Institute for Regional Earth System Science and Engineering, University of California, Los Angeles, CA, USA
[4]College of Surveying and Geo-Informatics, Tongji University, China
[5]Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, MD, USA
[6]Earth System Science Interdisciplinary Center, University of Maryland, College Park, MD, USA
[7]College of Atmospheric and Geographic Sciences, University of Oklahoma, Norman, OK USA
[8]Laboratoire des Sciences du Climat et de l'Environnement/IPSL, CEA-CNRS-UVSQ, Université Paris-Saclay, 91191 Gif-sur-Yvette, France
[9]Research Institute of Agriculture and Life Sciences, Seoul National University, Seoul, South Korea
[10]Earth Systems Research Center, Institute for the Study of Earth, Oceans, and Space, University of New Hampshire, Durham, NH, USA
[11]College of Life and Environmental Sciences, University of Exeter, Exeter EX4 4RJ, UK
[12]Laboratoire de Géologie, Ecole Normale Supérieure/CNRS UMR8538, IPSL, PSL Research University, Paris, France
[13]Department of Physics, University of Toronto, Toronto, Ontario, Canada
[14]NASA Ames Research Center, Moffett Field, CA, USA
[15]Centre for Atmospheric Sciences, Indian Institute of Technology Delhi, New Delhi, India
[16]Department of Meteorology and National Centre for Atmospheric Science, University of Reading, Reading, UK

Correspondence: Brendan Byrne (brendan.k.byrne@jpl.nasa.gov)

Abstract. Site-level observations have shown pervasive cold season CO_2 release across Arctic and boreal ecosystems, impacting annual carbon budgets. Still, the seasonality of CO_2 emissions are poorly quantified across much of the high latitudes due to the sparse coverage of site-level observations. Space-based observations provide the opportunity to fill some observational gaps for studying these high latitude ecosystems, particularly across poorly sampled regions of Eurasia. Here, we show that data-driven net ecosystem exchange (NEE) from atmospheric CO_2 observations implies strong summer uptake followed by strong autumn release of CO_2 over the entire cold northeastern region of Eurasia during the 2015–2019 study period. Combining data-driven NEE with satellite-based estimates of gross primary production (GPP), we show that this seasonality implies less summer heterotrophic respiration (R_h) and greater autumn R_h than would be expected given an exponential relationship between respiration and surface temperature. Furthermore, we show that this seasonality of NEE and R_h over northeastern Eurasia is not captured by the TRENDY v8 ensemble of dynamic global vegetation models (DGVMs), which estimate that ~~only 52%~~ 47–57% (interquartile range) of annual R_h occurs during Aug-Apr, while the data-driven ~~estimate suggests 64–70~~estimates suggest 59–76% of annual R_h occurs over this period. We explain this seasonal shift in R_h by respiration from soils at depth during the zero curtain period, when sub-surface soils remain unfrozen up to several months after the surface has frozen. Additional impacts of physical processes related to freeze-thaw dynamics may contribute to the seasonality of R_h. This study confirms a significant and spatially extensive early cold season CO_2 efflux in the permafrost rich region of northeast Eurasia, and suggests that autumn R_h from subsurface soils in the northern high latitudes is not well captured by current DGVMs.

1 Introduction

Boreal and Arctic ecosystems hold vast quantities of soil carbon and play an important role in the global carbon cycle (Schuur et al., 2015). These ecosystems are also experiencing the most rapid climate change (Overland et al., 2018), driving major changes in the carbon cycle, including: greening trends (Park et al., 2016), permafrost thaw (Schuur et al., 2015; Turetsky et al., 2019, 2020), and increased fire frequency and intensity (Veraverbeke et al., 2017, 2021). Yet, the impact of these changes on the carbon budget of the region remains uncertain (Schuur et al., 2015; McGuire et al., 2018; Miner et al., 2022). In part, this is due to sparse site level observations in boreal and Arctic ecosystems, while the limited available observations of high latitude ecosystems are providing surprises.

A ~~meta-analysis~~ synthesis of Arctic and boreal site-level flux measurements from the literature found pervasive CO_2 release during the cold season (Natali et al., 2019), such that the cold season is not a dormant period but strongly impacts annual carbon budgets (Zimov et al., 1993; Björkman et al., 2010; Natali et al., 2019). Particularly strong releases of CO_2 have been observed during the early cold season (Commane et al., 2017; Mastepanov et al., 2013; Jeong et al., 2018). This has been linked to the "zero-curtain effect", wherein the air and surface temperatures drop below 0 °C but deeper soils remain unfrozen for an extended period due to latent heat release (Outcalt et al., 1990; Romanovsky and Osterkamp, 2000; Hinkel et al., 2001; Zona et al., 2016). The result is an "active layer" of unfrozen soil that can persist for months, resulting in greater respiration than would be expected based on air temperature. Both aircraft (Commane et al., 2017) and site-level (Mastepanov et al., 2013; Jeong et al., 2018) measurements have found substantial CO_2 release during the zero-curtain period over Alaska (Sep–Dec) that is not well captured by our current generation of Earth System Models (Commane et al., 2017). Similarly, CO_2 mole fractions enhancements within soils have been observed during the zero-curtain period (Wilkman et al., 2021; Raz-Yaseef et al., 2017). Mechanistically, both biological and physical processes likely contribute to the enhanced early cold season CO_2 release. Physically, freezing forces dissolved CO_2 out of solution (Bing et al., 2015), which may then be released through mechanical channels and fissures in the soil that form during freezing (Mastepanov et al., 2013; Pirk et al., 2015; Wilkman et al., 2021). Enhanced CO_2 effluxes (release to the atmosphere) have also been observed during the spring thaw (Raz-Yaseef et al., 2017; Arndt et al., 2020). This spring signal has been linked to a delayed release of CO_2 production from the previous early cold season (Raz-Yaseef et al., 2017), while a rapid warming and introduction of oxygen during snow melt has also been proposed as a contributor to this signal (Arndt et al., 2020). Finally, observed CO_2 effluxes during the middle of the cold season

(Natali et al., 2019) have been mechanistically linked to microbial respiration that persists at subzero bulk soil temperatures (Rivkina et al., 2000; Panikov et al., 2006; McMahon et al., 2009; Drotz et al., 2010), with a possible additional contribution from the diffusion of stored CO_2 that is produced during the non-frozen season (Natali et al., 2019).

Still, the full spatial extent and magnitude of cold season CO_2 release is not well characterized due to ~~sparsity of~~ sparse site-level observations~~, particularly over much of north Eurasia~~. Here, we employ a “top-down” approach to estimate the seasonal cycle of data-driven carbon fluxes using space-based observations during the period 2015–2019. This approach complements previous site-level analyses by providing CO_2 flux constraints on large continental-scale regions. We utilize these data to investigate carbon cycle dynamics over three large regions within Eurasia (Fig. 1), which are defined based on the east-west temperature gradient (see Sec. 2.1), with the coldest region in the east and warmest region in the west. We focus on Eurasia, as much of this region has particularly sparse site-level observations, yet is experiencing rapid change (Liu et al., 2020; Bastos et al., 2019) . We further compare the observationally-constrained seasonality of CO_2 fluxes to a suite of dynamic global vegetation models (DGVMs) from the TRENDY ensemble (Sitch et al., 2015) version 8 as used in the Global Carbon Budget 2019 (Friedlingstein et al., 2019) (Sec. 3.1). Our study addresses two main questions: (1) Do large-scale observational constraints support enhanced CO_2 effluxes during the shoulder seasons at high-latitudes? And if so, (2) what are the underlying mechanisms driving this behaviour?

We first examine the seasonality of net ecosystem exchange (NEE) constrained by atmospheric inversions of retrieved column-averaged dry-air mole fractions of CO_2 (X_{CO_2}) from the Orbiting Carbon Observatory 2 (OCO-2) (Crisp et al., 2017; Eldering et al., 2017) and by flask and in situ CO_2 measurements (Sec. 2.2). Monthly NEE is obtained from version 9 of the OCO-2 Model Inter-comparison Project (v9 OCO-2 ~~MIPv9~~MIP) (Peiro et al., 2021). In addition, we perform a set of three higher temporal resolution inversions using the CAMS, TM5-4DVar and CMS-Flux inversion systems to examine sub-monthly variability in CO_2 fluxes.

~~We then look to~~ We then decompose NEE into component fluxes to better understand the processes driving the seasonality of NEE. In particular, we decompose the data-driven NEE fluxes into net primary production (NPP) and heterotrophic respiration (R_h):

$$\mathrm{NEE} = \mathrm{R_h} - \mathrm{NPP}. \quad (1)$$

To do this, we ~~utilize~~ use four data-driven gross primary production (GPP) products: FLUXCOM (Jung et al., 2020), FluxSat (Joiner and Yoshida, 2020), the Vegetation Photosynthesis Model ~~(VPM) (Zhang et al., 2017)~~(VPM, Zhang et al., 2017), and the Global OCO-2-based SIF product ~~(GOSIF) (Li and Xiao, 2019)~~(GOSIF, Li and Xiao, 2019). These datasets ~~utilize MODIS~~use MODerate-resolution Imaging Spectroradiometer (MODIS) reflectances, OCO-2 solar induced fluorescence and reanalysis data to infer GPP, and thus provide an ensemble of global estimates of GPP to inform its uncertainty. NPP is estimated from GPP using the monthly carbon use efficiency (CUE) from the TRENDY models (Sec. 2.4) using the relationship:

$$\mathrm{NPP} = \mathrm{CUE} \times \mathrm{GPP}. \quad (2)$$

We then combine the data-driven estimates of NEE and NPP to recover a data-driven seasonal cycle of R_h (Sec. 2.5).

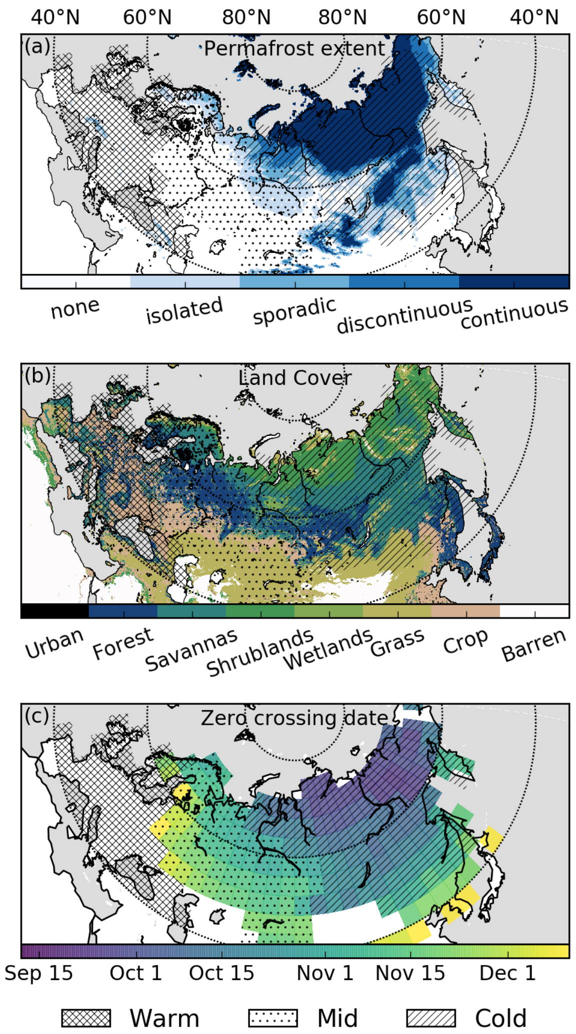

Figure 1. (a) Permafrost extent over 2000-2016 (Obu et al., 2019; Obu et al., 2018) (b) MODIS IGBP (MOD12C1 v6) land cover for urban areas, forest (tree cover >60% and height>2 m), savanna (tree cover 10–60% and height>2 m), shrublands (woody perennials cover >10% and height<2m), grasslands, croplands, and barren land. (c) zero-crossing date (date when the mean soil temperature drops below 0 °C) for the top 0.5 m of soil from the MERRA-2 Land dataset at $4^\circ \times 5^\circ$ spatial resolution. Gridcells with no shading do not have a zero-crossing date. Three regions are shown by different hatching patterns. The "Warm" (cross hatching) region does not have a zero-crossing date, the "Mid" (dots) region has a zero-crossing date after Oct 27, and the "Cold" (diagonal hatching) region has a zero-crossing date before Oct 27. Note that some adjustments from these definition are made so that the regions are contiguous. The Warm, Mid, and Cold regions have land areas of 5.66×10^6 km, 8.66×10^6 km, and 12.65×10^6 km, respectively.

~~We perform this analysis~~ This analysis is performed at two temporal resolutions. First, we leverage the large ensembles from TRENDY and the v9 OCO-2 ~~MIPv9~~ MIP that provide fluxes at monthly temporal resolution (Sec. 3.1). However, because phenological changes can be significant on shorter timescales ~~(e.g., weekly, Parazoo et al. (2018a))~~(e.g., weekly, Parazoo et al., 2018a), we perform a second analysis at ~~14 day~~ 14-day temporal resolution using three inversion analyses that optimize weekly or ~~14 day~~ 14-day NEE fluxes (Sec. 3.2). For these ~~14 day~~ 14-day fluxes, we further examine mechanistic explanations for data-model differences in $\mathrm{R_h}$ using a range of models (Sec. 3.3). Finally, we discuss the results (Sec. 4~~provides a discussion of the results and~~) and summarize our conclusions (Sec. 5~~summarizes the conclusions~~).

2 Data and methods

2.1 Environmental data and region definitions

We utilize MERRA-2 Land soil temperature data (Reichle et al., 2011, 2017; Gelaro et al., 2017) to define three large regions within Eurasia (Fig. 1). These data were downloaded from the Goddard Earth Sciences Data and Information Services Center at monthly temporal resolution and $4° \times 5°$ spatial resolution (regridded from model horizontal resolution of $\sim$50 km). Three regions are defined based on the date at which the top 0.5 m of MERRA-2 Land soil temperature falls below 0°C, referred to as the "zero-crossing date", for a mean seasonal cycle averaged over four years (2015, 2016, 2018 and 2019). The "Cold" region has a zero-crossing date before Oct 27, the "Mid" region has a zero-crossing date after Oct 27, and the "Warm" region does not have a zero-crossing date. This date was chosen as a cutoff to create two similarly sized Mid and Cold regions. Some adjustments from these definitions are made so that the regions are contiguous. We aggregate the CO_2 fluxes described below to these regions by (1) interpolating the Warm, Mid and Cold regions from $4° \times 5°$ spatial resolution to the grid of the CO_2 flux datasets (both GPP and NEE), and (2) calculate the ~~area weighted~~ area-weighted net fluxes over the regions. We also obtain the downward shortwave flux from the MERRA-2 Land dataset.

Several datasets are also used for supplementary evaluation of the MERRA-2 Land soil temperature seasonality (Text S2). For that analysis, we use ERA5-Land reanalysis soil temperature data (Munoz Sabater, 2019), generated using Copernicus Climate Change Service Information 2020. We also examined monthly soil temperature from seven models from the Coupled Model Intercomparison Project Phase 6 (CMIP6) (Eyring et al., 2016) for the historical and Shared Socioeconomic Pathway 585 (ssp585) simulations, which is the highest emission scenario. The CMIP6 simulations were included to compare with MERRA-2 simulated soil temperature over 2010–2019, and to examine possible trends in soil temperature under a high emission scenario. The model runs are: CanESM5 (r1i1p2f1), MIROC ES2L (r1i1p1f2), ACCESS EMS1 (r1i1p1f1), MRI ESM2 0 (historical r1i1p1f1, ssp585 r1i2p1f1), CNRM ESM2 1 (r1i1p1f2), E3SM 1 1 (r1i1p1f1), and UKESM1 0 LL (r4i1p1f2). These models were chosen because they participated in the Coupled Climate–Carbon Cycle Model Intercomparison (C_4MIP) (Jones et al., 2016). Finally, we compare the MERRA-2 Land soil temperature to borehole soil temperature measurements over the period 1998–2020, which were downloaded from the Global Terrestrial Network for Permafrost (GTN-P) borehole database (http://gtnpdatabase.org/boreholes).

2.2 Atmospheric flux inversions

The OCO-2 Model Inter-comparison Project (OCO-2 MIP) provides standardized experimental set-ups for assimilating atmospheric CO_2 to estimate net biosphere exchange (NBE), defined as

$$\mathrm{NBE} = \mathrm{NEE} + \mathrm{BB}, \tag{3}$$

where BB is biomass burning, across a range of inversion systems. ~~Version 9 of the~~ The v9 OCO-2 MIP ~~(MIPv9, (Peiro et al., 2021))~~(Peiro et al., 2021), provides ensembles of nine inversion systems that assimilated a standardized set of in situ and flask CO_2 measurements for one experiment (referred to as "IS") and OCO-2 ACOS b9 land nadir and land glint X_{CO_2} retrievals for a second experiment (referred to as "LNLG"). We estimate NEE fluxes from ~~MIPv9~~ v9 OCO-2 MIP NBE fluxes by ~~removing biomass burning emissions~~ subtracting biomass burning emission estimates from the Global Fire Emissions Database version 4 (GFED4.1s) (van der Werf et al., 2017). GFED4.1s provides estimates of biomass burning using MODIS burned area (Giglio et al., 2013), thermal anomalies, and surface reflectance observations (Randerson et al., 2012). Note that biomass burning is a relatively small contribution to NBE over the regions examined here during the study period (2015–2019) (Fig. S1). The NEE fluxes produced by each ensemble member over northern Eurasia are shown in Fig. S2.

To examine variability in fluxes at the sub-monthly time step, we examine three other inversion NEE estimates that optimize sub-monthly NEE fluxes: TM5-4DVAR$_{14\mathrm{day}}$ LNLGIS, CAMS$_{14\mathrm{day}}$ LNLGIS, and CMS-Flux$_{14\mathrm{day}}$ LNLGIS. These inversions assimilated both in situ and flask CO_2 in addition to OCO-2 ACOS b10 land nadir and land glint retrievals~~(we refer to this experiment as LNLGIS)~~. Note that the ACOS b10 retrievals are updated from the b9 retrievals employed in ~~MIPv9~~v9 OCO-2 MIP. The prior and posterior NEE fluxes produced by each ensemble member are shown in Fig. S3, and the inversion set-ups are described below.

TM5-4DVAR is a variational inversion framework based on the TM5 atmospheric tracer transport model (Meirink et al., 2008; Basu et al., 2013). The TM5-4DVAR$_{14\mathrm{day}}$ LNLGIS inversion assimilated 10 s averages of OCO-2 ACOS b10 land nadir and land glint measurements concurrently with in situ measurements to optimize weekly NEE and ocean fluxes. The OCO-2 10 s averages were constructed analogous to the b9 10 s averages assimilated by models in ~~MIPv9~~ v9 OCO-2 MIP (Peiro et al., 2021). The in situ measurements assimilated were updated from Peiro et al. (2021), specifically ObsPack NRT 5.0 was replaced by NRT 5.2. The flux inversion set-up was identical to the set-up of "TM5-4DVAR" in Peiro et al. (2021), except (i) the inversion was run from 2014-06-01 to 2021-02-01 ~~(instead of 2014-09-01 to 2019-06-01 in Peiro et al. (2021))~~(instead of 2014-09-01 to 2019-06-01 in Peiro et al., 2021), (ii) ECMWF ERA5 meteorology was used to drive the model instead of ERA Interim, (iii) a $1° \times 1°$ transport grid over North America was nested inside the global $3° \times 2°$ grid to take advantage of the higher in situ data density, and (iv) prior CO_2 fluxes were constructed following Weir et al. (2021).

The CAMS$_{14\mathrm{day}}$ LNLGIS inversion utilizes the CAMS greenhouse gases inversion system (Chevallier et al., 2005, 2010; Chevallier, 2013), and assimilates OCO-2 land nadir and land glint X_{CO_2} 10 s averages and in situ CO_2 measurements concurrently. A variational system is employed to optimize day-time and night-time NEE at 8-day temporal resolution of $1.875° \times 3.75°$ model grid. Tracer transport is performed using the Laboratoirede Météorologie Dynamique (LMDz) general circulation model version LMDz6A (Remaud et al., 2018). These data were downloaded from https://atmosphere.copernicus.eu/. Note that CAMS reports NBE - as explained earlier, we estimate NEE using GFED4.1s biomass burning emissions, as was done for the ~~MIPv9~~v9 OCO-2 MIP.

The CMS-Flux$_{14\mathrm{day}}$ LNLGIS flux inversions are performed using the set-up of Byrne et al. (2020b), which uses the Carbon Monitoring System-Flux (CMS-Flux) inversion system that has been developed under the NASA Carbon Monitoring System Flux project (https://cmsflux.jpl.nasa.gov) (Henze et al., 2007; Liu et al., 2014). These flux inversions optimize 14-day NEE and ocean fluxes by assimilating OCO-2 ACOS b10 land nadir and land glint "buddy" super-obs concurrently with in situ and flask measurements from version 6.0 of the GLOBALVIEW plus package (Masarie et al., 2014; Schuldt et al., 2020). OCO-2 "buddy" super-obs are obtained by averaging individual soundings into super-obs at $0.5° \times 0.5°$ spatial resolution (within the same orbit) with equal weighting following Liu et al. (2017). We assimilate surface-based in situ and flask measurements between 11 AM and 4 PM local time. These data were also pre-filtered to remove observations that were not well simulated by the model (based on a posterior data-model χ^2 mismatches greater than three for a preliminary flux inversion). For these inversions, ODIAC fossil fuel emissions (Oda and Maksyutov, 2011; Oda et al., 2018) and GFED 4.1s biomass burning emissions (van der Werf et al., 2017), including small fires (Randerson et al., 2012), are prescribed but not optimized.

2.3 Dynamic global vegetation models (DGVMs)

We use CO_2 flux estimates from an ensemble of 15 dynamic global vegetation models (DGVMs) from TRENDY v8 (Sitch et al., 2015). We utilize fluxes simulated by the CABLE-POP, CLASS-CTEM, CLM5.0, DLEM, ISAM, ISBA-CTRIP, JSBACH, JULES, LPJ, LPX-Bern, OCN, ORCHIDEE, ORCHIDEE-CNP, SDGVM and VISIT DGVMs. We exclude LPJ-GUESS because monthly output on $\mathrm{R_h}$ was not available. We utilize monthly GPP, autotrophic respiration ($\mathrm{R_a}$) and $\mathrm{R_h}$ fluxes from the "S3" experiment that employs time-varying CO_2, climate and land use forcing data. We further calculate NPP from the simulated GPP and $\mathrm{R_a}$ data ($\mathrm{NPP} = \mathrm{GPP} - \mathrm{R_a}$) at the models native resolution. The NEE, NPP, and $\mathrm{R_h}$ fluxes produced by each ensemble member are shown in Fig. S4 for the same three year period as the data-driven estimates (2015, 2016, and 2018).

We also utilize TRENDY v8 model output to estimate an ensemble of Carbon Use Efficiency ($\mathrm{CUE} = \mathrm{NPP}/\mathrm{GPP}$) from each DGVM. CUE can become negative during the winter and spring, when GPP is approximately zero but R_a is non-zero. However, we limit CUE values to a range between zero and one. These CUE estimates are then employed to estimate data-driven NPP estimates from the data-driven GPP data (see Sec. 2.4). Figure S5 shows the CUE estimates derived from the TRENDY v8 models.

2.4 GPP datasets and NPP estimates

We utilize four data-driven GPP estimates in this analysis: FluxSat, FLUXCOM, VPM, and GOSIF. These datasets differ in inputs and approach.

FluxSat Version 2 (Joiner and Yoshida, 2020) estimates GPP based on Nadir BRDF-Adjusted Reflectances (NBAR) from the ~~MODerate-resolution Imaging Spectroradiometer (MODIS)~~ MODIS MYD43D product (Schaaf et al., 2002). The GPP estimates are calibrated with the FLUXNET2015 GPP derived from eddy covariance flux measurements at Tier 1 sites (Pastorello et al., 2020).

FLUXCOM upscales CO_2 fluxes from flux tower observations using a variety of machine learning methods and forcing datasets (Jung et al., 2020). We examine the ensemble mean of the nine remote sensing (RS) learning algorithms.

VPM is a light use efficiency model that estimates GPP globally using MODIS surface reflectances and NCEP Reanalysis-2 PAR and temperature data (Xiao et al., 2004; Zhang et al., 2017). The native spatiotemporal resolution of the dataset is 500-m and 8-days. VPM has been shown to agree well with FLUXNET eddy covariance site-level data (Zhang et al., 2017) and with TROPOMI SIF at the global scale (Doughty et al., 2021).

The GOSIF GPP product estimates GPP based on OCO-2 SIF, MODIS EVI, and reanalysis data from MERRA-2 (Li and Xiao, 2019). To generate GPP estimates, first, ~~8 day~~ 8-day globally gridded $0.05° \times 0.05°$ SIF is estimated from the input data using machine-learning algorithms. GOSIF GPP is then estimated from the GOSIF SIF estimates using eight SIF-GPP relationships with different forms (universal and biome-specific, with and without intercept). In this analysis we utilize the mean GPP estimate across the eight SIF-GPP estimates.

These four data-driven GPP estimates are shown in Fig. S6. For this analysis, we estimate NPP from these data using the CUE from the TRENDY models. We perform this calculation differently for the monthly analysis and biweekly analysis. For the monthly analysis, we calculate 60 NPP seasonal cycles for each possible combination of the four GPP and 15 CUE seasonal cycles. We then calculate the ~~median~~ mean as our best estimate and interquartile range as a metric of uncertainty. For the biweekly analysis, we calculate the best estimate using the ~~median~~ mean GPP and CUE seasonal cycles, and calculate the uncertainty using the full range of GPP estimates and interquartile range of CUE estimates. This is done differently to match the NEE analysis, which leverages the larger ensemble from the ~~MIPv9~~ v9 OCO-2 MIP to examine the ~~median~~ mean and interquartile spread for the monthly analysis, but employs the full range for the smaller biweekly ensemble of three models.

2.5 Data-driven R_h estimates

We calculate the seasonal cycle of R_h by combining the data-driven estimates of NPP and NEE ~~,~~

$$\cancel{R_h = NEE + NPP.}$$

using Eq. 1 ($R_h = NEE + NPP$). We perform this calculation differently for the monthly analysis and biweekly analysis. For the monthly v9 OCO-2 ~~MIPv9~~ MIP IS- and LNLG-based estimates, we calculate 540 R_h seasonal cycles by combining the nine data-driven IS or LNLG NEE estimates with the 60 NPP estimates. We then take the ~~median~~ mean and interquartile spread as the best estimate and uncertainty. For the biweekly analysis, we calculate the best estimate of R_h from the best (~~median~~mean) estimates of NPP and NEE. We then ~~take~~ specify the uncertainty to be the full range of R_h estimates calculated from the three biweekly NEE estimates and the NPP range.

2.6 Soil carbon decomposition model

We use the soil carbon decomposition model developed in Yi et al. (2015, 2020) to simulate the contribution of soil at different depths to total $\mathrm{R_h}$ and NEE fluxes. The soil decomposition model uses multiple litter and ~~SOC~~ soil organic carbon (SOC) pools to characterize the progressive decomposition of fresh litter to more recalcitrant materials, which include three litterfall pools, three SOC pools with relatively fast turnover rates, and a deep SOC pool with slow turnover rates. The litterfall carbon inputs were first allocated to the three litterfall pools depending on the substrate quality of litterfall component and then transferred to the SOC pools through progressive decomposition. We then model the profile of the carbon pools through accounting for the vertical carbon transport (Yi et al., 2020). A constant diffusivity rate was assigned to permafrost (5.0 $\mathrm{cm^2\,yr^{-1}}$) and non-permafrost (2.0 $\mathrm{cm^2\,yr^{-1}}$) regions within the top 1 m soil, and then linearly decreased to 0 at the 3 m below surface (Koven et al., 2013). The boundary conditions at the soil surface were set as the carbon input rate to the three surface litterfall pools. A zero-flux was assigned at the bottom of the soil carbon pool, which was set as 3 m. This accounts for the upper permafrost layer, while carbon in deeper layers (e.g., 3-10 m) is largely insulated from climate variability and ignored in this study. The decomposition rate (day^{-1}) for each carbon pool was derived as the product of a theoretical maximum rate constant and dimensionless multipliers for soil temperature and liquid water content constraints (Yi et al., 2015)~~to decomposition indicated~~. In this study, the decomposition was driven by the MERRA2 soil temperature data. For simplicity, the soil saturation was assumed as 1.0 when soil temperature is above 0 °C, while the maximum liquid soil water fraction was used for below freezing (Schaefer and Jafarov, 2016).

2.7 FLUXNET data and processing

We examine 15 high latitude FLUXNET2015 sites to confirm the seasonality of carbon fluxes inferred from the atmospheric CO_2 and remote sensing datasets. These sites are listed in ~~table~~Table S1. For this, we utilize monthly data with the quality flag greater than 0.75. We calculate NPP and $\mathrm{R_h}$ for each site from the NEE and GPP datasets by applying the CUE from the TRENDY DGVMs at the gridcell containing the FLUXNET site.

3 Results

3.1 Differences between data-driven and DGVM carbon fluxes

We first examine the mean seasonal cycle of monthly NEE from the ~~MIPv9~~ v9 OCO-2 MIP inversions and TRENDY v8 DGVMs over the three north Eurasian regions (mean over 2015, 2016, and 2018; 2017 is excluded due to an OCO-2 data gap during Aug-Sep). The objective of this initial analysis is to identify the seasonal features of NEE over northern Eurasia, and identify how the data-driven and simulated estimates differ. The spread among the TRENDY v8 models is large and encompasses the data-driven estimates (Fig. S4). Thus, to identify data-model differences, we focus on differences in the ensemble ~~median~~ mean estimates and adopt the interquartile spread across ~~MIPv9~~ v9 OCO-2 MIP and TRENDY to quantify uncertainty in this estimate.

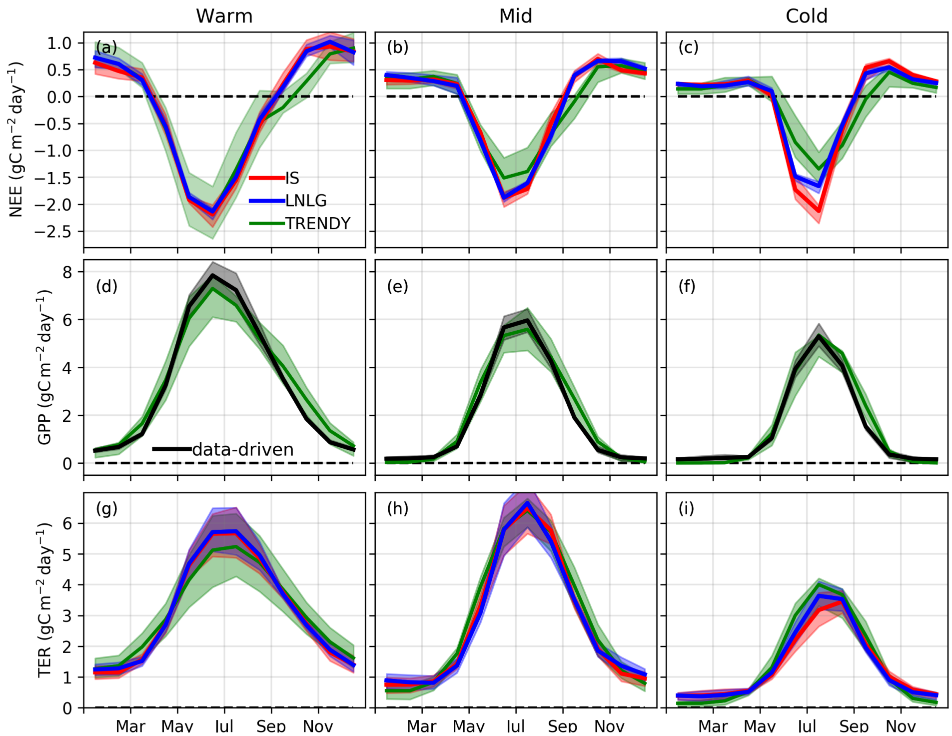

Figure 2. Monthly carbon cycle fluxes (average of 2015, 2016 and 2018; 2017 is excluded due to an OCO-2 data gap). (a-c) ~~Median~~ Mean (solid line) and interquartile range (shaded area) of NEE for the ensemble of IS (red) and LNLG (blue) v9 OCO-2 ~~MIPv9~~ MIP and for the TRENDY ensemble (green). (d-f) ~~NPP~~ GPP for the TRENDY ensemble (green) and ~~estimated from~~ data-driven ~~GPP~~ datasets (black). (g-i) ~~R_h~~ TER simulated by the TRENDY ensemble (green) and calculated from combining the data-driven GPP with the IS (red) and LNLG (blue) v9 OCO-2 ~~MIPv9~~ MIP NEE constraints.~~(j-l) Cumulative fraction of R_h over the growing season. Figure S7 shows these fluxes per unit area.~~

Figure ~~3~~2(a-c) shows the NEE fluxes for the ~~MIPv9~~ v9 OCO-2 MIP and TRENDY DGVMs for three regions over Eurasia. The two v9 OCO-2 ~~MIPv9~~ MIP ensembles (IS and LNLG) generally show close agreement and coherent differences from the TRENDY models (and prior NEE estimates, Fig. S2). The largest differences between the IS and LNLG ensembles occur over the Cold region, where the IS ensemble suggests increased uptake during July and somewhat increased release during October. Still, the coherent differences between the data-driven fluxes (both IS and LNLG) relative to the TRENDY ensemble gives us increased confidence that these inversions are precisely capturing the seasonality of NEE. The comparatively good agreement between the IS and LNLG inversions (relative to TRENDY) also suggests that artifacts related to observational coverage (Byrne et al., 2017; Basu et al., 2018) and data/model biases (Schuh et al., 2019) do not strongly impact the results. Although, we note that both datasets have spatial and seasonal gaps over northern Eurasia (e.g., Fig. S8–S9) as discussed in Sec. 4.2. The accuracy of the ~~MIPv9~~ v9 OCO-2 MIP fluxes is supported through an evaluation of the posterior CO_2 fields

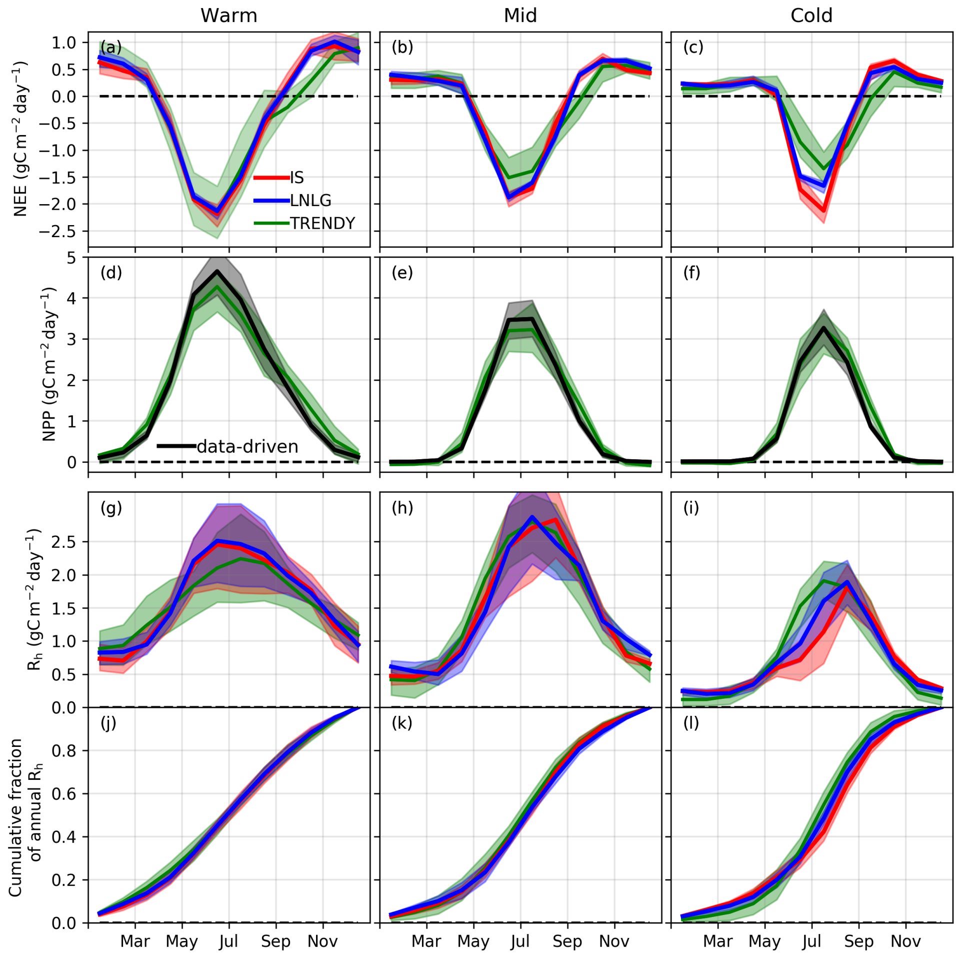

Figure 3. Monthly carbon cycle fluxes (average of 2015, 2016 and 2018; 2017 is excluded due to an OCO-2 data gap). (a-c) Mean (solid line) and interquartile range (shaded area) of NEE for the ensemble of IS (red) and LNLG (blue) v9 OCO-2 MIP and for the TRENDY ensemble (green). (d-f) NPP for the TRENDY ensemble (green) and estimated from data-driven GPP (black). (g-i) R_h simulated by the TRENDY ensemble (green) and calculated from combining the data-driven NPP with the IS (red) and LNLG (blue) v9 OCO-2 MIP NEE constraints. (j-l) Cumulative fraction of R_h over the growing season. Figure S7 shows these fluxes per unit area.

against independent atmospheric CO_2 measurements by Peiro et al. (2021), and a supplementary comparisons of the CMS-Flux$_{14\mathrm{day}}$ inversions with aircraft data over Alaska (Text. S1, Fig. S10–S11).

Comparing the ~~MIPv9~~ v9 OCO-2 MIP and TRENDY NEE estimates, good agreement is found for the Warm and Mid regions, while larger differences are found for the Cold region. In the Warm and Mid regions, systematic differences exceed the interquartile range during Sep–Oct, when the TRENDY models suggest a weaker efflux of CO_2 to the atmosphere (~~2.19–2.47 TgC day^{-1}~~ 0.30–0.51 $\mathrm{gC\,m^{-2}\,day^{-1}}$). The TRENDY models also tend to show weaker uptake by land during

June–July in the Mid region (~~2.32–2.77 $\mathrm{Tg\,C\,day^{-1}}$~~0.05–0.35 $\mathrm{g\,C\,m^{-2}\,day^{-1}}$). For the Cold region, the TRENDY models produce weaker carbon uptake during June–July (~~7.17–11.21 $\mathrm{Tg\,C\,day^{-1}}$~~0.48–0.83 $\mathrm{g\,C\,m^{-2}\,day^{-1}}$) but stronger uptake (or reduced efflux) during Aug–Oct (~~4.86–5.33 $\mathrm{Tg\,C\,day^{-1}}$~~0.31–0.36 $\mathrm{g\,C\,m^{-2}\,day^{-1}}$). This large amplitude of the NEE seasonal cycle contributes to the large seasonality in X_{CO_2} observed over eastern Eurasia (Jacobs et al., 2021).

To further investigate the causes of differences in NEE between the TRENDY and ~~MIPv9~~ v9 OCO-2 MIP ensembles, we separately examine component primary productivity and respiration fluxes. For the most direct decomposition, we employ the data-driven GPP estimates to decompose NEE into GPP and terrestrial ecosystem respiration fluxes (TER) (Fig. 2). This comparison shows that the TRENDY ensemble mean GPP tends to overestimate the data-driven GPP during the autumn (Sep–Nov), largely explaining the mismatch in NEE during this season. For TER, we find good agreement for over the Warm region except for an underestimate of TER for the TRENDY ensemble mean during the summer (mirroring GPP). For the Mid regions, agreement is found between the TRENDY and data-driven TER estimates throughout the growing season. For the Cold region, we find that the TRENDY ensemble mean suggested greater TER during May–Aug, which drives the mismatch found in NEE.

We next decompose NEE into component NPP and R_h fluxes. These estimates require an additional assumption about the CUE in comparison to the GPP/TER decomposition, but also have the potential to provide more process understanding. As described in Sec. 2.4, we employ the monthly CUE estimates from the ensemble of TRENDY models. This both allows an "apples-to-apples" comparison with the TRENDY models as the CUE estimates are consistent between the data-driven and TRENDY estimates, and allows us to propagate uncertainty in CUE from the ensemble spread. The data-driven NPP and TRENDY NPP are shown in Fig. 3(d-f). The seasonality in NPP between the data-driven and TRENDY estimates show good agreement for all regions. In the Mid and Warm regions, the TRENDY model ~~median~~ mean NPP tends to be lower than the data-driven estimates during Jun-Jul (~~1.94–2.01 $\mathrm{Tg\,C\,day^{-1}}$~~0.26–0.37 $\mathrm{g\,C\,m^{-2}\,day^{-1}}$). However, the largest differences are for the Cold region, where the TRENDY ensemble ~~median~~ mean shows increased NPP during Aug–Sep (~~5.64 $\mathrm{Tg\,C\,day^{-1}}$~~0.38 $\mathrm{g\,C\,m^{-2}\,day^{-1}}$). This largely accounts for the lower NEE during Aug-Sep (86-95%). Thus, despite previously reported deficiencies in model representation of photosynthesis over high latitudes (Rogers et al., 2017, 2019), we find that TRENDY NPP estimates largely capture the data-driven seasonality and do not drive NEE differences against the data-driven seasonal cycle.

Finally, we compare TRENDY R_h to data-driven R_h (Fig. 3(g-i)). In the Warm region, the TRENDY model ~~median~~ mean R_h is lower than the data-driven estimates during May–Sep (~~2.28–2.47 $\mathrm{Tg\,C\,day^{-1}}$~~0.21–0.26 $\mathrm{g\,C\,m^{-2}\,day^{-1}}$), but the seasonality is similar. In the Mid region, the data-driven and TRENDY R_h seasonal cycles show good agreement throughout the growing season. The largest differences between data-driven and TRENDY R_h seasonal cycles are found for the Cold region. The TRENDY model ~~median~~ mean shows increased R_h during May–Jul (~~4.05–7.07 $\mathrm{Tg\,C\,day^{-1}}$~~0.32–0.58 $\mathrm{g\,C\,m^{-2}\,day^{-1}}$) but show reduced R_h during the rest of the year (~~1.37–1.69 $\mathrm{Tg\,C\,day^{-1}}$~~0.29–0.51 $\mathrm{g\,C\,m^{-2}\,day^{-1}}$). As a result, the seasonality of data-driven R_h is shifted later in the year relative to TRENDY ensemble. This can be seen in the cumulative fraction of annual R_h, which quantifies the fraction of total R_h released as the season progresses (Fig. 3(j-l)). The percentage of total annual Rh released during ~~May-Jul is 48~~May–Jul is 46% for the TRENDY ensemble ~~median but 36~~mean but 37% (30%) for the LNLG (IS) data-driven R_h ensemble ~~median~~mean.

We independently confirm a shift in the seasonality of data-driven R_h relative to TRENDY for 15 high latitude FLUXNET sites (Fig. S12). Due to the sparsity of FLUXNET sites over northeastern Eurasia, we include sites outside of the "Cold" domain but that have early zero-crossing dates (estimated by a mean October air temperature less than 2 °C). The observed ~~median~~ mean R_h peaks across these sites during September, in agreement with the data-driven R_h seasonality. In contrast, the TRENDY ~~median~~ mean R_h peak occurs during July (consistent with the regional scale analysis). This phase shift is also evident in the cumulative fraction of annual R_h, which shows that the percentage of total annual Rh released during May-Jul is 46% for the TRENDY ensemble ~~median~~ mean but 35% for the FLUXNET-based ensemble ~~median~~mean.

Overall, these results indicate good agreement between the TRENDY ensemble and data-driven estimates for the Warm and Mid regions, but show marked differences over the Cold region. In particular, we find that the data-driven estimates suggest a seasonal redistribution of R_h with a reduction during May-Jul but an increase for the remainder of the year. Further, these results show that differences in R_h largely account for the differences between the data-driven and TRENDY NEE fluxes over the Cold region, except in Aug-Sep when NPP differences are large. In the remaining sections, we will characterize the data-driven seasonal cycle of NEE, NPP, and R_h at a higher (biweekly) temporal resolution and investigate mechanistic explanations for the data-model differences found over the Cold region.

3.2 Data-driven biweekly CO_2 fluxes

We now investigate the data-driven seasonal cycle of NEE, NPP, and R_h with biweekly (~~14 day~~14-day) temporal resolution. This higher resolution better resolves temporal changes in CO_2 fluxes throughout the growing season, particularly during the shoulder seasons, when week-to-week changes in CO_2 fluxes are large (Parazoo et al., 2018a). For this analysis, we utilize a set of three flux inversions that assimilate both in situ and OCO-2 land nadir and glint data (ACOS v10) to estimate sub-monthly CO_2 fluxes (TM5-4DVAR$_{14day}$, CAMS$_{14day}$, and CMS-Flux$_{14day}$; individual model fluxes shown in Fig. S3). These inversions give similar NEE seasonality to the ~~MIPv9~~ v9 OCO-2 MIP monthly fluxes (e.g., Fig S10) and have seasonality similar to the v9 OCO-2 ~~MIPv9~~ MIP LNLG inversions for the Cold region. For NPP, we utilize the same datasets as Sec. 3.1 but at 14-day temporal resolution. We examine the ensemble ~~medians~~ means for a best estimate and take the full range of model estimates as an illustration of the uncertainty.

Figure 4 shows the four-year-mean (2015, 2016, 2018, and 2019; 2017 is excluded due to an OCO-2 data gap in summer) seasonal cycle of NEE, NPP, and R_h for the three regions of Eurasia. NEE largely tracks the inverted seasonality of NPP, although peak NPP is slightly delayed relative to peak drawdown in NEE (by 0–2 weeks). Both NEE and NPP generally follow the seasonal cycle of insolation, but are somewhat delayed in the Mid and Cold regions, likely due to temperature limitation (Liu et al., 2020). Peak R_h is found to be delayed relative to peak NPP by 0–8 weeks. For the Warm and Mid regions, R_h follows the seasonal cycle of surface temperature, with 48% and 51% of the annual total R_h occurring after the peak in surface temperature, respectively. In contrast, the Cold region shows a substantial delay relative to surface temperature, with 63% of the total R_h occurring after the peak in surface temperature. The ~~median~~ mean R_h seasonal cycle is also found to have a double peak in this Cold region: a smaller peak of ~~9.8 TgC day^{-1}~~ 0.77 $\mathrm{gC\,m^{-2}\,day^{-1}}$ occurs during late May followed by a larger peak of ~~21.5 TgC day^{-1}~~ 1.70 $\mathrm{gC\,m^{-2}\,day^{-1}}$ at the beginning of September. This May peak roughly aligns with the spring

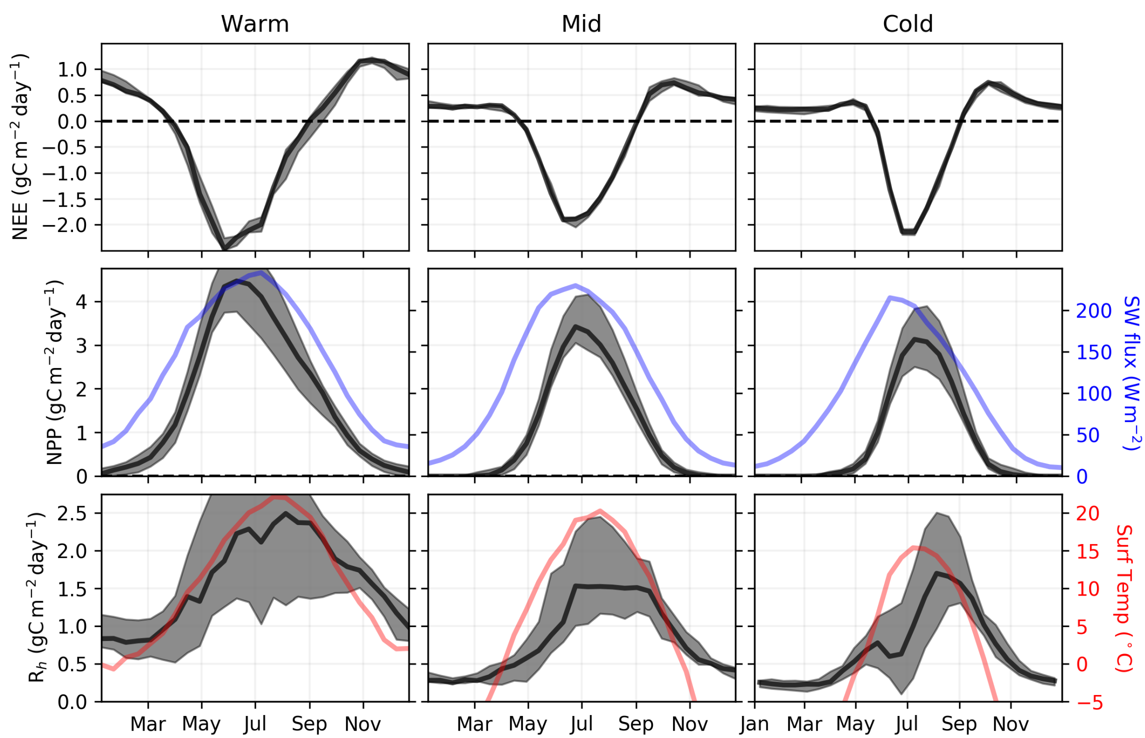

Figure 4. Data-driven four-year-mean (2015, 2016, 2018, and 2019) 14-day NEE, NPP, and Rh. (a-c) ~~Median~~ mean and ensemble spread of NEE for the CMS-Flux$_{14\mathrm{day}}$, TM4-4DVar$_{14\mathrm{day}}$, and CAMS$_{14\mathrm{day}}$ flux inversions. (d-f) ~~Median~~ mean and ensemble spread for the data-driven NPP. MERRA-2 Land net downward shortwave flux over land is shown in blue. (g-i) Ensemble estimate of $\mathrm{R_h} = \mathrm{NEE} + \mathrm{NPP}$ estimated from the three NEE and GPP estimates. MERRA-2 Land surface temperature is shown in red.

thaw and positive zero-crossing at the beginning of May. A potential mechanistic explanation for a spring pulse of $\mathrm{R_h}$ could be due to thawing soils that release CO_2 that has been trapped within subsurface soil layers over the winter (see Sec. 4). Another plausible mechanism could be the timing of snow melt, which may insulate the soil over winter (Yu et al., 2016). However, the signal from this first peak is small relative to the uncertainties.

3.3 Mechanistic drivers of late-season $\mathrm{R_h}$

Data-driven $\mathrm{R_h}$ for the Cold region indicates a delayed peak relative to surface temperature and the TRENDY model ~~median~~mean. Here we examine possible mechanistic explanation for this late season peak in $\mathrm{R_h}$ using models. We investigate two factors that could potentially contribute to the delay in $\mathrm{R_h}$: (1) Seasonal variations in the labile carbon pool. Leaf and fine root litter carbon pools tend to increase over the growing season as carbon is sequestered through photosynthesis (Randerson et al., 1996). Thus, increased substrate availability in the autumn relative to the spring will act to shift the seasonal cycle of R_h later in the year. (2) $\mathrm{R_h}$ from subsurface soil layers that have a delayed seasonal cycle driven by a lag in soil temperature. Heating and cooling at the surface slowly diffuses through the soil column resulting in a lagged seasonal cycle of temperature with depth (Parazoo et al., 2018b). Figure 5(a-c) shows the seasonal cycle in soil temperature from the MERRA-2 Land dataset. The phase shift in soil temperature seasonality with depth can be up to several months, and is largest for the colder regions. Note that we verify the fidelity of the MERRA-2 Land soil temperature against borehole measurements and against simulated soil temperature from ERA-5 reanalysis and the CMIP6 models (see Text. S2, Fig. S13–S14).

To test the impact of these factors, we consider a single layer model that represents $\mathrm{R_h}$ using a exponential relationship with temperature:

$$\mathrm{R_h} = \alpha e^{\beta \mathrm{T}}, \tag{4}$$

where α represents the labile carbon pool size, β is a constant, and T is the temperature of the carbon pool. To investigate the impact of seasonal and vertical variations in labile carbon, we consider three cases:

1. $\mathrm{R_h}(\alpha_c, \mathrm{T_{surf}}) = \alpha_c e^{\beta \mathrm{T_{surf}}}$: The carbon pool is constant in time (α_c) and the surface temperature ($\mathrm{T_{surf}}$) drives $\mathrm{R_h}$,
2. $\mathrm{R_h}(\alpha_c, \mathrm{T_{1m}}) = \alpha_c e^{\beta \mathrm{T_{1m}}}$: The carbon pool is constant in time (α_c) and the average top meter soil temperature ($\mathrm{T_{1m}}$) drives $\mathrm{R_h}$,
3. $\mathrm{R_h}(\alpha_t, \mathrm{T_{surf}}) = \alpha_t e^{\beta \mathrm{T_{surf}}}$: The carbon pool is dynamic in time (~~$\alpha = f(t)$~~$\alpha_t = f(t)$, described in the Appendix) and the $\mathrm{T_{surf}}$ drives $\mathrm{R_h}$. We assume seasonal variations in the carbon pool are within $\pm 15\%$ of the mean ($\gamma = 0.15$, Fig. S15 shows the seasonal variation in the labile carbon pool).

Table 1. Statistics on the data-model fits for the single layer models.

Region	Experiment	Slope	Intercept ($\mathrm{gC\,m^{-2}\,day^{-1}}$)	$\mathrm{R^2}$	Standard Error (SE) ($\mathrm{gC\,m^{-2}\,day^{-1}}$)
Warm	$\alpha e^{\beta \mathrm{T_{surf}}}$	0.94	0.11	0.89	0.067
Warm	$\alpha e^{\beta \mathrm{T_{1m}}}$	0.91	0.14	0.93	0.051
Warm	$\alpha(t) e^{\beta \mathrm{T_{surf}}}$	0.94	0.10	0.92	0.057
Mid	$\alpha e^{\beta \mathrm{T_{surf}}}$	1.03	0.00	0.84	0.091
Mid	$\alpha e^{\beta \mathrm{T_{1m}}}$	0.88	0.09	0.94	0.046
Mid	$\alpha(t) e^{\beta \mathrm{T_{surf}}}$	1.04	-0.01	0.89	0.074
Cold	$\alpha e^{\beta \mathrm{T_{surf}}}$	1.08	-0.01	0.66	0.16
Cold	$\alpha e^{\beta \mathrm{T_{1m}}}$	1.04	-0.01	0.88	0.08
Cold	$\alpha(t) e^{\beta \mathrm{T_{surf}}}$	1.11	-0.03	0.77	0.13

Figure 5(d-f) shows linear regressions for each one-layer model against the ~~median~~ mean biweekly estimate of $\mathrm{R_h}$. In each case, the parameters α and β are optimized (note linear regressions are performed on $\ln(\mathrm{R_h}) = \ln(\alpha) + \beta \mathrm{T}$). Statistics on the model fits are provided in Table 1. For the Warm region, all models are able to fit the data well ($\mathrm{R^2}$=0.89–0~~.94~~.93,

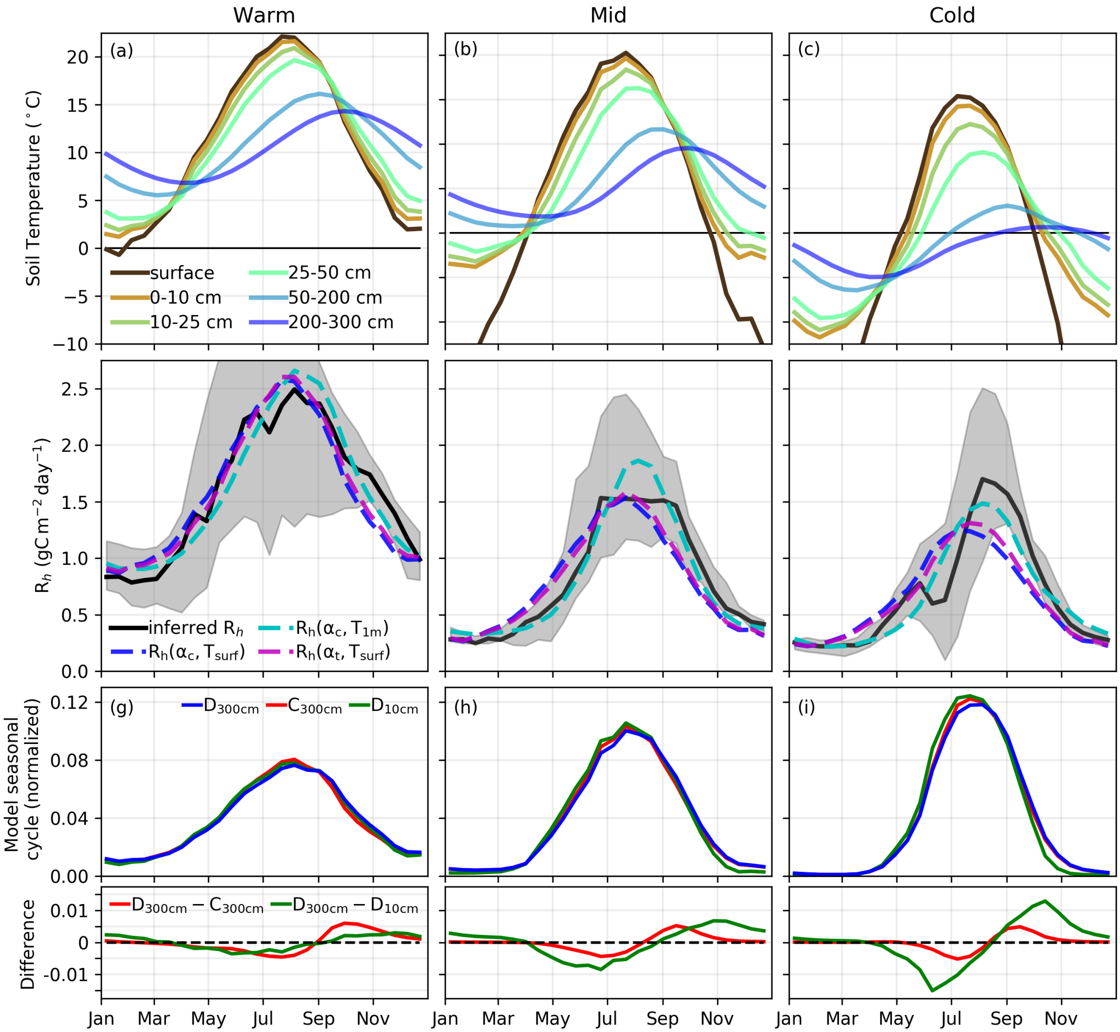

Figure 5. Impact of temporal and vertical variations in carbon pools on R_h. (a-c) MERRA-2 Land soil temperature over five intervals for the (a) Warm, (b) Mid, and (c) Cold regions. (d-e) ~~Median~~ Mean and range in inferred 14-day R_h with fits for single-layer R_h models that employ (navy dash) T_{surf} dependence and no seasonal variations in the carbon pool, (cyan dash) T_{1m} dependence and no seasonal variations in the carbon pool and (magenta dash) T_{surf} dependence and seasonal variations in the carbon pool. (g-i, top) Normalized seasonal cycle of R_h simulated ~~for D_{300cm}~~ by the soil decomposition model (Sec. 2.6). The different lines show different model simulations: D_{300cm} employs a dynamic carbon pool over 0-300 cm depth~~)~~, C_{300cm} ~~(~~employs a constant carbon pool over 0-300 cm depth~~)~~, D_{10cm} ~~(~~employs a dynamic carbon pool over 0-10 cm depth~~)and~~. (g-i, bottom) ~~the difference~~ Differences in simulated R_h between experiments.

$SE = 0.051–0.067\ gC\,m^{-2}\,day^{-1}$). Similarly, all models are able to largely capture the seasonal cycle in the Mid region (R^2=0.84–0.~~92~~94), although ~~the model driven by T_{1m}~~ $R_h(\alpha_c, T_{1m})$ appears to better capture the shoulder seasons and gives a smaller standard error ($SE = 0.046\ gC\,m^{-2}\,day^{-1}$) than the other models ($SE = 0.074–0.091\ gC\,m^{-2}\,day^{-1}$). The models diverge the most for the Cold region. ~~The model driven by T_{surf} with a constant carbon pool~~ $R_h(\alpha_c, T_{surf})$ gives the poorest performance (R~~$^2 = 0.65$~~$^2 = 0.66$, $\mathrm{SE} = 0.16\,\mathrm{gC\,m^{-2}\,day^{-1}}$), as the driving temperature data peaks too early to capture the seasonality of R_h. ~~The model with a dynamic carbon pool~~ $R_h(\alpha_t, T_{surf})$ performs somewhat better as the peak in model R_h is delayed relative to surface temperature ($R^2 = 0.77$, $\mathrm{SE} = 0.13\,\mathrm{gC\,m^{-2}\,day^{-1}}$). Still, ~~the model driven by T_{1m}~~ $R_h(\alpha_c, T_{1m})$ performs the best ($R^2 = 0.88$, $\mathrm{SE} = 0.08\,\mathrm{gC\,m^{-2}\,day^{-1}}$), and best captures the delayed R_h seasonality relative to surface temperature.

To further confirm that $R_h(\alpha_c, T_{1m})$ best captures the seasonality of R_h, we fit these same models to seasonal FLUXNET R_h averaged over cold sites. This is a rather rough comparison as we drive the models with soil temperatures averaged over the Cold region rather than site specific datasets (due to absence of soil temperature data). Figure S16 shows the resulting fits and Table S2 gives the statistics of the fits. We find that $R_h(\alpha_c, T_{1m})$ performs best ($R^2 = 0.96$, $\mathrm{SE} = 0.08\,\mathrm{gC\,m^{-2}\,day^{-1}}$), while $R_h(\alpha_t, T_{surf})$ performs second best ($R^2 = 0.85$, $\mathrm{SE} = 0.19\,\mathrm{gC\,m^{-2}\,day^{-1}}$) and $R_h(\alpha_c, T_{surf})$ gives the poorest performance ($R^2 = 0.75$, $\mathrm{SE} = 0.25\,\mathrm{gC\,m^{-2}\,day^{-1}}$), consistent with the regional-scale data-driven analysis.

This analysis demonstrates that the seasonality of R_h in the Warm and Mid regions are reasonably explained by seasonal variations in T_{surf}, but that inclusion of seasonal variations in the labile carbon pool and the impact of soil temperature with depth still improve the seasonal fit. However, for the Cold region, the seasonality of R_h is not well captured by T_{surf} and additional factors, particularly the impact soil temperature with depth, are required to explain the delayed seasonality of R_h over the Cold region.

We further investigate these mechanisms using a soil carbon decomposition model that can ~~resolve~~ simulate seasonal and vertical variations in carbon pools (Sec. 2.6). This allows for a prognostic simulation of mechanisms driving the seasonality, in contrast to the diagnostic one-layer models. We examine the seasonality of the R_h simulated down to a depth of 300 cm using a constant carbon pool (C_{300cm}), simulated within the top 10 cm of soil using a dynamic carbon pool (D_{10cm}), and simulated to a depth of 300 cm using a dynamic carbon pool (D_{300cm}, due to dynamic litterfall inputs and R_h outputs). We compare these seasonal cycles after normalizing by the annual total R_h. Figure 5(g-i) shows that incorporating seasonal and vertical variations in the carbon pool results in a phase shift in R_h to later in the year, consistent with the one-layer model results. The simulated impact of these factors is found to be quite small, possibly due to underestimation of the impact of seasonal and vertical variations in the carbon pools on R_h in the model. Still, these model simulations can inform the R_h tendencies of these carbon pool variations. Comparing the regions, the impact of seasonal variations in the labile carbon pool are quite similar, with reduced R_h in the summer and increased R_h during the autumn relative to a constant carbon pool. In contrast, the impact of vertically resolved R_h shows differences between the regions, with a small impact for the Warm region but a comparatively large impact for the Cold region. The larger impact over the Cold region is likely due to larger carbon pools at depth (Fig. ~~S16~~S17), with a possible contribution from regional differences in the thermal gradient with depth (Fig. 5). Similarly, we find that the fractional contribution of subsurface soils to total R_h has larger seasonal variation over the Cold region (Fig. ~~S17~~S18). Thus, these results support a substantial contribution of subsurface soil R_h, and suggests that an underestimation of this quantity by the DGVMs could explain the data-model differences.

4 Discussion

4.1 Implications

Over the cold northeastern region of Eurasia, our data-driven R_h seasonal cycle allocates 64–70% of annual CO_2 emissions to outside of the summer (August - April) compared to only 52% of annual R_h emissions allocated by the TRENDY DVGMs to this period. The reason that the TRENDY models do not capture this seasonality is unclear. A plausible explanation is that the TRENDY models do not capture the contribution of subsurface layers to R_h, especially during the zero curtain period. This is clearly the case for the subset of TRENDY models that drive R_h with air temperature. However, it is unclear if this is an important factor for models with more sophisticated soil modules. Surprisingly, a preliminary analysis did not find a relationship between model complexity and agreement with the data-driven estimate. The drivers of differences from the data-driven estimate may differ between models, and be impacted by the interplay of litterfall phenology, R_h formulation (Peylin et al., 2005), and number of soil layers, among other factors. Some potential areas of focus for improving models may be gleaned from recent studies. Seiler et al. (2021) suggest that the TRENDY models may systematically underestimate soil organic carbon at high latitudes, which could contribute to an underestimate of subsurface R_h across the models. Endsley et al. (2021) found a similarly phased bias in simulated R_h by the Terrestrial Carbon Flux (TCF) model against flux tower R_h to that reported here. They show that this bias could be largely mitigated by adding seasonally varying litterfall phenology, an O_2 diffusion limitation on R_h and a vertically resolved soil decomposition model, suggesting these may be foci for model improvements.

Differences between the data-driven and TRENDY R_h seasonal cycles suggest that DGVMs may be deficient in simulating the response of permafrost rich ecosystems to climate change, particularly in terms of subsurface R_h. Improving DGVM skill in these ecosystems is critical given the rapid northern high latitude warming and lengthening of the zero curtain period (Euskirchen et al., 2017; Parazoo et al., 2018b; Chen et al., 2021). The rapid changes in northern Eurasia are illustrated in Fig. 6, which shows the number of months per year that soil temperatures are greater than 0 °C as simulated by a set of CMIP6 models. Soils in the permafrost-rich Cold region are undergoing the most dramatic lengthening of the unfrozen period, particularly at depth (50–200 cm). Under scenario ssp585 (highest emission scenario), these soils are predicted to go from $\sim$5 months per year with a monthly mean soil temperature above 0 °C during the 20th century to $\sim$11 months per year by 2100. The impact is largest for the Cold region at depth because of the reduced seasonality relative to the surface, such that a warming of $\sim$7 °C shifts nearly the entire seasonal cycle above 0 °C at a depth of 50–200 cm (Fig. ~~S18~~S19). Such warming would drive the widespread formation of talik, a subsurface layer of perennial thawed soil (Parazoo et al., 2018b), and further enhance R_h at depth.

R_h from sub-surface layers may already be increasing substantially in permafrost regions. Examining the ~~41 year~~ 41-year record of CO_2 at Barrow tower, Commane et al. (2017) find that early cold NEE efflux (Oct-Dec) has increased 73.4%±10.8% over the 1975–2015 period. The standard CAMS IS inversion product similarly suggests an increase in the Sep-Oct NEE efflux of $\sim$80% over Siberia for the 2013–2017 period relative to the 1980–1984 period (see Fig. S20 of Lin et al. (2020)). In agreement, Hu et al. (2021) identified a strong increase ($\sim$10%) in Aug–Oct R_h over the North America Arctic-boreal region between

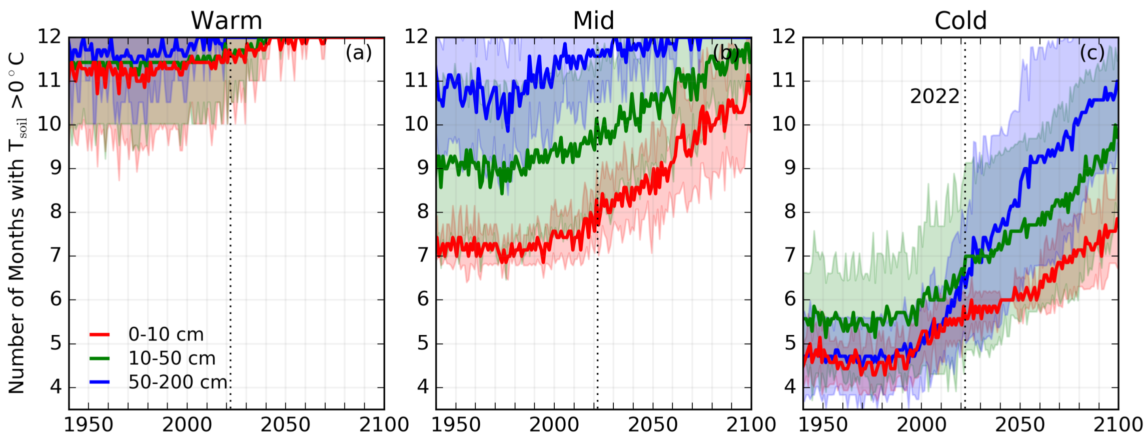

Figure 6. Number of months per year with monthly mean soil temperatures above 0 °C at depths of 0–10 cm (red), 10–50 cm (green), and 50–200 cm (blue) simulated by seven CMIP6 models under ssp585 for the (a) Warm, (b) Mid, and (c) Cold regions. The solid lines show the model mean and shading shows ±1 standard deviation.

1979–1988 and 2010–2019 based on measurements of atmospheric CO_2 and carbonyl sulphide. These inferred changes in R_h may in part be related to warming-induced changes in the seasonality of GPP (Liu et al., 2020; Kwon et al., 2021), but more research is needed to determine the impact of these different drivers.

4.2 Limitations

Atmospheric CO_2 measurements are relatively sparse over Northern Eurasia (Byrne et al., 2017). In situ and flask CO_2 measurements are spatially sparse over Mid and Cold regions (Fig. S9), with only a handful of sites assimilated over Russia as part of Japan-Russia Siberian Tall Tower Inland Observation Network (JR-STATION) of nine tower sites (Sasakawa et al., 2010, 2013). The OCO-2 coverage is seasonally variable (Fig. S8). Due to the fact that X_{CO_2} retrievals are performed on reflected sunlight, the coverage across Eurasia is quite good during the growing season (May-Sep). However, low signal and the inability to perform retrievals over snow limits the data coverage during the shoulder seasons and winter, resulting in few X_{CO_2} retrievals across the Mid and Cold regions during Nov-Feb. Ongoing research to both improve X_{CO_2} quality control filtering at high latitudes (Jacobs et al., 2020; Mendonca et al., 2021) and in retrievals X_{CO_2} over snow and ice surfaces (Mikkonen et al., 2021) may reduce these data gaps in the future. Despite this sparsity of measurements, we find that the LNLG and IS flux inversions show consistent differences from the TRENDY and prior fluxes. Furthermore, these data show good agreement with withheld in~~site~~ situ data (Peiro et al., 2021) and independent aircraft measurements over Alaska (Fig S11). Thus, we believe the results presented here to be robust despite data gaps. Still, this sparsity of data leads to some limitations. There are few sources of independent CO_2 measurements over the Mid and Cold regions to evaluate the inversion posterior CO_2 fields~~against~~. Independent measurements (possibly aircraft campaigns) would provide a valuable additional data set for validation. Similarly, increasing the number of year-round eddy-covariance sites across the Mid and Cold regions would provide a valuable independent dataset to compare against flux inversion estimated NEE. For example, Byrne et al. (2020a) were able to confirm top-down estimate of east-west differences in NEE interannual variability across North America against the dense network of eddy-covariance sites.

We also note that there are challenges in estimating data-driven GPP during the shoulder season due to reduced reflected radiance and snow cover, which impacts the spectral features of the vegetation canopy. Poor quality data, such as snowy and noisy samples, contributes to uncertainty in the timing of shoulder seasons (Wang et al., 2017; Zhang, 2015). In this analysis, we attempted to mitigate this issue though the use of an ensemble of data-driven GPP estimates, but we acknowledge that remaining biases may be present.

Furthermore, the partitioning of NEE into NPP and $\mathrm{R_h}$ could be biased if CUE estimates were seasonally biased. We employed TRENDY model CUE to translate data-driven constraints on NEE and GPP into estimates of NPP and $\mathrm{R_h}$. Thus, systematic errors across the TRENDY ensemble in CUE could impact conclusions about the relative contributions of errors in NPP and $\mathrm{R_h}$. A potential source of bias in CUE could be due to an underestimate of the impact of inhibition of leaf respiration by light (Wehr et al., 2016; Byrne et al., 2018; Keenan et al., 2019; Oikawa et al., 2017). This would result in greater CUE and NPP during June-July relative to the rest of the year, shifting the inferred $\mathrm{R_h}$ seasonal cycle earlier, with $\mathrm{R_h}$ increased during June-July but decreased elsewhere (Byrne et al., 2018). However, the magnitude of this impact on the ecosystem scale is uncertain, making accounting for this phenomena challenging. Recently, Endsley et al. (2021) found that the inhibition of leaf respiration by light to have a relatively modest impact on the seasonality of NPP and $\mathrm{R_h}$, suggesting that the results presented here are robust.

There are also remaining challenges in relating the inferred fluxes to underlying processes. Space-based flux constraints do not discriminate between biological and physical processes driving carbon cycle fluxes. It is currently unclear whether the substantial cold season CO_2 effluxes across permafrost regions are driven primarily by biological activity or physical processes (Natali et al., 2019; Arndt et al., 2020; Raz-Yaseef et al., 2017). Yet, isolating the primary driver of these fluxes is critical for inferring the sensitivity of $\mathrm{R_h}$ to climate change. If the cold season $\mathrm{R_h}$ comes from the metabolism of old permafrost carbon, then $^{14}CO_2$ measurements could help differentiate biological from physical CO_2 production.

5 Conclusions

Space-based and in situ atmospheric CO_2 measurements revealed strong summer uptake and early cold season release of CO_2 over the cold northeastern Eurasia region, implying a late summer peak in $\mathrm{R_h}$ with substantial early cold season respiration. Based on model simulations of $\mathrm{R_h}$, we suggested that this seasonality is driven by a large contribution of subsurface soils to the total $\mathrm{R_h}$. These results are consistent with site-level observations identifying substantial CO_2 release in permafrost regions outside the growing season (Natali et al., 2019), and in particular, reported spikes in early cold season respiration associated with the zero curtain period in Arctic ecosystems (Commane et al., 2017; Jeong et al., 2018).

The data-driven seasonality of $\mathrm{R_h}$ over the Cold region was generally not captured by the TRENDY DGVMs, which showed greater $\mathrm{R_h}$ during May-Jul and lower $\mathrm{R_h}$ during the rest of the year. The underlying cause of this discrepancy is unclear, but may be linked to an underestimate of the contribution of sub-surface soils to total $\mathrm{R_h}$. Given the rapid warming of permafrost soils (Euskirchen et al., 2017; Chen et al., 2021), talik formation (Parazoo et al., 2018b), and increasing early cold season CO_2 effluxes (Commane et al., 2017; Lin et al., 2020; Hu et al., 2021), improving DGVM simulations in permafrost regions should be a focus of future studies.

This analysis demonstrates the utility of space-based observations for studying carbon cycle dynamics at high latitudes, where in situ measurements are sparse. Although currently limited by a short observing record (2014 - present), the estimates of NEE inferred from the OCO-2 $\mathrm{X_{CO_2}}$ retrievals suggest that these data will provide a powerful tool for detecting change in seasonal cycle of NEE across northern Eurasia.

Data availability. TRENDY v8 gridded data can be accessed through the website https://sites.exeter.ac.uk/trendy. v9 OCO-2 MIP fluxes were downloaded from https://gml.noaa.gov/ccgg/OCO2_v9mip/. GFED data were downloaded from https://www.globalfiredata.org/. GFAS data were downloaded from https://apps.ecmwf.int/datasets/. We downloaded version 10 of the ACOS OCO-2 lite files from the CO_2 Virtual Science Data Environment (https://CO$_2$.jpl.nasa.gov/). OCO-2 data were produced by the OCO-2 project at the Jet Propulsion Laboratory, California Institute of Technology, and obtained from the OCO-2 data archive maintained at the NASA Goddard Earth Science Data and Information Services Center. FluxSat data were downloaded from https://avdc.gsfc.nasa.gov/pub/tmp/FluxSat_GPP/. The GOSIF data product is available at http://data.globalecology.unh.edu/. ERA5-Land data are obtained from the Climate Data Store (https://cds.climate.copernicus.eu).

Appendix A: Appendix 1

We estimate seasonal variations in labile carbon by estimating a litterfall flux of carbon. Litterfall seasonality is assumed to ~~be a fraction of NPP, following~~ follow the same pattern as Randerson et al. (1996) (Fig. S15)~~, and~~. We assume that the labile carbon pool is in steady state on annual timescales~~:~~, such that the annual total literfall is equal to the annual total $\mathrm{R_h}$:

$$\mathrm{Litterfall}(t) = f_{\mathrm{NPP}}(t)\cancel{\cdot\mathrm{NPP}(t)\cdot}\int_0^{365}\cancel{\frac{\mathrm{R_h}(t)}{f_{\mathrm{NPP}}(t)\cdot\mathrm{NPP}(t)}}\mathrm{R_h}(t)\,dt, \cancel{figure} \tag{A1}$$

where t is the day of the year and f_{NPP} is the ~~frac~~ fraction of annual total NPP that is converted to litterfall. The seasonal variation in the labile carbon pool ($\Delta\mathrm{C_{pool}}$) is defined as the difference in flux between litterfall and Rh:

$$\Delta\mathrm{C_{pool}}(t) = \int_0^{t}\left(\mathrm{Litterfall}(t) - \mathrm{R_h}(t)\right)dt \tag{A2}$$

Finally, we assume a fractional variation in the total carbon pool amount, γ, and calculate $\alpha(t)$:

$$\alpha(t) = \left(\frac{\mathrm{C_{pool}}(t)}{\max(|\mathrm{C_{pool}}(t)|)}\gamma + 1\right)\alpha_0, \tag{A3}$$

where α_0 is the mean carbon pool size, and is optimized in the regression in Sec. 3.3.

Author contributions. BB, JL, YY, AC, KWB, NCP, DC, and CEM conceived of the study. BB, JL, YY, AC, and SB designed the experiments. YY performed the soil carbon decomposition model runs. BB, SB and FC performed inversions for this study. BB performed the analysis and prepared the manuscript with contributions from all co-authors.

Competing interests. no competing interests are present

Acknowledgements. BB and JL were supported by the NASA OCO2/3 science team program NNH17ZDA001N-OCO2. AC, BB, JL and SB were also supported by the NASA OCO Science Team Grant #80NSSC21K1068. CEM was supported by NASA's Arctic Boreal Vulnerability Experiment (ABoVE) under NNH18ZDA001N-TE. JX was supported by the National Science Foundation (NSF) (Macrosystem Biology & NEON-Enabled Science program: DEB-2017870). MSJ acknowledges the internal funding from NASA's Earth Science Research and Analysis Program. SP acknowledges financial support of the NASA Academic Mission Services by Universities Space Research Association at NASA Ames Research Center. FC was funded by the Copernicus Atmosphere Monitoring Service, implemented by the European Centre for Medium-Range Weather Forecasts on behalf of the European Commission (grant no. CAMS73). The research carried out at the Jet Propulsion Laboratory, California Institute of Technology, was under a contract with the National Aeronautics and Space Administration. Resources supporting this work were provided by the NASA High-End Computing (HEC) Program through the NASA Advanced Supercomputing (NAS) Division at Ames Research Center. FC was granted access to the HPC resources of TGCC under the allocation A0110102201 made by GENCI. ODIAC project is supported by Greenhouse Gas Observing SATellite (GOSAT) project, National Institute for Environmental Studies (NIES), Japan.

References

Arndt, K. A., Lipson, D. A., Hashemi, J., Oechel, W. C., and Zona, D.: Snow melt stimulates ecosystem respiration in Arctic ecosystems, Global Change Biology, 26, 5042–5051, 2020.

Bastos, A., Ciais, P., Chevallier, F., Rödenbeck, C., Ballantyne, A. P., Maignan, F., Yin, Y., Fernández-Martínez, M., Friedlingstein, P., Peñuelas, J., Piao, S. L., Sitch, S., Smith, W. K., Wang, X., Zhu, Z., Haverd, V., Kato, E., Jain, A. K., Lienert, S., Lombardozzi, D., Nabel, J. E. M. S., Peylin, P., Poulter, B., and Zhu, D.: Contrasting effects of CO_2 fertilization, land-use change and warming on seasonal amplitude of Northern Hemisphere CO_2 exchange, Atmospheric Chemistry and Physics, 19, 12 361–12 375, https://doi.org/10.5194/acp-19-12361-2019, 2019.

Basu, S., Guerlet, S., Butz, A., Houweling, S., Hasekamp, O., Aben, I., Krummel, P., Steele, P., Langenfelds, R., Torn, M., Biraud, S., Stephens, B., Andrews, A., and Worthy, D.: Global CO_2 Fluxes Estimated from GOSAT Retrievals of Total Column CO_2, Atmospheric Chemistry and Physics, 13, 8695–8717, https://doi.org/10.5194/acpd-13-4535-2013, 2013.

Basu, S., Baker, D. F., Chevallier, F., Patra, P. K., Liu, J., and Miller, J. B.: The impact of transport model differences on CO_2 surface flux estimates from OCO-2 retrievals of column average CO_2, Atmospheric Chemistry & Physics, 18, 2018.

Bing, H., He, P., and Zhang, Y.: Cyclic freeze–thaw as a mechanism for water and salt migration in soil, Environmental Earth Sciences, 74, 675–681, 2015.

Björkman, M. P., Morgner, E., Cooper, E. J., Elberling, B., Klemedtsson, L., and Björk, R. G.: Winter carbon dioxide effluxes from Arctic ecosystems: An overview and comparison of methodologies, Global Biogeochemical Cycles, 24, 2010.

Byrne, B., Jones, D. B. A., Strong, K., Zeng, Z.-C., Deng, F., and Liu, J.: Sensitivity of CO_2 Surface Flux Constraints to Observational Coverage, J. Geophys. Res.-Atmos, 112, 6672–6694, https://doi.org/10.1002/2016JD026164, 2017.

Byrne, B., Wunch, D., Jones, D., Strong, K., Deng, F., Baker, I., Köhler, P., Frankenberg, C., Joiner, J., Arora, V., et al.: Evaluating GPP and Respiration Estimates Over Northern Midlatitude Ecosystems Using Solar-Induced Fluorescence and Atmospheric CO_2 Measurements, Journal of Geophysical Research: Biogeosciences, 123, 2976–2997, 2018.

Byrne, B., Liu, J., Bloom, A. A., Bowman, K. W., Butterfield, Z., Joiner, J., Keenan, T. F., Keppel-Aleks, G., Parazoo, N. C., and Yin, Y.: Contrasting regional carbon cycle responses to seasonal climate anomalies across the east-west divide of temperate North America, Global biogeochemical cycles, 34, e2020GB006 598, 2020a.

Byrne, B., Liu, J., Lee, M., Baker, I. T., Bowman, K. W., Deutscher, N. M., Feist, D. G., Griffith, D. W., Iraci, L. T., Kiel, M., Kimball, J., Miller, C. E., Morino, I., Parazoo, N. C., Petri, C., Roehl, C. M., Sha, M., Strong, K., Velazco, V. A., Wennberg, P. O., and Wunch, D.: Improved constraints on northern extratropical CO_2 fluxes obtained by combining surface-based and space-based atmospheric CO_2 measurements, Journal ofGeophysical Research: Atmospheres, 125, https://doi.org/10.1029/2019JD032029, 2020b.

Chen, L., Aalto, J., and Luoto, M.: Significant shallow–depth soil warming over Russia during the past 40 years, Global and Planetary Change, 197, 103 394, 2021.

Chevallier, F.: On the parallelization of atmospheric inversions of CO_2 surface fluxes within a variational framework, Geoscientific Model Development, 6, 783–790, https://doi.org/10.5194/gmd-6-783-2013, 2013.

Chevallier, F., Fisher, M., Peylin, P., Serrar, S., Bousquet, P., Bréon, F.-M., Chédin, A., and Ciais, P.: Inferring CO_2 sources and sinks from satellite observations: Method and application to TOVS data, Journal of Geophysical Research: Atmospheres, 110, 2005.

Chevallier, F., Ciais, P., Conway, T., Aalto, T., Anderson, B., Bousquet, P., Brunke, E., Ciattaglia, L., Esaki, Y., Fröhlich, M., et al.: CO_2 surface fluxes at grid point scale estimated from a global 21 year reanalysis of atmospheric measurements, Journal of Geophysical Research: Atmospheres, 115, 2010.

Commane, R., Lindaas, J., Benmergui, J., Luus, K. A., Chang, R. Y.-W., Daube, B. C., Euskirchen, E. S., Henderson, J. M., Karion, A., Miller, J. B., et al.: Carbon dioxide sources from Alaska driven by increasing early winter respiration from Arctic tundra, Proc. Natl. Acad. Sci., 114, 5361–5366, https://doi.org/10.1073/pnas.1618567114, 2017.

Crisp, D., Pollock, H. R., Rosenberg, R., Chapsky, L., Lee, R. A., Oyafuso, F. A., Frankenberg, C., O'Dell, C. W., Bruegge, C. J., Doran, G. B., et al.: The on-orbit performance of the Orbiting Carbon Observatory-2 (OCO-2) instrument and its radiometrically calibrated products, Atmospheric Measurement Techniques, 10, 59–81, 2017.

Doughty, R., Xiao, X., Köhler, P., Frankenberg, C., Qin, Y., Wu, X., Ma, S., and Moore III, B.: Global-scale consistency of spaceborne vegetation indices, chlorophyll fluorescence, and photosynthesis, Journal of Geophysical Research: Biogeosciences, p. e2020JG006136, 2021.

Drotz, S. H., Sparrman, T., Nilsson, M. B., Schleucher, J., and Öquist, M. G.: Both catabolic and anabolic heterotrophic microbial activity proceed in frozen soils, Proceedings of the National Academy of Sciences, 107, 21 046–21 051, 2010.

Eldering, A., O'Dell, C. W., Wennberg, P. O., Crisp, D., Gunson, M. R., Viatte, C., Avis, C., Braverman, A., Castano, R., Chang, A., Chapsky, L., Cheng, C., Connor, B., Dang, L., Doran, G., Fisher, B., Frankenberg, C., Fu, D., Granat, R., Hobbs, J., Lee, R. A. M., Mandrake, L., McDuffie, J., Miller, C. E., Myers, V., Natraj, V., O'Brien, D., Osterman, G. B., Oyafuso, F., Payne, V. H., Pollock, H. R., Polonsky, I., Roehl, C. M., Rosenberg, R., Schwandner, F., Smyth, M., Tang, V., Taylor, T. E., To, C., Wunch, D., and Yoshimizu, J.: The Orbiting Carbon Observatory-2: first 18 months of science data products, Atmos. Meas. Tech., 10, 549–563, https://doi.org/10.5194/amt-10-549-2017, 2017.

Endsley, K. A., Kimball, J., and Reichle, R. H.: Soil respiration phenology improves modeled phase of terrestrial net ecosystem exchange in northern hemisphere, Earth and Space Science Open Archive, p. 31, https://doi.org/10.1002/essoar.10507940.1, 2021.

Euskirchen, E., Bret-Harte, M., Shaver, G., Edgar, C., and Romanovsky, V.: Long-term release of carbon dioxide from arctic tundra ecosystems in Alaska, Ecosystems, 20, 960–974, 2017.

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E.: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, Geoscientific Model Development, 9, 1937–1958, https://doi.org/10.5194/gmd-9-1937-2016, 2016.

Friedlingstein, P., Jones, M. W., O'Sullivan, M., Andrew, R. M., Hauck, J., Peters, G. P., Peters, W., Pongratz, J., Sitch, S., Le Quéré, C., Bakker, D. C. E., Canadell, J. G., Ciais, P., Jackson, R. B., Anthoni, P., Barbero, L., Bastos, A., Bastrikov, V., Becker, M., Bopp, L., Buitenhuis, E., Chandra, N., Chevallier, F., Chini, L. P., Currie, K. I., Feely, R. A., Gehlen, M., Gilfillan, D., Gkritzalis, T., Goll, D. S., Gruber, N., Gutekunst, S., Harris, I., Haverd, V., Houghton, R. A., Hurtt, G., Ilyina, T., Jain, A. K., Joetzjer, E., Kaplan, J. O., Kato, E., Klein Goldewijk, K., Korsbakken, J. I., Landschützer, P., Lauvset, S. K., Lefèvre, N., Lenton, A., Lienert, S., Lombardozzi, D., Marland, G., McGuire, P. C., Melton, J. R., Metzl, N., Munro, D. R., Nabel, J. E. M. S., Nakaoka, S.-I., Neill, C., Omar, A. M., Ono, T., Peregon, A., Pierrot, D., Poulter, B., Rehder, G., Resplandy, L., Robertson, E., Rödenbeck, C., Séférian, R., Schwinger, J., Smith, N., Tans, P. P., Tian, H., Tilbrook, B., Tubiello, F. N., van der Werf, G. R., Wiltshire, A. J., and Zaehle, S.: Global Carbon Budget 2019, Earth System Science Data, 11, 1783–1838, https://doi.org/10.5194/essd-11-1783-2019, 2019.

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C. A., Darmenov, A., Bosilovich, M. G., Reichle, R., et al.: The modern-era retrospective analysis for research and applications, version 2 (MERRA-2), J. Climate, 30, 5419–5454, 2017.

Giglio, L., Randerson, J. T., and Van Der Werf, G. R.: Analysis of daily, monthly, and annual burned area using the fourth-generation global fire emissions database (GFED4), Journal of Geophysical Research: Biogeosciences, 118, 317–328, 2013.

Henze, D. K., Hakami, A., and Seinfeld, J. H.: Development of the adjoint of GEOS-Chem, Atmos. Chem. Phys., 7, 2413–2433, 2007.

Hinkel, K., Paetzold, F., Nelson, F., and Bockheim, J.: Patterns of soil temperature and moisture in the active layer and upper permafrost at Barrow, Alaska: 1993–1999, Global and Planetary Change, 29, 293–309, 2001.

Hu, L., Montzka, S. A., Kaushik, A., Andrews, A. E., Sweeney, C., Miller, J., Baker, I. T., Denning, S., Campbell, E., Shiga, Y. P., et al.: COS-derived GPP relationships with temperature and light help explain high-latitude atmospheric CO_2 seasonal cycle amplification, Proceedings of the National Academy of Sciences, 118, 2021.

Jacobs, N., Simpson, W. R., Wunch, D., O'Dell, C. W., Osterman, G. B., Hase, F., Blumenstock, T., Tu, Q., Frey, M., Dubey, M. K., Parker, H. A., Kivi, R., and Heikkinen, P.: Quality controls, bias, and seasonality of CO_2 columns in the boreal forest with Orbiting Carbon Observatory-2, Total Carbon Column Observing Network, and EM27/SUN measurements, Atmospheric Measurement Techniques, 13, 5033–5063, https://doi.org/10.5194/amt-13-5033-2020, 2020.

Jacobs, N., Simpson, W. R., Graham, K. A., Holmes, C., Hase, F., Blumenstock, T., Tu, Q., Frey, M., Dubey, M. K., Parker, H. A., Wunch, D., Kivi, R., Heikkinen, P., Notholt, J., Petri, C., and Warneke, T.: Spatial distributions of X_{CO_2} seasonal cycle amplitude and phase over northern high-latitude regions, Atmospheric Chemistry and Physics, 21, 16 661–16 687, https://doi.org/10.5194/acp-21-16661-2021, 2021.

Jeong, S.-J., Bloom, A. A., Schimel, D., Sweeney, C., Parazoo, N. C., Medvigy, D., Schaepman-Strub, G., Zheng, C., Schwalm, C. R., Huntzinger, D. N., et al.: Accelerating rates of Arctic carbon cycling revealed by long-term atmospheric CO_2 measurements, Science advances, 4, eaao1167, 2018.

Joiner, J. and Yoshida, Y.: Satellite-based reflectances capture large fraction of variability in global gross primary production (GPP) at weekly time scales, Agricultural and Forest Meteorology, 291, 108 092, https://doi.org/https://doi.org/10.1016/j.agrformet.2020.108092, 2020.

Jones, C. D., Arora, V., Friedlingstein, P., Bopp, L., Brovkin, V., Dunne, J., Graven, H., Hoffman, F., Ilyina, T., John, J. G., Jung, M., Kawamiya, M., Koven, C., Pongratz, J., Raddatz, T., Randerson, J. T., and Zaehle, S.: C4MIP – The Coupled Climate–Carbon Cycle Model Intercomparison Project: experimental protocol for CMIP6, Geoscientific Model Development, 9, 2853–2880, https://doi.org/10.5194/gmd-9-2853-2016, 2016.

Jung, M., Schwalm, C., Migliavacca, M., Walther, S., Camps-Valls, G., Koirala, S., Anthoni, P., Besnard, S., Bodesheim, P., Carvalhais, N., Chevallier, F., Gans, F., Goll, D. S., Haverd, V., Köhler, P., Ichii, K., Jain, A. K., Liu, J., Lombardozzi, D., Nabel, J. E. M. S., Nelson, J. A., O'Sullivan, M., Pallandt, M., Papale, D., Peters, W., Pongratz, J., Rödenbeck, C., Sitch, S., Tramontana, G., Walker, A., Weber, U., and Reichstein, M.: Scaling carbon fluxes from eddy covariance sites to globe: synthesis and evaluation of the FLUXCOM approach, Biogeosciences, 17, 1343–1365, https://doi.org/10.5194/bg-17-1343-2020, 2020.

Keenan, T. F., Migliavacca, M., Papale, D., Baldocchi, D., Reichstein, M., Torn, M., and Wutzler, T.: Widespread inhibition of daytime ecosystem respiration, Nature ecology & evolution, 3, 407–415, 2019.

Koven, C. D., Riley, W. J., Subin, Z. M., Tang, J. Y., Torn, M. S., Collins, W. D., Bonan, G. B., Lawrence, D. M., and Swenson, S. C.: The effect of vertically resolved soil biogeochemistry and alternate soil C and N models on C dynamics of CLM4, Biogeosciences, 10, 7109–7131, https://doi.org/10.5194/bg-10-7109-2013, 2013.

Kwon, M. J., Ballantyne, A., Ciais, P., Bastos, A., Chevallier, F., Liu, Z., Green, J. K., Qiu, C., and Kimball, J. S.: Siberian 2020 heatwave increased spring CO2 uptake but not annual CO2 uptake, Environmental Research Letters, 16, 124 030, 2021.

Li, X. and Xiao, J.: Mapping photosynthesis solely from solar-induced chlorophyll fluorescence: A global, fine-resolution dataset of gross primary production derived from OCO-2, Remote Sensing, 11, 2563, 2019.

Lin, X., Rogers, B. M., Sweeney, C., Chevallier, F., Arshinov, M., Dlugokencky, E., Machida, T., Sasakawa, M., Tans, P., and Keppel-Aleks, G.: Siberian and temperate ecosystems shape Northern Hemisphere atmospheric CO_2 seasonal amplification, Proceedings of the National Academy of Sciences, 117, 21 079–21 087, 2020.

Liu, J., Bowman, K. W., Lee, M., Henze, D. K., Bousserez, N., Brix, H., Collatz, G. J., Menemenlis, D., Ott, L., Pawson, S., et al.: Carbon monitoring system flux estimation and attribution: impact of ACOS-GOSAT X_{CO_2} sampling on the inference of terrestrial biospheric sources and sinks, Tellus B, 66, 22 486, https://doi.org/10.3402/tellusb.v66.22486, 2014.

Liu, J., Bowman, K. W., Schimel, D. S., Parazoo, N. C., Jiang, Z., Lee, M., Bloom, A. A., Wunch, D., Frankenberg, C., Sun, Y., O'Dell, C. W., Gurney, K. R., Menemenlis, D., Gierach, M., Crisp, D., and Eldering, A.: Contrasting carbon cycle responses of the tropical continents to the 2015–2016 El Niño, Science, 358, https://doi.org/10.1126/science.aam5690, 2017.

Liu, J., Wennberg, P. O., Parazoo, N. C., Yin, Y., and Frankenberg, C.: Observational constraints on the response of high-latitude northern forests to warming, AGU Advances, 1, e2020AV000 228, 2020.

Masarie, K., Peters, W., Jacobson, A., and Tans, P.: ObsPack: a framework for the preparation, delivery, and attribution of atmospheric greenhouse gas measurements, Earth Syst. Sci. Data, 6, 375–384, 2014.

Mastepanov, M., Sigsgaard, C., Tagesson, T., Ström, L., Tamstorf, M. P., Lund, M., and Christensen, T. R.: Revisiting factors controlling methane emissions from high-Arctic tundra, Biogeosciences, 10, 5139–5158, https://doi.org/10.5194/bg-10-5139-2013, 2013.

McGuire, A. D., Lawrence, D. M., Koven, C., Clein, J. S., Burke, E., Chen, G., Jafarov, E., MacDougall, A. H., Marchenko, S., Nicolsky, D., et al.: Dependence of the evolution of carbon dynamics in the northern permafrost region on the trajectory of climate change, Proceedings of the National Academy of Sciences, 115, 3882–3887, 2018.

McMahon, S. K., Wallenstein, M. D., and Schimel, J. P.: Microbial growth in Arctic tundra soil at -2 C, Environmental Microbiology Reports, 1, 162–166, 2009.

Meirink, J. F., Bergamaschi, P., and Krol, M. C.: Four-Dimensional Variational Data Assimilation for Inverse Modelling of Atmospheric Methane Emissions: Method and Comparison with Synthesis Inversion, Atmospheric Chemistry and Physics, 8, 6341–6353, https://doi.org/10.5194/acp-8-6341-2008, 2008.

Mendonca, J., Nassar, R., O'Dell, C. W., Kivi, R., Morino, I., Notholt, J., Petri, C., Strong, K., and Wunch, D.: Assessing the feasibility of using a neural network to filter Orbiting Carbon Observatory 2 (OCO-2) retrievals at northern high latitudes, Atmospheric Measurement Techniques, 14, 7511–7524, https://doi.org/10.5194/amt-14-7511-2021, 2021.

Mikkonen, A., Lindqvist, H., Peltoniemi, J., and Tamminen, J.: Satellite-based remote sensing of carbon dioxide over snow-covered surfaces, in: EGU General Assembly Conference Abstracts, EGU General Assembly Conference Abstracts, pp. EGU21–8207, 2021.

Miner, K. R., Turetsky, M. R., Malina, E., Bartsch, A., Tamminen, J., McGuire, A. D., Fix, A., Sweeney, C., Elder, C. D., and Miller, C. E.: Permafrost carbon emissions in a changing Arctic, Nature Reviews Earth & Environment, 3, 55–67, 2022.

Munoz Sabater, J.: Copernicus Climate Change Service (C3S) Climate Data Store (CDS), https://doi.org/10.24381/cds.68d2bb30, 2020-08-11, 2019.

Natali, S. M., Watts, J. D., Rogers, B. M., Potter, S., Ludwig, S. M., Selbmann, A.-K., Sullivan, P. F., Abbott, B. W., Arndt, K. A., Birch, L., et al.: Large loss of CO_2 in winter observed across the northern permafrost region, Nature Climate Change, 9, 852–857, 2019.

Obu, J., Westermann, S., Kääb, A., and Bartsch, A.: Ground Temperature Map, 2000-2016, Northern Hemisphere Permafrost, https://doi.org/10.1594/PANGAEA.888600, 2018.

Obu, J., Westermann, S., Bartsch, A., Berdnikov, N., Christiansen, H. H., Dashtseren, A., Delaloye, R., Elberling, B., Etzelmüller, B., Kholodov, A., Khomutov, A., Kääb, A., Leibman, M. O., Lewkowicz, A. G., Panda, S. K., Romanovsky, V., Way, R. G., Westergaard-Nielsen, A., Wu, T., Yamkhin, J., and Zou, D.: Northern Hemisphere permafrost map based on TTOP modelling for 2000–2016 at 1 km^2 scale, Earth-Science Reviews, 193, 299–316, https://doi.org/https://doi.org/10.1016/j.earscirev.2019.04.023, 2019.

Oda, T. and Maksyutov, S.: A very high-resolution (1 km$\times$ 1 km) global fossil fuel CO 2 emission inventory derived using a point source database and satellite observations of nighttime lights, Atmos. Chem. Phys., 11, 543–556, 2011.

Oda, T., Maksyutov, S., and Andres, R. J.: The Open-source Data Inventory for Anthropogenic CO_2, version 2016 (ODIAC2016): a global monthly fossil fuel CO_2 gridded emissions data product for tracer transport simulations and surface flux inversions, Earth Syst. Sci. Data, 10, 87–107, https://doi.org/10.5194/essd-10-87-2018, 2018.

Oikawa, P., Sturtevant, C., Knox, S., Verfaillie, J., Huang, Y., and Baldocchi, D.: Revisiting the partitioning of net ecosystem exchange of CO_2 into photosynthesis and respiration with simultaneous flux measurements of $^{13}CO_2$and CO_2, soil respiration and a biophysical model, CANVEG, Agricultural and Forest Meteorology, 234, 149–163, 2017.

Outcalt, S. I., Nelson, F. E., and Hinkel, K. M.: The zero-curtain effect: Heat and mass transfer across an isothermal region in freezing soil, Water Resources Research, 26, 1509–1516, 1990.

Overland, J., Hanna, E., Hanssen-Bauer, I., Kim, S.-J., Walsh, J., Wang, M., Bhatt, U., Thoman, R., et al.: Surface air temperature, Arctic report card, 2018.

Panikov, N., Flanagan, P., Oechel, W., Mastepanov, M., and Christensen, T.: Microbial activity in soils frozen to below -39 C, Soil Biology and Biochemistry, 38, 785–794, 2006.

Parazoo, N. C., Arneth, A., Pugh, T. A., Smith, B., Steiner, N., Luus, K., Commane, R., Benmergui, J., Stofferahn, E., Liu, J., et al.: Spring photosynthetic onset and net CO_2 uptake in Alaska triggered by landscape thawing, Global change biology, 24, 3416–3435, 2018a.

Parazoo, N. C., Koven, C. D., Lawrence, D. M., Romanovsky, V., and Miller, C. E.: Detecting the permafrost carbon feedback: talik formation and increased cold-season respiration as precursors to sink-to-source transitions, The Cryosphere, 12, 123–144, https://doi.org/10.5194/tc-12-123-2018, 2018b.

Park, T., Ganguly, S., Tømmervik, H., Euskirchen, E. S., Høgda, K.-A., Karlsen, S. R., Brovkin, V., Nemani, R. R., and Myneni, R. B.: Changes in growing season duration and productivity of northern vegetation inferred from long-term remote sensing data, Environmental Research Letters, 11, 084 001, 2016.

Pastorello, G., Trotta, C., Canfora, E., Chu, H., Christianson, D., Cheah, Y.-W., Poindexter, C., Chen, J., Elbashandy, A., Humphrey, M., Isaac, P., Polidori, D., Ribeca, A., van Ingen, C., Zhang, L., Amiro, B., Ammann, C., Arain, M. A., ArdÃ¶, J., Arkebauer, T., Arndt, S. K., Arriga, N., Aubinet, M., Aurela, M., Baldocchi, D., Barr, A., Beamesderfer, E., Marchesini, L. B., Bergeron, O., Beringer, J., Bernhofer, C., Berveiller, D., Billesbach, D., Black, T. A., Blanken, P. D., Bohrer, G., Boike, J., Bolstad, P. V., Bonal, D., Bonnefond, J.-M., Bowling, D. R., Bracho, R., Brodeur, J., BrÃ¼mmer, C., Buchmann, N., Burban, B., Burns, S. P., Buysse, P., Cale, P., Cavagna, M., Cellier, P., Chen, S., Chini, I., Christensen, T. R., Cleverly, J., Collalti, A., Consalvo, C., Cook, B. D., Cook, D., Coursolle, C., Cremonese, E., Curtis, P. S., Dâ™Andrea, E., da Rocha, H., Dai, X., Davis, K. J., De Cinti, B., de Grandcourt, A., De Ligne, A., De Oliveira, R. C., Delpierre, N., Desai, A. R., Di Bella, C. M., di Tommasi, P., Dolman, H., Domingo, F., Dong, G., Dore, S., Duce, P., DufrÃªne, E., Dunn, A., DuÅ¡ek, J., Eamus, D., Eichelmann, U., ElKhidir, H. A. M., Eugster, W., Ewenz, C. M., Ewers, B., Famulari, D., Fares, S., Feigenwinter, I., Feitz, A., Fensholt, R., Filippa, G., Fischer, M., Frank, J., Galvagno, M., Gharun, M., Gianelle, D., Gielen, B., Gioli, B., Gitelson, A., Goded, I., Goeckede, M., Goldstein, A. H., Gough, C. M., Goulden, M. L., Graf, A., Griebel, A., Gruening, C., GrÃ¼nwald, T., Hammerle, A., Han, S., Han, X., Hansen, B. U., Hanson, C., Hatakka, J., He, Y., Hehn, M., Heinesch, B., Hinko-Najera, N., HÃ¶rtnagl, L., Hutley, L., Ibrom, A., Ikawa, H., Jackowicz-Korczynski, M., JanouÅ¡, D., Jans, W., Jassal, R., Jiang, S., Kato, T., Khomik, M., Klatt, J., Knohl, A., Knox, S., Kobayashi, H., Koerber, G., Kolle, O., Kosugi, Y., Kotani, A., Kowalski, A., Kruijt, B., Kurbatova, J., Kutsch, W. L., Kwon, H., Launiainen, S., Laurila, T., Law, B., Leuning, R., Li, Y., Liddell, M., Limousin, J.-M., Lion, M., Liska, A. J., Lohila, A., LÃ³pez-Ballesteros, A., LÃ³pez-Blanco, E., Loubet, B., Loustau, D., Lucas-Moffat, A., LÃ¼ers, J., Ma, S., Macfarlane, C., Magliulo, V., Maier, R., Mammarella, I., Manca, G., Marcolla, B., Margolis, H. A., Marras, S., Massman, W., Mastepanov, M., Matamala, R., Matthes, J. H., Mazzenga, F., McCaughey, H., McHugh, I., McMillan, A. M. S., Merbold, L., Meyer, W., Meyers, T., Miller, S. D., Minerbi, S., Moderow, U., Monson, R. K., Montagnani, L., Moore, C. E., Moors, E., Moreaux, V., Moureaux, C., Munger, J. W., Nakai, T., Neirynck, J., Nesic, Z., Nicolini, G., Noormets, A., Northwood, M., Nosetto, M., Nouvellon, Y., Novick, K., Oechel, W., Olesen, J. E., Ourcival, J.-M., Papuga, S. A., Parmentier, F.-J., Paul-Limoges, E., Pavelka, M., Peichl, M., Pendall, E., Phillips, R. P., Pilegaard, K., Pirk, N., Posse, G., Powell, T., Prasse, H., Prober, S. M., Rambal, S., Rannik, U., Raz-Yaseef, N., Reed, D., de Dios, V. R., Restrepo-Coupe, N., Reverter, B. R., Roland, M., Sabbatini, S., Sachs, T., Saleska, S. R., SÃ¡nchez-CaÃ±ete, E. P., Sanchez-Mejia, Z. M., Schmid, H. P., Schmidt, M., Schneider, K., Schrader, F., Schroder, I., Scott, R. L., SedlÃ¡k, P., Serrano-OrtÃz, P., Shao, C., Shi, P., Shironya, I., Siebicke, L., Å igut, L., Silberstein, R., Sirca, C., Spano, D., Steinbrecher, R., Stevens, R. M., Sturtevant, C., Suyker, A., Tagesson, T., Takanashi, S., Tang, Y., Tapper, N., Thom, J., Tiedemann, F., Tomassucci, M., Tuovinen, J.-P., Urbanski, S., Valentini, R., van der Molen, M., van Gorsel, E., van Huissteden, K., Varlagin, A., Verfaillie, J., Vesala, T., Vincke, C., Vitale, D., Vygodskaya, N., Walker, J. P., Walter-Shea, E., Wang, H., Weber, R., Westermann, S., Wille, C., Wofsy, S., Wohlfahrt, G., Wolf, S., Woodgate, W., Li, Y., Zampedri, R., Zhang, J., Zhou, G., Zona, D., Agarwal, D., Biraud, S., Torn, M., and Papale, D.: The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data, Scientific Data, 7, 225, https://doi.org/10.1038/s41597-020-0534-3, 2020.

Peiro, H., Crowell, S., Schuh, A., Baker, D. F., O'Dell, C., Jacobson, A. R., Chevallier, F., Liu, J., Eldering, A., Crisp, D., Deng, F., Weir, B., Basu, S., Johnson, M. S., Philip, S., and Baker, I.: Four years of global carbon cycle observed from OCO-2 version 9 and in situ data, and comparison to OCO-2 v7, Atmospheric Chemistry and Physics Discussions, 2021, 1–50, https://doi.org/10.5194/acp-2021-373, 2021.

Peylin, P., Bousquet, P., Le Quéré, C., Sitch, S., Friedlingstein, P., McKinley, G., Gruber, N., Rayner, P., and Ciais, P.: Multiple constraints on regional CO2 flux variations over land and oceans, Global Biogeochemical Cycles, 19, 2005.

Pirk, N., Santos, T., Gustafson, C., Johansson, A. J., Tufvesson, F., Parmentier, F.-J. W., Mastepanov, M., and Christensen, T. R.: Methane emission bursts from permafrost environments during autumn freeze-in: New insights from ground-penetrating radar, Geophysical Research Letters, 42, 6732–6738, 2015.

Randerson, J., Chen, Y., Van Der Werf, G., Rogers, B., and Morton, D.: Global burned area and biomass burning emissions from small fires, Journal of Geophysical Research: Biogeosciences, 117, 2012.

Randerson, J. T., Thompson, M. V., Malmstrom, C. M., Field, C. B., and Fung, I. Y.: Substrate limitations for heterotrophs: Implications for models that estimate the seasonal cycle of atmospheric CO_2, Global Biogeochem. Cy., 10, 585–602, https://doi.org/10.1029/96GB01981, 1996.

Raz-Yaseef, N., Torn, M. S., Wu, Y., Billesbach, D. P., Liljedahl, A. K., Kneafsey, T. J., Romanovsky, V. E., Cook, D. R., and Wullschleger, S. D.: Large CO_2 and CH_4 emissions from polygonal tundra during spring thaw in northern Alaska, Geophysical Research Letters, 44, 504–513, 2017.

Reichle, R. H., Koster, R. D., De Lannoy, G. J., Forman, B. A., Liu, Q., Mahanama, S. P., and Touré, A.: Assessment and enhancement of MERRA land surface hydrology estimates, J. Climate, 24, 6322–6338, 2011.

Reichle, R. H., Draper, C. S., Liu, Q., Girotto, M., Mahanama, S. P., Koster, R. D., and De Lannoy, G. J.: Assessment of MERRA-2 land surface hydrology estimates, J. Climate, 30, 2937–2960, 2017.

Remaud, M., Chevallier, F., Cozic, A., Lin, X., and Bousquet, P.: On the impact of recent developments of the LMDz atmospheric general circulation model on the simulation of CO_2 transport, Geoscientific Model Development, 11, 4489–4513, https://doi.org/10.5194/gmd-11-4489-2018, 2018.

Rivkina, E., Friedmann, E., McKay, C., and Gilichinsky, D.: Metabolic activity of permafrost bacteria below the freezing point, Applied and Environmental Microbiology, 66, 3230–3233, 2000.

Rogers, A., Serbin, S. P., Ely, K. S., Sloan, V. L., and Wullschleger, S. D.: Terrestrial biosphere models underestimate photosynthetic capacity and CO2 assimilation in the Arctic, New Phytologist, 216, 1090–1103, 2017.

Rogers, A., Serbin, S. P., Ely, K. S., and Wullschleger, S. D.: Terrestrial biosphere models may overestimate Arctic CO 2 assimilation if they do not account for decreased quantum yield and convexity at low temperature, New Phytologist, 223, 167–179, 2019.

Romanovsky, V. E. and Osterkamp, T.: Effects of unfrozen water on heat and mass transport processes in the active layer and permafrost, Permafrost and Periglacial Processes, 11, 219–239, 2000.

Sasakawa, M., Shimoyama, K., Machida, T., Tsuda, N., Suto, H., Arshinov, M., Davydov, D., Fofonov, A., Krasnov, O., Saeki, T., et al.: Continuous measurements of methane from a tower network over Siberia, Tellus B, 62, 403–416, https://doi.org/10.1111/j.1600-0889.2010.00494.x, 2010.

Sasakawa, M., Machida, T., Tsuda, N., Arshinov, M., Davydov, D., Fofonov, A., and Krasnov, O.: Aircraft and tower measurements of CO_2 concentration in the planetary boundary layer and the lower free troposphere over southern taiga in West Siberia: Long-term records from 2002 to 2011, J. Geophys. Res.-Atmos., 118, 9489–9498, https://doi.org/10.1002/jgrd.50755, 2013.

Schaaf, C. B., Gao, F., Strahler, A. H., Lucht, W., Li, X., Tsang, T., Strugnell, N. C., Zhang, X., Jin, Y., Muller, J.-P., et al.: First operational BRDF, albedo nadir reflectance products from MODIS, Remote sensing of Environment, 83, 135–148, 2002.

Schaefer, K. and Jafarov, E.: A parameterization of respiration in frozen soils based on substrate availability, Biogeosciences, 13, 1991–2001, https://doi.org/10.5194/bg-13-1991-2016, 2016.

Schuh, A. E., Jacobson, A. R., Basu, S., Weir, B., Baker, D., Bowman, K., Chevallier, F., Crowell, S., Davis, K. J., Deng, F., et al.: Quantifying the impact of atmospheric transport uncertainty on CO_2 surface flux estimates, Global Biogeochem. Cy., 33, 484–500, 2019.

Schuldt, K., Mund, J., Luijkx, I. T., Jacobson, A. R., Aalto, T., Abshire, J. B., Aikin, K., Andrews, A., Aoki, S., Apadula, F., Baier, B., Bakwin, P., Bartyzel, J., Bentz, G., Bergamaschi, P., Beyersdorf, A., Biermann, T., Biraud, S. C., Bowling, D., Brailsford, G., Chen, G., Chen, H., Chmura, L., Clark, S., Climadat, S., Colomb, A., Commane, R., and van den Bulk, P.: Multi-laboratory compilation of atmospheric carbon dioxide data for the period 1957-2019; obspack_co2_1_GLOBALVIEWplus_v6.0_2020-09-11; NOAA Earth System Research Laboratory, Global Monitoring Laboratory, https://doi.org/10.25925/20200903, 2020.

Schuur, E. A., McGuire, A. D., Schädel, C., Grosse, G., Harden, J., Hayes, D. J., Hugelius, G., Koven, C. D., Kuhry, P., Lawrence, D. M., et al.: Climate change and the permafrost carbon feedback, Nature, 520, 171, 2015.

Seiler, C., Melton, J. R., Arora, V., Sitch, S., Friedlingstein, P., Arneth, A., Goll, D. S., Jain, A., Joetzjer, E., Lienert, S., and et al.: Are terrestrial biosphere models fit for simulating the global land carbon sink?, Earth and Space Science Open Archive, p. 48, https://doi.org/10.1002/essoar.10508892.1, 2021.

Sitch, S., Friedlingstein, P., Gruber, N., Jones, S. D., Murray-Tortarolo, G., Ahlström, A., Doney, S. C., Graven, H., Heinze, C., Huntingford, C., Levis, S., Levy, P. E., Lomas, M., Poulter, B., Viovy, N., Zaehle, S., Zeng, N., Arneth, A., Bonan, G., Bopp, L., Canadell, J. G., Chevallier, F., Ciais, P., Ellis, R., Gloor, M., Peylin, P., Piao, S. L., Le Quéré, C., Smith, B., Zhu, Z., and Myneni, R.: Recent trends and drivers of regional sources and sinks of carbon dioxide, Biogeosciences, 12, 653–679, https://doi.org/10.5194/bg-12-653-2015, 2015.

Turetsky, M. R., Abbott, B. W., Jones, M. C., Anthony, K. W., Olefeldt, D., Schuur, E. A., Koven, C., McGuire, A. D., Grosse, G., Kuhry, P., et al.: Permafrost collapse is accelerating carbon release, 2019.

Turetsky, M. R., Abbott, B. W., Jones, M. C., Anthony, K. W., Olefeldt, D., Schuur, E. A., Grosse, G., Kuhry, P., Hugelius, G., Koven, C., et al.: Carbon release through abrupt permafrost thaw, Nature Geoscience, 13, 138–143, 2020.

van der Werf, G. R., Randerson, J. T., Giglio, L., van Leeuwen, T. T., Chen, Y., Rogers, B. M., Mu, M., van Marle, M. J. E., Morton, D. C., Collatz, G. J., Yokelson, R. J., and Kasibhatla, P. S.: Global fire emissions estimates during 1997–2016, Earth Syst. Sci. Data, 9, 697–720, https://doi.org/10.5194/essd-9-697-2017, 2017.

Veraverbeke, S., Rogers, B. M., Goulden, M. L., Jandt, R. R., Miller, C. E., Wiggins, E. B., and Randerson, J. T.: Lightning as a major driver of recent large fire years in North American boreal forests, Nature Climate Change, 7, 529, 2017.

Veraverbeke, S., Delcourt, C. J., Kukavskaya, E., Mack, M., Walker, X., Hessilt, T., Rogers, B., and Scholten, R. C.: Direct and longer-term carbon emissions from arctic-boreal fires: a short review of recent advances, Current Opinion in Environmental Science & Health, p. 100277, 2021.

Wang, C., Chen, J., Wu, J., Tang, Y., Shi, P., Black, T. A., and Zhu, K.: A snow-free vegetation index for improved monitoring of vegetation spring green-up date in deciduous ecosystems, Remote sensing of environment, 196, 1–12, 2017.

Wehr, R., Munger, J., McManus, J., Nelson, D., Zahniser, M., Davidson, E., Wofsy, S., and Saleska, S.: Seasonality of temperate forest photosynthesis and daytime respiration, Nature, 534, 680, 2016.

Weir, B., Crisp, D., O'Dell, C. W., Basu, S., Chatterjee, A., Kolassa, J., Oda, T., Pawson, S., Poulter, B., Zhang, Z., Ciais, P., Davis, S. J., Liu, Z., and Ott, L. E.: Regional Impacts of COVID-19 on Carbon Dioxide Detected Worldwide from Space, Science Advances, 7, eabf9415, https://doi.org/10.1126/sciadv.abf9415, 2021.

Wilkman, E., Zona, D., Arndt, K., Gioli, B., Nakamoto, K., Lipson, D. A., and Oechel, W. C.: Ecosystem scale implication of soil CO_2 concentration dynamics during soil freezing in Alaskan Arctic tundra ecosystems, Journal of Geophysical Research: Biogeosciences, 126, e2020JG005 724, 2021.

Xiao, X., Zhang, Q., Braswell, B., Urbanski, S., Boles, S., Wofsy, S., Moore III, B., and Ojima, D.: Modeling gross primary production of temperate deciduous broadleaf forest using satellite images and climate data, Remote sensing of environment, 91, 256–270, 2004.

Yi, Y., Kimball, J. S., Rawlins, M. A., Moghaddam, M., and Euskirchen, E. S.: The role of snow cover affecting boreal-arctic soil freeze–thaw and carbon dynamics, Biogeosciences, 12, 5811–5829, https://doi.org/10.5194/bg-12-5811-2015, 2015.

Yi, Y., Kimball, J. S., Watts, J. D., Natali, S. M., Zona, D., Liu, J., Ueyama, M., Kobayashi, H., Oechel, W., and Miller, C. E.: Investigating the sensitivity of soil heterotrophic respiration to recent snow cover changes in Alaska using a satellite-based permafrost carbon model, Biogeosciences, 17, 5861–5882, https://doi.org/10.5194/bg-17-5861-2020, 2020.

Yu, Z., Wang, J., Liu, S., Piao, S., Ciais, P., Running, S. W., Poulter, B., Rentch, J. S., and Sun, P.: Decrease in winter respiration explains 25% of the annual northern forest carbon sink enhancement over the last 30 years, Global ecology and biogeography, 25, 586–595, 2016.

Zhang, X.: Reconstruction of a complete global time series of daily vegetation index trajectory from long-term AVHRR data, Remote Sensing of Environment, 156, 457–472, 2015.

Zhang, Y., Xiao, X., Wu, X., Zhou, S., Zhang, G., Qin, Y., and Dong, J.: A global moderate resolution dataset of gross primary production of vegetation for 2000–2016, Scientific data, 4, 1–13, 2017.

Zimov, S., Semiletov, I., Daviodov, S., Voropaev, Y. V., Prosyannikov, S., Wong, C., and Chan, Y.-H.: Wintertime CO_2 Emission from Soils of Northeastern Siberia, Arctic, pp. 197–204, 1993.

Zona, D., Gioli, B., Commane, R., Lindaas, J., Wofsy, S. C., Miller, C. E., Dinardo, S. J., Dengel, S., Sweeney, C., Karion, A., et al.: Cold season emissions dominate the Arctic tundra methane budget, Proceedings of the National Academy of Sciences, 113, 40–45, 2016.