In blue: Reviewer's comments. [ ] = Numbering

In black: Answers to referees.
P=Page; L=Line; Track change version
In black and italic: Modification added to text.

Reviewer #1:

Synopsis:

The manuscript by Mavrovic et al. is a nice investigation on the controls of CO2 emissions from high latitude soils during the cold season. Although the study is not showing many things that are particularly new, there are a few insights here that I like. And since there's little data from the cold season in these environments, the data presented in this study is going to quite valuable to a wider audience. Therefore, I think it can be suitable for publication after some adjustments and clarifications.

Thank you for this assessment of our study.

First of all, one of the main conclusions is that non-growing season fluxes during the zero curtain are controlled by liquid water content, and not by temperature. That's not particularly surprising since soil temperatures do not vary during the zero curtain period, so other variables should be more important. Likewise, the authors say that liquid water content is less important than soil temperature when liquid water content is low, but that's pretty much the same thing: it's a subset of the data where there's little change in liquid water content. The authors acknowledge these aspects briefly in their results, but it should be discussed more extensively in the discussion. What is the relative importance of liquid water content? Figure 7 suggests that it's responsible for a huge change in the fluxes.

Based on this comment, Sect. 4.1 (Controls of non-growing season CO_2 fluxes) has been fleshed out to discuss further the respective impact of soil temperature and liquid water content on CO_2 fluxes. The Random Forest model generated in Sect. 3.2.1 (Fig. 4) is used to estimate the predictors' relative importance (i.e., environmental controls).

P16, L441-448: The RF model *predictors' relative importance showed* that *during the non-growing season, T_{soil} emerged as the dominant predictor of F_{CO2} when the soil was frozen. Nevertheless, in the closed-boreal forest site (i.e., MM) where zero-curtain conditions persisted throughout the non-growing season, soil LWC took precedence as the dominant predictor as there was minimal variation in T_{soil} under these conditions.* Our results corroborate the strong non-growing season F_{CO2} dependency on T_{soil} shown by Natali et al. (2019), although we observed fluxes lower than reported by Natali et al. (2019) at T_{soil} < -5°C and mostly higher fluxes at T_{soil} > -5°C. *Considering the two regressions of the relationship between T_{soil} and F_{CO2} have large uncertainties attached to them, the difference between them falls inside the uncertainty margin (Fig. 5).*

General comments:

[1] Otherwise, it's strange to see so few parameters being part of the random forest analysis. Why weren't more soil properties, like C:N ratios part of this analysis? Or calculated available pore space? There's a huge variation in the fluxes, and there are few soil or site-specific differences that are addressed in this manuscript. Other causes for the peaks seen at below-zero temperatures are not discussed, even though it has been discussed in the past that soil-freezeup reduces available pore space, which leads to bursts of greenhouse gases in permafrost environments (see e.g. Mastepanov et al 2013).

All soil variables available were included in the analysis (i.e., suface soil temperature and soil liquid water content). In this study, we aimed to get several study sites and sampling locations to obtain a good overview of the spatial variability. This contrasted with most other studies, which focused on a limited number of study sites (Pork et al., 2016; Webb et al., 2022). Considering the workload associated with winter data collection during harsh field conditions at the selected sites, we had to limit our data collection to the presented data of CO_2 soil gases, snow and environmental measurements. Further, since we conducted our CO_2 flux measurements during winter at most study sites, the soil was too frozen to collect soil samples for laboratory analysis. Consequently, it is acknowledged in the discussion that soil properties and biogeochemistry was not included in our analysis and might explain the unaccounted variability in our results.

In regards to peaks seen at below-zero temperatures, see answer to specific comment [10].

P17, L493-495: The unexplained variance (16%) suggests that non-growing season CO_2 fluxes might have been controlled by other environmental variables such as soil physical-chemical properties regulating soil biogeochemistry and soil redox conditions, which were not addressed *nor measured* in this study.

[2] Finally, the authors compare their results to those from Natali et al. (2019), and then discuss how their regression is different below and above -5 degrees C. However, both regressions have large uncertainties attached to them, so this difference is most likely non-significant. If you think your result is different, please show that with a statistical test.

It is true that the difference between our exponential regression and the one from Natali et al. (2019) falls inside the uncertainties of both regressions. This point was spelled out more clearly in the manuscript.

P16, L447-448: *Considering the two regressions have large uncertainties attached to them, the difference between them falls inside the uncertainty margin (Fig. 5).*

Specific comments:

[1] The title does not show that these are only soil fluxes. Please add that detail. Also, I would say 'cold season' or 'winter' rather than 'non-growing season' since you clearly measured in the middle of winter.

The title was modified to specify that we are looking at soil carbon fluxes. The term non-growing season was chosen instead of winter as it is commonly used in carbon science and more clearly defined. By clarifying that this study focused on soil carbon fluxes, it also implies that we remove the influence of photosynthesis to focus on soil respiration. The terminology was uniformized to non-growing season throughout the manuscript.

Modified title: Environmental controls of non-growing season *soil* carbon dioxide fluxes in boreal and tundra environments

P3-4, L113-115: Spatio-temporal measurements of snowpack CO_2 diffusion gradients were performed at several locations in four sites during the 2020-2021 and 2021-2022 *non-growing seasons* (December to May).

P7, L192-193: All data were collected during the 2020-21 and 2021-22 *non-growing seasons* between December and May (Table 1).

P8, L238-239: Randomly distributed gas samples collected during the 2020-21 *non-growing season* were analyzed with a Picarro G2201-I CRDS gas analyzer (Picarro, Santa Clara, Californie; $\sigma < 0.1\%$; N = 26).

P9, L267-269: LWC was only monitored at the MM site since it was the only site where T_{soil} *remained around 0°C for the whole non-growing season*, allowing liquid water in the soil throughout the *non-growing season*.

P11, Fig. 3: CO_2 flux (F_{CO2}) uncertainty relationship to F_{CO2} for the four study sites and two *non-growing seasons* 2020-2021 and 2021-2022. Specifications of the linear fit can be found in the upper left. The data dots color indicates the study site and its symbol (i.e., circle or x-shaped) indicates the *non-growing season* during which it was collected.

P14, Fig. 6: CO_2 flux (F_{CO2}) as a function of soil temperature (T_{soil}) at the Montmorency Forest study sites where soil liquid water content (LWC) was greater than 0 m^3/m^3 through the *non-growing season*. An exponential regression was fitted to the data (black line).

P15, L420-421: Higher F_{CO2} can be explained by warmer mean annual average temperature, a deeper snowpack and *non-growing season* T_{soil} around 0°C (See Sect. 3.4).

P16, L441-444: The RF model supported that *during the non-growing season, T_{soil} emerged as the dominant predictor of F_{CO2} when the soil was frozen. Nevertheless, in the closed-boreal forest site where zero-curtain conditions persisted throughout the non-*

growing season, soil LWC took precedence as the dominant predictor as there was minimal variation in T_{soil} under these conditions.

P16, L458-459: Soil LWC was observed only at the MM site, where T_{soil} was around 0°C throughout *the non-growing season*.

P17, L479-484: Our study shows that abiotic variables related to T_{soil}, LWC, and physical snowpack properties explain the majority of variance in *non-growing season* CO_2 fluxes. It should be noted that we did not incorporate variables related to temporal dynamics such as the previous days' soil temperature and LWC, which have been shown by Harel et al. (2023) to be of importance during the growing season. However, *non-growing season* soil variables are not expected to be as dynamic as during the growing season because of the snowpack insulating properties.

P17, L493-495: The unexplained variance (16%) suggests that *non-growing season* CO_2 fluxes might have been controlled by other environmental variables such as soil physical-chemical properties regulating soil biogeochemistry and soil redox conditions, which were not addressed in this study.

P18, L523-524: However, we found that at our site maintaining zero-curtain conditions throughout *the non-growing season*, LWC becomes the main control of non-growing season F_{CO2}.

[2] Page 6, Equation 2: it's a minor detail but this is only true if there's no liquid water in the snowpack. Looks like your sites did not experience melt events, so this is probably not an issue. Could be mentioned though.

There are some measurements at the Montmorency Forest site where there was liquid water in the snowpack. The more general equation for snow porosity was used in this instance. It was corrected in the manuscript.

P6, Eq. 2: $\varphi = 1 - \frac{\rho_{snow}}{\rho_{ice}} + \theta \cdot \left(\frac{\rho_{snow}}{\rho_{ice}} - 1\right)$

P6, L169-170: where ρ represents the density of snow and pure ice ($\rho_{ice} = -0.0001 \cdot T_{ice} + 0.9168$ with T_{ice} as ice temperature in °C and ρ_{ice} in g cm^{-3}; Harvey et al., 2017) *and θ is the snow liquid water content*.

P8, L222-226: Snow properties *were measured at every 5 cm* including snow temperature (Snowmetrics digital thermometer; Fort Collins, Colorado; tenth of a degree resolution), snow density (Snowmetrics digital scale, 100 and 250 cm^3 snow cutters *used to weigh snow samples*; $\sigma(\rho_{snow}) \approx 9\%$; Proksch et al., 2016), *snow liquid water content (hand test from Fierz et al., 2009)* and snow stratigraphy.

P22, L691-693: *Fierz, C., R. L., A., Durand, Y., Etchevers, P., Green, E., McClung, D., Nishimura, K., Satyawali, P., and Sokratov, S.: The International Classification for*

Seasonal Snow on the Ground, IHP-VII Technical Documents in Hydrology N83, IACS Contribution N1, UNESCO-IHP, Paris, 2009.

[3] Line 217-218: did you determine average snow density for the snowpack or did you make a profile? I wonder how the density differences between the depth hoar and wind slab affects your calculations.

The snowpack density was measured every 5 cm, as we specified in the manuscript. Although the density of the depth hoar and wind is typically different, the diffusion gradient was still linear. Therefore, the snowpack average density was used for the calculation.

P8, L222-229: Snow properties *were measured at every 5 cm* including snow temperature (Snowmetrics digital thermometer; Fort Collins, Colorado; tenth of a degree resolution), snow density (Snowmetrics digital scale, 100 and 250 cm^3 snow cutters *used to weigh snow samples*; σ(ρsnow) ≈ 9%; Proksch et al., 2016), *snow liquid water content (hand test from Fierz et al., 2009)* and snow stratigraphy. T_{soil} was measured at 1 cm depth under the soil/snow interface *(Snowmetrics digital thermometer; Fort Collins, Colorado; tenth of a degree resolution)*, three measurements of T_{soil} were averaged. Snow depth measurements were done with a ruler graduated every 1cm ($\sigma(d_{snow}) \approx 0.5$cm).

[4] Line 219-221: why only measure Tsoil at such a shallow depth? I would expect respiration to be relatively high across the root zone, and temperatures may differ with depth, affecting your correlations.

Soil temperature was measured at a shallow depth because it was not possible to go deeper in frozen soil and no permanent sensors were installed at the individual sampling locations. It is now mentioned un the manuscript.

P8, L226-229: T_{soil} was measured at 1 cm depth under the soil/snow interface *as it was not possible to go deeper in frozen soil and no permanent sensors were installed* (Snowmetrics digital thermometer; Fort Collins, Colorado; tenth of a degree resolution), three measurements of T_{soil} were averaged.

P16, L454-456: *It should be reminded that the T_{soil} used in his study refer to near-surface temperature, deeper T_{soil} may vary and affect the correlation with F_{CO2}.*

[5] Line 223: why only 86%?

The rest of the gas samples were processed by an independent lab (Groupe de recherche interuniversitaire en limnologie, Université de Montréal) to validate the method that was used at the Université du Québec à Trois-Rivières lab for most of the samples. It was clarified in the manuscript.

P8, L238-242: Randomly distributed gas samples collected during the 2020-21 non-growing season were analyzed with a Picarro G2201-*i* CRDS gas analyzer (Picarro, Santa

Clara, Californie; σ < 0.1%; N = 26) *to validate the method used with the LI-7810 to determine CO_2 concentration*. CO_2 concentrations estimated from the LI-7810 and Picarro G2201-*i* gas analyzers were not significantly different in their concentration range and distribution (Fig. A2; $R^2 = 0.92$).

[6] Line 241: these are only random errors related to your calculations, assuming that the method is perfect. Which systematic errors may have affected your measurements?

It was clarified that Sect. 2.2.3 focuses on random errors while systematic errors are discussed at the end of Sect. 2.2.1.

P6, L183-190: The diffusion gradient method assumes that gas fluxes are the result of simple, linear, gradient-induced diffusion in uniform porosity through snow cover (McDowell et al., 2000). A snowpack with strongly heterogeneous density (i.e., vertical stratification) can induce a bias when gas flow is altered by dense layers or ice crusts, typically leading to F_{CO2} overestimation (Seok et al., 2009). Such layers were rarely found in our study sites. The diffusion gradient assumption also does not hold when strong wind events occur, decreasing snowpack CO_2 concentration through wind-pumping and inducing a negative bias on CO2 fluxes (Seok et al., 2009). Consequently, $d[CO_2]/dz$ was not measured in days following a strong wind event.

P8, L249-250: *The uncertainty assessment focuses on random errors, as systematic errors are discussed at the end of Sect. 2.2.1.*

[7] Line 248-249: the uncertainty analysis is not explained well. What is the min-max uncertainty propogation method? Please elaborate, and give a reference.

Details were added on the uncertainty analysis as follow:

P9, L258-261: *F_{CO2} uncertainty was estimated by propagation of the uncertainties of $d[CO_2]/dz$ and snow density using Eq. 1 (Taylor, 1997). The uncertainty of ρsnow was fixed at 9% (Proksch et al., 2016) while the uncertainty of $d[CO_2]/dz$ was estimated based on the root mean squared error of the linear regression for each snowpack concentration gradient measurement.*

P29, L960-961: *Taylor, J.R.: An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements. 2nd Edition, University Science Books, Sausalito, United States, 343 pages, ISBN-10: 093570275X, 1997.*

[8] Line 256-258: this is written a bit strangely. Zero curtain happens at all your sites, you simply were not there to measure it.

The point we wish to convey is that Montmorency Forest is the only site where zero-curtain conditions last the whole non-growing season, while it only occurs during shoulder season freezing and thawing at the other sites. The sentence was rephrased for clarity.

P9, L267-270: LWC was only monitored at the MM site since it was the only site where T_{soil} *remained around 0°C for the whole non-growing season*, allowing *the presence of* liquid water in the soil throughout the *non-growing season*.

[9] Line 354: which lack of measurements between -0.5 and -6? There a few. Or do you mean the number is too low?

We meant that the number of measurements is low, it was clarified in the manuscript as follows:

P12, L371-372: Note that the *low number* of F_{CO2} measurements with T_{soil} between -6°C to -0.5°C restrict the capacity to evaluate the regression within this range.

[10] Line 362-363: like I mentioned before, this may be due to changes in available porosity. See also Pirk et al. (2015) for how this works with methane fluxes (but same principal holds for CO2). Other useful papers are Zona et al. (2016) and Raz-Yaseef et al. (2017), who showed similar bursts with eddy covariance. Otherwise, I wonder whether changes in air pressure may have played a role, which may affect the storage in deep snow packs.

Unlike Pirk at al. (2015), who observed CH_4 bursts during autumn freeze-in, we observed a few high F_{CO2} during winter when T_{soil} was between -25°C and -10°C. It seems unlikely that much ice formation occurs at those T_{soil}, but maybe some soil cracking. This hypothesis was added to the manuscript.

The challenge of air pressure change is addressed by Seok et al. (2009). Change in air pressure is strongly correlated with the wind speed. Changes in air pressure and wind speed decrease snowpack CO_2 concentration inducing a negative bias on CO_2 fluxes. There are still uncertainties on the impact of changes in air pressure and wind on the gas diffusion method, Seok et al. (2009) proposed corrections that can be applied to account for strong winds. In our study, we did not conduct measurements close to strong wind events to account for this uncertainty. No disparities were observed in the range and average of our measurements over time during the course of our two-week campaigns, which indicates that no drastic changes in snowpack CO_2 concentration seem to have occurred during our measurements.

P13, L380-384 *It has been suggested that gas bursts during autumn freeze-up in permafrost environments might be due to gas compression by ice formation and ground cracking (Pirk et al. 2015). This hypothesis can be considered to explain the high FCO2 observed in this study, although the high F_{CO2} observed occurred at a near-surface T_{soil} between -25°C and -10°C so the freeze-up would have to occur at lower depths in the soil.*

P26-27, L876-879: *Pirk, N., Santos, T., Gustafson, C., Johansson, A., Tufvesson, F., Tamstorf, Parmentier, F.-J., Mastepanov, M., and Christensen, T.: Methane emission*

bursts from permafrost environments during autumn freeze-in: New insights from ground-penetrating radar. Geophysical Research Letters, 42(16), 6732-6738, doi: 10.1002/2015GL065034, 2015.

[11] Line 425-427: but Natali et al. also had many more datapoints, so their estimate is better constrained. Anyway, I doubt there is a statistically significant difference between the regressions in your two studies.

It is true that the difference between our exponential regression and the one from Natali et al. (2019) falls inside the uncertainties of both regressions. This point was spelled out more clearly in the manuscript.

P16, L447-448: *Considering the two regressions have large uncertainties attached to them, the difference between them falls inside the uncertainty margin (Fig. 5).*