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 #2:

Synopsis:

This manuscript reports data on non-growing season CO_2 fluxes from 4 different sites in the Arctic-boreal region. Three sites are Arctic while one site is Boreal. Measurements of CO_2 concentrations down through the snowpack were done over two consecutive years. Snow samples were used to infer the diffusion coefficient of the snow in order to calculate CO_2 fluxes based on the concentration gradient using Fick's law of diffusion.

As the authors point out, there is still today a lack of data and understanding of what governs CO_2 flux rates during the non-growing season in the Arctic. Based on the measurements performed the authors show that soil temperature is the dominant predictor of resulting CO_2 fluxes at sub-zero temperatures across sites, whereas during zero-curtain conditions liquid water content becomes the primary predictor of CO_2 flux. These two variables dominated also over e.g. vegetation type. This is an interesting result warranting publication. However, some issues should be resolved before the manuscript is ready for publication.

Thank you for this assessment of our study.

General comments:

[1] As such, the applied field methodology seems sound and well described except for a few critical details that should be explained further, including how many and in which depths snow samples were done for obtaining information about the snowpack conditions. The authors explain how the snow properties change down through the profile but it is unclear if this is reflected in the snow sampling procedure – e.g. how are changes down through the profile in terms of changes in diffusivity incorporated into the flux calculations.

The snowpack density was measured every 5 cm, it was 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*; $\sigma(\rho snow) \approx 9\%$; 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).

[2] More critically, it is clear that liquid water content was only measure at one site, but unclear if the liquid water content was then estimated at all the other sites except for MM, based on soil temperature and soil properties. It seems to me that this is what was done but it is unclear and this disturbs the overall understanding also of the RF modelling, where there may or may not be missing LWC estimates from 3 of the 4 sites. The authors should clarify explicitly which LWC data were available from all sites for the RF model. If data on LWC were indeed not estimated for the 3 other sites, and therefore missing in the RF modelling, the authors should explain how the RF model handles these missing values and how this affects the interpretation of the RF model result for the importance of LWC when all data are included in the analysis. Except for this lack of clarity, the analyses performed also seems sound.

LWC was only calculated at MM and was estimated as negligible (i.e., LWC ≈ 0) for all other sites. This assumption is supported by the model from Zhang et al. (2010) and served as input for the Random Forest model. This assumption was spelled out in the manuscript.

P9, L267-273: 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*. The Zhang et al. (2010) empirical soil liquid water and ice mixing model was used to calculate soil liquid water content (m_{uw}) *(Eq. 5 to 8). LWC was estimated to be negligible at the CB, TVC and HPC sites since T_{soil} was in-between -5°C and -25°C. The model from Zhang et al. (2010) supports that at T_{soil} colder than -5°C, LWC is negligible.*

[3] An important finding reported is the shift from temperature-dependent CO2 fluxes at sub-zero temperatures to liquid-water dependent fluxes at zero curtain conditions. The authors should here also reflect on the potential bias that this was only observed at one of the 4 sites and consequently this result may be affected also by site-specific differences. Particularly because this is also the warmer and boreal site, where the other sites are all arctic.

In general, the authors focus a lot on RMSE of the different models, but reflect less on the other estimated parameters. E.g. the temperature dependency parameter (B) in figure 5 and 6 differ strongly but a discussion of the potential impact of these differences is lacking. I also suggest relating RMSE, e.g. also in the abstract to mean/median or e.g. seasonal fluxes in order to be able to judge the error in more relative terms.

It is correct that caution should be taken about overgeneralizing our findings from a single site with measurements in zero-curtain conditions. We clarified it in the manuscript. It is also correct that the temperature-dependency of F_{CO2} to T_{soil} changed significantly between the two regimes (i.e., T_{soil} < 0^{o}C and zero-curtain conditions). Following this comment, we pointed out that it is clearly displayed by the disparities in the temperature-dependency parameter B. This is the reason why it was not possible to fit a single regression over the measurements at T_{soil} < 5^{o}C and T_{soil} ≈ 0^{o}C. RMSE was added as a percentage of mean F_{CO2}.

P11, L337-340: *While the first regime mostly corresponds to Arctic study sites, the second regime only includes one study site (MM) located in the southern boreal forest. Therefore, conclusions from the second regime should be less generalized than those from the first regime.*

P13, L386-393: F_{CO2} increases *more* rapidly with T_{soil} around freezing point than at T_{soil} < 5^{o}C, which is *shown by the higher temperature-dependency parameter (B = 2.82 $^{o}C^{-1}$) of the MM site* exponential regression (RMSE = 0.286 gC m^{-2} day^{-1}) *compared to the exponential regression of Fig. 5 (B = 0.18 $^{o}C^{-1}$). This discrepancy in temperature-dependency creates a discontinuity between the measurements at T_{soil} < 5^{o}C and T_{soil} ≈ 0^{o}C that did not allow for a continuous temperature-dependency regression across all the study sites. The lower RMSE of the exponential regression of Fig. 5 (RMSE = 0.024 gC m^{-2} day^{-1}; 70.3% of mean F_{CO2}) compared to the exponential regression of the MM site (RMSE = 0.286 gC m^{-2} day^{-1}; 112.4% of mean F_{CO2}) might be due to the impact of soil LWC at the MM site (see Sect. 3.2.3).*

P14, L404-406: The relationship between LWC and F_{CO2} during the non-growing season at MM (RMSE = 0.137 gC m^{-2} day^{-1}*; 49.1% of mean F_{CO2}*) was stronger than between T_{soil} and F_{CO2} (RMSE = 0.286 gC m^{-2} day^{-1}*; 112.4% of mean F_{CO2}*) …

P16, L460-464: *It should be noted that it would be ill-advised to generalize the relationship between soil LWC and F_{CO2} as it is only based on data from one study site, and it cannot be ruled out that this relationship is site-specific depending on soil and vegetation composition. Nevertheless,* our study highlighted the important impact of LWC on F_{CO2} around soil freezing point when there is a mixed state of ice and free water in soils.

P16, L469-471: *Further research on non-growing season F_{CO2} in zero-curtain conditions should investigate different sites to assess if the relationship between F_{CO2} and soil LWC is site-specific or dependent on soil properties.*

[5] In addition, as a suggestion, the authors could consider if a combined model, taking into account both soil T and LWC at the same time could be even better than the presented alternative models only taking one or the other variable into account. After all, both temperature and liquid water are co-limiting the CO2 fluxes but to different extent in the two temperature regimes. If such a combined model could work for both sub-zero and zero curtain conditions it would be a robust model to use for winter conditions in general in the boreal/arctic region.

This is a good suggestion. However, measurements under zero-curtain conditions were only collected at a single study site, a combined model looking at T_{soil} and soil LWC would only represent that one site. To avoid overgeneralizing the results from this site, we did not produce a combined model for both T_{soil} and LWC as it would not take into account potential site-specific differences. To create such a model with confidence, measurements during zero-curtain conditions from other sites would be required. It should be noted that the Random Forest model generated in Sect. 3.2.1 is a model that combined the T_{soil} and LWC effects, although we only exploit it to get an estimate of the predictors' relative importance (i.e., environmental controls) as the authors (and also as proposed by the reviewers) do not want to overgeneralize results from the MM study site.

Specific comments:

[1] L29-30: Exponential relationship with temperature is expected – but they differ in the two situations. Why do you report only RMSE here? Yes, it indicates that it is a better model but why not a line on what was the effect of liquid water availability (i.e. how was the model characterized)?

RMSE was used as a metric to evaluate and compare the regressions as correlation coefficient cannot be used with exponential regressions. More details about the results were added in the abstract.

P1, L29-33: We observed exponential regressions between CO_2 fluxes and soil temperature in *fully* frozen soils *(RMSE = 0.024 gC* m^{-2} day^{-1}*; 70.3% of mean* F_{CO2}*)* and soils around freezing point *(RMSE = 0.286 gC* m^{-2} day^{-1}*; 112.4% of mean* F_{CO2}*).* F_{CO2} *increases more rapidly with* T_{soil} *around freezing point than at* $T_{soil} < 5^oC$*. In zero-curtain conditions, the strongest regression was found with soil liquid water content (RMSE = 0.137 gC* m^{-2} day^{-1}*; 49.1% of mean* F_{CO2}*).*

[2] L119: Is a more natural order of the sections here to switch to have data collection before CO2 flux calculations? Consider the same in the result section.

It was chosen to put the theoretical framework of the snowpack diffusion gradient method before discussing data collection to allow the reader to understand why we are measuring certain environmental parameters that might not be obvious otherwise. We understand that there are divergences among researchers about the order of presenting the theoretical framework and data collection, but we think that in the framework of our study, the proposed order is adequate.

[3] L129: NWT not explained

NWT is now spelled out.

P4, L131-133: Trail Valley Creek (TVC), *Northwest Territories*, situated just north of the treeline in the transitional zone between the boreal and Arctic biomes close to the

Mackenzie delta, is dominated by erect-shrub tundra with remaining tree patches (Martin et al., 2022).

[4] L152: In Jones et al 1999, d[CO2]/dz has the unit of ppmv m-1 and you are shifting from ppmv to g C m-3. Also you do not explain that z is in meters and that is how you get to this unit. Please use one line to explain this.

Details were added on how to convert gas concentration units (i.e., ppm to gC m^{-3}) and how interpret a vertical CO_2 diffusion gradient.

P5, L154-157: Consequently, a vertical CO_2 diffusion gradient is maintained through the snowpack ($d[CO_2]/dz$; gC m^{-4}), *with CO_2 concentration ($[CO_2]$; gC m^{-3}) decreasing with snow height from the soil surface (z; m) (Jones et al., 1999). Hereafter, $[CO_2]$ is expressed in gC m^{-3} but units of concentration could also be expressed in relative units (i.e., ppm) using the ideal gas law.*

[5] L176: D usually should have the unit of m2 time-1 (often seconds but in your case recalculated to daily). You should state the unit, so that the reader can follow how the resulting unit for FCO2 arises.

The units of D were spelled out next to the equation. Note that the unit is also spelled out a few lines earlier.

P6, L177-179: Standard diffusion coefficients of CO_2 *(unit: m^2 day^{-1})* are available in literature but must be corrected for temperature and pressure (Marrero and Mason, 1972; Massman, 1988):

[6] L189: Did you sample from the top of the snow pack first and pushed the sampling rod deeper? Please explain in detail.

Indeed, the gas samples were collected from top to bottom by pushing the sampling rod downward. It was clarified.

P7, L196-199: Gas present in snow pores was collected with a thin, hollow stainless-steel rod (50-120 cm long, 4 mm outer diameter and 2 mm inner diameter) *starting with gas samples in the upper snowpack and then pushing the sampling rod downward to collect gas samples deeper in the snowpack* to minimize snow disturbance (Fig. 2a).

[7] L218: How many different snow densities were measured in the different profiles? And how exactly did you sample snow for density estimation? Please give more info on that.

The snowpack density was measured every 5 cm, it was 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).

[8] L243: – snowpit measurements – more precisely snow density measurements, right? Or do you mean all three (density, porosity, tortuosity?) please clarify.

It was clarified that the snowpit measurement uncertainty sources come from snow density and temperature. Afterward, snow density uncertainty propagates to porosity and tortuosity (Eq. 2 and 3) and snow temperature uncertainty propagates to the diffusion coefficient (Eq. 4).

P8, L251-253: Uncertainties can be subdivided into four sources: gas concentration estimates, gas transfer/transport/storage, evaluation of the snowpack d[CO_2]/dz and snowpit measurements *(i.e., snow density and temperature)*.

[9] L251: Zero-curtain conditions – it's a little unclear here whether you are defining this term here – which I believe you are. I suggest rephrasing to "Zero-degree Celsius curtain conditions exist when the soil temperature is around freezing point (0°C) for longer periods of time and a mix…" – and maybe even shortly explain in the abstract too as not all readers will be familiar with this term.

It was clarified that the first sentence of Sect. 2.3 is aiming at defining the Zero-curtain conditions.

P1, L25-27: We identified that soil temperature as the main control of non-growing season CO2 fluxes with 68% of relative model importance, except when soil liquid water occurred during zero-degree Celsius curtain conditions (*i.e.*, Tsoil ≈ 0°C and liquid water coexist with ice in soil pores).

P9. L263-265: Zero-degree Celsius curtain conditions *exist when* the soil temperature is around freezing point (0°C) and a mix of ice and liquid water coexists in the soil pore space because the phase transition between water and ice is slowed due to latent heat (Outcalt et al., 1990).

[10] L258: It is unclear if you mean that you estimated LWC at all the other sites this way. This is critical to clarify.

LWC was only calculated at MM and was estimated as negligible (i.e., LWC ≈ 0) for all other sites. This assumption is supported by the model from Zhang et al. (2010) and served as input for the Random Forest model. This assumption was spelled out in the manuscript.

P9, L267-273: LWC was only monitored at the MM site since it was the only site where T_{soil} *in upper layers remained* around 0°C for the whole *non-growing season*, allowing liquid water in the soil throughout the *non-growing season*. The Zhang et al. (2010) empirical soil liquid water and ice mixing model was used to calculate soil liquid water content (m_{uw}) *(Eq. 5 to 8). LWC was estimated to be negligible at the CB, TVC and HPC sites since T_{soil} was in-between -5°C and -25°C. The model from Zhang et al. (2010) supports that at T_{soil} colder than -5°C, LWC is negligible.*

[11] L281: "…based on a **multitude of** decision trees".

Corrected.

P10, L295-296: Random forest (RF) is an ensemble machine learning method based on a *multitude of* decision trees (Breiman, 2001).

[12] L298: Since a major part of the uncertainty is related to density, then it is even more important to know exactly how it was sampled.

See answer to specific comment [7].

[13] L305: in principle your figure legend symbol indications are not entirely correct, but I see the point. Maybe add to figure text that colors indicate site and symbol type indicate year…

Clarifications about the legend were added to the figure text.

P11, Fig. 3: *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.*

[14] L320: But these regimes coincide with regime 1 being the northern sites and regime 2 being MM – seems like a major confounding factor of this cross site analysis, as site-specific conditions may

See answer to general comment [2].

[15] L328: SWE only explained first time a few lines later…

The definition of SWE was moved to its first mention.

P11-12, L344-349: T_{soil} was the F_{CO2} predictor with the highest relative importance (68%) when using the complete dataset (Fig. 4a), followed by LWC (17%). Snowpack characteristics, ρ_{snow} (11%) and *snow water equivalent (SWE)* (2%), had a lower relative importance in the RF model. Contrary to what might be expected, the vegetation type had near-negligible relative importance (1%) in F_{CO2} prediction. The RF model was developed starting with all environmental variables available: T_{soil}, LWC, vegetation type, *SWE*, snow depth, mean ρ_{snow}, max ρ_{snow}, φ, τ, wind slab fraction and wind slab thickness.

[16] Also – how can you include LWC in a full model across all sites when it was only measured at the MM site? Connected to my question above if you estimated LWC at the other sites with the method described above.

See answer to specific comment [10].

[17] L410: Figure 8 – shrubs, not schrubs.

It was corrected in Fig. 8.

[18] L421: confirms not corroborates

Corrected.

P16, L444-446: Our results *confirm* 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 $T_{soil} < -5^oC$ and mostly higher fluxes at $T_{soil} > -5^oC$.

[19] L445: soil pores not soil porosities

Corrected.

P16-17, L474-477: It is well known that anaerobic conditions created by high soil moisture (at least > 50%) constrain soil CO_2 respiration rates during the growing season because many microorganisms require oxygen for organic matter decomposition which they lack if soil *pores* are filled with water (Linn and Doran, 1984; Davidson and Janssens, 2006).

[20] L463: availability and **quality** of labile C

Added.

P17, L495-498: CO_2 production is governed by the availability *and quality* of labile C compounds regulating the decomposition of soil organic matter (Michaelson et al., 2005; Wang et al., 2011), and the activity and composition of the soil microbial community (Monson et al., 2006).

[22] L990: Figure A1 + A3: should be ‘shrubs’ – not ‘schrubs’

It was corrected in Fig. A1 and A3.