Overview

This study examines the impact of fire severity and logging/silviculture on soil greenhouse gas fluxes in the Scots pine-dominated boreal forests, which fits within the scope of BG. It utilizes the static chamber method to measure gas fluxes from forest soils over a four-year period following a large natural wildfire in Sweden (2018). The data it presents help to fill the knowledge gap about greenhouse gas fluxes in the first years after wildfire in the European boreal forests of Sweden; the emphasis is placed on CO_2, while CH_4 results are not discussed in such detail. The main findings are explained well by existing work and data, but in general, the analyses in this paper do not deepen our overall understanding of these biogeochemical processes. However, this paper has a generally good structure and appropriate references.

I see that some opportunities were missed in the data analyses that could provide insight into the drivers behind the fluxes and give more depth to the findings and discussion included here. Some more minor comments pertaining to writing style are also included below. I also have some concerns about possible sources of error.

Main concerns

No unmanaged control areas were included adjacent to salvage-logged and regenerated sites. At a scale of kilometers (3 km as stated in the manuscript), it allows for variation in stand and soil properties (including chemistry and biology) due to a combination of natural influences and stand management history. In this case, where burned unmanaged forests were not adjacent to salvage-logged ones and stand ages differ, some site-level description (some landscape-level details are provided, which suggests there is variation in ecology) of pre-fire ecology and management history (thinning, fertilization) to show that the sites were comparable prior to treatments would support the work. I appreciate that <u>unburned</u> controls are not available adjacent to burned, managed sites.

By using different gas analyzers in different years with no overlap, you are unable to separate differences between years from differences due to the analyzer (Los Gatos vs LICOR). This is a concerning source of error that has not been addressed. Any difference in the flux measured by each device should be accounted for and discussed as a source of error.

The use of mixed models to analyze fluxes is clearly described. However, when building mixed-effects models, why were the only environmental data included soil temperature (for R_{ff}) and moisture (for CH_4)? Expanding the models to include soil chemistry measurements and vegetation cover would allow you to use your own measurements to address sources of variation in the measured fluxes, which is included in the discussion. At least in L441-442 and L497-502 it is mentioned that there is a causal effect between vegetation, soil chemistry and fluxes, so why not test it with your own data? This would improve the interpretation of the results and the depth of the discussion. If these were tested, but no significant impact was found, it also would provide value for the reader to know this.

While the use of beta regression is appropriate for the proportional cover data, why were multivariate techniques not used to assess changes in vegetation over time? This could strengthen the understanding of the nature of the changes and differences in plant communities between sites. For example, NMDS and adonis could be applied to test differences in communities between sites and years. Why was vegetation measured at species-level if only overall cover was considered?

On the SLM site, half the collars were on furrows, while the other half were on ridges. By pooling these to represent the whole site, it assumes that each of these surface types covers exactly half of the sites. Is this accurate, or would weighting the fluxes by actual precent cover of each surface type for pooling the data provide a more accurate analyses of the site? Please at least report the actual percent cover of each surface type (furrow, ridge) on the site. If this is much different than half-and half, the fluxes of each surface type should be weighted by proportional cover for pooling. This comment is not relevant if there is no significant difference between the fluxes on ridges and furrows – was this tested? If so, it should be reported. Addendum: I have now found this in the supplement (please reference this so the reader knows where to look), and it looks like CH_4 differs significantly between scarified and unscarified surfaces, which raises the question of whether these differences are coming from the removal of the vegetation or from disturbing the soil.

Figure captions should provide enough information for figures to be easily interpreted without reference to the text. Please go through the captions (including supplement) and ensure that this is the case for all the figures. number of replicates? years? sensors?

Figure 5. What year are these data from? Please update the caption to include this and which sensors the data are from. Why are the data from only one year presented if you have data from the whole study period? Data from multiple years would help to understand the variation between the years, which may have impacted fluxes.

Section 4.5. Your aims and hypotheses focus on fluxes, so why are you reporting and discussing this in such detail? I know that vegetation can help to explain fluxes, but here, vegetation is not included as a predictor of fluxes in your models so it becomes ancillary, not directly adding to the main story. Somehow, this section doesn't quite fit with the way the research is structured and the data are currently analyzed.

The conclusions section is redundant.

Minor comments

Beware capitalization rules (e.g. L96 podzol, gas analyzer)

The word "data" is almost always plural, i.e. "data **are,**" unless you are talking about a single point.

Remove unnecessary gridlines in tables.

Be aware of correct hyphenation. "Mixed-effects models" or "mixed models." When used as an adjective, "low-severity" and "high-severity" is correct, be consistent.

When introducing the sites, I don't like the usage of the wording "groups," which makes it seem like these are actual groups of unique sites. Consider a different term. These "groups" seem to be more akin to comparison sets than actual groups of sites (L106-115; L167-169). Comparisons were only done within groups, not between them, making them comparison sets, i.e. group 2 contains a site from group 1, but groups 1 and 2 are not compared.

The methods section lacks a parallel structure. Most data analyses are grouped in the same section as the methods for collecting those data. However, flux analyses are given their own section. The separate section for flux analyses can be eliminated (other data analyses are described under the same section as the relevant data collection) or consider putting all data analyses in a separate section.

Tables 2,3, and 4 could be moved to supplementary information. The results of these ANOVA/Chi-squared analyses (significance) are already indicated by the letters on the boxes in Figs. 2 and 6.

Lines

L48: Currently reads that soil respiration leads to high tree mortality? Please fix this sentence.

L87-88: you don't actually use these data to explain fluxes - they are analyzed separately.

L98: there should be a clearer definition of severity as it relates to soils, especially for high-severity areas (was there any complete consumption of the upper organic layers?). tree mortality ok.

Table 1 (L128): is very informative for detailing the experimental design, but format to remove unnecessary grid lines. Please show *n* (in the caption at least) for forest floor organic/charred depth measurements.

Fig. 1 (L134): align UM pic with other 2019 photos

L147-148: beware of capitalization of common nouns. What is the model of the Los Gatos analyzer? GLA132-GGA?

L156: min = minute, use consistent unit abbreviations (see "second" on L157)

L 203: What were the 4 composite samples per site? Was each composite sample a different layer (i.e. four layers, and all the material from that layer in each sample was pooled)? Or did the layers get pooled together? It's not clear to me. Or were they pooled so that each transect had 2 pooled samples at the end: one organic/forest floor, the other mineral?

L208: reference? data?

L225: The way the methods are currently worded, it is unclear if the vegetation surveys took place annually or took 3 years to complete. This can be inferred as annual surveys from Fig. 6 but is not immediately evident.

L247-248: "lower but not as low as" is awkward. Suggestion: "intermediate."

L 263-265: Reporting each value is tedious. These values are already presented in Fig. 2 (d-f). A supplementary table with exact values could be included if it is necessary to include precisely these.

L268-269: Does this also refer to Fig 2?

L273: Fig.2 "...where letters above the **boxes**..."

L275: This period should be a comma? add the word "and"

Fig 3 (L281): The legend needs to be simplified. Sites are listed multiple times.

Fig. 4. (L321-344): Please check the calculations for these points in regards to the error. Why are the SE error bars so small and nearly non-existent for some data points? Also, reference to the wrong supplementary figure.

Fig. 6. (L417): Why is the unburned control not included here? You need to include this, otherwise there is no "baseline" control. Panels d-f show that the change in cover over time for some sites appears to be significant. In this case, is it meaningful to group several years together in panels a-c (panels d-f already show which sites have higher cover in addition to annual changes)? Even better, these analyses can be replaced with multivariate methods.

L428, L458 etc., no need to refer to results figures in the discussion

467-468: Was this relationship tested and found not to be causal (i.e. that you didn't find any controls of temperature/moisture on Rff?), or was this an observation that temp/moist was higher, but not Rff?

L504-512: I'm not sure that this is necessary – how does it serve to discuss/explain the result that was found?

Supplement

I don't understand how Fig. S1, Table S1, and Table S2 improve our understanding of the data compared to what is already presented in the results. I see it covers a different time period (July/Aug vs entire growing season).

Fig. S1 Also, the legend needs to be simplified; sites are listed multiple times.

Fig S4 - year? sensor?