A review of Soil CO2 flux across a permafrost transition zone: spatial structure and environmental correlates.

Alfred Stein, University of Twente

The paper is an interesting and important combination of a statistically well rooted soil study. A strong point is that a very careful analysis has been carried out on a timey and important problem. Another strong point is that a geostatistical analysis could have been carried out, thanks to the large number of relevant data that were collected. Further, the manuscript is fairly well written, and the scientific logic could be followed throughout.

There are some choices that need a better justification, though.

1. I was somewhat surprised by the sampling design. It appears that the transects of the 51 soil collars are not equi-distant; how is this choice made, and why did you deviate from equal distances? In figure 1 it is also clear that some of the soil collars were removed. In the four transects it are always groups of collars that were removed. What was the reason for this choice? Needless to say, this choice could have an effect on the final outcomes. It would be good if a discussion paragraph could be added on this point.

We used a cyclic sampling design, which has previously been shown to provide a more efficient approach for spatial sampling and also provides a more robust variogram due to more evenly distributing sample pairs across spatial distances. This is discussed on page 3. The missing soil collars were due to another study collecting soil cores at those locations between the two seasons sampled in the current manuscript. To evaluate the influence of removing these locations we performed analyses with these locations removed for both seasons; removing those locations did not qualitatively alter our results or change any inferences. These analyses are presented in Supplementary Table S1 and Fig. S2 and are noted throughout the Results section.

2. What surprised me in the end are the large differences. In the paper the terms 'summer' and 'fall' are mentioned, but the observations are just a few weeks apart (August vs. September). Looking at the tables 2 – 4 , however, we notice substantial differences. Even a change in sign occurs (for Soil temperature in the permafrost-free stratum). Maybe it has to do with the direction of the fluxes because of the weather conditions, or the expansion of the frozen soil a few weeks later. The manuscript requires a better definition of the 'summer' and 'fall' terms which would make it more likely that relatively large differences occur that are more than just coincidences.

In boreal Alaska conditions change very rapidly in September, which was the timing of the Fall sampling. In mid-September, deciduous trees generally drop their leaves and air and soil temperatures decline precipitously. While we did not track terrestrial vegetation physiological dynamics, we did measure soil temperature, which was significantly lower during the Fall sampling, relative to Summer sampling (see Figure 2). As discussed in section 4.4, we infer that lower temperatures in the Fall placed a strong constraint on soil respiration that overrode other constraints that were likely more important during the Summer season. In support of this inference, the seasonal shift in temperature was the dominant driver of the shift in flux rates between seasons; this is indicated in Table 4. In

addition, during the Fall sampling the soil has not yet begun to freeze, so expansion of the frozen soil is not an explanatory factor in our study.

3. I have little information, if any, on the Autokrig function in 'automap'. I have no idea whether the routine is reliable, neither which choices are made by the authors and which by the software. It is of some concern, as quite some conclusions are drawn from the fitted parameters. I am also somewhat doubtful whether the spherical and the Matèrn models can be compared in a straightforward way. The Matèrn model is a hybrid between the exponential and the Gaussian model, and has one more parameter, but it is unclear to which degree the range parameters are comparable.

Thank you for these suggestions. To improve quantitative comparability we elected to fit Matern models and compare the fitted parameters (see page 5 for a justification of this choice). With regard to the automap package being robust, we added a sentence to the manuscript noting that this package has been validated through heavy use in the peer reviewed literature (see page 5).

4. On page 6 it is stated that the data were 'transformed to improve normality'. That is odd, and maybe not even necessary. Distribution of the data is an inherent property of the underlying variable, and that may reflect itself through the collected sample. Normality is just a specific kind of distribution. This distribution is useful when statistical testing comes into view – which is not the case in this paper. Technically, kriging does not require normality, or even continuity in the response variable. Also, the GLS modeling may not require it. The transformation should be better justified and it should also be specified in more detail which transformation exactly was carried out. Further, a standardization is reported; would the results then in the end be interpretable and understandable? In particular interpretation of the sized of the estimated coefficients in tables 3 and 4 may have a difficult interpretation.

We log-transformed variables to improve normality because statistical testing using generalized least squares (GLS) does assume normally distributed errors. One can use generalized linear models to evaluate non-normal error distributions, but that method assumes independent errors. We do not have independent errors due to spatial autocorrelation, and we selected GLS to account for this non-independence. For our study, it was therefore useful to improve normality of the error distributions.

Our goal in Tables 2-4 was to indicate the relative degree to which each explanatory variable was associated with soil respiration. Differences in regression coefficients derived from explanatory variables that are not standardized are primarily due to among-variable differences in numerical scale and range. Coefficients derived from standardized explanatory variables, however, can be directly compared. The interpretation we aimed for was identifying variables that were most likely important to spatial and temporal variation in soil respiration. As such, we standardized all explanatory variables, which greatly improved interpretation of the relative importance of each variable.

5. On page 6 it is reported that variogram fit to SR are consistent with the CV results. A better explanation is required here.

This statement was removed to streamline the Results section, leaving interpretation to the Discussion.

Details

- Table 1 would benefit from including the number of samples (n) as a separate column

This information is now included.

- The story on the SR variance (page 6, Results, second paragraph) reads somewhat awkwardly. I think that in the end it is correct, but the problem comes when the variances are reported in m (or cm), whereas one would expect then to be expressed in squared units. Possibly some rephrasing would be helpful.

We believe the confusion is due to the writing not being clear that spatial positions are being reported in meters, not the SR variance. We added text to clarify this point. It now reads:

"We plotted SR variance within a given spatial domain against the position of the Western boundary of that sampled domain (see Methods). Doing so revealed a strong threshold at spatial positions near ~40-45m (moving East to West). At positions beyond this threshold SR variance increased rapidly and then stabilized at ~55-60m (Fig. 4a,b). ALD at the Western boundary of the sampled domain also showed threshold behavior, increasing rapidly at spatial positions near ~30-35m and then reached its maximum value (150cm) at about 50m (Fig. 4a,b). These patterns in SR variation and ALD were found in both seasons even though the SR variance was much lower in the Fall (cf. Fig. 4a,b; Table 1). Also in both seasons, SR variance increased rapidly beyond an ALD threshold of ~140cm (Fig. 4c,d)."

################

Comments on Stegen et al. 'Soil CO2 flux across a permafrost transition zone: spatial structure and environmental correlates' General perspectives

--- In general the manuscript is clear and well-written

--- In general it is a good experimental design except lack of soil moisture recording

--- In general authors give proper data analysis

Authors addressed spatial heterogeneity of soil C efflux in Alaska permafrost with six 72 m-long transacts. In general, it is a meaningful study given the unsolved problem on spatial structure of soil C efflux and the research priority of permafrost. My major concern is lack of water condition monitoring both at temporal and spatial scale for this study. As shown in the method section, measurements covers nearly a half month both during summer (7.31~8.13) and fall (9.10~9.24). During this period, the soil water condition might be changed which caused either by rainfall or evaporation, or both. This is especially important when authors suggest 'one potential explanation is that SR associated with thinner ALDs is constrained by relatively high soil moisture – likely due to facilitation of anaerobic conditions.'

We agree it would have been very helpful to have soil moisture data through space and time during our SR sampling. Unfortunately this was not logistically feasible. To do so would have significantly disturbed the ecosystem (e.g., digging a hole to install moisture probes), thereby undermining our repeated measurements of SR. Following the final SR measurements we did collect soil cores, however, and found a significant difference in moisture between the permafrost and permafrost-free spatial domains. As discussed in section 4.3, this difference is expected due to permafrost impeding vertical water movement.

Nonetheless, we recognize that the lack of spatiotemporal moisture data is a limitation of our study. As such, we added a paragraph to the end of section 4.3 that points out this limitation. It reads: "It should be recognized, however, that we were not able to characterize moisture patterns through space and time such that our interpretations of moisture impacts are speculative. In addition, field observations alone are insufficient to resolve underlying mechanisms leading to the observed thresholding behavior, especially because multiple environmental variables co-varied in space. In particular, multiple variables changed across the permafrost-to-permafrost-free transition such that multiple factors likely contributed simultaneously to the observed thresholding behaviour. We strongly encourage future manipulative experiments designed to resolve the governing mechanisms."

On the other, the most significant finding in the study as suggested by authors is the thresholding behavior of the soil C efflux (see figure 4). Nevertheless, authors did not provide strong direct supports to explain this finding instead of providing some possible discussions.

We believe the reviewer is highlighting a fundamental limitation of field observational studies, which can be used to reject hypotheses, but cannot be used to confidently assign causality. This limitation applies broadly to field observations, not just our study. The new paragraph at the end of section 4.3 now indicates the proper level of interpretation.

I suggest author to reduce hypothesis in the introduction section. Some of the hypotheses do not have strong significance and some of the hypothesis authors did not give explicitly testing conclusion. Focus on the threshold finding and give solid evidence to support it.

We went carefully through the stated hypotheses and agree that some were too vague to be useful. We would like to retain a hypothesis-evaluation framework and have edited the hypotheses to be more direct and more clearly tied to the analyses that were performed. The stated hypotheses now read: "…we test the following qualitative hypotheses: (i) Given previous work showing that spatial variability in SR is only weakly related to temperature and soil moisture (e.g., Song et al., 2013;Yim et al., 2003), we hypothesize that SR will be predominantly influenced by carbon inputs such that variability in SR throughout the spatial domain will be best explained by tree-stand variables (e.g., basal area); (ii) Within the drier, permafrost-free domain we hypothesize that SR will again be best explained by tree-stand variables, but within the wetter permafrost-associated domain SR will be decoupled from carbon inputs and will therefore be poorly explained by tree-stand variables; and (iii) Cold temperatures will fundamentally constrain SR such that we hypothesize a smaller coefficient of variation and less well defined spatial gradients (i.e., weaker spatial structure) in SR under colder temperatures."

The first two hypotheses were evaluated with our GLS regression analyses and the third hypothesis was evaluated with our CV and variogram/kriging analyses.

Overall, I would like to recommend it be accepted by BG finally even though some of the points might need some revision.

Thank you.

Specific points

P1L13-15: I suspect the permafrost depth and tree basal area is highly correlated. Authors could try to find which one is the major driver and the other is just a correlation.

These two variables are correlated with each other, though there is significant scatter in the relationship. A linear regression revealed an R.sq of 0.41, indicating that these variables are not redundant from a statistical perspective. Our GLS modeling was designed to identify which variables were the most likely to be major drivers. One element of this was standardizing all the explanatory variables to be z-scores (i.e., mean of 0 and standard deviation of 1). This standardization allows the regression coefficients in Tables 2-4 to be directly compared across explanatory variables. Examining these Tables indicates that across the full spatial domain and within the permafrost domain, basal area is far more likely to be a driver of SR than is ALD. In the full and permafrost domains (within a given season) ALD was not significantly correlated with SR (Tables 2,3). Within the permafrost-free domain, however, the situation was less clear, with ALD and stem density having similar regression coefficients (Tables 2,3). While ALD is not specifically called out, these points are discussed in the last paragraph of page 8, which also connects the GLS analyses to the CV analyses to support the statements in the Abstract referred to by the reviewer.

P1L16-18: be specifying here. Spatial variation and scaling contains a lot of information, please point out in detail in which aspect or aspects Boreal forests is similar to other biomes.

Specificity has been added. The sentence now reads: "Our analyses further show that spatial variation (the coefficient of variation) and mean-variance power-law scaling of soil respiration in our boreal system are consistent with previous work in other ecosystems (e.g., tropical forests) and in population ecology, respectively."

P1L18-19: This has been stressed in L13-15.

We believe the reviewer is indicating there is redundancy within the Abstract and refers to these sentences from lines 13-15: "We also find that within each season tree basal area is a dominant driver of soil respiration regardless of spatial scale…"

And lines 18-19: "Comparing our results to those in other ecosystems suggests that temporally-stable features such as tree stand structure are often primary drivers of spatial variation in soil respiration."

It is true that both sentences call out the influence of trees on SR, but in our opinion they convey different information. In the first sentence we refer only to a result from our field system, which specifically calls out basal area. In the second sentence we place that result in

context of other ecosystems, and point more broadly to the influence of 'temporally stable features.' In this case tree stand structure, which includes more than just basal area, is only one example of a feature that is relatively stable through time.

Given these differences, we would like to retain both sentences as they are.

P1L19-20: If remote sensing implication presented in the abstract, it is better to show it in discussion.

We feel this is an important implication, but one that is also relatively simple and straightforward. As such, we included only a brief statement related to remote sensing in the Discussion. It is at the top of page 9 and reads: "Furthermore, if spatial variation in SR is broadly influenced by tree-stand structure, it suggests an opportunity to use remote sensing techniques to characterize stand structure (van Leeuwen and Nieuwenhuis, 2010) and, in turn, the spatial structure of SR in boreal and other high-latitude ecosystems (Kushida et al., 2004)."

More could certainly be said, but we feel the sentence we provide captures the most important point and it is not clear to us that adding additional text will contribute additional insights.

P2L23-24: the range here might be related to sampling scale and is not comparable here.

We agree and have added a sentence near the bottom of page 2 to reflect this. This set of sentences now reads: "Previous work in forests has revealed that SR spatial autocorrelation occurs across a broad range of length scales (referred to as the 'range' in variogram models), from <1m to >40m (Foti et al., 2014;Russell and Voroney, 1998;Song et al., 2013;Singh et al., 2008;Rayment and Jarvis, 2000), but none of these estimates come from boreal forests. In addition, among-study variation in the length scale of spatial autocorrelation may be partially due to differences in the spatial scale of sampling. It is therefore difficult to generate quantitative a priori expectations for the parameters of SR variograms (functions that describe spatial continuity or variability) in our study system."

P2L30: weak spatial structure or weak heterogeneity?

This has been clarified with the revision of our hypotheses and now reads: "(iii) Cold temperatures will fundamentally constrain SR such that we hypothesize a smaller coefficient of variation and less well defined spatial gradients (i.e., weaker spatial structure) in SR under colder temperatures."

P2L33: I am wondering why not including soil moisture during field work even if authors want to test the idea that soil moisture play slight role in driving spatial pattern of soil C efflux.

The reason is that soil moisture measurements would have been very disruptive given the need to obtain an integrated measure throughout the soil column. We worked very hard to minimize disturbance to the sampled locations to enable sampling across seasons. As discussed above, we do have soil moisture data that show significant differences between the permafrost-associated and permafrost-free spatial domains.

P3L16: How long will it last before measurements but after collar installed. You know, there will be certain kind of disturbance to soil when a soil collar insert to a depth of 5 cm. It might cut some of the surface root and may change soil structure.

Soil collars were installed before the Summer sampling and remained in place for the Fall sampling. The minimum time between installation and measurement was 2 days, this information is now provided at the bottom of page 3. Soil respiration was measured in most soil collars in both seasons, providing an opportunity to evaluate any systematic bias introduced by the shorter time-for-recovery in the Summer season. If the shorter time-for-recovery in the Summer decoupled soil respiration from factors governing its spatial structure, we would expect a poor correlation between Summer and Fall soil respiration measured in the same soil collar. In contrast, we found a strong correlation between the two seasons, which indicates that the shorter time-for-recovery in the Summer did not mask the primary drivers of soil respiration.

P3L32: How air and soil temperature was measured? Which sensor was used? How it collected?

At the top of page 4 we added the following text that describes these measurements: "For temperature measurements we used an analog thermometer at the time of sampling for each soil collar. For air temperature we shaded the thermometer to avoid elevating the temperature due to solar inputs. For soil temperature, we placed the thermometer's stem into the ground to the specified depth."

P4L19: in examine soil C efflux spatial pattern vary across season, authors might be better provide some basic information on forest phenology, i.e. leaf area index or normalized vegetation index from satellite. This is important when authors suggest carbon input by forest have strong impact on soil C efflux. Soil C efflux is the sum of soil heterotrophic respiration and respiration that contributed by plant root. During the summer period, root respiration might be a large proportion in total Soil C efflux.

A very interesting idea, we thank you for the suggestion. We looked into NDVI data that are publically available (MODIS, https://modis-land.gsfc.nasa.gov/vi.html) and found the highest spatial resolution to be 250m. Our field domain was ~75m and spans strong spatial gradients in balance between deciduous and coniferous trees. A time series of remotely sensed NDVI could potentially give a rough idea of forest phenology, but the pattern will be very difficult to interpret given the mismatch in scale and unknown mixture of deciduous/coniferous trees across the whole 250m pixel that would contain our field site.

On the other hand, we observed first-hand that during our Fall SR measurements the deciduous trees were in dropping their leaves, which happens very quickly in our field system. We did not quantify this dynamic, but there is no question that the deciduous trees were shutting down metabolically during our Fall sampling.

We find it very interesting that despite the seasonal shift in tree physiological status, there was consistency in the primary driver of SR. Specifically, we found that tree basal area was consistently the best predictor of SR across the full and permafrost-associated spatial domains. Furthermore, in the permafrost-free spatial domain there were 3 significant

variables in the Summer and all 3 were also significant in the Fall, though the relative magnitudes of their regression coefficients were shifted.

On the other hand, between the two seasons there was a shift in how strongly SR increased moving from the permafrost domain into the permafrost-free domain (see Figure 5). We inferred that lower temperatures in the Fall led to this weaker spatial structure in the Fall. However, the change in spatial structure may also be due to a larger decrease in the contribution of root respiration in the permafrost-free domain, which is dominated by deciduous trees. Even if we had perfect NDVI data we could not parse the relative contributions of thermal constraints from shifts in the contribution of root respiration. To do so would require SR measurements from root-excluding soil collars paired with SR measurements from non-excluding soil collars. That is a very interesting direction to pursue, but was far beyond the scope of what we could accomplish.

Given that lower temperatures and shifts in tree physiology could both have contributed to weaker spatial structure in the Fall, we added modified the text in the 1st paragraph of section 4.4. It now reads: "These patterns indicate that SR has much weaker spatial structure during colder periods and across permafrost-associated spatial domains. There are a number of mechanisms potentially contributing to these observed patterns. For example in the Fall, colder temperatures may place an upper constraint on microbial and root respiration, and senescence of deciduous leaves—which occurred during our Fall sampling—may indicate decreases in root respiration in the deciduous-dominated permafrost-free spatial domain. Both mechanisms could lead to weaker spatial structure and we look forward to future studies that parse their relative contributions, potentially using root-excluding soil collars. The spatial structure of SR in the permafrost-associated domain may be further influenced by high soil moisture placing an upper constraint on microbial respiration in soils with a thin ALD."

P9L17-30: I am not so convinced by the explanation on the threshold finding. Not only because there lack the soil moisture data, but also other variables, such as soil temperature, aboveground vegetation change dramatically in space at the same time. Soil temperature and total basal area have some major influence on causing the pattern (Table 3 and table 4).

This is related to previous comments and as noted above, we editing the end of the associated section (4.3) to indicate the appropriate level of interpretation and caution. For reference, the edited text reads: "It should be recognized, however, that we were not able to characterize moisture patterns through space and time such that our interpretations of moisture impacts are speculative. In addition, field observations alone are insufficient to resolve underlying mechanisms leading to the observed thresholding behavior, especially because multiple environmental variables co-varied in space. In particular, multiple variables changed across the permafrost-to-permafrost-free transition such that multiple factors likely contributed simultaneously to the observed thresholding behaviour. We strongly encourage future manipulative experiments designed to resolve the governing mechanisms."

P10L5-10: The absolute value of soil C efflux is higher in the summer than that in the autumn. It is might be one of the reasons to find higher heterogeneity in summer.

We agree and feel that this is another way of stating our interpretation that lower temperatures in the Fall constrained soil respiration to low rates such that other spatially structured variables had less influence in the Fall.

P17Figure 3: This figure might could be moved to support material.

We prefer to retain this figure in the main manuscript because it offers a bridge to a body of literature—specifically, population ecology—that we feel has insights to offer SR studies (and vice versa).

################

Stegen et al., soil CO2 flux across a permafrost transition zone: spatial structure and environment correlates.

I agree with authors' arguments that soil respiration of permafrost is important in global change studies. Thus, the work has its significance for publication. The topic is covered in the BGs' scope. The most important finding in the study as stressed by authors is the non-linear threshold phenomena. I recommend a major revision because the following points need to be addressed.

1. in the title authors' called 'soil CO2 flux' while in the article they used 'soil respiration'. Please try to make them consistent throughout the whole manuscript.

This has been changed, thank you.

2. whether the PP system is commercial or self-made? If it is commercial, please provide the manufacture information.

It is commercial, and this information is now included on page 3.

3. How long was last after installation of collar before measurements? One week or one month? You know when you insert a collar into ~ 5 cm depth, it will disturb the soil and cut down the plant root which grows in the top soil. Need time to make it stable.

Soil collars were installed before the Summer sampling and remained in place for the Fall sampling. The minimum time between installation and measurement was 2 days; this information is now provided at the bottom of page 3. Soil respiration was measured in most soil collars in both seasons, providing an opportunity to evaluate any systematic bias introduced by the shorter time-for-recovery in the Summer season. If the shorter time-for-recovery in the Summer decoupled soil respiration from factors governing its spatial structure, we would expect a poor correlation between Summer and Fall soil respiration measured in the same soil collar. In contrast, we found a strong correlation between the two seasons, which indicates that the shorter time-for-recovery in the Summer did not mask the primary drivers of soil respiration.

4. The sampling detail need to be described. For example, how long was last for one measurement? How could you make a well-mixing of chamber air? How many data points remained in your special segmental regression?

We have added additional sampling details, as requested, on page 4.

5. for the section 4.3 page 9 line 8~21. Have this threshold phenomenon been revealed in other studies? Is this the first reporting on this phenomenon? Will other soil variables, such as carbon density, soil water content, root biomass and soil microbes, varied in the same way? Authors attribute this to soil moisture, but some more evidences will be needed to support the remark.

A similar comment was made by another reviewer. In summary, our opinion is that an approach based on field observations cannot truly assign causality. We added a paragraph to the end of the associated section (4.3) that indicates the appropriate level of interpretation and caution. This paragraph also highlights the need for controlled experiments in order to reveal governing mechanisms. The new paragraph reads: "It should be recognized, however, that we were not able to characterize moisture patterns through space and time such that our interpretations of moisture impacts are speculative. In addition, field observations alone are insufficient to resolve underlying mechanisms leading to the observed thresholding behavior, especially because multiple environmental variables co-varied in space. In particular, multiple variables changed across the permafrost-to-permafrost-free transition such that multiple factors likely contributed simultaneously to the observed thresholding behaviour. We strongly encourage future manipulative experiments designed to resolve the governing mechanisms."