A review of Soil CO_2 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)."