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.

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.

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.

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.

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

Details

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