Note: in the following document, the original comments made by the reviewer are copied in black, while the authors' responses to these comments follow in blue.

Authors' response to comments submitted by Reviewer #1

Summary: The study "Shifts in permafrost ecosystem structure following a decade-long drainage increase energy transfer to the atmosphere, but reduce thaw depth" by Göckede et al. investigates the impact of a drainage experiment within a wet tundra landscape on components of the surface energy balance. The study seeks to quantify differences in microclimate between a drainage site and a control site which show marked differences in vegetation cover and soil moisture. The authors, therefore, compare sensible and latent heat fluxes which were measured simultaneously at both sites by eddy covariance. In addition, differences in radiation, near surface soil temperatures, and active layer depths are presented. The found differences between the drained site and the control site are discussed in the context of differences in vegetation cover and soil moisture. The study reports increased sensible heat fluxes at the drained site, whereas much small differences are reported for the latent heat flux. Only minor differences are also reported for the net radiation. The authors argue that the increased sensible heat flux is related to a decreased ground heat flux which is reflected in decreased soil temperatures and lower active layer depths. The study focuses on a very important issue in permafrost research that addresses possible hydrological changes in the Arctic due to climate warming and permafrost degradation. This study is, therefore, within the scientific scope of the journal The Cryosphere. The study is written in a clear and understandable manner and is well structured. There are, however, some major issues that should be addressed before publication.
The authors thank the reviewer for her/his overall supportive evaluation of the findings presented in this manuscript. We would like to remark at this point that due to a minor adjustment in the quality assessment protocol for the eddy-covariance datasets, the numbers of the vertical turbulent heat fluxes presented herein have slightly changed since we submitted the first version of the manuscript to The Cryosphere. The overall trends remain the same, i.e. we find a shift of the energy partitioning from latent to sensible heat fluxes, leading to a significant increase in Bowen Ratios following drainage. With the new numbers, however, the absolute increase in sensible heat H has been reduced, while the latent heat LE is actually slightly reduced, leading to a budget H+LE that is about the same at drainage and control sites, respectively.

General comments: The measured differences in the turbulent heat fluxes and the radiation are essential to the argumentation of this study since the ground heat flux is not determined independently. Thus, it would be highly recommended to perform a comprehensive uncertainty analysis and quality assessment of the fluxes. Uncertainty ranges should be calculated in order to evaluate whether the measured differences are significant.
In the revised version of the manuscript, we will add a comprehensive uncertainty assessment of our flux datasets, as requested by the reviewer.
It must be excluded that the differences in the heat fluxes are subject to instrument biases and/or site related methodological biases.
We used exactly the same instrumentation at both sites. Moreover, we also applied the same data processing protocol for datasets from both towers, and the quality assessment protocol includes a direct comparison of data elements from both sites. Therefore, we believe that systematic methodological bias can be excluded here. We will extend the documentation of these checks in the revised manuscript version (see below).
Due to the importance of reliable measurements, it seems to be inadequate to just refer to another study for quality assessment.
We will add a new appendix in the revised version of the manuscript that will present the core elements of our eddy-covariance data processing protocol, which will include the details of

the quality assessment protocol applied for the flux and meteorological data used within the context of this study.

The authors state that they do not see any differences in the long wave radiation budget. This is surprising since increased sensible heat fluxes are often related to increased surface temperatures. I think a discussion of this point would be highly interesting for understanding the reasons for the increased sensible heat fluxes.
We decided to leave out details on this part of the study because we believe the longwave radiation results do not provide enough additional insight to justify the required extra text passages. 'Virtually unchanged' refers to a shift in longwave radiation budgets of -0.2% (summers 2014/15) between drainage and control sites. However, there is considerable interannual variability, with higher deficits (LW_{down}-LW_{up}) in 2014 at both sites. Since this interannual variability turned out to be more pronounced at the control site, compared to the drainage site, we measured offsets of opposite signs in both data years, which largely cancelled each other. This pattern in interannual variability agrees well with the patterns in both sensible and latent heat fluxes observed at the two study sites. We will include links to longwave radiation patterns for the revised discussion of sensible and latent heat fluxes in the new paper version

The authors present differences in friction velocity as indicator for differences in vegetation structure (height/density). In this context, I think it would be more instructive to calculate roughness lengths during neutral conditions. In addition, a footprint analysis might be useful since vegetation seems to vary within the footprint area of the eddy covariance tower (Figure 1). It might be interesting for the process understanding to further investigate the impact of changed surface roughness on flux partition.
The author preferred to make this point based on the friction velocity since it provides a broader and thus more representative data basis (no filtering for stability of stratification). However, the same can of course be done based on the roughness length, derived from wind profile relationships during neutral stratification, as suggested by the reviewer. Also in this case, we find a systematic enhancement of roughness lengths (mean increase: 0.035m), with aerodynamically rougher surfaces in all wind sectors for the drainage site, compared to the control. Since this is rather a side aspect of the analysis presented within the context of this manuscript, we do not think a proper footprint analysis will be required to make this point, also because we lack spatially distributed information on vegetation height (a highest resolution (~2m) WorldView land cover map would be available). We will include information on differences in roughness length between both sites broken up by wind sectors to include a spatial context of the differences observed.

In general, the data analysis presented in this study is very limited and should be expanded. Besides footprint analysis, the analysis of the diurnal and seasonal signals can provide deeper insights into the processes behind flux partition. Figure 9b, for example, clearly demonstrates that more than just average differences can be observed in the time series of sensible heat fluxes. I think that the very strong and general statements made in the conclusions and the abstract are not possible based on the very limited data analysis presented here and, thus, should be toned down. Without further process analysis leading to a better understanding of heat transfer processes within and below shrubs, the presented data and results must be considered site specific and are not directly transferable to other tundra landscapes.
As mentioned in the comment above, the authors believe that a footprint analysis in this context will not provide further insights that help interpreting the differences in energy flux signals. However, a directional effect will be added. The general seasonal trends in fluxes and their differences between the two sites have been covered in Section 3.6 already, but we will refine this description in the revised version of the manuscript. This description could in theory be extended by also comparing typical diurnal courses between the treatments, but since the meteorological forcing is virtually the same at both of them, and more influential factors such as e.g. water levels, vegetation structure and soil thermal regimes vary at timescales much longer than diurnal, the authors believe that analyses on such short timescales will not

strengthen the study. We are certainly aware that we can only provide results for a single site, and the derivation of process understanding that can be transferred also to other Arctic sites will be associated with uncertainties. As suggested by the reviewer, we will therefore tone down all statements that may hint at a general applicability of our findings.

The authors do not present any data or results related to snow. However, several paragraphs in the discussion include speculations on snow including the impact of vegetation on snow, and the impact of ground heat fluxes on snow melt. As these subjects are essentially not part of the presented investigations, I strongly recommend to exclude them from the discussion and to focus on the presented results.
We discuss differences in the snow cover periods between the two study sites in Section 3.4, which are reflected in the radiation data shown in Figure 6, as well as in the numbers on albedo and SWnet summarized in Table 2. These results emphasize that snow cover dynamics are influenced by drainage disturbance, and that shifts in snow cover may cause systematic secondary effects on both the radiative budget and the soil thermal regime. We do not measure any snow properties directly at both sites in parallel, snow depth is only monitored at a single location currently. We therefore agree with the reviewer that we do not present any direct observational evidence on snow characteristics herein. Still, we have the indirect evidence on systematic shifts in the snow cover period obtained through the albedo record. The passages in the discussion that the reviewer refers to aim at pointing out mechanisms, mostly links between vegetation and snow cover, that can explain these patterns in albedo. Therefore, we believe that these discussion sections are required to interpret the complex interactions between secondary disturbance effects.

Specific comments:
p.8,l.1: Sections 0 and 0?
This was a typo based on a broken automated cross-reference, and has been updated.

p.8, l.2: Why not including the LW budget in the analysis, in particular the outgoing LW radiation?
As mentioned in more detail above, there is virtually no difference to be seen in the net LW radiation budget, so the authors believe that including more details here would provide only very little information.

p.8, l.6: What is the absolute accuracy of the used radiation sensor? What is the footprint of the radiation sensor? Does the sensor measure a representative area of the drained site?
We use Kipp & Zonen CNR4 as radiation sensors at both observation sites. These instruments are officially classified as 'first class' instruments (for shortwave radiation e.g. a resolution of +/-5 Wm^{-2}, and a stability of +/-2%). However, based on the direct comparison of data from both instruments, we find cumulative differences below 1% of the total incoming radiation, therefore our sensors would even qualify for the next highest quality rating (secondary standard, e.g. a resolution of +/-1 Wm^{-2}, and a stability of +/-1%).
The CNR4 instruments have downward looking opening angles of 150 degrees. For sensor heights of 4.5m (drainage tower) and 4.66m (control tower), this translates into circular footprints with a radius of ~16.8m and 17.4m, or footprint areas of 886m^2 and 950m^2. We believe that the area covered for the drained site is indeed representative for the drainage area overall, but of course it is obvious that we can only cover a small fraction of the total area.

p.9, l.1: Here and throughout the manuscript it should be thermal conductivity.
This has been corrected throughout the manuscript.

p.9,l.3-5: This is a very general statement which does not provide any quantitative information on the heat transfer processes within the soil. Please note that increased soil moisture means a higher content of latent heat which besides thermal conductivity and heat capacity determines the duration of the zero curtain.

We obviously chose a misleading term here, since with ‘heat capacity linked to higher water content’ we actually wanted to refer to the latent heat. This will be corrected in the revised version of the manuscript.

p.9,l.25-27: Besides the ground heat flux, the thaw depth is determined by the soil ice content. Are there any information on differences in soil composition between both sites?
We took soil profiles across the treatment areas during the growing season, and analyzed for e.g. carbon content and nutrients. Regarding ice content, we also have a limited number of cores taken in November, i.e. with frozen ground, where relative ice content was assessed for wet and dry microsites in both drainage and control areas. As a general trend, we found higher ice content in the wet microsites within the top ~30cm of the soil profile, while relative ice content was higher in the dry microsites between ~30-50cm below the surface. Further down, no systematic differences were observed.

p.10,l.6: This is a very small number (1,8%), is this within the accuracy of the sensor?
Please refer to the comments on p.8, l.6 above. The relative deviations in incoming radiation between both sensors were found to be <1% of the total signal. Therefore, the deviations between net energy input described in this section are higher then the accuracy level of the employed sensors.

p.10,l.13: Is it possible to distinguish between the impact of drainage and changed vegetation cover within this study? Or is it more likely that the vegetation cover has adapted to drainage which then has modified the surface energy balance? I think it would be important to separate these things since there drainage can happen on relatively short time scales while changes in vegetation cover requires some time.
Based on the data that we have available at this time, we cannot cleanly separate between a direct impact of water level changes (drainage) and the associated effect of changes in vegetation structure as a secondary disturbance. To be able to do so, we would need more data years with differences in water levels, so that we could do proper statistics on the impact of these year-to-year changes, while vegetation stays more or less constant. The importance of covering the long-term effect of vegetation changes has been included in the discussion section.

p.11,l.13: Does ’slightly lower’ mean a more negative sensible heat flux at the drained site?
As mentioned in the text, the mean sensible heat flux for both towers during wintertime is -3.4 Wm^{-2}. Here, the drainage tower has a mean value of -4.3 Wm^{-2}, and the control tower a mean of -2.5 Wm^{-2}. Accordingly, ‘slightly lower’ referred to a more negative sensible heat flux at the drainage site. This will be clarified in the revised version of the text. Due to the revision of quality flagging as mentioned in the first comment above, these values have slightly changed (drainage: -5.6 Wm^{-2}; control: -4.4 Wm^{-2}; difference: -1.2 Wm^{-2}), but the overall trend remains the same.

p.12,l.6-15: See general comment on snow cover discussion.
As mentioned above, we believe that this interpretation, fortified by the references given in this paragraph, is necessary to provide an interpretation why we see a consistently earlier buildup of a snow cover at the drained site, compared to the control (Figure 6, Table 2).

p.12,l.23-30: Why so much emphasis on discussing the impact of mosses if mosses are absent at the study site?
We decided to include this paragraph to point out a feature of the site characteristics that is different from many other places in the Arctic. As mentioned in the text, we believe that a higher abundance of mosses would significantly alter the findings we observed at our sites. Accordingly, our results may not be representative for sites at other locations that feature a higher moss coverage.