General Comments:

This paper synthesizes available eddy covariance data in the Arctic and attempts to identify areas which are well and not well represented by the current distribution of eddy covariance stations. The topic is interesting and important, the paper is generally well written, and the analysis appears sound though the spatial representativeness assessment lies well outside my own competence. A major source of impact and novelty in the present paper is the method used to identify specific new locations/upgrades to existing locations which would result in the greatest relative increase in biome representativeness. This advance is particularly valuable since it could greatly improve the effectiveness of strategic research planning for future flux sampling efforts. I have only fairly minor suggestions for improvements (see below).

Specific Comment:

All figures: It is difficult to distinguish the "no data" and poorly represented areas with the current color scheme. Also, it may be a good idea to explicitly exclude areas with permanent ice (ie: much of Greenland), unless you think these areas should be represented with EC data?

We will change the color scheme for the no data and ice sheet regions. The final figures will be provided in higher resolution, which will also aid in distinguishing between the different cases.

Lines 136-139) I miss a bit more detailed discussion and justification in the intro, results and discussion of the broader implications/considerations of the 18 variables chosen to represent environmental variability. How/why were these particular variables chosen? Are there any other significant variables that could have been interesting to include? How might the variables selected in turn impact your estimates of spatial variability in representativeness? For example, are there some important variables which remain poorly represented even in Alaska/Fennoscandia? How were the variables combined together to create a single metric of environmental variability? Many of the 18 variables seem likely they would be strongly autocorrelated with each other, so any procedure that treats all individual variables as "equal" in weight to each other may be flawed…

The combination of these variables into a single metric is discussed in the first paragraph of methods section 2.2. The representativeness value is the Euclidian distance in n-dimensional space between each location in our domain and the closest EC site. Here the n-dimensional space consists of the normalized bioclimatic variables.

This issue of the 18 variables was also raised by reviewer 2, and in the response to his comments we discuss it in detail in our answer A1. We copied this response below.

We agree that indeed we are estimating the variability of these 18 variables and not the GHG fluxes themselves. And while we do not discuss this limitation in 4.3, we did in 4.2:
"*However, the assessment of specific fluxes provided by the eddy covariance tower network based on these data layers must largely remain qualitative, since no clear quantitative linkage between the bioclimatic controls and the fluxes for CO_2 and CH_4 can be considered. In other words, assigning equal weights to all 18 data layers allows us to assess a general level of*

similarity between pixels within the domain, but does not necessarily reflect how strongly potential differences will influence greenhouse gas flux rates (see e.g. Tramontana et al.(2020))"

The application of this type of method of using ecological land cover classifications for EC network analysis has been successfully demonstrated in earlier references (e.g. Hargrove and Hoffman, 2004; Hoffman et al., 2013). However, we agree with the reviewer that this core assumption to our study needs to be presented more prominently in the text. Therefore, we plan to change the second last paragraph of the introduction to:

"Building on a study by Hoffman et al. (2013) that presented an analysis of the Alaskan EC network, in this study we will provide a first in-depth evaluation of the current and past pan-Arctic EC flux observation infrastructure. Our method uses quantitative multivariate clustering which has many uses from creating maps of geological regions (Harff and Davis, 1990), to watershed delineation (Hessburg et al., 2000) and ecoregion classification (Zhou et al., 2003). Hargrove and Hoffman (2004) give an extensive overview of these applications, which are all based on the concept of mapping normalized ecosystem variables such as topography, precipitation and temperature in an n-dimensional data space, using one axis for each variable. The closer two points are in this variable space, the more alike they are, and the more likely they are to be classified as belonging to the same ecoregion when clustered by a k-means algorithm. Thus, the distance can be interpreted as a metric of variability. Aiming at assessing the representativeness of the EC network in the US, Hargrove and Hoffman (2004) then calculated the distances between each constructed ecoregion without an EC site to the closest ecoregion with an EC site. Hoffman (2013) later extended this method to map the Alaska EC network, this time, instead of aggregating the distances between ecoregions, calculating the distance between each pixel in the map and the closest EC site. This approach thus preserves the fine scale variability that is lost when aggregating to the ecoregion level.. In our implementation we will also perform this analysis on an individual pixel scale."

Hill et al. (2017)explain the statistical requirements for upscaling fluxes from site levels to the ecosystem level. No actual fluxes are required with their typical example, just the ecoregion/ecosystems and the number of towers present. We will add according information to the last paragraph (more details in the next answer A2):

"Our analysis aims to quantify representativeness in the pan-Arctic domain based on this similarity in key ecosystem characteristics of any location in our domain to those of the EC sites. We further use the analysis by Hill et al. (2017) on the statistical power of EC systems to put these representatives measures into perspective regarding the general potential to upscale fluxes from sparse EC networks. Moreover, we use the results from the representativeness analyses to identify the most suitable locations for new observation sites, and upgrades to existing infrastructure, that would optimally enhance the performance of the Arctic EC network as a whole. Finally, this manuscript and its corresponding online tool aim at providing an easily accessible source of information on Arctic flux monitoring infrastructure and literature for scientists working on the carbon cycle."

Regarding the selection of variables, and to which degree they are relevant for EC flux network, we will add the following explanation and addition to the paper as part of the new third paragraph of the introduction:

"Knowing the current and past spatiotemporal distribution of EC sites is not enough to fully understand to which degree this network represents the Arctic domain. The reason for this is that EC towers have a field of view that typically does not extend further than a kilometer from the tower, often less (Leclerc and Thurtell, 1990; Horst and Weil, 1992; Schmid, 1994, 2002; Vesala et al., 2008). Accordingly, with currently about 120 terrestrial EC towers situated within the Arctic domain, only a very small fraction of the region gets directly observed, while most of its expanse remains unsampled. Larger footprints would not solve this problem, as the greater heterogeneity would still be hard to capture. Meteorology, vegetation, above/below ground conditions, and topography are critical drivers of hydrological and biogeochemical processes at landscape scale and of GHG fluxes, and their variability across the Arctic therefore also causes variability in flux rates. For upscaling purposes (i.e., when fluxes are predicted over larger areas), typically a tower is held as representative for the ecosystem and the region where it is stationed (Desai, 2010; Jung et al., 2011; Xiao et al., 2012; Chu et al., 2021); however, except when using a very coarse classification of ecosystem types, the existing EC network still cannot cover all ecosystems across the Arctic, and a courser classification would increase heterogeneity within the ecosystem and reduce the representation within the ecosystem. Still, a number of published studies have successfully demonstrated the effectiveness of using meteorological and environmental variables as explanatory variables for estimating GHG fluxes at regional to global scales (e.g. Jung et al., 2020; Knox et al., 2019)."

To provide further insights into the rationale behind choosing the 18 bio-climatic variables used in our study, we plan to add the following text to Appendix 1:

"The set of 18 variables used in our study was carefully selected to capture the broad environmental conditions that are the important drivers of hydro-biogeochemical processes, and GHG fluxes, in the Arctic ecosystem. The selected variables cover meteorological and bioclimatic conditions, soil properties, topographic, and permafrost conditions. Meteorological and bioclimatic conditions are primary drivers of vegetation, biological and ecological processes, and a similar selection of variables as chosen herein have been demonstrated to perform well for upscaling purposes in past published studies (Schimel et al., 2007; Jung et al., 2009, 2011; Dengel et al., 2013; Knox et al., 2016, 2019; Jung et al., 2020; Malone et al., 2021). Complex microtopography is known to be an important driver of microclimate and vegetation in many parts of the Arctic, and is represented here by the compound topographic index (CTI), a parameter designed to capture the impact of topography on hydrological processes. In addition to surface processes and vegetation, subsurface biogeochemical processes play an important role in high latitude Arctic ecosystems. Soil properties used for this study, including e.g. bulk density, or carbon and nitrogen contents, were selected to capture the heterogeneous subsurface conditions. Ecosystems in the vast Arctic region span continuous, discontinuous and sporadic permafrost conditions, and varying seasonal permafrost thaw conditions that regulate the GHG fluxes. Three permafrost related variables were therefore selected to reflect these heterogeneous conditions. In conclusion, even though not all of the 18 variables selected for our study are directly connected to variability in GHG fluxes, their combination is, to our knowledge, the best

representation to capture the variability in environmental drivers that influence biogeochemical processes, and thus also the GHG fluxes across the Arctic."

A geostatistical method, while certainly a viable approach, was not performed since it requires a large dataset of EC fluxes on top of the raster datasets of bio-climatic variables. Collecting the actual flux data for the 120+ EC sites analyzed within our study was beyond the scope of this project, our concept builds on their metadata alone. We fully agree with the reviewer that the geostatistical approach would provide more certainty in its selection (or weighing) of variables; however, we do not see a practical way of building the required EC database with reasonable expenses. We therefore decided to apply the distance-based approach promoted by Hoffman et al. (2013), and accept (and discuss) the extrapolation uncertainty that comes with weighing all selected bio-climatic parameters equally.

Lines 509-501) Its probably true that more EC sites within an ecosystem would create greater certainty for that ecosystem but the key question is would this result in a greater % improvement in overall biome representativeness compared to installing an EC site following your optimized site selection protocol? I assume not, but your approach should be able to resolve this.

The answer to that question depends on the scale of analysis. When looking at variability within pixels, we clearly cannot differentiate with this method; however, having multiple towers located in pixels with very similar characteristics can affect the number of sites outside that pixel that fall within the ER5 metric. On an ecoregion scale, as described in lines 419-425 and 544-549 adding an additional site to ecoregions that already contain a site in certain cases can indeed be slightly more beneficial for the network-wide representativeness than adding a site to a new region. The decision where to place new sites also depends a bit on the objective, i.e. whether the representativeness, ER1 or ER5 metric is supposed to be the target. We see in our data that the representativeness and fraction of area in ER1 and ER5 are all highly correlated. This means that if we improve the overall representativeness as is the target in the current optimization setup, we will also increase the ER5 value which is a measure of high confidence in similar ecoregions. For future research it would be interesting to also optimize directly for the ER5 (or similar) metric and compare this to the optimization of the representativeness as a whole.