Anonymous Referee #1, 19 Mar 2025

The authors would like to thank Reviewer 1 for the feedback on our manuscript and for providing valuable comments and suggestions that helped improve the quality of the manuscript. In this manuscript, we used long-term (14 years) eddy covariance CO_2 fluxes measured at 3 m over a steppe environment, which we consider as the "small scale" dataset for the studied ecosystem. In addition, we used 10 months of CO_2 fluxes measured at a height of 19 m, which served as more spatially integrated data for the steppe and is thus considered to be the "large-scale" dataset. Finally, the overlapping CO_2 flux data over a period 21 days was used for the direct comparison between small- and large-scale fluxes. The following responses address the comments and the issues raised by the Anonymous Referee 1 point-by-point. We have responded (in **black**) to each comment (in **blue**).

First and foremost, the period for which concurrent 2 and 19 m data are available is confined to merely 3 weeks. Given day-to-day variability in weather conditions and the short-term random uncertainty of eddy covariance flux measurements, this is way too short to achieve anything meaningful. The authors attempt to navigate around this key problem by analyzing the 2m fluxes which are available for multiple years (and published in a previous study) in comparison to 10 months of flux measurements from 19 m, but this approach is flawed as it ignores the influence of differences in biotic and abiotic conditions (at least in the way the authors attempt this analysis).

We appreciate the reviewer's observation regarding the limited duration of overlapping flux measurements at 3 m and 19 m. It is true that the simultaneous measurements span approximately three weeks, but it is also an issue we cannot change anymore given the overlapping realities of project lifetime as well as climatic, logistical, and political challenges when working on the Tibetan Plateau. However, while we acknowledge that a longer period of concurrent measurements would have strengthened our ability to draw more robust conclusions about the spatial representativeness of the 3 m flux data, we believe the data from the overlapping period still allow us to extract meaningful insights into the effects of footprint-induced heterogeneity on the measured fluxes. Within these limitations, we carefully examined diurnal patterns and footprint characteristics to identify plausible drivers and processes of the observed differences between the 3 m and 19 m NEE.

The primary intention of comparing 3 m fluxes available for multiple years with 19 m data was to assess if the 19 m fluxes fall within the multi-year envelope of the 3 m fluxes. The analysis helped to understand whether the observed differences persist across months and seasons, and under what environmental conditions they diverge. We believe this approach provides a useful framework for understanding the spatial and temporal variability in CO_2 fluxes, even within the constraints of the dataset.

To address the lack of 3 m data in the period 2018-2019, an XGBoost machine learning model was trained and tested using long-term NEE data collected at the 3 m measurement height. This approach aimed to simulate what the 3m tower would have measured during the time when the 19 m flux data were available. The observed NEE data at 3m was divided into two subsets: a training set (2006–2012), comprising 79% of the total data, and a testing set (2013–2018), comprising the remaining 21%. This division was made with careful consideration of data quality, variable distribution, and prevailing biotic and abiotic conditions. The final model for NEE at 3 m measurement height was built as a function of air temperature, soil temperature, soil moisture, vapour pressure deficit (VPD), relative humidity (RH), radiation, presence or absence of snow, and normalized difference vegetation index (NDVI).

Several of these key meteorological variables particularly soil temperature, soil moisture, radiation, RH, and VPD required for the model training contained gaps during winter months. To ensure continuity in the input dataset, these gaps were filled using downscaled ERA5 reanalysis data tailored to the site level. Downscaling was performed by developing models that relate ERA5 variables to observed site-level measurements using overlapping periods of valid data. For most variables, linear regression models were used. However, for variables where linear regression did not capture the relationships well, we applied random forest (RF) machine learning models to better handle nonlinearity and interactions. The downscaled estimates showed a high degree of agreement with observed site-level measurements, with R^2 values > 0.93. The RMSEs of the downscaled soil temperature, soil moisture, radiation, RH, and VPD were 0.58 °C, 0.002 %, 116 Wm^{-2}, 3.41 %, and 0.65 hPa, respectively. The presence or absence of snow was determined by applying a threshold of 0.4 (Muhammad and Thapa, 2020; Raghubanshi et al., 2023) on the daily snow cover product of MODIS (MOD10A1). The Values above 0.4 were classified as snow-covered, while values below this threshold were considered snow-free. The daily NDVI values were calculated from the MODIS surface reflectance product (MOD09GA). The daily NDVI was obtained by averaging the NDVI over the area corresponding to the 3 m flux footprint. To align with the half-hourly flux data, the daily NDVI and normalized difference snow index (NDSI) values were interpolated to a half-hourly timescale by assigning the same daily value to all half-hourly intervals within that day.

The XGBoost model demonstrated satisfactory performance, achieving RMSE, MAE, and R^2 of 0.86 gC m^{-2} s^{-1}, 0.56 gC m^{-2} s^{-1}, and 0.75, respectively (Figure S1). This corresponds to a normalized RMSE of approximately 7.4% and a normalized MAE of 4.8% relative to the observed NEE range. The trained model was then used to estimate NEE at 3 m for the period 2018–2019, which enabled a comparison with concurrent NEE measurements taken at 19 m (Figure 1). While we recognize that the model-based approach cannot fully eliminate uncertainty, it provides a useful estimate of the expected NEE behaviour at 3 m and accounts for the varying environmental and meteorological conditions. The comparison with the measured 19 m fluxes offers valuable insights into flux differences driven by spatial heterogeneity.

We clarified this rationale and the associated results in the revised manuscript, along with a discussion of the associated limitations and assumptions underlying this approach.

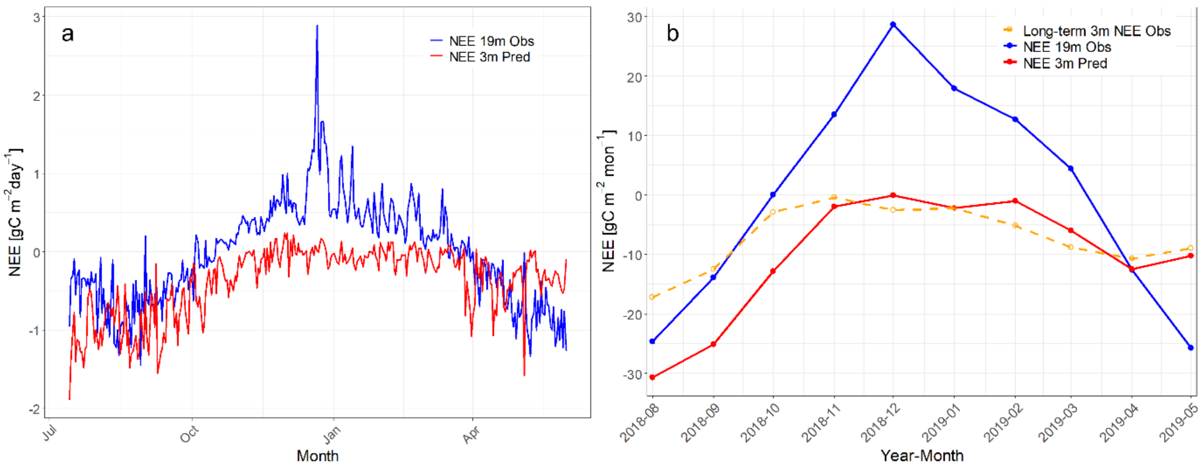

Figure 1: Predicted 3 m daily (a) and monthly (b) NEE plotted along 19 m NEE

Second, while the authors use some footprint model, the way the resulting data are used is highly non-transparent and confusing. It is unclear (in text and figures) when the authors address the 19 m fluxes as measured, i.e. inclusive of the lake contribution, or those restricted to times when almost all flux originates from land only. In addition, the authors appear to use the results of the footprint analysis in a quite simplistic fashion instead of as a powerful tool for exploring the spatial variability of the source area and the resulting effects on the measured fluxes.

Thank you for pointing this out. We acknowledge that the use and presentation of the footprint analysis in the current manuscript and its integration into our results and discussion lacked clarity.

To improve transparency and clarity, we revised the text and figures to more explicitly show how the footprint model was used to distinguish the contributions from land-only and land-lake conditions. In our analysis, we used a footprint-weighted land cover classification to identify the dominant source area of the fluxes measured at 19 m. For each half-hourly timestamp, we calculated the fractional contribution of land (alpine steppe) and lake surfaces based on the footprint model results. We then categorized the data into two subsets:

All 19 m fluxes (NEE_LL) – including measurements influenced by both land and lake surfaces.

Land-only 19 m fluxes (NEE_LO) – restricted to timestamps where the footprint model indicated >99% contribution from land (alpine steppe).

Both datasets underwent the same quality control and gap-filling procedure (Marginal Distribution Sampling) algorithm. For the land-only dataset, timestamps with any significant lake contribution were excluded prior to gap-filling. This ensured that the resulting gap-filled time series reflected fluxes from the alpine steppe only. We then compared the 3 m fluxes (representing fluxes from a smaller, more localized footprint over the alpine steppe) with the NEE_LL and NEE_LO. Our hypothesis was that the 3m and NEE_LO should be more similar than the NEE_LL. These distinctions are now explicitly stated in the revised manuscript to make the data processing steps and comparisons clear.

Regarding the reviewer's concern about the limited use of footprint data, we would like to clarify that the current manuscript includes several analyses based on footprint information (line number 410 – 445).

- We combine **footprint modelling with land use/land cover information** (LULC) to separate 19 m fluxes into land-only and land–lake categories. This allowed us to evaluate the influence of surface heterogeneity on NEE over 3 m and 19 m footprints.
- **Footprint-weighted NDVI** was used to characterize the vegetation activity within the respective source areas of the 3 m and 19 m land only fluxes, providing insights into the role of vegetation dynamics and their impact on flux differences between the two spatial scales.
- In the revised manuscript, we have expanded our analysis by incorporating **footprint-weighted land surface temperature (LST),** footprint**-weighted Normalized difference soil index (NDSI), footprint-weighted Normalized difference moisture index (NDMI),** and the contribution of individual land cover classes within the flux footprint. These analyses address the difference observed in the NEE between the two footprints. These additions help to identify the surface drivers and physiological processes influencing fluxes over the alpine steppe across spatial scales.

To investigate the spatial variation in NEE observed at 3 m and 19 m in May, we combined footprint modeling with remote sensing-derived indices. Specifically, we used footprint-weighted values of NDVI, NDMI, NDSI, LST, and a high-resolution land cover map (Figure S2) to characterize the biophysical properties within each footprint. Lake areas were masked in the vegetation and soil indices (NDVI, NDMI, NDSI) as well as LST to ensure values represented terrestrial surfaces only.

The 3 m footprint was associated with slightly higher NDVI (0.11 vs. 0.10), LST (32.5 °C vs. 28.6 °C), and NDSI (0.33 vs. 0.30) compared to the 19 m footprint, suggesting drier soils with greater microbial activity and less active vegetation. Conversely, the NDMI was lower at 3 m (-0.10 vs. -0.08), further indicating reduced moisture availability. These results imply that the 3 m tower footprint represented a more open, sparsely vegetated area with higher ecosystem respiration, while the 19 m footprint encompassed more productive vegetation patches. Furthermore, the correlation of NEE with NDVI supports this interpretation, capturing the influence of higher productivity in the 19 m footprint (Figure S3).

Light-response curve analysis over the land area showed slightly higher Amax at 3 m (8.57 μmol CO_2 m^{-2} s^{-1}) compared to 19 m (8.01 μmol CO_2 m^{-2} s^{-1}), while alpha remained similar, indicating comparable light-use efficiency. This suggests that even within land-dominated footprints, photosynthetic parameters derived from flux measurements can vary with sensor height. Although the lake contributed variably (0–60%) to the 19 m footprint at the half-hour scale, its average contribution during the overlapping period was ~11%. Neither NDVI nor lake contribution alone explained the more negative daytime NEE at 19 m; rather, we hypothesize that differences in autotrophic and heterotrophic respiration rates (Reco) across footprints played a more dominant role. This interpretation is supported by the footprint-weighted NDSI values, indicating more stressed vegetation and potential for higher respiration at 3 m. While LST was also higher at 3 m, its negative correlation with NEE during the early growing season suggests that GPP increased more rapidly than Reco with rising LST, particularly during the day. This interpretation is further corroborated by the partitioned NEE data (Figure S4).

Based on our finding, we have added a paragraph in the discussion:

"Determining the optimal spatial scale for estimating carbon sequestration depends on the intended application. The 3 m footprint provides high-resolution information on localized fluxes and is particularly sensitive to soil respiration and small-scale vegetation heterogeneity. This makes it useful for understanding fine-scale ecosystem processes. In contrast, the 19 m footprint integrates a broader range of land cover types and better represents landscape-scale patterns, making it more appropriate for comparisons with satellite observations and for upscaling carbon flux estimates. In our study, NEE at 3 m was more strongly influenced by Reco, while at 19 m it was more associated with GPP (Figure S5), suggesting scale-dependent process dominance. Although both scales provide valuable insights, the 19 m footprint may be more appropriate for estimating carbon sequestration at ecosystem or regional scales, particularly when aiming to integrate flux data with remote sensing products. However, further validation with coarser-resolution satellite-based estimates is needed to confirm this upscaling potential."

However, we recognize that the rationale and workflow behind these analyses may not have been clearly presented in the initial submission, which may have contributed to the perception of under-utilization. The revised manuscript now provides a clearer and more detailed explanation, which we believe better demonstrates the strength of our spatially explicit approach.

In the discussion the authors state that they did not do so because of "the inherent variability and potential noise of the flux measurements". This is quite odd as it would suggest doing so is not possible, which would invalidate the entire point of this study. As a consequence, we do not learn much apart from that the fluxes at 2 and 19 m are different at various temporal scales. Here I should add that I think there is tremendous potential if such an analysis would be done in a proper fashion that exploits the information content that the combination of footprint modeling and remote sensing data analysis offers.

We appreciate the reviewer's thoughtful comments and would like to clarify a potential misunderstanding regarding the statement in the discussion referring to "the inherent variability and potential noise of the flux measurements''. This statement was not intended to suggest that decomposition of fluxes by source area is impossible or meaningless. Rather, it referred specifically to our decision not to assign precise numerical flux values to each contributing surface.

However, contrary to any impression of underutilization, footprint and remote sensing data formed a core part of our spatial analysis. In the revised manuscript, we have clarified and expanded our use of these data to better highlight their role in explaining observed differences between 3 m and 19 m fluxes. Specifically, we used footprint-weighted land cover fractions to categorize fluxes into land-only and land–lake cases; applied footprint-weighted NDVI, NDMI, NDSI, and LST to characterize vegetation and surface conditions; and assessed how these spatial drivers influenced NEE at different measurement heights. These analyses revealed important differences in vegetation cover, soil moisture, and surface temperature between the footprints, which we interpret as key contributors to the observed flux variations. Also, this approach allowed us to assess spatial patterns without introducing the additional uncertainty associated with model-based partitioning of flux values.

We recognize that our original wording may have led to confusion, and we have revised this section of the manuscript to more accurately reflect the scope and limitations of our footprint-based analysis.

Third, many of the conclusions are not well supported by the results, for example the reasoning about the influence of snow cover is fully anecdotal.

Thank you for pointing this out . In the original manuscript, our conclusion referenced the extreme 2018–2019 snow event but did not clearly tie this to the supporting analyses presented in the results section. As noted, we had reported statistically significant snow depth differences ($p < 0.001$), satellite-derived snow cover maps, and the results of the Random Forest model, which identified snow depth (29.1% increase in MSE) and snow density (31.4%) as the most important drivers of winter Reco. In the revised manuscript, we have now clarified this connection in the conclusion and emphasized that our interpretation of the snow cover impact is supported by these analyses. Additionally, we provided a Random-Forest variable-importance figure (Figure S6) demonstrating the dominant role of snow parameters in driving winter Reco. The revised version of the conclusion is given below:

"*This study contributes to the growing body of literature on carbon fluxes by using a dual-tower approach to investigate the dynamics of carbon release and sequestration in an alpine steppe–lake landscape. Notably, the sensor at 19 m height recorded higher absolute net ecosystem exchange (NEE) values compared to 3 m long term average monthly NEE budget, with particularly pronounced differences during the winter of 2018–2019. This period was marked by unusually deep and persistent snow cover, as confirmed by snow depth records (average 26 cm, significantly higher than other years; $p < 0.001$) and satellite imagery. Random Forest analysis further identified snow density and snow depth as key drivers of winter Reco, with relative importance (%IncMSE) values of 31.4% and 29.1%, respectively. These findings underscore the substantial role of snow cover in regulating winter carbon dynamics*."

Fourth and finally, the manuscript is poorly structured with entire sections (e.g. 3.6) not being supported by any display items (table or figure), which makes it near impossible to follow the text.

Thank you for pointing this out. We acknowledge that the material related to section 3.6 was originally included as part of (Figure 1 (E)), rather than as a separate display item, which may have made it harder to follow. In the revised manuscript, we have removed Section 3.6. While the wind sector analysis offered some qualitative insight into spatial heterogeneity, we concluded that it did not substantially strengthen the core results and conclusions. The primary findings regarding spatial variation are more effectively supported by footprint-weighted

NDVI, LST, and LULC analyses. To improve the overall structure and clarity of the manuscript,

- we have added a workflow diagram (Figure 2) that illustrates how the different data sources were integrated into the analysis, which we believe enhances clarity and guides the reader through the analytical process more effectively.
- Included a new subsection titled 'Overview of Methodological Workflow' in the Methods section, which introduces the analytical approach and contextualizes the workflow diagram.
- Reorganized and renamed several sections and subsections to better reflect the sequence of the analysis.

These revisions aim to enhance both the focus and readability of the manuscript.

2.2.1 Overview of Methodological Workflow

The comprehensive overview of the methodological workflow is presented in Figure 1. It outlines the different data we used and the sequential steps undertaken in the study. We used:

- *site-specific eddy covariance (EC) and meteorological data (Observation data),*
- *remote sensing (RS) data, and (RS data)*
- *ERA5 reanalysis products(Reanalyis data)*

EC flux data measured at 3 m and 19 m were subjected to standard quality control procedures, gap filling, and flux partitioning procedures. The ERA5 Land dataset at hourly temporal resolution and ~9 km spatial resolution was obtained for the NamCo region. It included variables such as soil temperature, soil water content, radiation, RH, precipitation, snow cover, snow density (rsn), snow temperature (tsn), and snow depth (sde). RS datasets included Landsat 8, Sentinel-2, and MODIS, which were used to derive vegetation and surface indices such as normalized difference vegetation index (NDVI), land surface temperature (LST), normalized difference moisture index (NDMI), and normalized difference soil index (NDSI), normalized difference snow index and to generate the land use/land cover (LULC) map.

ERA5 meteorological variables were downscaled to the site level using linear regression fits with in situ measurements to produce a continuous, site-representative meteorological dataset for the study period (2018–2019). An XGBoost model was trained using site meteorology, MODIS-derived NDVI and NDSI, and observed NEE at 3 m to reconstruct 3 m NEE for the full study period (2018-2019).

We compared the monthly and seasonal NEE at 19 m against the long-term 3 m NEE to examine whether the monthly and seasonal NEE at 19 m fall within the multi-year variability of NEE at 3 m. Additionally, we compared the observed 19 m NEE with modelled 3 m NEE for the period 2018-2019 to account for different meteorological and environmental conditions between the years of measurement.

To characterize the spatial footprint of the EC towers, the Kormann and Meixner (KM) footprint model was applied to estimate the probability density matrices for both 3 m and 19 m measurement heights. These modeled footprints were combined with RS-derived products to *compute footprint-weighted indices (LULC, NDVI, NDSI, NDMI, and LST). To isolate land-originating fluxes from mixed (land + lake) contributions, footprint-weighted LULC was used. Half-hourly flux data with >99% land contribution were classified as land fluxes. The land-only fluxes were further gap-filled and portioned into component fluxes.*

The NEE measurements in the overlapping period (May 2019) were used for the direct comparison to assess the footprint-induced heterogeneity. Comparisons were conducted between (i) 3 m and 19 m total fluxes, and (ii) 3 m and footprint-filtered 19 m land fluxes. Finally, spatial drivers of observed NEE differences were investigated by analyzing differences in footprint-weighted surface indices and land cover distributions.

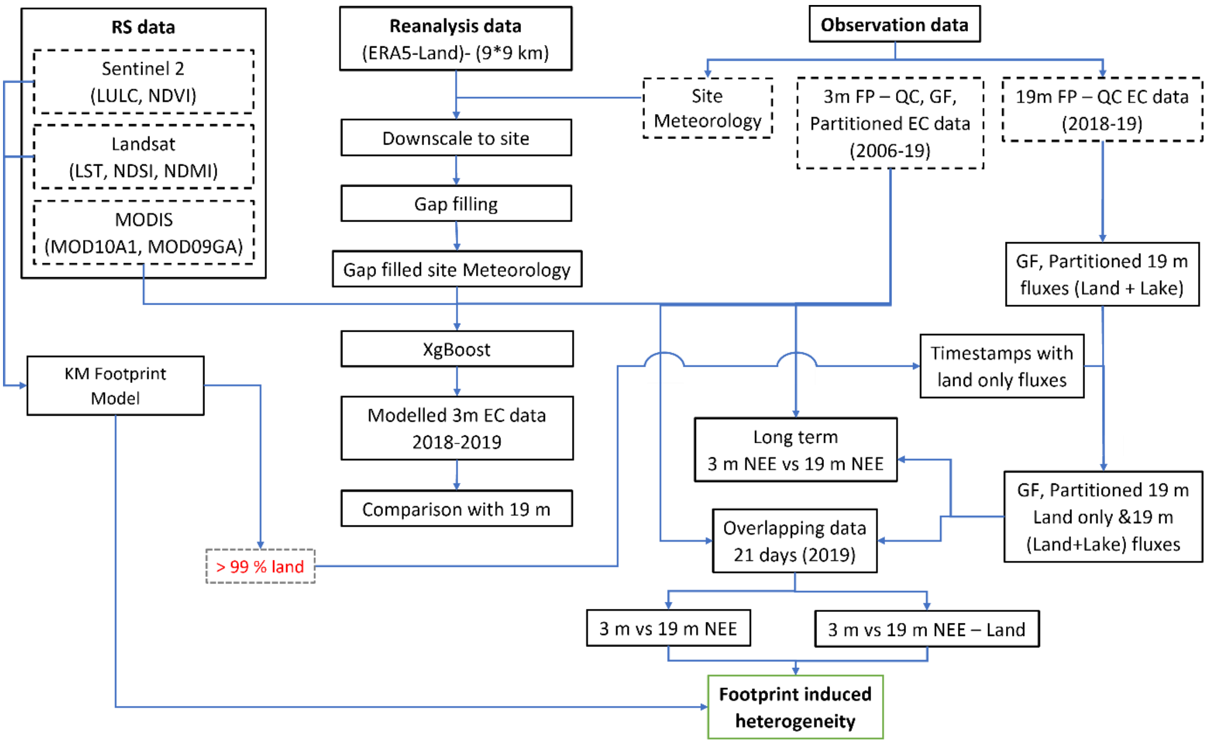

Figure 2: Overview of Methodological Workflow. GF- Gap filled; QC – Quality controlled, FP – Footprint, EC – Eddy covariance

Note: We appreciate the reviewer's constructive suggestions. The additional analysis combining footprint-weighted remote sensing indices with flux measurements provided new insights into the scale-dependent controls on NEE. Based on these findings, we have made corresponding revisions to the abstract and conclusion. The updated abstract and conclusion will be included in the revised manuscript.

Supplementary material

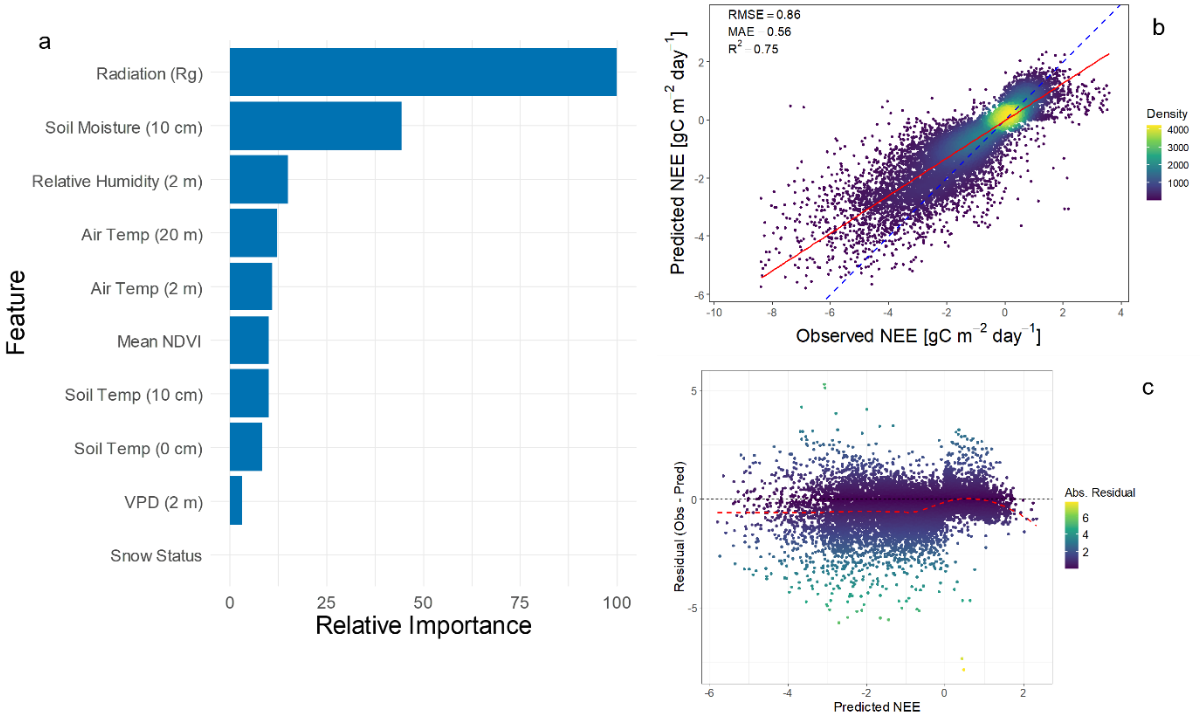

Figure S1: XgBoost results. Feature importance (a); scatterplot of observed and predicted NEE at 3 m with regression line (red) and 1:1 line (blue) (b); residual for the predicted NEE (c)

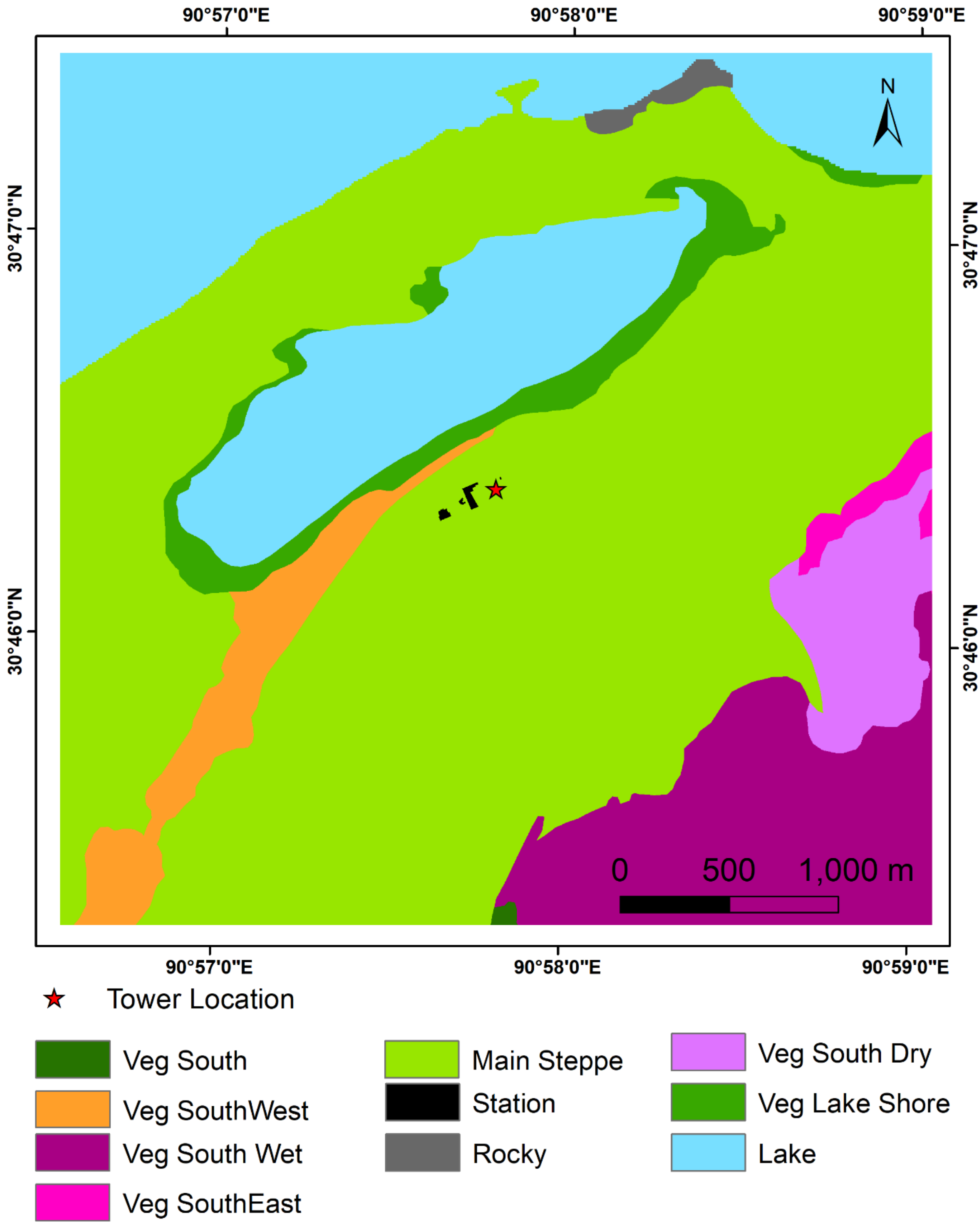

Figure S2: Land use land cover map Namco

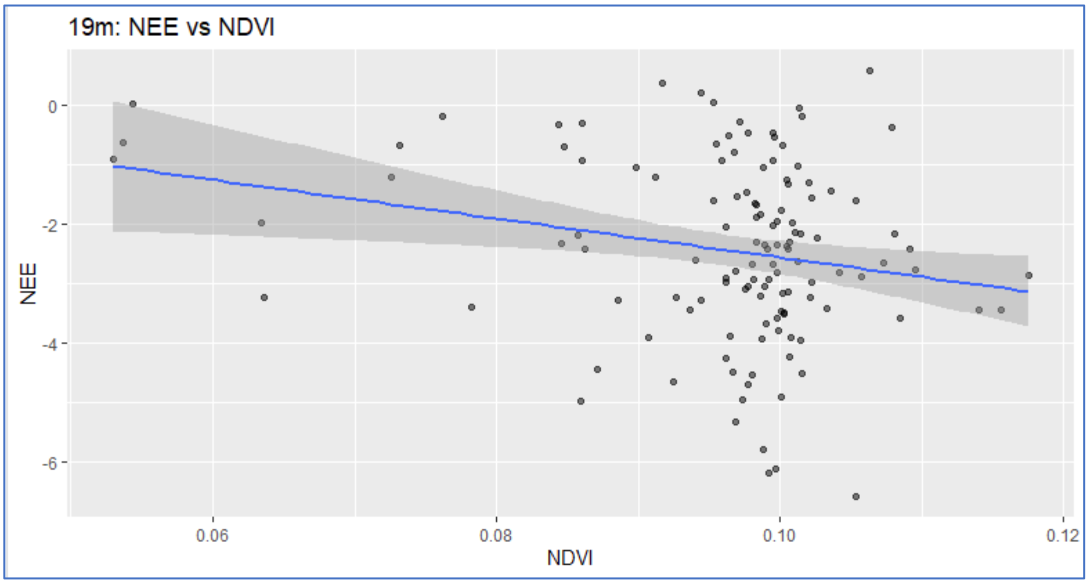

Figure S3: NDVI plotted against NEE at 19 m - May

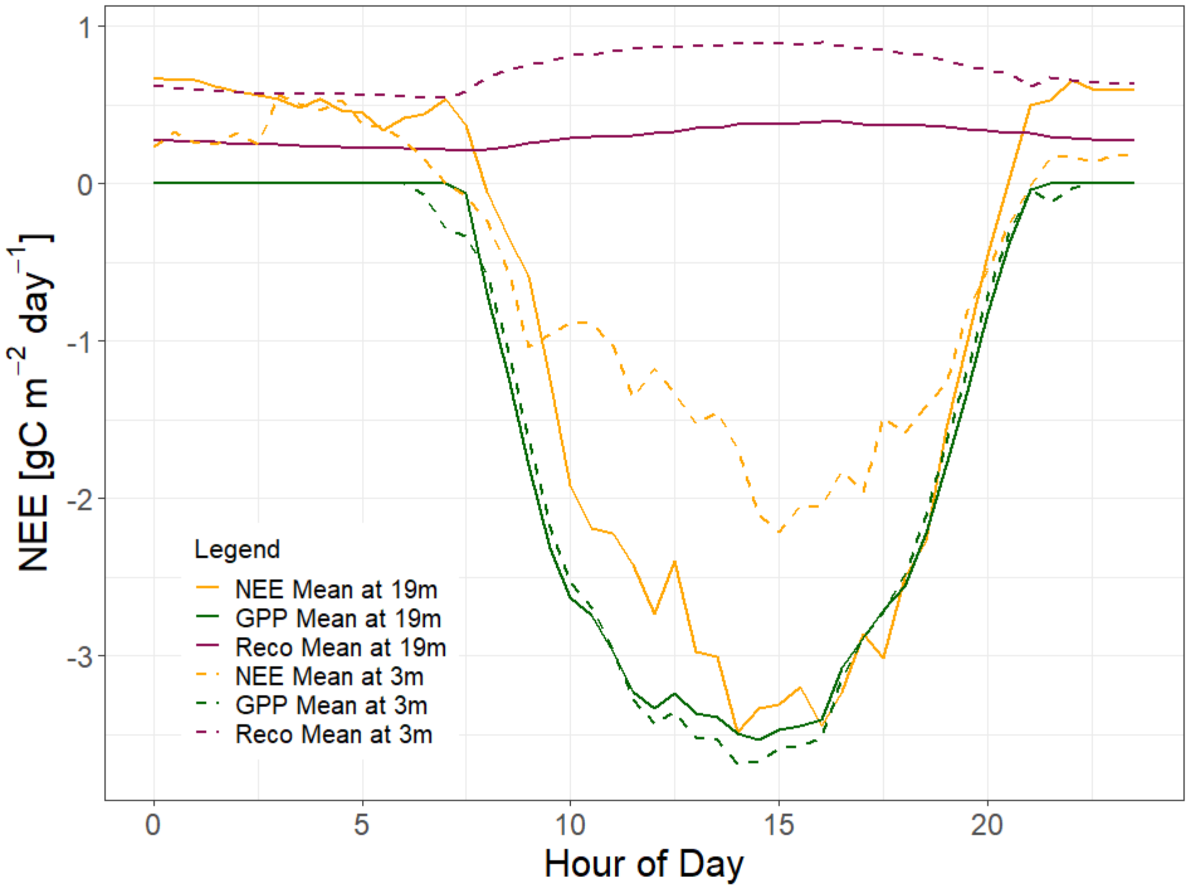

Figure S4: Diurnal cycle showing NEE, GPP and Reco over the overlapping period at 3 m and 19 m

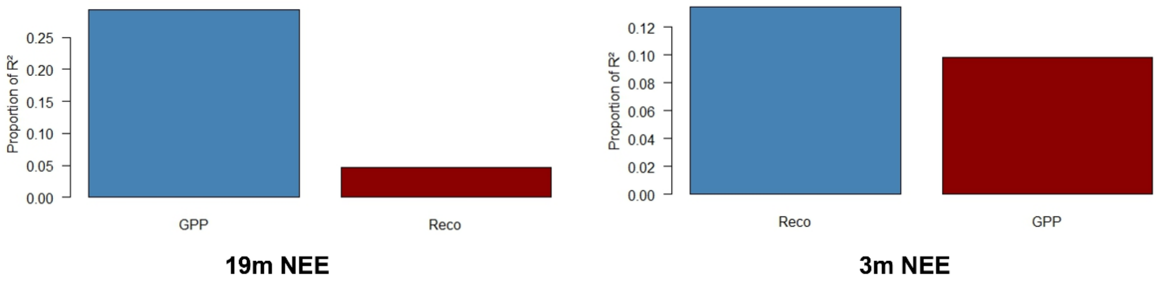

Figure S5: Relative effect of portioned GPP and Reco components on NEE at 3 m and 19 m for May 2019

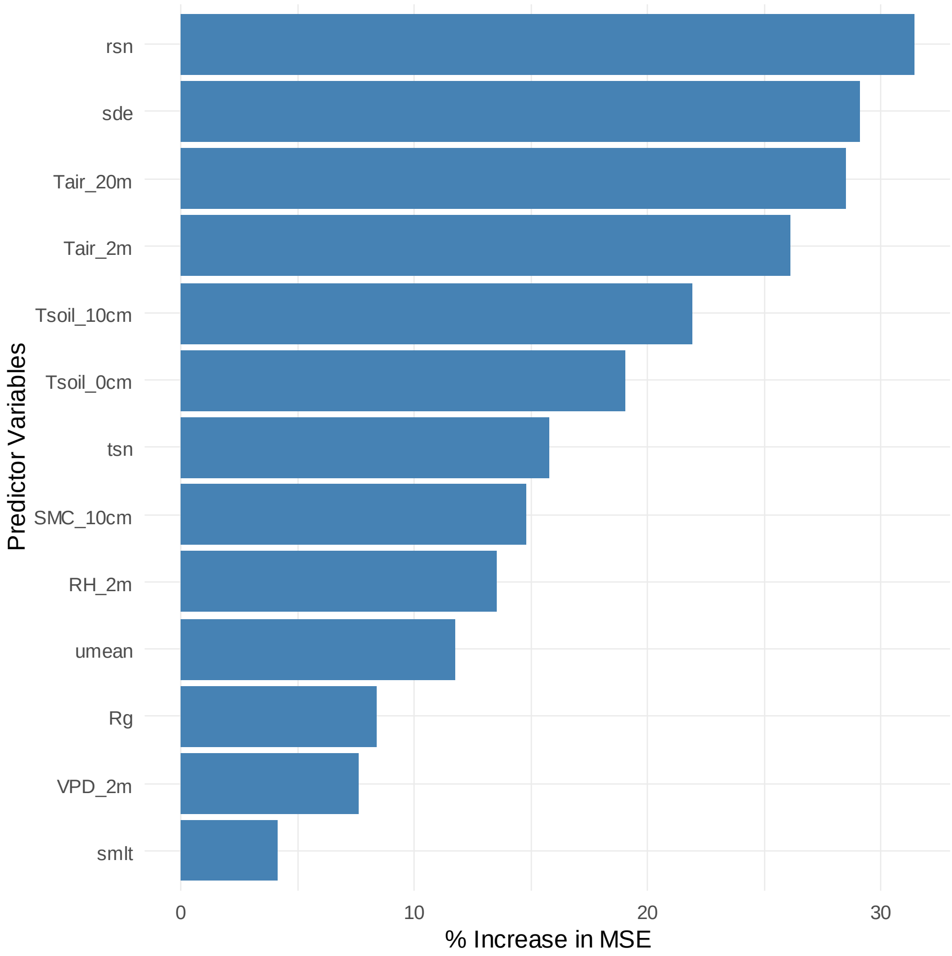

Figure S6: Feature importance of Reco in winter based on RF