Anonymous Referee #2, 20 Jun 2025

The authors would like to thank Reviewer 2 for the thoughtful and constructive feedback provided on our manuscript. We appreciate the valuable comments and suggestions, which have helped us improve the clarity and overall quality of the work.

Note: *We would like to highlight that substantial revisions have been made to the manuscript beyond direct responses to the reviewers' comments. In addition to addressing all reviewer feedback point-by-point, we have incorporated new analyses to further strengthen the study. These include footprint-weighted evaluations of biophysical indices (NDVI, NDMI, NDSI, LST), and land use/land cover classification.*

These analyses were added to better explain the observed flux variability between measurement heights (or the different spatial scales) and to provide a more comprehensive interpretation of the results. As a result, several sections of the manuscript—particularly the Results and Discussion—have been extensively revised and restructured. Some content originally in the Discussion has been moved to the Results to maintain a clear distinction between data presentation and interpretation, and to improve manuscript coherence.

We believe these additions and restructuring efforts substantially improve the clarity, depth, and overall quality of the manuscript.

The updated abstract and conclusion will be included in the revised manuscript.

The following responses address the comments and the issues raised by the Anonymous Referee 2 point-by-point. We have responded (in **black**) to each comment (in **blue**).

In general, I find the manuscript well written and easy to read from a language point of view, figures are all understandable and figure captions sufficient. However, I find the structure of the manuscript at times hard to follow.

We thank the reviewer for the valuable feedback. In response to the comment regarding the structure of the manuscript, we have made several revisions to improve its clarity and organization.

- Specifically, we have added a workflow diagram (Figure 1) 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 improve both the focus and readability of the manuscript.

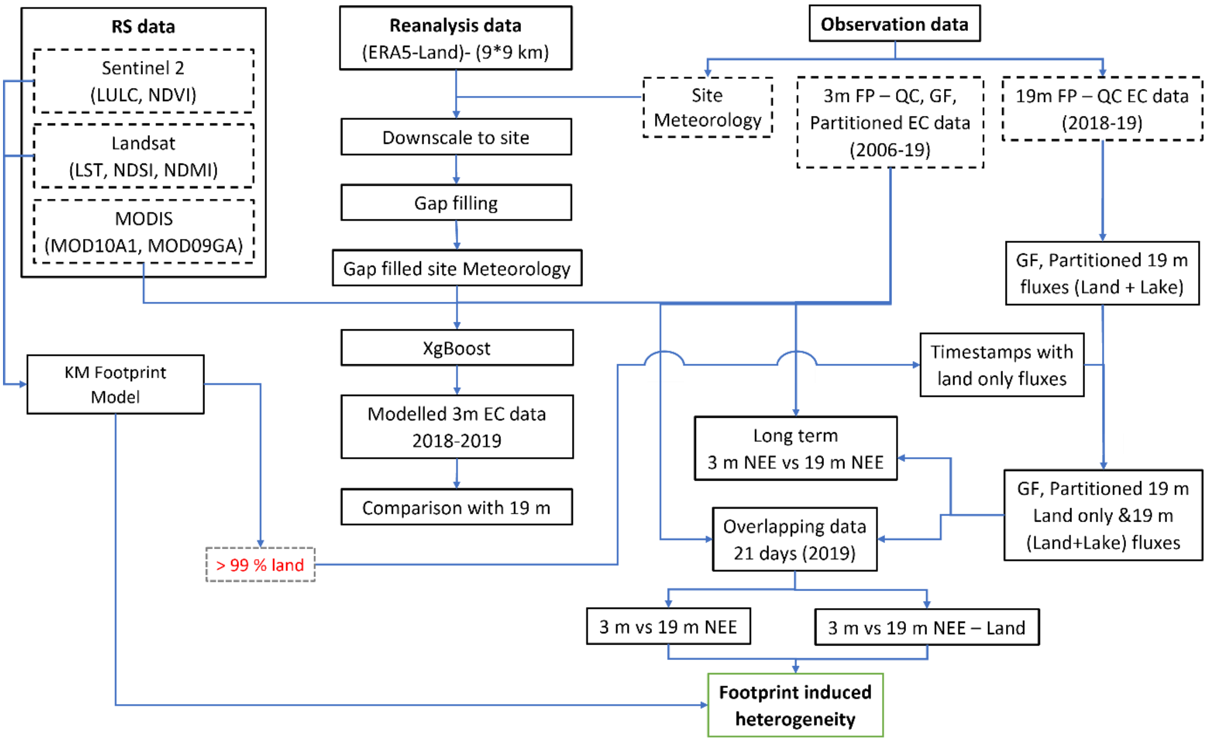

Figure 1: Overview of Methodological Workflow. GF- Gap filled; QC – Quality controlled, FP – Footprint, EC – Eddy covariance

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.

One of my major points about this manuscript is that the information about available and used data is not consistent. There are miss matches in the description of the data availability of the two systems. The data availability of the 3m system is stated as 2006-2019, with gaps 2012-2015 and 2018-2019, while in Figure 2 it is stated that data from 2006-2007 and 2016-2017 is used for the comparison. This makes it unnecessarily confusing for the reader. Could you please look through the manuscript and fix those mismatches? An easier understandable presentation of the data availability would also help the reader to follow.

Further you state that the overlapping time of the two systems was from May 13th to June 3rd 2019 when talking about the processing of the data from the 3m system. However, when talking about data from the 19m system you state that data was analysed from August 01st 2018 to May 31st 2019. This would only leave an overlap of data from May 13th till 31st. While you treat the long-term data set more as a reference data set, I still find the comparison period of less than 3 weeks too short. This is particularly true when comparing the fluxes monthly, especially at the early stage of the vegetation period, before onset of monsoon when fluxes are rather small and uncertainties high. I see a vast improvement if the dataset could be extended to include more overlapping data from both systems.

We acknowledge that the inconsistency in the information on the available data and the used data might have created confusion for the reader. In the study, we used EC data measured at 3 m height and EC data measured at 19 m height. The data availability for both heights was different. The 3 m data spanned from December 2005 to September 2019, with data gaps in several years (Nieberding et al., 2020). The 19 m EC data spanned from mid-July 2018 to June

2019. For a meaningful comparison, we needed a continuous period of data. So, we restricted the main analysis for the 19 m system to 10 months between 1st August 2018 and 31st May 2019, to preserve full calendar months for seasonal and monthly analyses. However, for direct comparison with the 3 m system, which had data available only from May to September in 2019, we included the first three days of June. This extended the overlapping period between the two systems to May 13 to June 3, 2019, providing just over three weeks of data for flux comparison between the two heights.

3 m data were grouped into 10-month periods for the years where data were available. The word period in the manuscript refers not to a year, but rather the 10-month period between August and to following May. The data used in the study consists of the periods 2006-2007, 2007-2008, 2008-2009, 2009-2010, 2010-2011, 2011-2012, 2012-2013, 2013-2014, 2014-2015, 2015-2016, 2016-2017, 2017-2018, and 2018-2019. Out of these 13 periods, only six periods had continuous data without any gaps after gap-filling. These periods were: 2006–2007, 2007–2008, 2008–2009, 2009–2010, 2010–2011, and 2016–2017. For meaningful comparison of the carbon budget at 3 m with 19 m at monthly, seasonal, and 10-month scales, data with a complete period were only used (Table 1). This processing step is now clearly explained in the revised manuscript.

Table 1: Data used in the study

	Available vs used data for long term comparison (Periods highlighted in green were used)													**Direct comparison**
3 m	2006-2007	2007-2008	2008-2009	2009-2010	2010-2011	2011-2012	2012-2013	2013-2014	2014-2015	2015-2016	2016-2017	2017-2018	2018-2019	13th May to 3rd June 2019
19 m													2018-2019	13th May to 3rd June 2019

You explain the treatment of data with contribution from the lake and how a land only and land lake flux dataset was derived from your overall dataset.

Why did you choose this the specific height and was the lake contribution was specifically targeted? This does not become clear when reading the manuscript.

Thank you for this observation. The 19 m sensor was installed to extend the observational range over more undisturbed steppe areas and to explore how larger spatial integration can inform carbon flux upscaling efforts. While the primary aim was not to specifically target the lake contribution, the sensor's broader footprint naturally included both land and lake surfaces.

Do I understand it correctly that the lake contribution was the difference of the total NEE and the pure land NEE? Considering that it is very common to have a strong land-sea breeze situation at Nam Co with quite distinct times when the EC system sees a lake contribution in the fluxes, does this lead to a large uncertainty the lake contribution?

Yes, your understanding is correct; the lake contribution was approximated by taking the difference between 19 m NEE (NEE_LL-which includes both **l**and and **l**ake signals) and the 19 m land-only NEE (NEE_LO).

We acknowledge that the land–lake breeze dynamics at Nam Co can significantly influence the flux footprint, introducing uncertainty due to temporal variability in footprint contributions across the diurnal cycle and months. However, Nonetheless, we believe this approach provides valuable, though indicative rather than definitive insight into the seasonal influence of the lake on net carbon fluxes.

In response to your comment, we have revised the approach and compared it with the approach used in the initial submission. We used NEE_LL, which includes both land and lake signals. To isolate NEE_LO, we applied a footprint filter (>99% contribution from land) to exclude mixed signals. The resulting data was gap-filled to get a complete timeseries for the NEE_LO and was used for calculating the monthly budget for land. The lake contribution was then estimated as the difference between the NEE_LL and NEE_LO.

We acknowledge that even this approach involves trade-offs: footprint filtering removes a large portion of data, which increases reliance on gap-filling and can introduce uncertainty. To evaluate this, we also tested the alternative approach (used in the initial submission), where we calculated monthly NEE from land-only periods based on the average NEE calculated using the data without complete diurnal coverage. Both methods carry uncertainty, either from incomplete temporal coverage or from gap-filling of sparse data. Our bias analysis comparing non-gap-filled land-only fluxes and footprint-filtered gap-filled fluxes revealed substantial and inconsistent relative biases ranging from approximately -108% to +82% across months. Relative bias was calculated as the difference between the monthly NEE budget from non-gap-filled NEE_LO data and the corresponding footprint-filtered, gap-filled value, expressed as a percentage of the latter. This large variation suggests that incomplete diurnal sampling in the non-gap-filled data leads to significant under- or overestimation of monthly NEE budgets. Given that these biases exceed the typical uncertainties associated with gap-filling, we conclude that the gap-filled flux method provides a more accurate and reliable estimate of NEE_LO in this study. Based on this, we have selected gap-filled NEE_LO data for the main analysis, as it provides more complete, consistent coverage while acknowledging the uncertainty introduced through gap-filling.

In the revised manuscript, we now present the results based on the gap-filled NEE_LO only. This has been clarified in both the methods and results subsections of the manuscript.

Revised portion of manuscript based on the above analysis– line 335-340, section 3.5 of the manuscript in the initial submission)

"To understand the influence of the lake on the NEE budget measured at the 19 m system, the NEE budget was calculated separately for the whole system (0.23 g C m^{-2}) and land alone (0.80 g C m^{-2}). The relative contribution of the lake to the total NEE displayed seasonal variation across the study period from August to May. During the snow-free months, in August (4 %), September (6 %), and April (21 %), the lake contributed positively to net CO_2 uptake, with the highest influence in April, likely driven by early ice melt and increased lake-atmosphere exchange. From October to March, the entire area was snow-covered, and the lake became fully ice-covered by late December. During this period, the overall NEE of the system was positive, indicating net CO_2 emission to the atmosphere. Interestingly, the lake contributed to reducing those emissions in November (4 %), December (14 %), February (11 %), and March

(7 %), acting relatively as a sink compared to the land surface under snow and ice cover. In contrast, January (-4 %) and May (-2 %) showed negative relative contributions, suggesting the lake slightly enhanced net CO_2 release during those months. October presented a unique case where total NEE was near zero (~0.07 gC m^{-2} $month^{-1}$), causing an unrealistically large and unstable relative contribution; therefore, this value was excluded from interpretation."

The estimation of the monthly NEE budget of the 19 m footprint area for land-only and land-lake systems also raises some questions. It might be my misunderstanding of the described method, so could you please explain how this was calculated in more detail?

We appreciate the opportunity to clarify our methodology for estimating the monthly NEE budget for both the 19 m (land+lake) system and the 19 m land-only system.

In our analysis, we used a footprint-weighted land cover classification to identify the dominant source area of the fluxes measured at the 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 (inclusive of lake contribution) and NEE_LO (exclusively from the alpine steppe).

These distinctions and the data processing workflow are now explicitly stated in the revised manuscript and illustrated in Figure 1 to improve clarity.

How were lake contributions to the total NEE considered for periods the EC system measured only land fluxes?

To specifically address how lake contributions were considered during periods when the EC system measured only land fluxes:

The lake's contribution to the NEE_LL was estimated indirectly by subtracting the monthly NEE_LO budget from the monthly NEE_LL budget. The NEE_LL timeseries from the 19 m tower includes both land and lake fluxes, while the NEE_LO timeseries obtained through footprint filtering and gap-filling represents fluxes exclusively from the alpine steppe. By calculating the difference between these two monthly time series, we derived an estimate of the lake contribution to the NEE_LL budget. Because this estimation is performed monthly, any concerns related to periods when only land fluxes were measured are minimized.

These distinctions are now clearly stated in the revised manuscript.

I think I can follow the classification into land-land and lake-land due to the contribution derived from the footprint analysis. You describe that you created the land-land and land-lake system dataset by including fluxes form the specific surfaces according to the overall contribution of the flux from this surface to the overall flux. In my understanding this was done for each of the 8 wind sectors you defined and then a composite flux was estimated according to the land use contribution derived from the footprint model. Then monthly budgets were calculated. This is the point I must admit that I can't follow how this was done. Could you please elaborate more on how it is done? How were fluxes gap filled/modelled for specific land cover for periods when there were no contributions from this direction.

We appreciate the opportunity to clarify these aspects to improve understanding and transparency.

Regarding the wind sector analysis: in the original manuscript, our focus was on comparing NDVI and NEE variations across the defined wind sectors to identify dominant wind sectors for carbon uptake. However, we did not calculate or gap-fill fluxes separately for land and lake contributions within each wind sector, nor did we generate composite fluxes based on land cover proportions per sector.

In line 240 you describe that that the portioning was performed after gap filling. Does this mean lake fluxes were gap filled as if they were land fluxes?

In the revised manuscript, the limitation due to gap-filling the time series prior to classifying the footprints was addressed by first applying the footprint filter to isolate land-only fluxes and then performing gap-filling on this filtered dataset. This updated approach ensured that the resulting land-only NEE time series had continuous coverage, including all diurnal periods, thereby reducing uncertainty and improving the reliability of the budget estimates.

You go on to state that the datasets might have missing half hour value. Can you please explain? Does this mean each dataset was only sorted according to the classification but no continuous dataset for each classification was derived? Are the monthly budgets, growing season and non-growing season then calculated on unequal amount of data? Or did you include the relative contribution of the lake fluxes into each half hour based on your footprint calculations? This is a point that needs to be clarified for future publication. If it is rather my lack of understanding, please consider the way how you describe this rather complex but very important part of the data handling.

As for the budget calculations: monthly and seasonal budgets were not calculated from datasets of unequal temporal coverage. All comparisons of NEE_LO and NEE_LL were based on datasets of equal length and consistent time coverage, ensuring a comparable and meaningful comparison.

These revisions are clearly described in the updated Methods section.

The subsequent comparison of the two datasets is based on a dataset of an averaged ten-month period resembling the month the 19m system was deployed. While I see some value in this approach, I don't see how this can lead to the rather in detail comparison the authors present. Considering the rather rapid trend in warming on the TP, coinciding with shifts in precipitations pattern, available soil moisture and potential vegetation changes, I would have expected to see the in-detail comparison of the overlapping date before discussing the differences of the reference period with the 19m system.

We understand your concern. We have revised the manuscript to prioritize a detailed comparison of the datasets during their overlapping deployment period, as suggested. To analyze the overlapping period and address the variability observed in May,

- We used **footprint-weighted land use/land cover** (LULC) data to separate 19 m fluxes into land-only and land–lake categories. This allowed us to evaluate the spatial 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 NEE_LO, providing insights into the role of vegetation dynamics and their impact on flux differences between the two heights.
- 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 address the difference observed in the NEE over the two footprints. These additions help to better explain the observed differences in NEE between the two measurement heights and to identify the surface drivers and physiological processes influencing fluxes over the alpine steppe.

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) 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, more stressed vegetation with greater microbial activity. 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 and less stressed vegetation patches. Furthermore, the correlation of NEE with NDVI supports this interpretation, capturing the influence of higher productivity in the 19 m footprint.

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 S3).

Based on our findings, we have added a paragraph to the discussion in the revised manuscript:

"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 S4), 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."

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

The purpose of the paragraph of wind sector discussion in Section 3.6 is not entirely clear to me since it on one hand talks about NDVI but quantify the differences in NDVI only briefly and on the other hand it discussed the different contributing wind directions, namely the land sea (lake) breeze system found at NAMORS due to Nam Co Lake. Considering that the main aim of the study was to relate CO_2 fluxes to landscape heterogeneity I would suggest including an overview of the differences found in NDVI also in a map.

Thank you for pointing this out. In the revised manuscript, we have removed Section 3.6. on the wind sector analysis. To address landscape heterogeneity and its relation to CO_2 fluxes more effectively, we expanded our analysis using footprint-weighted NDVI, LST, and LULC data, which offer more direct and quantitative insights into spatial variation. Although we have not included an additional NDVI map as suggested, we provide a comprehensive table (Table 2) summarizing the values of these indices across different footprints.

Table 2: Statistics of different remotely sensed indices for 3 m and 19 m footprints

Indices	**3 m**				**19 m**			
	Min	**Max**	**Mean**	**Median**	**Min**	**Max**	**Mean**	**Median**
NDVI	0.10	0.12	0.11	0.11	0.06	0.12	0.10	0.10
LST	29.6	34.08	32.4	32.4	12.1	34.1	28	29
NDSI	0.29	0.6	0.33	0.33	0.14	0.36	0.30	0.30
NDMI	-0.11	-0.08	-0.095	-0.095	-0.11	-0.04	-0.08	-0.08

Detailed remarks:

Ll 21: Here in the abstract, you mention a 10-month long average of -78 gC m-2 derived from data collected 2006-2018. This does not fit to the data-availability described in the Method section. Did I misunderstand something? Could you please elaborate how this number was calculated? If it is indeed derived from a non-continuous dataset, I suggest you mention this here. Otherwise, it confuses the reader later.

Thank you for pointing this out. We appreciate the opportunity to clarify. The 3 m data spanned from December 2005 to September 2019, but includes gaps in several years. For a meaningful comparison with the 19 m dataset, which spanned from mid-July 2018 to June 2019, we needed a continuous period of overlapping data. Therefore, we limited our analysis to the 10-month period from 1 August 2018 to 31 May 2019, which was the longest continuous period available for the 19 m dataset.

To compare the carbon budget of this 10-month period with long-term data from the 3 m dataset, we selected all available 10-month periods between August and the following May for the years where data were continuous. These periods were: 2006–2007, 2007–2008, 2008–2009, 2009–2010, 2010–2011, and 2016–2017 (Table 1). Periods with incomplete data coverage were excluded.

The average value of –78 gC m^{-2} reported in the abstract was calculated as the mean carbon budget across these six 10-month periods from the 3 m data. In the revised manuscript, we have clarified this point to avoid any confusion.

L 24: Please define extreme in this context.

L 25: for which period? 2006-2018 or 2019?

Thank you for pointing these out. We appreciate the opportunity to clarify. In this context, “extreme” refers to a statistically significant snow event based on historical snow depth records. Specifically, the winter of 2018–2019 experienced prolonged snow cover from October 2018 to March 2019, with unusually high snow depth.

We compared snow depth during the 2018–2019 period with other periods for which 3 m flux data were available and found that 2018–2019 was significantly different ($p < 0.001$), showing elevated snow depth. To further assess whether this constituted an extreme event, we analyzed long-term snow depth data spanning the past 74 years (1950–2024). For consistency, we grouped each year into winter periods covering October to April (e.g., 1950–1951 refers to October 1950 through April 1951). This analysis confirmed that the 2018–2019 snow depth anomaly exceeded 2 to 3 standard deviations above the long-term mean, thereby meeting the criteria for an extreme event.

We revised the abstract to define "extreme" in this context and specify the reference period to avoid ambiguity.

Ll28: If one of your main conclusions is that the 3m system is representative for alpine steppe, but that the lake has a rather large influence on CO2 exchange on a landscape level, wouldn't it make more sense to propose a lake flux system to have continuous fluxes from both surfaces?

Thank you for this insightful comment. In our initial analysis, we found that fluxes measured over land-dominated footprints at both 3 m and 19 m heights were largely consistent, which led us to conclude that the 3 m system was representative of alpine steppe carbon fluxes. However, in response to your comment and based on further detailed analysis, we now recognize important differences in NEE between the two systems, even under land-only conditions. When separating daytime and night-time NEE, we found statistically significant differences between the 3 m and 19 m land-only fluxes ($p < 0.001$ in both cases). These differences were not apparent when analyzing combined 24-hour NEE data ($p = 0.107$), likely due to offsetting effects between photosynthetic uptake and night-time respiration.

To better understand the drivers behind these differences, we analyzed the footprint characteristics during overlapping periods using NDVI, LST, NDMI, NDSI, and land cover data. Additionally, we assessed the relative contributions of GPP and Reco to NEE at both tower heights. Our results show that NEE at 3 m was more strongly influenced by Reco, whereas at 19 m it was more closely associated with GPP, indicating a scale-dependent shift in process dominance in May.

Although the footprint-weighted land cover analysis indicated variable lake influence at the half-hourly scale (0–70%, as noted in the original manuscript), the average lake contribution during the overlapping period was ~11%. Correlation analysis further showed a slight positive relationship between lake contribution and NEE in May, suggesting that the lake may actually dampen CO_2 uptake rather than enhance it. This indicates that the lake is unlikely to be the main driver of the more negative daytime NEE observed at 19 m. Instead, we conclude that the observed differences are more plausibly linked to coarse-scale heterogeneity in vegetation productivity and respiration patterns across the broader footprint.

While our current data suggest that the lake's influence is relatively modest during the study period, we agree with your broader point of establishing a dedicated flux measurement system over the lake would be valuable for fully capturing the lake–land interactions and better quantifying their role in landscape-scale carbon dynamics. We have revised the manuscript accordingly to reflect the updated conclusions from this additional analysis.

Introduction:

Covers the most important background both for the TP and EC, well written and easy to read. Leads into the problem of the spatial differences between in situ observations and larger scale model domains.

Ll 82: I like this paragraph as introduction to the measurements done in this study; however, I am not sure if it is the correct place. It already includes quite a lot of details regarding the 19m system and differences in the properties of the fetch. However, I think that for most readers some more details are missing to fully understand what is going on. I suggest generalizing a bit more here and move some of this paragraph into the method section.

Thank you for your helpful suggestion. We agree that the original paragraph contained a blend of conceptual framing and methodological specifics, which could be challenging for readers unfamiliar with flux footprint dynamics. In response, we have revised the paragraph to retain its role in introducing the study rationale and objectives, while generalizing the description of the measurement systems and spatial heterogeneity. The revised portion is as follows:

"Unlike a mixed, disturbed landscape (Krasnova et al., 2022), the TP alpine steppe presents a relatively undisturbed, rather homogeneous ecosystem, allowing the study of complex interactions at varying spatial scales. Nieberding et al., (2020, 2021) studied in-situ CO2 fluxes from the alpine steppe ecosystem at Nam Co, spanning 14 years of flux observations along with meteorological variables and plant cover estimations. The authors revealed a long-term increase in CO_2 uptake, mainly caused by increased atmospheric temperatures during winter and a shift in summer monsoon precipitation. Extending this research, the present study compares net ecosystem exchange (NEE) measurements of CO_2 collected at two sensor heights (3 m and 19 m) at the same site. These two heights represent different spatial scales, with the lower sensor integrating fluxes from a more homogeneous steppe surface and the higher sensor capturing a broader, more heterogeneous footprint. Even a seemingly homogeneous alpine steppe ecosystem can exhibit substantial spatial heterogeneity in vegetation distribution, soil moisture regimes, and proximity to features such as lakes, all of which can influence flux measurements (Biermann et al., 2014). By evaluating how fluxes differ between these measurement heights, our study emphasizes the value of shifting from ecosystem-specific to landscape-level carbon assessments. A broader spatial perspective allows us to better account for heterogeneity and interactions across ecosystems, supporting more accurate upscaling and model parameterization in high-altitude regions."

Methods:

2.1 gives a good overview over the site setup, location and used equipment. I think that some of the information from the paragraph ll82 would fit well here. Potentially as a new sub chapter.

Thank you for the helpful suggestion. In response to your recommendation, we moved the content (formerly in the Introduction) related to the respective footprint area of 3 m and 19 m sensor height to section 2.1. Rather than creating a new subsection, we integrated this material as a concise paragraph. The added text is as follows:

"The source area of the 3 m measurements is relatively homogenous, and exhibits minimal variation in vegetation types and density, making it a suitable reference for understanding the baseline patterns in the fluxes. In contrast, the source area of the 19 m measurement covers a broader area that includes alpine steppe with subtle vegetation density variations, more productive and wetter patches near the lakeshore, and occasional inclusion of the adjacent lake surface. The 19 m sensor was installed to extend the spatial scale of the footprint to a more diverse steppe area and to explore how larger spatial integration can inform carbon flux upscaling efforts. These spatial differences in vegetation and surface cover within the 19 m footprint are integral to understand how these spatial components in the footprint affect the overall carbon dynamics within the alpine steppe ecosystem".

ll 230: You describe that you calculated the land cover contribution for each of their 8 wind sectors, but the result of this is not presented adequately in my opinion. This information would help a lot to interpret the overall results.

The wind sector analysis has been removed from the revised manuscript.

Results:

Ll250: are these the results of land-land or overall NEE? In my opinion a figure showing the NEE data would make it easier for the reader to follow than just numbers in the text.

We confirm that, unless otherwise stated, the NEE values presented in the manuscript refer to the overall fluxes from the 19 m footprint (NEE_LL). The sections that discuss land-only fluxes (NEE_LO) explicitly mention this distinction in the text. To improve clarity, we have now added a sentence in the Results section to explicitly state that NEE_LL is the default throughout the manuscript unless otherwise indicated.

While we appreciate the reviewer's suggestion, we have chosen not to include an additional standalone figure showing the NEE_LL trends at 19 m, as these patterns are already clearly described in the text and further figures would increase redundancy and length. However, we have added new figures in the revised manuscript that present NEE_LL in the context of comparisons between fluxes predicted at 3 m and measured at 19 m (Figure 2 below).

3.3 When comparing the single monthly values to the long-term mean from previous years. It would be good to show also how the meteorological conditions compare and how vegetation cover was developing over the time period.

We thank the reviewer for this constructive suggestion. While Figure 2 in the manuscript compares single monthly NEE values with long-term means using box plots, we have already addressed the observed deviations in the text by discussing relevant meteorological anomalies (e.g., soil moisture and snow cover) and spatial heterogeneity during the study period.

To further strengthen this, 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).

The XGBoost model demonstrated satisfactory performance, achieving RMSE, MAE, and R2 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.

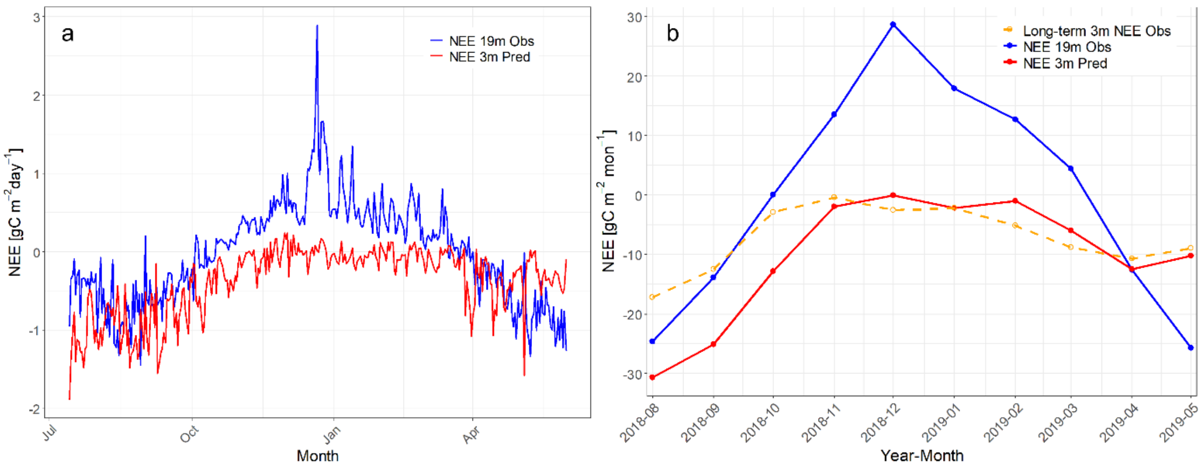

Figure 2: Predicted 3 m daily (a) and monthly (b) NEE plotted against 19 m NEE

Figure 3: There seems to be a large shift in the magnitude of NEE for all seasons from 2010 on at the 3m system. Is there an explanation for this?

We thank the reviewer for this thoughtful observation. This pattern is consistent with findings from our earlier study (Nieberding et al., 2021), which reported a phase of pronounced ecological and climatic changes in the region, including increased plant cover and stronger CO_2 uptake. As the current study extends that earlier work, the observed trend is well supported by the previously documented ecosystem response.

Where was the snow depth measured at? Or is this from a remote sensing product?

The snow depth data reported in the manuscript were not collected through direct field measurements. Instead, they were obtained from the ERA5 reanalysis land data product, covering the study region. Details of this data source and its processing are provided in the methodology section.

Ll 350: How large was the difference? What do you mean with practical importance?

The median difference for the land exclusive fluxes ($NEE_{3m} - NEE_{19m}$) between the 3 m and 19 m measurements during the overlapping period was 0.58 gC m^{-2} day^{-1} (IQR = 1.36). We clarified both the magnitude and the potential practical relevance of the difference in the revised manuscript. The Wilcoxon signed-rank test yielded $p < 0.001$, indicating a statistically significant difference between NEE measured at 3 m and NEE_LO during the overlapping period. The rank-biserial correlation was 0.57, reflecting a large effect size. This suggests that the measurement heights and corresponding footprints can substantially influence NEE estimates, even when restricted to land-only fluxes.

The revised portion of the results (line 350 of the manuscript in the initial submission) is as follows:

"The fluxes originating exclusively from the land area during the overlapping period showed a statistically significant difference (Wilcoxon signed-rank test, $p < 0.001$) between the NEE measured at 3 m and 19 m. The median difference ($NEE_{3m} - NEE_{19m}$) was 0.58 gC m^{-2} day^{-1} (IQR = 1.36). The rank-biserial correlation was 0.57, reflecting a large effect size. This suggests that the measurement heights and corresponding footprints can substantially influence NEE estimates, even when restricted to land-only fluxes".

This whole section discusses a lot of statistics of comparison of the fluxes measured from the different levels but does only briefly discuss the results of the comparison on a flux level.

NDVI data: Also, here a figure or a table would make it easier for the reader to follow and the manuscript much more attractive. What were the actual NDVI values?

We acknowledge the reviewer's point and have revised the manuscript accordingly. A clearer discussion of the actual flux differences between measurement heights has been added, including their ecological relevance and potential causes. To better understand the spatial variation in NEE observed at 3 m and 19 m during May, we combined flux footprint modelling with remote sensing-derived indices. Specifically, we used footprint-weighted values of NDVI, NDMI, NDSI, land surface temperature (LST), and a high-resolution land cover map to characterize the biophysical properties within each footprint. To improve clarity and support interpretation, we have included a summary table (Table 1) presenting key results from these analyses, including both footprint-weighted biophysical indices and parameters derived from the light-response curves.

Discussion:

Your discussion provides a lot of interesting facts and puts your findings in context. However, I would have liked to see some numbers a various places eg. Ll386 (slightly more release), ll387 (percentage contribution). Overall, the discussion touches a lot of topics, that have not been explored earlier in this study. I think by looking more into detail of the differences in vegetation cover, composition etc in the wind sectors and the different footprint areas, and showing this, some of this discussion points would be more rooted in the manuscript. The figure 6 and snow cover result description should in my opinion be rather moved up into the results.

We thank the reviewer for the valuable feedback on the discussion section. In response, we have revised the manuscript to better integrate key numerical values into the discussion. Specifically, while the percentage contribution of non-growing season respiration to annual totals (35% at 3 m and 57% at 19 m) was already presented in the Results section (Line 314), we have now ensured this is also referenced in the Discussion where relevant, to strengthen the connection between the data and interpretation.

We agree that integrating results into the discussion must be done carefully. Our intention in this section was to interpret the observed differences in fluxes, and thus we believe that content such as the snow cover description and the footprint-based vegetation analysis is most appropriate in the Discussion, where it helps explain the variability in NEE and ecosystem respiration between measurement heights. However, we acknowledge that some aspects (such as Figure 6 and associated snow cover data) present empirical findings and have therefore moved them into the Results section for improved clarity and manuscript structure.

Overall, we have revised the manuscript with careful attention to the reviewer's suggestions and aimed to balance clarity, structure, and scientific reasoning throughout.

Ll409: this should have been shown in detail before.

We thank the reviewer for this valuable observation. In response, we have expanded the manuscript to include detailed analyses of the spatial variability within the respective flux footprints. Specifically, we now present footprint-weighted values of land surface indices (NDVI, NDMI, NDSI, LST) and land use/land cover (LULC) classifications, which highlight the heterogeneity in biophysical conditions captured at 3 m and 19 m measurement heights. These additions are now included in the Results section (Table 1). As a result of this expanded analysis, the corresponding section of the Discussion has been revised to better reflect the new findings and their implications for interpreting NEE variability.

Ll420: this is information missing in the result part.

We thank the reviewer for pointing this out. In response, we have moved the quantitative findings regarding footprint composition (i.e., land vs. water contribution) into the Results section, where they are now presented as part of the footprint-weighted land cover analysis. These values are subsequently interpreted in the Discussion section in the context of their influence on flux variability.

Ll 425: From here on this is more results than discussion as you discuss only your own results and not with other literature. And foremost this is information that would have helped to follow the described results.

Thank you for this helpful observation. In response, we have extensively revised and restructured the manuscript to better separate Results and Discussion sections, ensuring that data presentation and interpretation are clearly distinguished.

Supplementary material

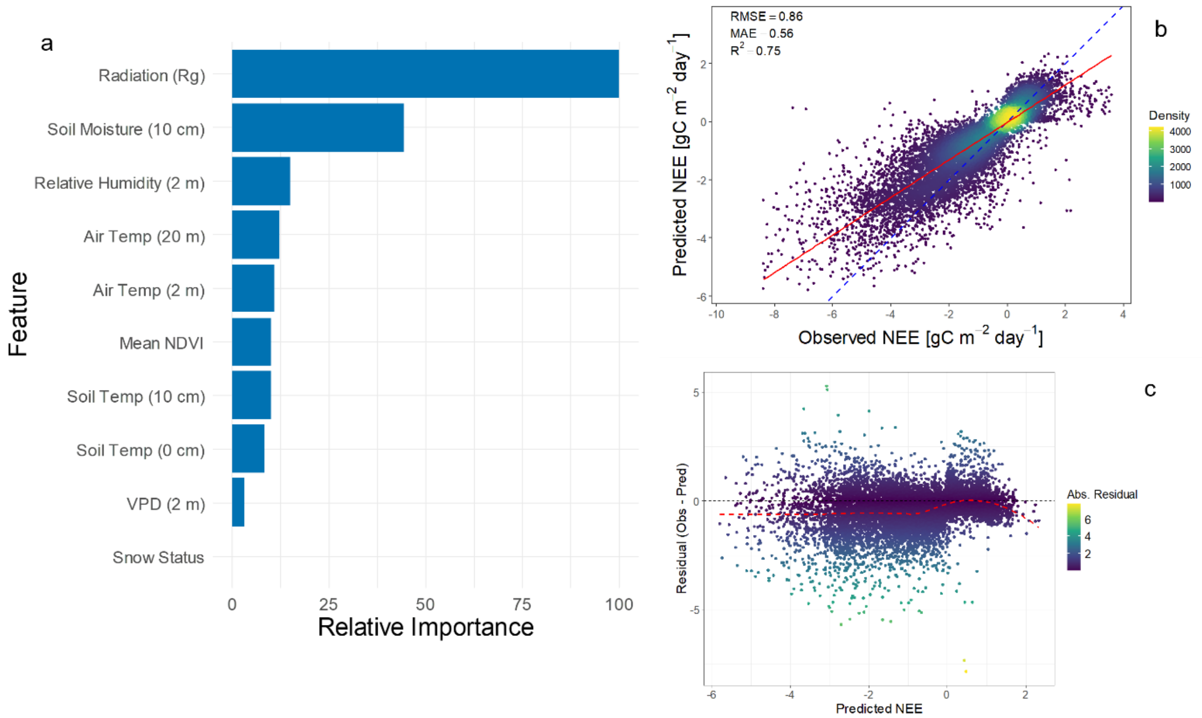

Figure S1: XgBoost results. Feature importance (a); scatterplot of observed and predicted NEE at 3 m with regression line (red) and 1:1 line (blue) (b); residual for the predicted NEE (c)

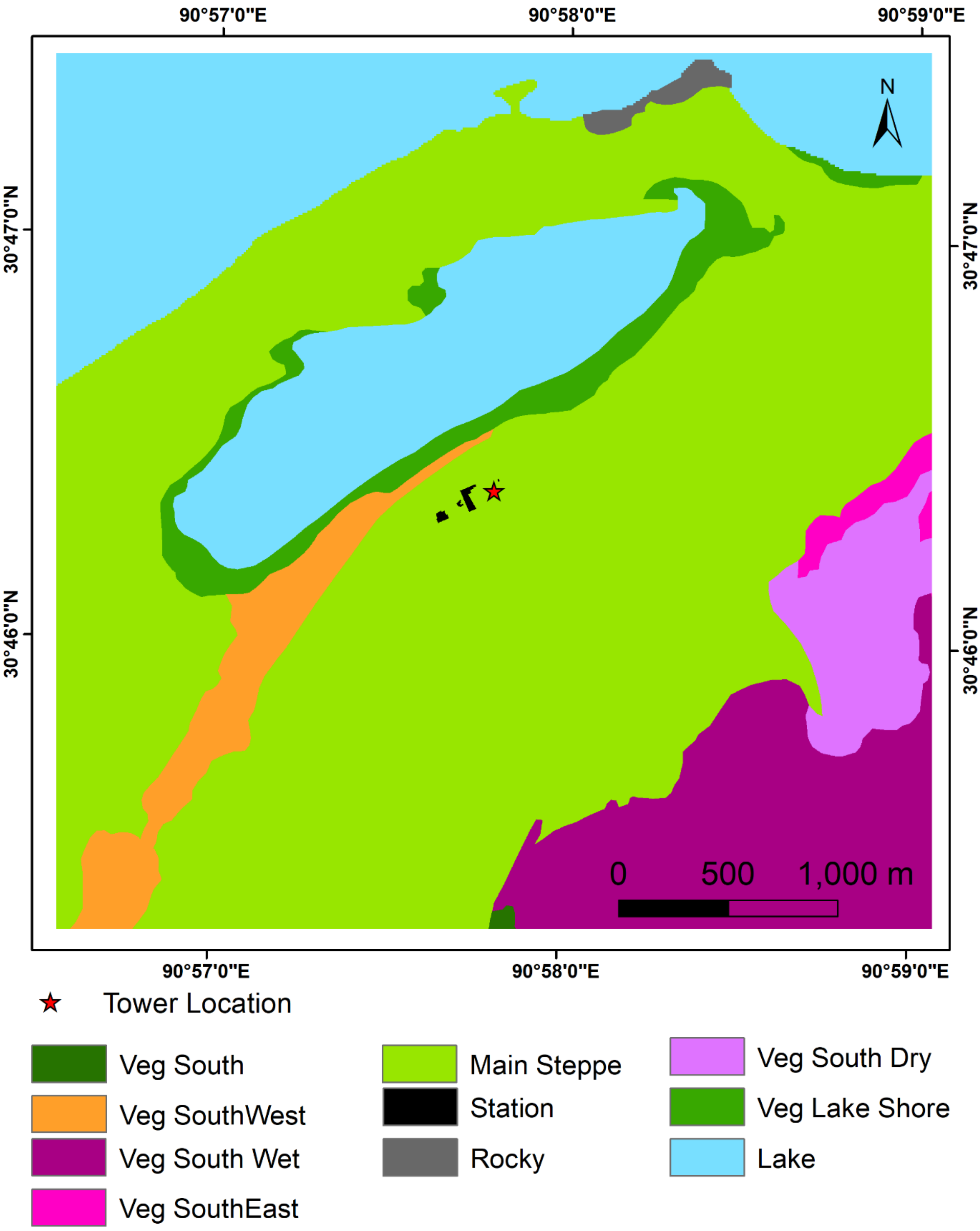

Figure S2: Land use land cover map Namco

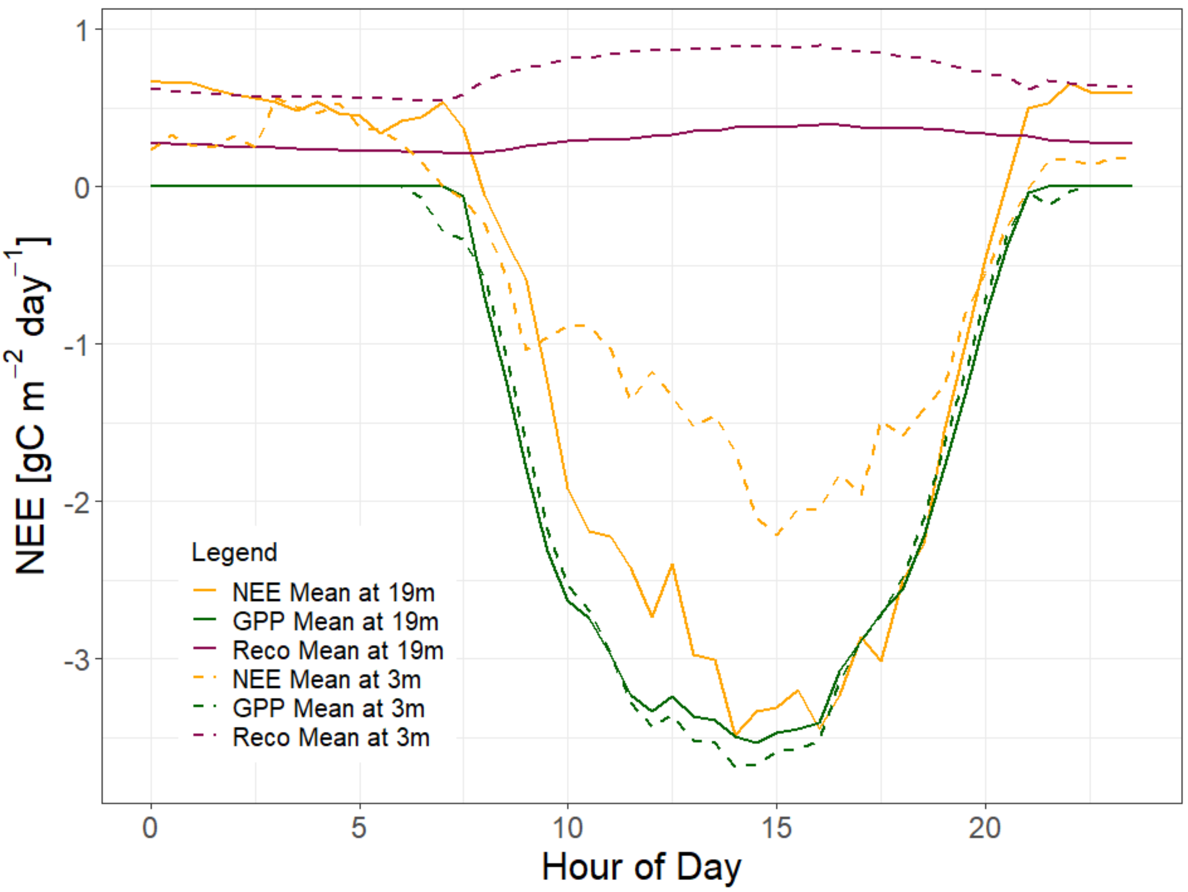

Figure S3: Diurnal cycle showing NEE, GPP and Reco over the overlapping period at 3 m and 19 m

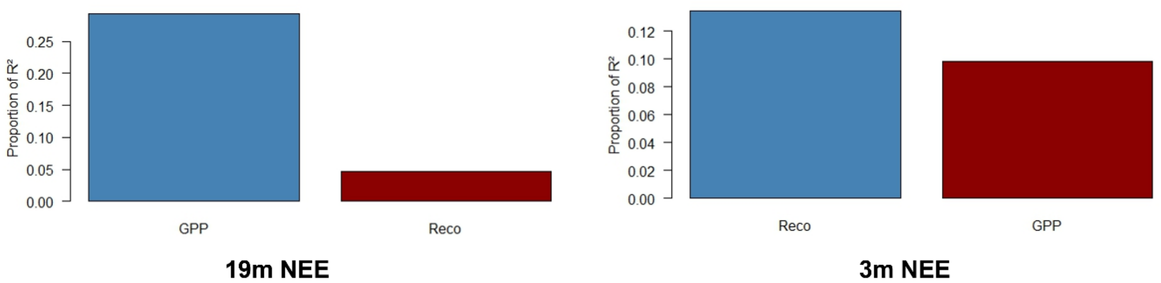

Figure S4: Relative effect of portioned GPP and Reco components on NEE at 3 m and 19 m for May 2019