Response to Referee #1

Assessing evapotranspiration dynamics across central Europe in the context of land-atmosphere drivers

Comments from Reviewer 1	
The manuscript titled **"Assessing evapotranspiration dynamics across central Europe in the context of land-atmosphere drivers"** evaluates evapotranspiration (ET) products derived from satellite remote sensing, modeling, and reanalysis data in conjunction with in-situ observations from Integrated Carbon Observation System (ICOS) stations in central Europe from 2017 to 2020. The study investigates the effects of varying land cover types, soil moisture (SM), and vapor pressure deficit (VPD) on ET dynamics, including the severe drought of 2018. It uses extended triple collocation methods to assess the accuracy of ET products, revealing notable differences among products under heterogeneous land cover conditions and during drought years. The findings highlight that ET variability is strongly influenced by VPD under non-limiting soil moisture conditions and demonstrate that SEVIRI, ERA5-land, and GLEAM products show superior performance. The research provides insights into the suitability of various ET products for capturing land-atmosphere interactions and drought impacts across diverse land cover types. This manuscript presents valuable insights into evapotranspiration dynamics across central Europe, but significant revisions are needed before it can be considered for publication.	Thank you very much for all efforts and help to improve the manuscript. We tried our best to answer all comments satisfactorily.
Major comments: 1. Limitations in the Performance Analysis of ET Products	We discuss performance differences among ET products regarding varying climatic conditions, landcover types, and retrieval methods throughout the manuscript and especially in sec. 4. However, we can certainly add more information about the practical implications and objective of this study in sec. 4. and 5. of the manuscript.

The manuscript reveals significant performance differences among ET products, but it does not fully address how these discrepancies might affect the conclusions drawn, particularly under specific land cover types or climatic conditions. A deeper analysis of how these differences impact the study's practical implications is crucial.	In order to provide additional deeper analyses as suggested by the reviewer, we want to suggest two potential analyses, which we could conduct: 1. Provide statistics (e.g., variance) for figures 10 & 11 of the manuscript, in order to analyze in more detail the impact of the drought year. 2. Provide power spectrum analyses for different time frames for every ET product (e.g., daily, bi-weekly, monthly) to analyze performance and biases of every product along time.
3. In-depth Analysis of Drought Year Effects The analysis of the 2018 drought is valuable, but the study does not sufficiently explore how different ET products capture the effects of drought. A more detailed comparison of how products perform under drought conditions is needed, including which models and parameterizations are more sensitive to such extremes.	We discussed the impacts of the drought year on ET products in sec. 4.2., e.g., lines 518-526 or lines 538-546. However, related to the comment above, we suggest to add statistics to the binning plots (Figs. 10 & 11) in order to provide deeper analyses of drought year effects.
1. Improvement of Remote Sensing and Ground Observation Matching There is an acknowledged mismatch between point-scale ICOS data and the coarser-resolution remote sensing products. The manuscript could benefit from a discussion on methods to address this issue, such as spatial downscaling or data fusion techniques, to improve the alignment of ground and remote sensing observations.	In the field of remote sensing, the spatial mismatch between satellite data and ground measurements is a well-known problem and somehow accepted assumption in order to be able to validate remote sensing products. In that sense, we agree on providing some discussion. We will add an explanation why we kept coarse-scale ET products at their original spatial resolution after line 174: 'The MODIS product with nominal spatial resolution of 500 m is aggregated to the 3 km footprint, while the SEVIRI, ERA5-land, GLDAS-2, and GLEAM products are maintained at their original spatial resolutions of 3 km, 9 km and 25 km, respectively. **Although there exist several downscaling methods and data fusion techniques for improving the spatial resolution of remote sensing products (Ha et al., 2013; Mahour et al., 2017; Peng et al., 2017), we decided to keep ET products with a spatial resolution lower than 3 km at their original resolution (i.e., GLDAS-2 and GLEAM at 25 km), since, for one, the intention of this study is a comparison of well-known and established ET products and not an optimization of rescaled comparisons. Second, because we did not want to include additional uncertainties which potentially originate from the employed downscaling method or auxiliary datasets. Every downscaling approach has its advantages and drawbacks as it intends to statistically correlate coarse-scale data and fine-scale auxiliaries or develop physically-based models to enhance the spatial resolution (Peng et al., 2017).**'

	Further, we will add a discussion addressing this comment after line 470: 'One reason for the mismatch between the ICOS product and SEVIRI, GLEAM and ERA5-land is surely the spatial mismatch between the point-scale ground-based EC tower measurements and the remote sensing (3 km) or reanalysis (9 km) products. To capture vegetation stress, ecosystem health, and fine-scale variability in ET globally, adequate spatial (and temporal) resolutions are necessary. **Here, detailed research regarding downscaling techniques as reviewed in, e.g. Mahour et al., 2017; Peng et al., 2017, that combine medium-scale ET data with fine-scale auxiliaries in order to improve the spatial resolution are needed regarding its uncertainties and impact on product comparisons**.' Lastly, we draw a conclusion regarding this topic after line 577: 'In summary, when considering all conducted analyses together (spatial and temporal resolutions, product dependencies, ETC results, SM and VPD controls on ET), the remote sensing products SEVIRI and GLEAM as well as reanalysis product ERA5-land seems to provide most reasonable results compared to all other ET products, with SEVIRI providing a higher temporal and spatial resolution compared to GLEAM and ERA5-land. **Hence, despite their coarse spatial resolution, GLEAM and ERA5-land are able to capture ET dynamics sufficiently even under drought conditions. Future research regarding data fusion techniques and downscaling approaches that combine coarse- or medium-scale ET data with fine-scale auxiliaries in order to improve the spatial resolution of certain ET products may help to decrease the spatial mismatch and optimize the comparison between point-scale field measurements and satellite remote sensing or modelling data.**' Even if we could largely extend the discussion on these methods, we prefer to limit it to the paragraphs written above since research regarding improved matching of remote sensing observations with ground measurements is out of the scope of this manuscript.
1. Considering extending the study period The study period appears to be too short, and the comparison between satellite-based ET and local measurements seems insufficient. The authors might consider extending the study period and incorporating a comparison across different timeframes. This would enable the use of the	We decided on purpose to analyze the years 2017 to 2020, for one, due to the availability of all datasets, and second, because we wanted to include detailed analyses regarding multiple ET products and varying climatic, landcover and site conditions. Further, we do not think that adding more years is necessary to draw conclusions on how well different ET products capture ET dynamics and drought events, as we already included varying climatic (one drought (2018) and one wet year (2017)) as well as landcover and site conditions (eight stations with different landcover and climate conditions as well as levels of landcover heterogeneity). Analyzes across different time frames and evaluation regarding the water balance principle is out of the scope of the manuscript, which

water balance principle for a more comprehensive evaluation.	intends to compare well-known and established ET products that are commonly used in climatic research studies.
Minor comments 1. At lines 34-35: please add explanation (why).	The explanation is found in the following sentence. We can rephrase the sentences to provide the explanation right away: 'The greatest deviations are found at the agricultural-managed sites Selhausen (Germany) and Bilos (France), with the former also showing the highest potential dependencies (error cross-correlation) between the ET products. **Hence, o**ur results indicate that ET products differ most at stations with spatio-temporal varying land cover conditions (varying crops over growing periods and between seasons).'
2. At line 34: please add the specific number to describe "the highest potential dependencies (error cross-correlation)".	Of course, we can add the specific number: 'The greatest deviations are found at the agricultural-managed sites Selhausen (Germany) and Bilos (France), with the former also showing the highest potential dependencies (error cross-correlation **(ECC)**) between the ET products **(up to 7.6 and outside the acceptable range of -0.5 < ECC < 0.5)**.'
3. At lines 36-38, please explain why.	The explanation is found in the following sentence. Due to the more complex heterogeneity of land covers at agricultural stations, the estimation of ET is more complex and challenging, while needle-leaved stations have less complexity due to rather stable vegetation conditions. We can rephrase the sentences to be more precise: 'This **is since** complex heterogeneity complicates the estimation of ET, while ET products agree **well** at evergreen needle-leaved stations with less temporal changes throughout the year and between years.'
4. Please add the longitude and latitude in Figure 1.	Done. We will add coordinates to the map in figure 1.
5. Please add a table that describes all the sites, including their longitude, latitude, land cover type, altitude, and other relevant details.	Done. We will add a table S1 in the supplement providing the coordinates, ecosystem type, altitude, climate zone and mean annual precipitation and air temperature for all stations. Details on percentages of different land cover types for every station are already given in supplement table S2 (formerly tab. S1).
6. At line 150: Please explain why you used the standardized precipitation-evapotranspiration index (SPEI).	Done. We can add some sentences to explain why we used the SPEI instead of SPI or other indices after line 153: '**We choose to use the SPEI to identify drought conditions instead of the standardized precipitation index (SPI) or other indices (i.e., Palmer drought severity index), since the SPEI additionally considers temporal changes in ET and hence, temperature, which showed to be evident for identifying abnormal (drought) conditions. Previous studies showed that not only the lack of precipitation defines drought events but also the level of temperature and consumption of rainfall by evaporation and/or transpiration (Vicente-Serrano et al., 2010).**'

7. If the time step is from hourly to 8-day, consider generated them at a shorter time step rather than monthly which loose too much information.	We agree that analyzing monthly time steps would be temporally too coarse. That is why in the manuscript, we are analyzing daily and 8-daily time steps, never monthly.
8. At lines 206-225: The spatial resolution differences among these ET products are quite significant. The authors should consider using methods to standardize all the ET products to a common resolution or use products with longer time periods. Otherwise, direct comparisons may not be valid or meaningful.	This comment is related to comments #3 and #4 above, where we provide our changes implemented in the manuscript to account for them. We agree that the change across spatial resolutions between different products are quite significant and we tried to account for it to certain degree as we upscaled high-scale ET products (e.g., MODIS at 500 m) to the baseline resolution of 3 km. And we decided to keep coarse-scale ET products (> 3 km, e.g., GLDAS-2 and GLEAM at 25 km) at their original resolutions instead of rescaling, since, for one, the downscaling from 25 km to 3 km would be a very big step and we did not want to include additional uncertainties from the employed rescaling method, and second, end users who would use only one ET product within their analyses alongside other variables would not rescale the ET product to one specific resolution and hence, loose the conclusions drawn from our analyses if we would compare all products at one rescaled resolution. We think the comparison is meaningful and valid as we compare well-known and employed ET products under varying climatic conditions and for multiple landcover classes at several stations all across Central Europe. And we were able to show that despite the coarse spatial resolution of GLEAM, the product is able to capture ET dynamics quite well compared to other products, which is an important conclusion for the scientific community and especially for studies that are based on the GLEAM ET product only.
9. Table 1: please change as three lines table format and added other dataset's information such as soil moisture and SPET etc.	Done, we included other products (VPD, SM, SPEI) in table 1. However, we did not change the format of the table as we like it this way.
10. At lines: 228-233 The removal of the seasonal signal may not be necessary, as the study period is too short for such adjustments.	We understand this point and agree that for longer time series, detrending has much more impact. We would like to keep the anomalies in order to compare detrended ET time series, e.g., figure 8 & 9, even if the time series considered is rather short compared to real climatological analyses.
11. At lines 295-296: Please provide the specific values for the highest R^2 and lowest RMSE and lowest percentage bias, PBIAS here.	We provide the detailed numbers for R^2, RMSE and PBIAS in the following sentences (lines 297-300). Since every station provides multiple statistics between all products (as shown in supplement figures S5-S7), we first wanted to highlight specific stations with outstanding statistics and then provide more details with actual numbers for the products. Otherwise, the sentences would get very long and complex, hence, we thought it might be easier for the reader to follow.

12. Figure 4. Please added some statices numbers in every panel.	Figure 4 shows the time series of all ET products and for all stations. Meaning, we are showing one single time series of every ET product and for every station individually. We are not sure what kind of statistics would make sense for single time series that would provide any useful information.
13. Same as Figure 4	We guess this comment is related to figure 5 and that we should add statistics in figure 5, same as in figure 4? Or to which figures is this comment related to? The statistics (correlation, error & percentage bias) for figure 5 as well as similar supplement figures S2-S4 are given in supplement figures S5-S7. Since the figures give already a lot of information, we provide the statistics in separate figures.