***Response to Reviewer #2***

***Summary and significance***

***<u>Reviewer:</u>*** *This manuscript fits well within the scope of Hydrology and Earth System Sciences. The authors introduce a new daily, 4km evapotranspiration (ET) dataset over continental US (CONUS) using the surface flux equilibrium (SFE) approach and compares it against other ET products (GLEAM, FluxCom, ERA5-Land). The authors present a careful statistical evaluation via triple collocation, giving random error and correlation to truth metrics. This manuscript is well written and conceptually clear. I particularly appreciate how the authors have put great effort and care in explaining how SFE compares to other ET estimation approaches and explains assumptions, strengths and limitations.*

*The demonstration that SFE has comparable performance to more complex approaches in many regions, particularly in the western US, is useful as it adds confidence in SFE as a practical alternative to estimate ET.*

*I recommend publication with minor revisions for clarity.*

**<u>Response:</u>** Thank you!

***Suggestions***

***<u>Reviewer:</u>*** *Figure 1: Clarify what panels b-g are showing by giving them a title and explicitly labeling the x-axis. The current x-axis was confusing, and I suggest writing out the month and year (e.g., Dec 2000).*

**<u>Response:</u>** We will relabel the x-axis, add a title ("example pixels") above the subplots, and edit the caption as follows, adding the bolded text:

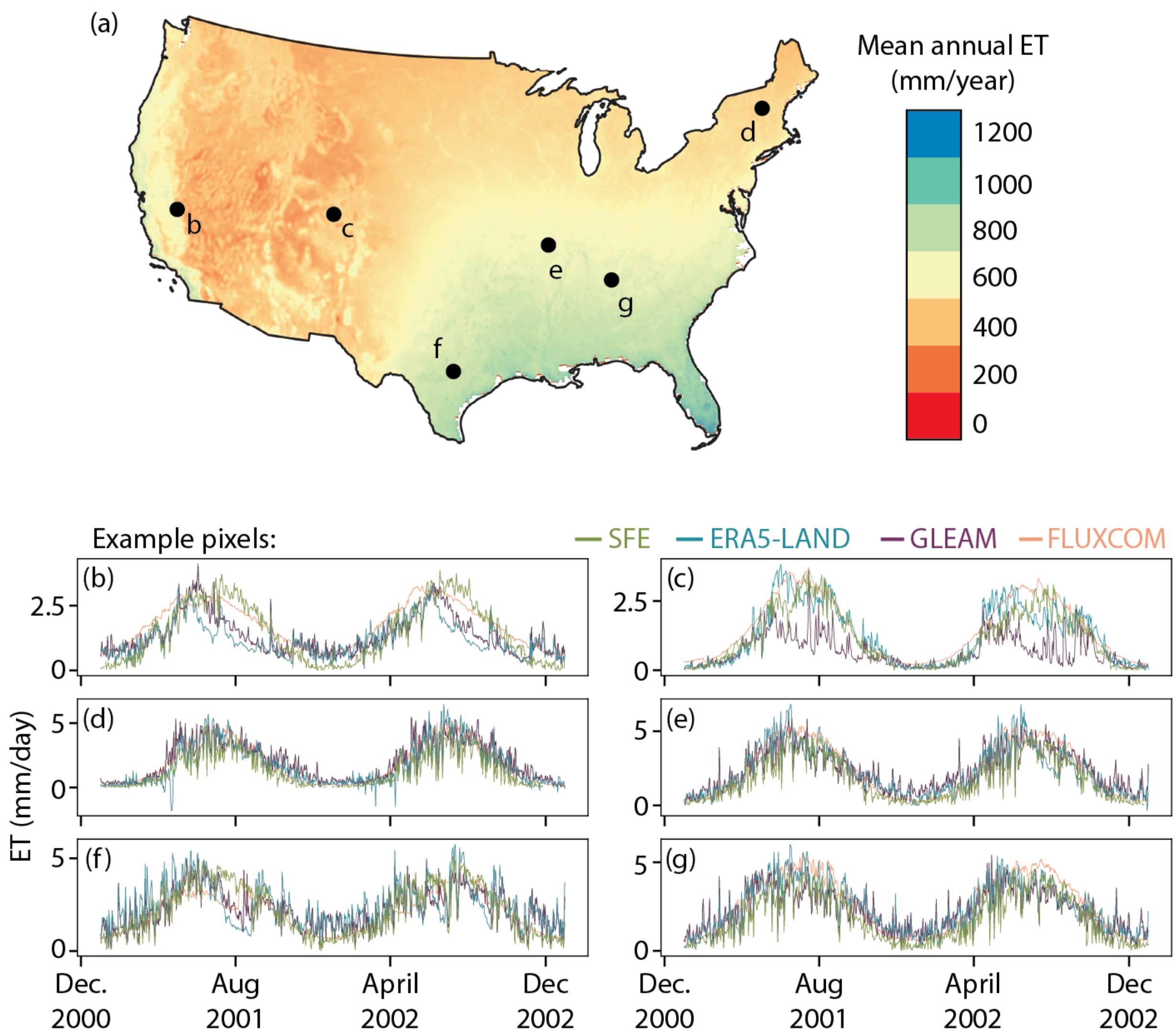


*Figure 1. Mean annual SFE ET across CONUS from 1979 to 2024. Points show timeseries for example pixels f**rom Dec. 2000 to Dec. 2002** for SFE (green), ERA5-Land (blue), GLEAM (purple), and FluxCom (pink).*

***Methods**:*

***Reviewer:*** *L141-144: I understand that the input data for SFE has been proven to be robust at the eddy covariance tower level (addressed in the introduction, Thakur et al., 2025). This may be extended when using gridMET and ERA5-Land data for this analysis, but can the authors directly tie that logic in Section 2.1? Can the authors address the biases of gridMET and ERA5-Land and how that may affect SFE ET?*

**Response:** We will add the following (changes in **bold**) to section 2.1:

> *"We choose gridMET **because it downscales output from the North American Land Data Assimilation System (NLDAS) with PRISM. This incorporation of statistically interpolated station data at a fine resolution helps gridMET achieve a high correlation with in situ stations, particularly for the variable of temperature, while maintaining** a relatively fine spatial resolution of 4 km at a daily timescale across CONUS **(Abatzoglou, 2013)**. Net radiation (Rn) allows conversion from the Bowen ratio to ET (Eq 2). We use Rn from ERA5-Land (Muñoz-Sabater et al., 2021) because of its high agreement with in situ measurements across CONUS (Yin et al., 2023). **However, we note that error in these input datasets will propagate to error in the resulting ET estimates."***

***Reviewer:*** *L145: Can the authors justify the 10% ground heat flux (G) assumption with a citation or provide a sensitivity analysis showing how varying G can affect σε and RT? The former is more reasonable to accomplish, but I would want to know the authors expect σε and RT to change if G is varied (e.g., 5-20%)*

**Response:** The ground heat flux can vary from around 10% of Rn to as much as 50% of Rn, depending on ground cover (Clothier et al., 1986, Santanello and Friedl, 2003). Previous SFE implementations have either neglected the ground heat flux entirely (e.g. Chen et al., 2021) or have calculated SFE with in situ data from FluxNet where estimates of the soil heat flux are available (e.g. Zhu et al., 2024).

We have performed a sensitivity analysis of SFE for values of G of 0%, 10%, 15%, and 20%. We find that, while the magnitude of daily ET is by definition impacted by the choice of G, the results from triple collocation (i.e. the error statistics of SFE) change very little. We will add the following two figures to the SI (and re-order the remaining SI figures) to show the change in mean annual ET across CONUS and the change in the mean $\sigma_\varepsilon$ and $R_T$ across triplets for various choices of G.

By definition, increasing the assumed percentage of net radiation that is partitioned to the ground heat flux reduces the magnitude of SFE ET (new Figure S1), which also reduces estimates of $\sigma_\varepsilon$. However, the choice of G has little impact on $R_T$ (new Figure S7).

These two figures will be referenced in section 3.1 and 3.2 with the following text:

> (Line 273) "*This spatial pattern exists regardless of the choice of parameter for the ground heat flux (G), although the magnitude of mean annual ET is altered (Figure S1).*"
>
> (Line 373) *"The triple collocation results are also relatively insensitive to the choice of the ground heat flux (G) parameter used in the calculation of SFE, although increases in G necessarily reduce ET estimates, and therefore also reduce $\sigma_\varepsilon$ (Figure S7). To the extent that uncertainty in G causes errors in the SFE ET estimate, it will also cause errors in estimates from other ET products, which must make similar assumptions or approximations for G."*

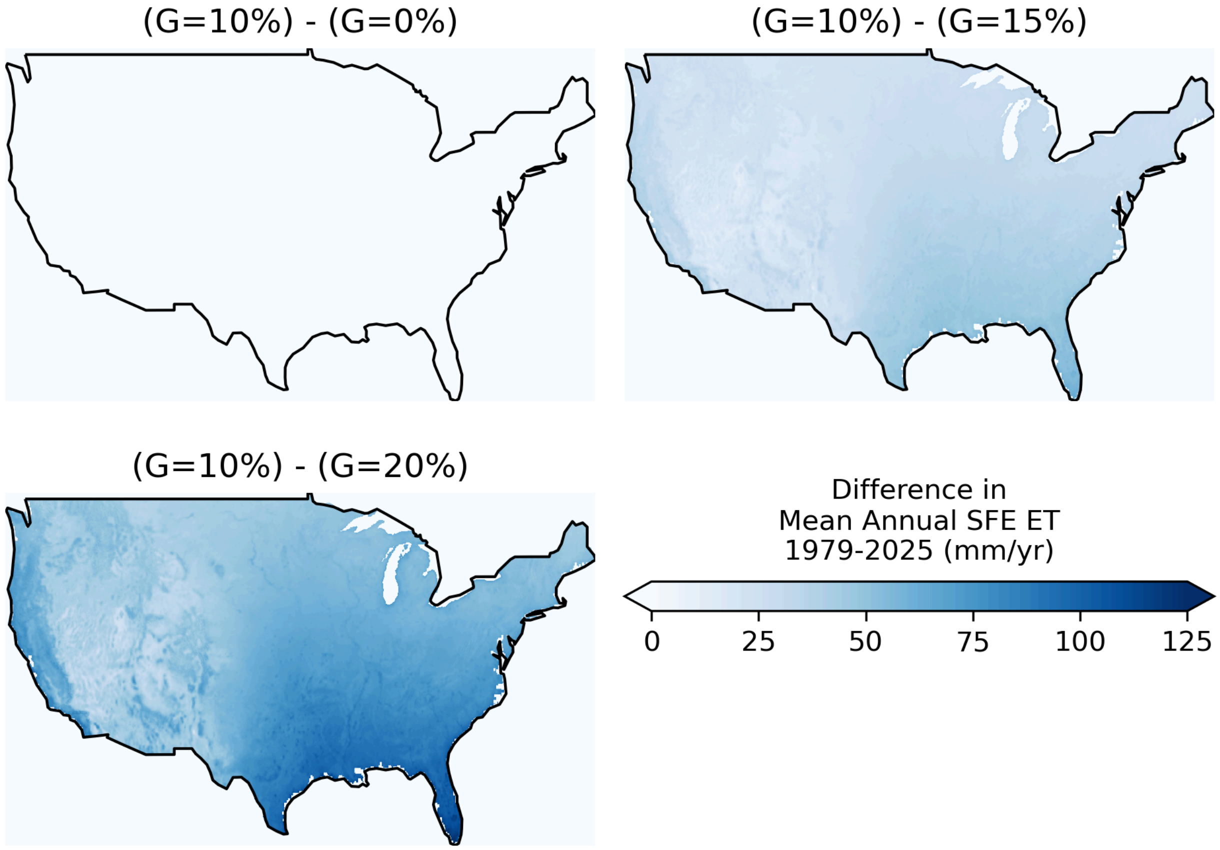


*Figure S1. The difference in mean annual SFE ET from 1979 to 2025 for different values of the ground heat flux (G).*

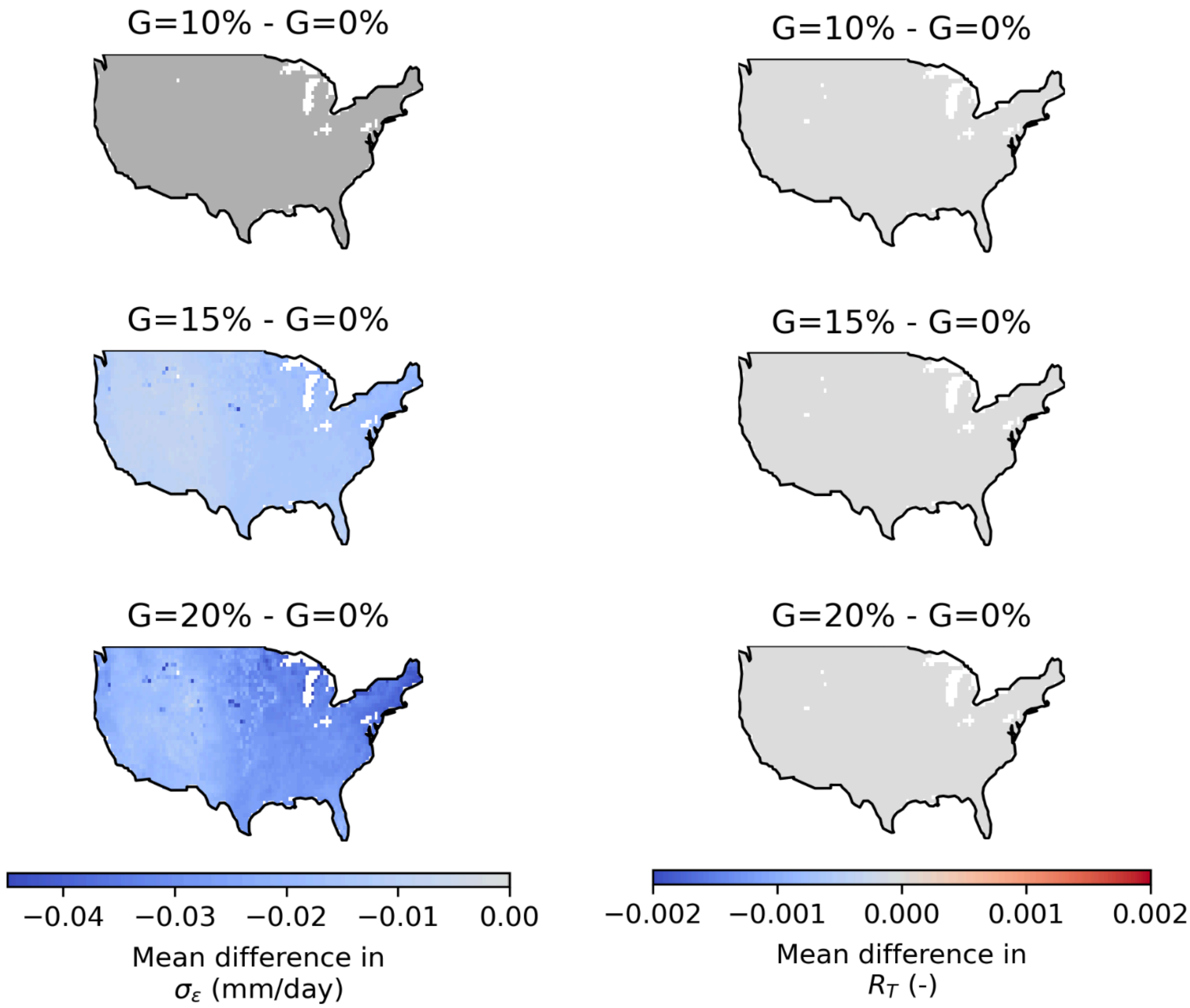


*Figure S5. The change in the standard deviation of the random error ($\sigma_{\varepsilon}$, left column) and the correlation coefficient ($R_T$, right column) averaged across all possible triplets for SFE calculated with different values of the ground heat flux (G), expressed as a percentage of total net radiation. Grey indicates no change.*

Additionally, while conducting this sensitivity analysis, we realized that the figures we show in the manuscript as submitted were actually calculated using a G of 0%, not a G of 10% as stated. We will update all of the figures and numerical results in the text to show results for G of 10%. Because the change in TC results is minimal (see Figure S5 above), changing the choice of G has no impact on the main findings of the paper. The change is visually detectable in Figure 1 (see new version in response above), Figure 2, and Figure 4 as well as in the specific pixel counts listed in Table 1. However, the main messages of these figures remain the same.

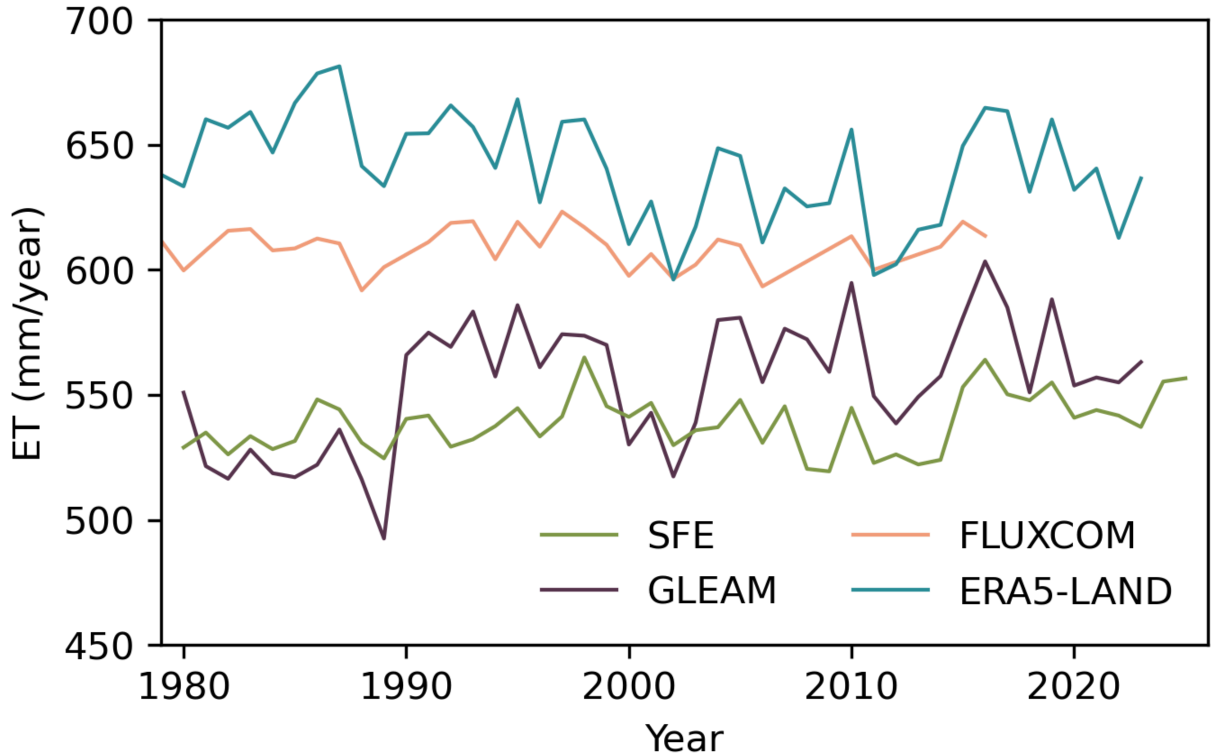


*Figure 2. Interannual variability in mean annual ET across CONUS from 1979 through the record length of each dataset.*

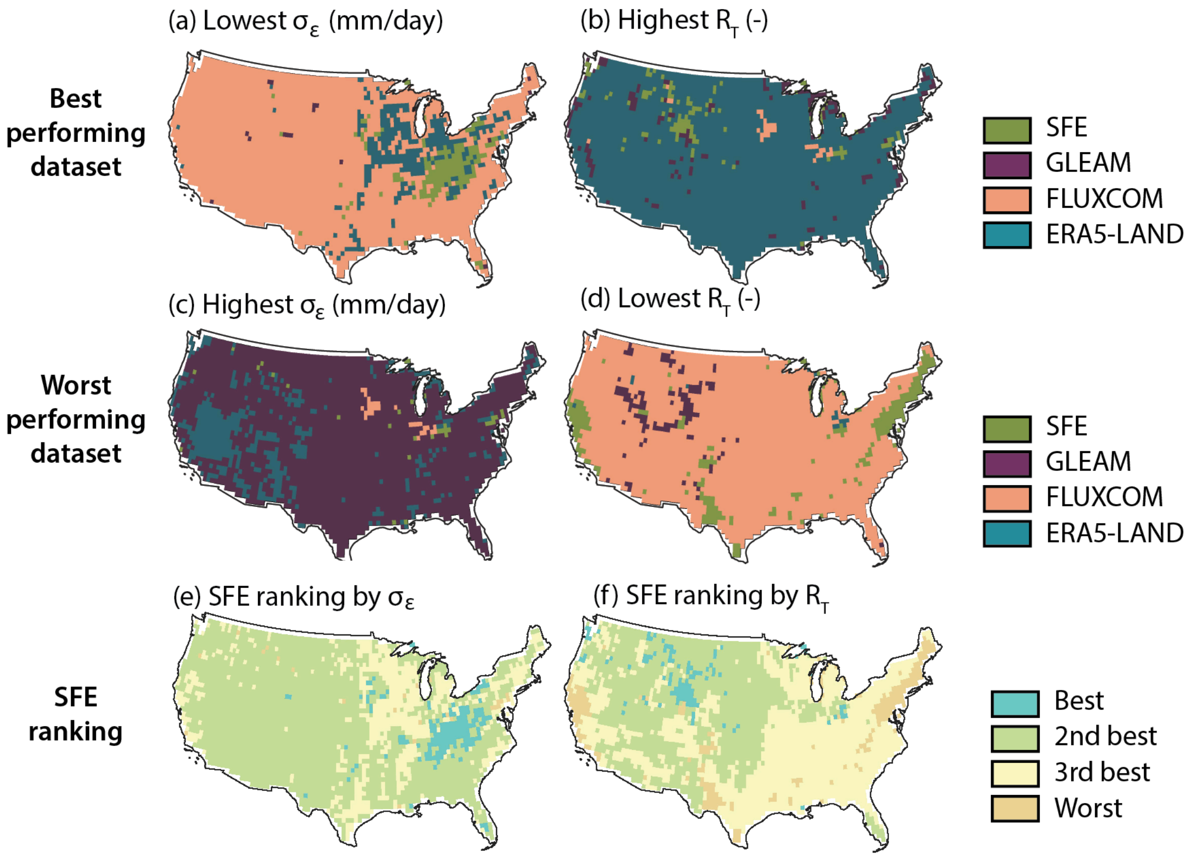


*Figure 4. Summary of relative performance of all four datasets. The dataset with highest performance for the standard deviation of the random error, $\sigma_\varepsilon$ (a) and the correlation coefficient with 'true' ET, $R_T$ (b) for each pixel. The worst performing datasets for $\sigma_\varepsilon$ (c) and $R_T$ (d). The relative ranking of SFE for $\sigma_\varepsilon$ (e) and $R_T$ (f). The total number of pixels (and relative percent of pixels) of each color are shown in Table S1. Pixels with centroids within 4 km of the border have been removed.*

*Table 1. (Top) The number of pixels where each dataset has the best performance according to the standard deviation of the random error, σε, and the correlation coefficient to the truth, $R_T$. (Bottom) The number of pixels by SFE ET ranking.*

| *Best dataset* | | | | |
|---|---|---|---|---|
| | *By $\sigma_\varepsilon$* | | *By $R_T$* | |
| | *Pixels* | *Percent* | *Pixels* | *Percent* |
| *SFE* | 164 | (5.4%) | 115 | (3.8%) |
| *GLEAM* | 17 | (0.6%) | 159 | (5.2%) |
| *FLUXCOM* | 2537 | (83.7%) | 33 | (1.1%) |
| *ERA5-Land* | 314 | (10.4%) | 2725 | (89.9%) |
| *Ranking of SFE* | | | | |
| | *By $\sigma_\varepsilon$* | | *By $R_T$* | |
| | *Pixels* | *Percent* | *Pixels* | *Percent* |
| *1st* | 111 | (3.7%) | 156 | (5.1%) |
| *2nd* | 1286 | (42.4%) | 2206 | (72.8%) |
| *3rd* | 1397 | (46.1%) | 646 | (21.3%) |
| *4th* | 238 | (7.8%) | 24 | (0.8%) |

References:

Clothier, B. E., Clawson, K. L., Pinter Jr, P. J., Moran, M. S., Reginato, R. J., & Jackson, R. D. (1986). Estimation of soil heat flux from net radiation during the growth of alfalfa. *Agricultural and forest meteorology*, 37(4), 319-329.

Purdy, A. J., Fisher, J. B., Goulden, M. L., & Famiglietti, J. S. (2016). Ground heat flux: An analytical review of 6 models evaluated at 88 sites and globally. *Journal of Geophysical Research: Biogeosciences*, 121(12), 3045-3059.

***Reviewer:*** *L145: Can the authors explain how including days with negative net radiation (Rn) can affect daily ET estimation and triple collocation statistics and justify their exclusion?*

**Response:** As stated in line 145, we do not evaluate SFE ET on any day with negative net radiation. We will further explain this by adding the following in **bold:**

> *"We assume a ground heat flux (G)* ***that is*** *10% of* $\boldsymbol{R_n}$***. Additionally,*** *we do not evaluate SFE ET on any days with negative* $R_n$ ***because doing so would result in a negative ET estimate, which is not physical.***"

We also thank the reviewer for helping us realize that we did not explicitly address how we deal with these no-data days in the triple collocation analysis. We will add the following text to Line 192 (section 2.3) to make it clear that we did not perform triple collocation on winter days, when net radiation is most likely to be negative:

> *"After removing the seasonal cycle, we choose only the months of March through October for the triple collocation analysis. This is because negative daily net radiation occurs for some pixels during the winter months, prohibiting the calculation of SFE. Because the number of days with negative net radiation varies for each pixel, we eliminate all winter months for all datasets to ensure a consistent number of data at each dataset and pixel."*

To remind the readers of this important detail, we will also reference it throughout the Results and Discussion sections, for example by adding the following text in **bold**:

> *Line 310 (Section 3.2): "SFE performance* ***during non-winter months*** *as estimated by triple collocation is comparable - and even exceeds - the performance of the comparison datasets across much of CONUS, despite its extreme simplicity, lack of tunable parameters, and relatively small number of assumptions (Figure 3)."*

> *Line 442 (Section 4.1): "While triple collocation reveals that SFE is rarely the highest performing dataset* ***for the non-winter months evaluated in this study****, it is the second-best performing dataset across much of CONUS for both* $\sigma_\varepsilon$ *and* $R_T$ *(Figure 4e,f)."*

***Discussion:***

***<u>Reviewer:</u>*** *I suggest adding a brief discussion about expected SFE performance outside CONUS and considerations for global implementation. How do the authors expect SFE to perform in regions with weaker land-atmosphere coupling (e.g., Southeast Asia)?*

**<u>Response:</u>** There is no reason to believe that SFE should not perform similarly at the global scale, particularly outside of regions with substantial influence from ocean dynamics (such as island or coastal regions). However, global implementation is dependent on input data quality (with corresponding choices for spatial scale, for example), which is why we chose to focus this analysis on CONUS.

With regards to your question about land-atmosphere coupling: If by 'land-atmosphere coupling' we mean the feedback of not just the land on the atmosphere, but also the atmosphere on the land, then the strength of this coupling should have no impact on SFE performance. SFE does not actually take advantage of land-atmosphere coupling, but rather relies on the fact that fluxes on the land surface *do* impact atmospheric conditions (regardless of 'coupling strength'). In other words, it is not necessary for atmospheric conditions to impact surface fluxes in order for the SFE method to work.

To reduce confusion about this, we will remove the phrase "In leveraging land-atmosphere coupling" from Line 556 in the Conclusion. We will also edit Lines 556-560 in the Conclusion (adding the text in **bold**) to reiterate the possibility of estimating ET beyond CONUS:

> *"**That** SFE estimates ET from atmospheric conditions alone **has several advantages: It can be calculated at a variety of scales and geographic domains and** it provides an opportunity to test hypotheses about vegetation response to environmental drivers without assuming that response a priori in the creation of the ET estimate itself."*

***<u>Reviewer:</u>*** *L572: The authors note that SFE bias in arid conditions needs further investigation. Can the authors please add specific recommendations for future investigation and/or what additional measurements may be needed (advocating for certain measurements?).*

**<u>Response:</u>** We will replace the sentence "*Additional investigation into this is necessary*" with "*Further in situ validation of SFE in arid ecosystems in particular would be beneficial*." (Line 578)