Responses to Reviewer 1:

It proposed a framework of estimating field scale ET using the GEE and DisALEXI model. Then produced 30 m ET were evaluated with flux tower observations over different land cover types. They were also compared with the existing dataset of OpenET and water balance ET. These datasets mostly performed similarly. So, it should show the advantage of this method.

We thank the reviewer for the helpful comments. We have added more information in the methodology and discussion sections to better describe the model and also highlighted the difference of DisALEXI comparing with other remote sensing-based ET models. Particularly, we have one section 4.1. in Discussion to focus on the uniqueness of GEE-DisALEXI model among OpenET models. One key difference is that DisALEXI explicitly solves the soil and canopy energy balance using land surface temperature as a primary constraint. In contrast, many other models either rely on empirical stress functions or treat the land surface as a single source, providing less direct information on soil and canopy fluxes.

1. Why were 11 different air temperature maps used? It is confused for selecting different air temperature maps. Who is the benchmark at 4-km, DisALEXI or ALEXI, why? They are close, does not mean they are reliable.

   We thank the reviewer for this important question. The benchmark at the 4 km scale is ALEXI, as DisALEXI is designed as a disaggregation framework that constrains fine-scale (30 m) ET estimates to be consistent with the coarse-scale ALEXI results.

   In the standard offline version of DisALEXI approach, air temperature is iteratively adjusted so that the aggregated 30 m ET matches the ALEXI ET at the 4 km scale. However, due to computational constraints within the Google Earth Engine environment, performing full iterative optimization is not efficient.

   To address this, we approximate the iterative process by generating 11 (this number doesn't have physical meaning, just making sure capturing the range of the air temperature) candidate air temperature fields and evaluating them at the pixel level. For each pixel, we select the air temperature that produces ET estimates most consistent with the corresponding ALEXI value at the coarse scale. This approach allows us to efficiently mimic the iterative solution while maintaining consistency with the ALEXI constraint.

   We have revised the manuscript to clarify this procedure, including the role of ALEXI as the benchmark and the rationale behind the use of multiple air temperature realizations. Please see “DisALEXI Module” section under 2.2.1 for the detailed changes.

2. How to fill Landsat daily ET when it is partly cloudy covered?

   We thank the reviewer for this important question. In the current implementation on Google Earth Engine, we do not perform explicit cloud gap-filling for Landsat overpass day ET estimates.

   We have developed an offline gap-filling approach (Yang et al., 2017) to address cloud contamination; however, the current algorithm is computationally intensive and not yet efficient for implementation within the GEE environment. We are working on the gap-filling algorithm on GEE and are preparing another manuscript for this.

   Following Volk et al. (2024), when computing monthly ET, all missing or masked overpass ET pixels are computed by linearly interpolating between the nearest unmasked, cloud free pixels in time window within ±32 days. We have added this information in “DisALEXI Module” section under 2.2.1 in the revised manuscript.

3. The comparison between the Landsat and flux tower daily ET showed that the relative error of Landsat daily ET could be more than 40 to 50%. This model needs to be refined at forest and shrub areas.

   We thank the reviewer for this important observation. Based on the evaluation, model performance in forest and wetland areas is relatively lower comparing with crop lands. In the manuscript, we discussed several potential factors contributing to the relatively larger errors in these land cover types, including high elevation variation of the forest sites, forest structure and the unique heat storage characteristics of wetlands.

   It is also important to note that these challenges are not unique to GEE-DisALEXI, but are common across many remote sensing-based ET models. We are actively working on improving model performance over these areas. We have further clarified these limitations and contributing factors in section 4.3 “Limitations and known issues” in the revised manuscript. We also want to note that this study establishes a baseline framework, which provides a foundation for future model development and refinement, particularly in more complex vegetation systems.

4. What about the terrains of over these flux tower areas? The comparisons are at point rather than regional.

We thank the reviewer for this important question. While flux tower measurements are often referred to as point observations, they in fact represent fluxes integrated over a source footprint, which typically extends over several hundred meters depending on atmospheric conditions and surface characteristics.

To better account for this spatial representativeness, we compared model results with flux tower observations using a window centered around each tower, rather than a single pixel. This approach is consistent with previous studies (e.g.,Volk et al., 2024) and helps reduce potential mismatch between tower observations and gridded model outputs.

Regarding terrain, terrain variability can influence both the flux tower footprint and remote sensing retrievals, particularly in mountains or topographically complex regions. We have revised the section 2.3.1 “Flux Tower Observations” to reflect the footprint of flux tower observations.

5. It cannot find these reductions in Table A1.

   Thanks for pointing this typo out. We have fixed it in the revised manuscript following the numbers in Table A1.

6. There is no OpenET in Figure 6

   We thank the reviewer for pointing this out. We meant to refer to Figure 7. We have corrected it in the revised manuscript.

7. Actually, the modeled monthly ET, both GEE-DisALEXI and ensemble, have significant differences with ground measurements at the forest sites. It need more details for these disagreement.

   Thank you for the comments. As noted in the manuscript, many of these sites are located in high-elevation regions with complex terrain and strong spatial heterogeneity, which can contribute to increased uncertainty in both model estimates and ground-based observations.

   We have further expanded the discussion in the revised manuscript to better explain the potential sources of these discrepancies and provide additional context on model performance over forested environments. These texts have been added in the discussion: “Forest and wetland represent another challenging environment for RS-based models, including GEE-DisALEXI. In forests, the relationship between radiometric surface temperature observed by satellites and the aerodynamic temperature governing turbulent heat exchange can become more complex due to canopy structure, multiple scattering, and within-canopy temperature gradients. Wetlands may exhibit unique energy balance characteristics associated with standing water and enhanced heat storage that are not fully represented in the current model framework. Future model updates improving representation of forest canopy structure and wetland energy balance processes are needed."

8. The DisALEXI ET overestimated the low values but underestimated the high values. But the ensemble ET does not have this issue. So what is the advantage of DisALEXI ET?

   We assume this comment refers to the watershed-scale validation shown in Figure 10. As discussed in the manuscript, the biases at the lower and higher ends are primarily associated with watersheds characterized by complex terrain and strong elevation variability, where model performance remains challenging.

   We have revised the section 3.3 to more clearly explain these patterns and their potential causes. At the same time, we have better articulated the advantages of the DisALEXI framework throughout the manuscript, including its physically based structure, ability to explicitly represent surface energy balance processes, and its capability for higher-resolution ET estimation.

   While ensemble approaches can reduce bias through model averaging, DisALEXI provides a process-based framework that is more directly linked to surface conditions and can be further improved through targeted model refinement. We have revised the introduction, methodology and discussion sections in the manuscript to better distinguish the strengths of the approach.

9. Why did the comparison between ET metrics and Drought ? They are different things, especially over the dryland area, lower ET doesn't mean drought.

   Thank you for bringing up this question. We agree that ET magnitude alone is not always a direct indicator of drought, particularly in dryland regions where water limitation is persistent.

   In this study, we used ET anomalies rather than absolute ET values to capture deviations from typical conditions. Because ET is closely linked to vegetation physiological processes, changes in ET can reflect vegetation stress and response to water availability. This relationship has been widely documented in previous studies (Anderson et al., 2013; Otkin et al., 2018; Yang et al., 2020).

We have revised the section 3.4. to more clearly explain the rationale for using ET anomalies as a drought indicator. The following text has been added at the beginning of section 3.4. to clarify the ET metric we are using for this analysis: "Because ET is closely linked to vegetation physiological processes, variations in ET can provide valuable information on vegetation water stress and response to water availability, making ET a useful indicator of drought conditions. However, ET is also strongly influenced by meteorological forcing. To minimize the effects of weather variability and better isolate vegetation response, drought indicators often use the ratio of actual ET to reference ET or its anomaly."

Reference

Anderson, M. C., Hain, C., Otkin, J., Zhan, X., Mo, K., Svoboda, M., Wardlow, B., and Pimstein, A.: An Intercomparison of Drought Indicators Based on Thermal Remote Sensing and NLDAS-2 Simulations with U.S. Drought Monitor Classifications, J. Hydrometeorol., 14, 1035–1056, https://doi.org/10.1175/JHM-D-12-0140.1, 2013.

Otkin, J., Svoboda, M., Hunt, E., Anderson, M., Hain, C. R., and Basara, J.: Flash droughts: A review and assessment of the challenges imposed by rapid onset droughts in the United States, Bull. Am. Meteorol. Soc., https://doi.org/10.1175/BAMS-D-17-0149.1, 2018.

Volk, J. M., Huntington, J. L., Melton, F. S., Allen, R., Anderson, M., Fisher, J. B., Kilic, A., Ruhoff, A., Senay, G. B., Minor, B., Morton, C., Ott, T., Johnson, L., Comini de Andrade, B., Carrara, W., Doherty, C. T., Dunkerly, C., Friedrichs, M., Guzman, A., Hain, C., Halverson, G., Kang, Y., Knipper, K., Laipelt, L., Ortega-Salazar, S., Pearson, C., Parrish, G. E. L., Purdy, A., ReVelle, P., Wang, T., and Yang, Y.: Assessing the accuracy of OpenET satellite-based evapotranspiration data to support water resource and land management applications, Nature Water, 2, https://doi.org/10.1038/s44221-023-00181-7, 2024.

Yang, Y., Anderson, M. C., Gao, F., Hain, C. R., Semmens, K. A., Kustas, W. P., Noormets, A., Wynne, R. H., Thomas, V. A., and Sun, G.: Daily Landsat-scale evapotranspiration estimation over a forested landscape in North Carolina, USA using multi-satellite data fusion, Hydrol. Earth Syst. Sci., 21, 1017–1037, https://doi.org/doi:10.5194/hess-21-1017-2017, 2017.

Yang, Y., Anderson, M., Gao, F., Hain, C., Noormets, A., Sun, G., Wynne, R., Thomas, V., and Sun, L.: Investigating impacts of drought and disturbance on evapotranspiration over a forested landscape in North Carolina, USA using high spatiotemporal resolution remotely sensed data, Remote Sens. Environ., 238, 111018, https://doi.org/https://doi.org/10.1016/j.rse.2018.12.017, 2020.

Responses to Reviewer 2:

This study presents a modeling tool based on Google Earth Engine that allows the well-published remote sensing-based evapotranspiration model, DisALEXI, to be efficiently applied at the continental US and beyond. The study examined the uncertainty of the tool by comparing eddyflux ET data measured at selected sites and ET derived from the watershed water balance method at the HUC8 scale. The authors claim that the modeling tool works well the study region and performs better the OpenET ensembled ET at multiple scales. The authors also suggest stress index derived for this study is comparable to U.S. Drought Monitor drought signal (especially in humid region) and thus can be used for monitoring drought impacts.

We thank the reviewer for the positive feedback and for recognizing the significance and potential impact of the proposed modeling framework!

I think the modeling tool developed by this study has a huge potential to provide high resolution ET (e.g., 30 m that is most useful for a few applications) products although currently the resolution is 4-5 km.

We thank the reviewer for highlighting the strong potential of our ET modeling framework. The model integrates two spatial scales: ALEXI at coarse scale (4-5 km) and DisALEXI at finer scale (30 m). ALEXI serves as the backbone of DisALEXI during the disaggregation process. In this manuscript, we focus on the implementation of 30-m DisALEXI on the cloud computing platform – Google Earth Engine. We have further refined the introduction and method sections to more clearly describe the multi-scale framework.

I have offered detailed comments (see attached) throughout the main text that needs authors to carefully address. Here are some highlights that I hope the authors can address before the work can be published.

We thank the reviewer for the detailed comments in the main text. We have carefully revised the manuscript accordingly. To facilitate the review process, we provide a point-by-point response to each comment below as minor comments and describe the corresponding revisions made in the manuscript.

1. Model descriptions are needed to give readers a high level overview about how the model works and how it differs from other RS based models.

We thank the reviewer for this helpful suggestion. We agree that a clearer high-level overview would improve readability. We have revised the section 2.1.1 “TSEB” in the manuscript to provide a more comprehensive description of the model framework and further clarified its key differences from other remote sensing–based ET models.

2. Alternative model validation at the watershed scale might be useful. Please note that not all HUC8 watersheds are gaged (no references are given about the source of streamflow and ET data) and therefore comparing ET at the watershed level is still model to model comparison. It would make a strong case if the authors compare modeled ET to derived ET from smaller gaging

referenced watersheds like what the following papers did for general hydrological model validations that use commonly used datasets published by USGS. The gaged watersheds offer more convincing ET values than the HUC8 (that are rather large and heterogenous). The lathe HUC8 can have issues of ET such as P

We thank the reviewer for this helpful suggestion and for pointing us to the relevant references. We agree that not all HUC8 watersheds are gaged and that watershed-scale comparisons may introduce uncertainty. In our analysis, we used USGS streamflow data for HUC8 watersheds with available gage stations to conduct the water balance validation.

We have revised the manuscript to clearly describe the data sources and methodology, including the use of USGS streamflow data, to avoid any confusion. The following texts have been added in section 2.3.4. "Water-balance ET": "The analysis was carried out over watersheds with available USGS streamflow observations. There are 1548 HUC8 watersheds over our study area, of which 1520 have stream flow data, leaving 28 without runoff data. After removing boundary watersheds, there are 1412 watersheds left."

Following the reviewer's suggestion, we have clarified the limitations associated with HUC8-scale validation in section 3.2 "WBET Evaluation of Spatial Patterns in GEE-DisALEXI ET" and have added the following texts: "At the low ET end, biases may be related to uncertainties in precipitation and streamflow data in water-limited basins, where small absolute errors can translate into larger relative differences. At the high ET end, the tendency toward underestimation may reflect limitations in representing complex and dense vegetation canopies. In addition, both models show higher biases in basins with high elevation variability and lower cropland proportions. While WBET provides an independent estimate of long-term basin scale ET, it is subject to uncertainties associated with precipitation, streamflow observations, groundwater exchanges, and inter-basin water transfers, which may also contribute to the biases in the scatterplot in Fig. 10."

Li, C., et al, 2020. Impacts of urbanization on watershed water balances across the conterminous United States. *Water Resources Research*, *56*(7), p.e2019WR026574.

Caldwell, P.V., et al. 2012. Impacts of impervious cover, water withdrawals, and climate change on river flows in the conterminous US. *Hydrology and Earth System Sciences*, *16*(8), pp.2839-2857.

Minor comments from the PDF file (all line numbers are from the original submission):

1. Line 41, how about canopy interception?

Thanks for pointing out this. Now we are using "the combined process of evaporation and transpiration", since the canopy intercepted water will return back the atmosphere through either evaporation or getting into soil and used by plant through transpiration.

2. Line 43, delete tremendous.

We placed it with “a large amount of ”.

3. Line 45, and water supplies?

We added “water supplies (Liu et al., 2025)”

4. Line 46, how about forests?

Thanks for this comment. We now replaced “agricultural water use (Gowda et al., 2008)” with “vegetation water use (Gowda et al., 2008; Yang et al., 2027)”

5. Line 53, “vegetation health” has been said earlier.

We revised it as “ET can also provide critical early warning signals regarding vegetation stress onset, which…”

6. Line 63 & 64, Isn’t PM method energy based method? PT model is a simplified version of PM model.

PM method is considered as energy balance and mass transfer-based ET model. The PT model can be viewed as a simplified form of PM under conditions where ET is primarily energy-limited and surface resistance is negligible. We added one sentence to further clarify this: “While Penman-Monteith and Priestley-Taylor-based methods are also physically based formulations, they are listed separately here because of their distinct representations of surface and atmospheric controls on ET.”

7. Line 135, Typo of LE

Thank you for catching the typo. It is corrected in the revised manuscript.

8. Line 136, how Trad is related to LE? Or H?

We have modified the paragraph to provide more information about Trad and the TSEB model. This paragraph now reads as: “Latent heat is calculated as a residual in the energy balance equation. A key input to TSEB is the land surface temperature (LST), which provides direct information on surface thermal conditions and serves as the primary constraint for partitioning available energy between sensible and latent heat fluxes. Because satellite observations measure a composite radiometric temperature from both soil and vegetation, the bulk directional surface radiometric temperature (T_{RAD}) is first partitioned between soil and canopy components using Eq. 2.”

9. Line 146, pls rewrite this sentence.

We have rewritten the sentence as: “TSEB uses an iterative adjustment procedure to account for canopy water stress. The model first assumes potential canopy transpiration and solves the energy balance. If the resulting soil latent heat flux is negative, canopy transpiration is iteratively reduced until soil latent heat flux is larger or equal to zero, reflecting the assumption that soil evaporation near midday should not be negative.”

10. Figure 1, 4-km and text in the figure to be consistent with the text

Thanks for the comment. We have corrected them.

11. Line 160, removing standard

Removed in the revised manuscript.

12. Line 163, TRAD format

We have corrected the format in the revised manuscript.

13. Line 166 and line 201, remove the duplicated citation

We have removed the duplicated citation.

14. line 213, replace “based on” with “of”

Thank you for the comment. We decided to keep “based on”, because the algorithm is not exactly the same as in Gao et al., 2012.

15. Line 282, remove “tower”

We added “eddy covariance” before tower to clarify it.

16. Line 285, replacing “gauge station” with “streamflow”

Replaced in the revised manuscript.

17. Figure 4. Not sure how important elevation is. I would think precip and PET are most important and deserve a fig to show the difference as a whole.

Precipitation and PET can provide important information about the climate zone condition for these sites. We have modified Figure 4 and added average precipitation from PRISM and average reference ET from Gridmet for different vegetation types.

18. Line 340, is the D0 need to add the multiply sign.

We added the multiply sign in the equation.

19. Line 346, Not all HUC8 watersheds are gaged. Explain how the streamflow is derived.

This is correct. We only used the HUC8 watersheds that have available streamflow observations. This information has been provided in the revised manuscript. The following texts have been added in the manuscript: “The analysis was carried out over watersheds with available USGS streamflow observations. There are 1548 HUC8 watersheds over our study area, of which 1520 have stream flow data, leaving 28 without runoff data. After removing boundary watersheds, there are 1412 watersheds left.”

20. Line 356, it is common Runoff/Precip >40% in forested areas.

We didn’t find many HUC8 watersheds were filtered out using this threshold. It is possible that the current study area doesn’t include the east coast region. We also modified the sentence in the manuscript to clarify the reason of using this filter: “Basins with annual run off larger than 40% precipitation were also removed from the analysis to reduce effects of regional groundwater flow.”

21. Line 370, do you set the intercept zero?

Yes, we have set the intercept as zero. We have also added clarification in section 3.2 in the revised manuscript.

22. Line 406, what is the unit of RMSE. Do you mean % error here

It is the normalized percentage of the weighted mean closed flux tower ET. We added this in the revised manuscript. To make this clear, we now use relative RMSE and relative MAE when referring to percentages in the manuscript.

23. Line 416, RMSE, MAE, MBE have unit of mm? What do you mean 79.3%?

Please see the response in 22. Thank you for pointing this out.

24. Line 420, add “comparing to eddy flux data”

We have added “comparing to eddy covariance flux observations”.

25. Line 449, What is closed ET?

This is referred to the energy balance closure corrected flux tower ET. We clarified this in the section 2.3.1. Flux Tower Observations in the revised manuscript. We also modified it as closed flux tower ET here at Line 449.

26. Line 478, what time frame?

The time frame is from 2016 to 2022. This information is provided at the beginning of section 3.2. “WBET Evaluation of Spatial Patterns in GEE-DisALEXI ET”.

27. Line 530, flux

We have replaced “flux” with “observed” in the revised manuscript.

28. Line 541, is the solar radiation measured variable?

This solar radiation we used in the ET interpolation is not from measured data. It is from the reanalysis data CFSR (Climate Forecast System Reanalysis). We have clarified this in the revised manuscript.

29. Line 612, replace “who” with “that”

We have replaced the word in the revised manuscript.

30. Line 630, also canopy interception

Thanks for pointing this out. We have removed the “soil” and “plant” to include the canopy interception and to be consistent with the introduction.