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 will add more information in introduction and discussion about the differences of DisALEXI comparing with other remote sensing based ET models and highlight the advantage of DisALEXI.

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 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 will revise 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.

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-based 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.

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 will clarify this 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. We agree that model performance in forest and shrub areas is relatively lower comparing with crop lands. In the manuscript, we discuss several potential factors contributing to the relatively larger errors in these land cover types, including high elevation variation of the forest sites.

   We are currently working on improving model performance over these areas. We will further clarify these limitations and contributing factors in the revised manuscript. We also 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 compare 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 point-scale observations and gridded model outputs.

We will clarify this approach in the revised manuscript and expand the discussion on how terrain and spatial heterogeneity may influence the comparison.

5. It cannot find these reductions in Table A1.

   Thanks for pointing this typo out. We will fix it in the revised manuscript.

6. There is no OpenET in Figure 6

   We thank the reviewer for this comment. The primary focus of this manuscript is to introduce the GEE-based implementation of the DisALEXI model and evaluate its performance. Accordingly, Figure 6 is designed to present the evaluation of DisALEXI independently.

   Comparisons with OpenET are included to provide additional context rather than as a central objective of the study. These comparisons are presented in Figure 7, where the OpenET ensemble is used as a reference to help interpret the performance of DisALEXI.

   We will revise the manuscript to clarify the roles of Figures 6 and 7 and better explain the rationale for separating these analyses.

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 will further expand 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.

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 will revise the Discussion section to more clearly explain these patterns and their potential causes. At the same time, we will better articulate the advantages of the DisALEXI framework, 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 will clarify these points in the revised 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 use 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 will revise the Introduction, Results, and Discussion sections to more clearly explain the rationale for using ET anomalies as a drought indicator, as well as to better clarify the limitations of this approach, particularly in dryland environments.

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.