To help the editor and the other reviewers better see both my comments and the perfunctory nature of the author' responses, I have had to copy my previous-round suggestions along with the author' replies and provide a further response. The author have rejected some of the reviewer's points, but I have provided specific evidence to refute their arguments. In addition to these issues, there are still errors in some of the figures, and parts of the text remain insufficiently rigorous. This is a manuscript that I have already reviewed through three rounds.

Reviewer comments: I previously pointed out there were some potential updates in next version, e.g. increasing ensemble member, updating the soil resistance, and canopy resistance. Unfortunately, the authors are reluctant to adopt my suggestions. They respond from budget and timeline. This is publication of scientific work, which should pass through a critical review. Otherwise, a product introduction document is enough. It is not necessary to publish on HESS.

Author Response: While we do not dispute the contribution of the studies mentioned by the reviewer in the previous review round to the science of modelling evapotranspiration, we also do not consider the inclusion of the specific methodological developments described in those publications as critical to the scientific significance of this manuscript at the current stage of this study. The focus of the study was not on developing or improving ET models but rather on evaluating the scientific quality of two initially selected methods (including input pre-processing and output post-processing) in the context of global and operational deployment. Algorithmic improvements and/or an extension of the approaches under consideration could be considered as a next step, following a thorough assessment of all the most promising options. Indeed, there are many studies looking at improving parts of the ET modelling process (e.g. the ET algorithms themselves, LST sharpening, ground heat flux estimation, net radiation estimation and improvement of other biophysical variables and parameterizations in general) but adding all those innovations to the initial version of a global and operational product is simply not feasible. Moreover, budget and timelines constraints are present in all scientific studies, and this is not a limitation specific to this study. This explains why the call for tenders on which this initial study is based (https://etendering.ted.europa.eu/cft/cft-display.html?cftId=13795) was limited to the two ET models proposed. We would therefore appreciate it if the reviewer could take into account the

context in which the study was conducted and consider an approach different from the one he might have adopted himself in a more conventional research project. We would also like to share with him our sense of pressure regarding his insistence on incorporating the approach he proposes. A pressure the reviewer may not have intended, but which would be contrary to the guidelines of the Committee on Publishing Ethics (https://publicationethics.org/guidance/guideline/ethical-guidelines-peer-reviewers/plain-text):

"Refrain from suggesting that authors include citations to your (or an associate's) work merely to increase citation counts or to enhance the visibility of your or your associate's work; suggestions must be based on valid academic or technological reasons."

Reviewer Re-response: I thank the authors for their detailed response, but I remain unconvinced by their reasoning, and I must reiterate my concerns.

First, my original suggestions—such as increasing ensemble members, updating soil and canopy resistance parameterizations—were not aimed at transforming the study into a model development paper. Rather, they were intended to enhance the scientific robustness and reproducibility of the evaluation itself. These are not peripheral "innovations"; they are standard good practices in evapotranspiration modeling, especially when claiming operational readiness.

Second, the authors repeatedly invoke budget and timeline constraints as reasons for not adopting these improvements. While I acknowledge that all studies operate under practical limitations, these constraints do not exempt a manuscript from scientific scrutiny. If a study is not designed to accommodate essential sensitivity checks or parameter updates that are widely recognized as influential, then the scope of the study may be too narrow for a journal like *HESS*, which expects methodological rigor and transparency over expediency.

Third, the authors' reference to the procurement tender (TED) as justification for limiting the study to two ET models is noted. However, the existence of contractual boundaries does not preclude the authors from acknowledging, in the discussion or conclusions, how their findings might change if the suggested improvements were implemented. A thoughtful limitations section would go a long way here—but the current response suggests resistance even to such acknowledgment.

Fourth, I take issue with the implication that my suggestions are driven by a desire for self-citation. My suggestions were based on well-established literature. The authors' reference to COPE guidelines appears defensive and misplaced. It is not unethical to recommend scientifically valid improvements, even if those improvements happen to appear in some of the reviewer's own publications—provided the recommendations are justified on scientific grounds, which they are.

Finally, I wish to address the broader issue of author–reviewer relations from the perspective of peer review ethics. Peer review is a voluntary, unpaid service that reviewers provide to the scientific community. I, like most reviewers, dedicate my time and expertise without any financial compensation, with the sole purpose of improving the quality of published research. When authors dismiss constructive, scientifically justified suggestions primarily on the grounds of "budget and timeline" — without engaging with the scientific substance or providing evidence that the suggestions would not affect their conclusions — they are, in effect, devaluing the free labor that reviewers contribute. If every author were to respond to critical feedback by simply citing project constraints, the peer review system would lose its purpose. The fact that reviewers serve without compensation does not obligate them to accept suboptimal work; rather, it places a greater moral responsibility on authors to take the review process seriously. I have invested my unpaid time to help improve this manuscript. I expect the authors to reciprocate that investment by engaging substantively with the recommendations, not by hiding behind administrative or contractual excuses

In summary, the authors seem unwilling to strengthen their study with relatively straightforward enhancements that would increase its value to the community. Instead, they hide behind project constraints and imply improper motive on my part. This is disappointing.

I therefore maintain my original position: unless the authors provide a clear, scientifically grounded explanation—supported by preliminary evidence or literature—of why the suggested modifications would *not* meaningfully affect their conclusions, or unless they agree to implement at least a subset of the recommended improvements (e.g., ensemble sensitivity or resistance parameter testing), I cannot recommend publication in *HESS* in its current form. A product introduction document may indeed suffice for the stated purpose, but that is not the standard expected by this journal.

Reviewer comments： Section 4.4 did not reflect my suggestions at all. The suggestions to soil resistance were not considered and discussed at all. Their excuses on non-available soil texture were not scientific reasoning.
Author Response： In the first round of reviews there was only one comment from the reviewer regarding Section 4.4 and that was to clarify if the temporal running mean for biophysical parameters "is planned for the initial operational release (end of 2025) or for a future reprocessing". We provided the requested clarification in the revised manuscript (L821-824) as also indicated in the response to reviewers. However, we have now added a sentence to Section 4.4 on adding additional ET models to the ensemble, including ETMonitor.
Reviewer Re-response： please follow my comments. When you have time to write these response, I understand you also have time to do sensitive analysis on the soil resistance. The sensitive analysis can be done at site scale or regional scale. I am not asking to redo the calculation at the global scale. The soil texture is available at some flux sites. I am quite sure about this.

Author Response： Regarding the soil resistance, in the first round of reviews the reviewer suggested to use equations 14-17 and Table 1 of Yuan, L. et al., 2024. Long-term monthly 0.05° terrestrial evapotranspiration dataset (1982–2018) for the Tibetan Plateau. Earth Syst. Sci. Data, 16(2): 775-801. Equation (14) requires two empirical parameters (a and b) the values of which vary based on soil texture and are listed in Table 1 but it is not described how they were derived, θsat which is estimated using equation (15) – (17) which themselves use other empirical parameters and constant values, and topsoil moisture which is obtained from GLEAM dataset with 0.25o spatial resolution and that contains data up to December 2024 and therefore is not suitable for near-real-time use.
Reviewer Re-response： a and b are derived from flux tower measurement over specific soil texture. Yuan et al. 2024 has written that "There are many parameters in Eqs. (7)–(17). Some parameters have already been assessed for their importance in ET estimation by Yuan et al. (2021)." Yuan et al .2021 has introduced how to obtain a and b in the Table. The author again create an obstacle for themselves that they are reluctant to overcome and pass the difficulty to others.
Yuan et al. 2024 use topsoil moisture from GLEAM, however there are other global soil moisture product, which are available to the author. How about the soil moisture from CLMS portfolio， and CAMS forecasts?

Author Response: Furthermore, neither TSEB nor in ETLook are "neglecting the influence of turbulent activity in the canopy" as stated by the reviewer. In both cases aerodynamic roughness is computed, in particular for tall canopies, based on the works of Raupach

(1994), considering not only the canopy (obstacle) height but also LAI (see https://github.com/un-fao/wapor-et-look/blob/13534beb778d6ed31074126d5c2e39bb00cc74a9/ETLook/roughness.py#L102-L107 for ETLook and https://github.com/hectornieto/pyTSEB/blob/master/pyTSEB/resistances.py#L124-L234 for TSEB-PT).

Reviewer Re-response： The authors believe that TSEB and ETLook consider the turbulent activity in the canopy. I copied the roughness calculation function below from the URL given by the author. It uses LAI, the turbulent activity in the canopy was not considered in its roughness estimation. The code clearly shows the facts:

d0 = z*(1-(1-exp(-sqrt(12*LAI)))/(sqrt(12*LAI))),

term1 = min(0.41**2/((log(z/(0.002*maximum obstacle height))+0.193)**2),1.0) + 0.35*LAI/2

z0m = exp(0.41/min(sqrt(term1), 0.3)- 0.193)

The original code is below:

```
    The roughness module contains all functions related to surface roughness
import math
def orographic_roughness(slope, dem_resolution=1000.0):
    r"""
    Computes the orographic roughness, which is a function of the slope of the
    terrain. Steeper slopes will increase the orographic roughness

    .. math ::
        z_{oro}=0.002\left(\frac{\left(x_{res}\frac{\Delta}{100}\right)}
        {x_{res}}^{2}\right)

    Parameters
    ----------
    slope : float
        slope
        :math:`\Delta`
        [rad]
    dem_resolution : float
        resolution of dem
        :math:`x_{res}`
        [m]

    Returns
    -------
    z_oro : float
        orographic roughness
        :math:`z_{oro}`
        [m]
    """
```

```
    slope_per = math.tan(slope)
    return 0.002*(((dem_resolution*slope_per)**2)/dem_resolution)
def roughness_length(lai, z_oro, z_obst, z_obst_max, land_mask=1):
    r"""
    Computes the surface roughness length. The roughness length is related to
    the roughness characteristics. For the logarithmic wind-profile the
    surface roughness length is the height at which the wind speed is zero.
    The roughness length is calculated differently for different types of land use
    Land use is classified as follows:
    0. no data
    1. land
    2. water
    3. urban
    .. math ::
        z_{0,m}=\begin{cases}
        \begin{array}{cc}
        0 & l=0\\
        z_{0,m} & l=1\\
        0.0001 & l=2\\
        \frac{1}{7}z_{obst,max}+z_{oro} & l=3
        \end{array}\end{cases}
    Parameters
    ----------
    lai : float
        leaf area index
        :math:`I_{lai}`
        [-]
    z_oro : float
        orographic roughness
        :math:`z_{oro}`
        [m]
    z_obst : float
        obstacle height
        :math:`z_{obst}`
        [m]
    z_obst_max : float
        maximum obstacle height
        :math:`z_{obst,max}`
        [m]
    land_mask : int
        land use classification
        :math:`l`
        [-]
    Returns
```

```
    -------
    z0m : float
        roughness length
        :math:`z_{0,m}`
        [m]
    Examples
    --------
    >>> import ETLook.roughness as roughness
    >>> roughness.roughness_length(0.4)
    0.34179999999999999
    """
    def veg_roughness(zo, d, zo_max, l, oro):
        veg = 0.193
        z_dif = zo-d
        term1 = min(0.41**2/((math.log(z_dif/(0.002*zo_max))+veg)**2), 1.0) + 0.35*l/2
        term2 = math.exp(0.41/min(math.sqrt(term1), 0.3)-veg)
        return z_dif/term2+oro

    # roughness length specific displacement height
    if lai == 0:
        disp = 0
    elif lai > 0:
        disp = z_obst*(1-(1-math.exp(-math.sqrt(12*lai)))/(math.sqrt(12*lai)))

    if land_mask == 0:
        z0m = 0
    elif land_mask == 1:
        z0m = veg_roughness(z_obst, disp, z_obst_max, lai, z_oro)
    elif land_mask == 2:
        z0m = 0.0001
    elif land_mask == 3:
        z0m = (1. / 7.) * z_obst_max + z_oro

    return z0m

def obstacle_height(ndvi, z_obst_max, ndvi_obs_min=0.25,
                    ndvi_obs_max=0.75, obs_fr=0.25):
    r"""
    Computes the obstacle height. The ndvi is used to limit the obstacle
    height.

    .. math ::
        z_{obst} =     \begin{cases}
        \begin{array}{cc}
```

```
    f_{obs}z_{obst,max} & I_{ndvi}\leq I_{ndvi,obs,min}\\
    z_{obst,max}\left(f_{obs}+\left(1-f_{obs}\right)\left
    (\frac{I_{ndvi}-I_{ndvi,obs,min}}
    {I_{ndvi,obs,max}-I_{ndvi,obs,min}}\right)\right) &
    I_{ndvi}>I_{ndvi,obs,min}\&I_{ndvi}<I_{ndvi,obs,max}\\
    z_{obst,max} & I_{ndvi}\geq I_{ndvi,obs,max}
    \end{array}\end{cases}

Parameters
----------
ndvi : float
    normalized difference vegetation index
    :math:`I_{ndvi}`
    [-]
ndvi_obs_min : float
    normalized difference vegetation index @ min obstacle height
    :math:`I_{ndvi,obs,min}`
    [-]
ndvi_obs_max : float
    normalized difference vegetation index @ max obstacle height
    :math:`I_{ndvi,obs,max}`
    [-]
obs_fr : float
    ratio of minimum and maximum obstacle height
    :math:`f_{obs}`
    [-]
z_obst_max : float
    maximum obstacle height
    :math`z_{obst,max}`
    [m]

Returns
-------
z_obst : float
    obstacle height
    :math:`z_{obst}`
    [m]

Examples
--------
>>> import ETLook.roughness as roughness
>>> roughness.obstacle_height(0.4, 2.0)
0.95
```

```
    """

    cond = [ndvi <= ndvi_obs_min, (ndvi > ndvi_obs_min) &
            (ndvi < ndvi_obs_max), ndvi >= ndvi_obs_max]

    def frac_func(n):
        return (obs_fr + (1-obs_fr)*(n-ndvi_obs_min)/
                (ndvi_obs_max-ndvi_obs_min))

    if ndvi <= ndvi_obs_min:
        obs_height = obs_fr*z_obst_max
    elif (ndvi > ndvi_obs_min) & (ndvi < ndvi_obs_max):
        obs_height = frac_func(ndvi)*z_obst_max
    elif ndvi >= ndvi_obs_max:
        obs_height = z_obst_max

    return obs_height

def displacement_height(lai, z_obst, land_mask=1, c1=1):
    r"""
    Computes the displacement height. The lai is used to limit the displacement
    height. It is defined differently for different types of landuse.

    Land use is classified as follows:

    0. no data
    1. land
    2. water
    3. urban

    .. math ::

        z_{disp}=\begin{cases}
        \begin{array}{cc}
        0 & l=0\\
        z_{obst}\left(1-\frac{1-\exp\left(-\sqrt{c_{1}l_{lai}}\right)}
        {\sqrt{c_{1}l_{lai}}}\right) & l=1\\
        0 & l=2\\
        \frac{2}{3}z_{obst} & l=3
        \end{array}\end{cases}

    Parameters
    ----------
    lai : float
```

```
    leaf area index
    :math:`l_{lai}`
    [-]
z_obst : float
    obstacle height
    :math:`z_{obst}`
    [m]
land_mask : int
    land use classification
    :math:`l`
    [-]
c1 : float
    exponential growth rate displacement height function
    :math:`c_1`
    [-]


Returns
-------
disp : float
    displacement height
    :math:`{disp}`
    [m]

Examples
--------
>>> import ETLook.roughness as roughness
>>> roughness.displacement_height(0.4, 2.0)
0.51779495

"""

def disp_func(l):
    return z_obst * (1-(1-math.exp(-math.sqrt(c1*l)))/math.sqrt(c1*l))

if (land_mask == 0) | (lai == 0):
    disp = 0
elif land_mask == 1:
    disp = disp_func(lai)
elif land_mask == 2:
    disp = 0
elif land_mask == 3:
    disp = (2./3.)*z_obst
```

return disp

Author Response: Furthermore, a closer look at Figure 12, shows that the systematic biases that the reviewer is claiming (and probably wrongly attributing to the calculation of aerodynamic resistances) is not that evident.

Reviewer Re-response: I copied Figure 12 in the blow. This figure generally shows that H lower-estimation and LE over-estimation. These are not systematic biases? You change the figure results without any explanation, then says reviewer`s claiming is not evident.

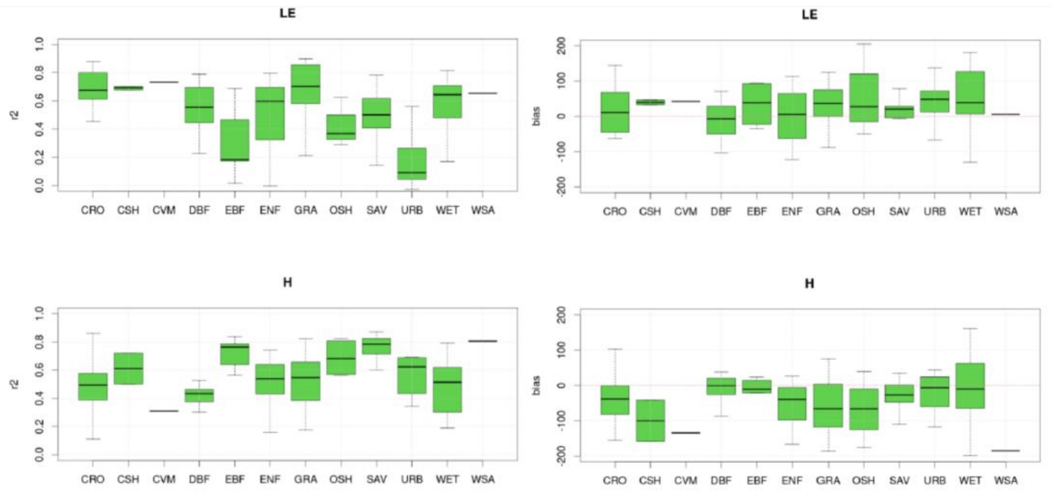

Figure 12. Boxplots of r^2 (left) and bias (right) values of LE and H at overpass time as modelled by the TSEB-PT model.

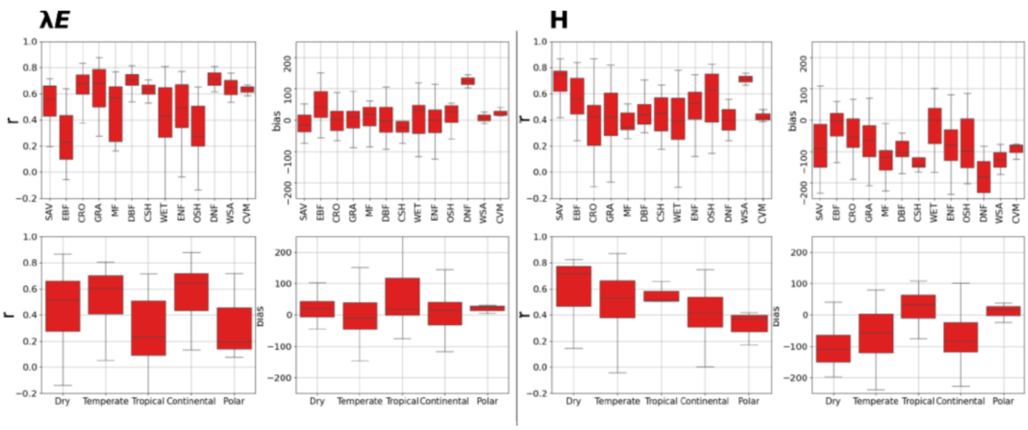

Figure 9. Boxplots of r (left panels) and bias (right panels) values of λE (top panels) and H (bottom panels) at overpass time as modelled by the TSEB-PT model. The horizontal line indicates the median value, the box limits show the first and third quartiles, while the whiskers give the minimum and maximum value.

Abstract says that CLMS ETa product consists of water flux (ETa, evaporation and transpiration) and energy flux (latent and sensible heat) sub-products, has a spatial resolution of 300 m, dekadal (10-day mean) temporal resolution for water fluxes and daily temporal resolution for instantaneous energy fluxes. This makes readers believe that CLMS ETa product include water flux and energy flux. Unfortunately, the energy flux was not evaluated at all. In addition, ETa accuracy was investigated, meanwhile, E and T was not assessed. There are many other ETa datasets, which has ETa component. To do E and T comparison among different datasets was not so difficulty. These evaluations are ***necessary*** for the publication.

Line 375, The resulting combination of 9 variables (called "DMS - WaPOR selected" in Section, which variables was combined? Please specify the names of the 9 variables. I did not find in the paper.

Line 658, In Figure 11 it can be observed that T is overestimated in the three sites located in arid conditions (AU-Whr, ES-LM1 and US-Var) which could point to problems with the root-zone SM estimation, whether the overestimation was caused by the problems of SM could be verified. Please use SM observation and SM estimation from CLMS method to check whether the T overestimation was associated with the bias in SM estimation. Please show the facts not by guessing.

Line 660, The evaporation estimated by ETLook tends to display both low variability and low values. This again could be linked to uncertainties in the estimation of top-soil moisture and to the use of constant parameterization (regardless of soil type) when calculating soil resistance (eq. (8)). “could be” comes again. The model can be quickly run at the site scale, forced with the meteorological data measured by the flux tower, to check ETa sensitivity to the SM and parameters used for the rs calculation. This method can help to answer whether the low values of E was associated to the soil moisture or parameters.

Figure 4, the coverage of the 9 regions shown in the global map include ocean. This paper only produced land ET product. The bar in the figure is result of land area in each region. The region definition is not accurate. The figure caption is not accurate. How can a dekad has 11 observations? The maximum of x-axis is 10 not 11. In addition, this figure only shows the first decade of July 2021, how is it representative? I suggest to give a result at least for one year.

The spatial map comparison between CLMS ETa and WaPor global ETa is necessary to demonstrate its performance on ETa spatial distribution. Evaluation at flux sites does not support its robust behavior at spatial representation. CLMS ETa can be upscaled to coarse resolution to compare with other global ETa product. Figure 10 only compare CLMS`s two members does not make sense. The two members use similar forcing datasets.

Figure 17, what is ETA? The left panel is SW irradiance, why ETA is set as the xlabel? The colorbar was not shown. Both ylabel of the two panels were “EC”, This figure is organized in a very confusing manner, requiring the reader to make considerable effort to infer its intended meaning.