Major comments

This scientific quality of this note is overall okay, and its contents are moderately interesting and marginally useful for the kind of research or context presented in the paper (micrometeorological or water balance research). It may be useful for irrigation planning or garden water use monitoring, but I am not at all convinced that it marks a “watershed moment in hydrology”.

My reservations relate to the fact that:

- generally instantaneous values of weather variables and ET estimates are less useful than hourly values integrated to e.g. daily values. No operator is going to stand in the field for 24/7 to take these measurements.

We fully agree with this point. In many practical applications, one would like an estimate of the daily ET sum. The conversion from instantaneous to daily values was however not our focus, since this is a topic that is already well covered in the literature on satellite estimation of ET. These methods are equally valid for our approach. We will acknowledge this in a revised manuscript by including a references to Jiang et al. (2021) in which several upscaling methods are compared. It should be noted that in contrast to satellite estimation of ET, which depends on available overpasses and cloud-free conditions, our method can be used for single or multiple measurements per day. More estimates at various points in the diurnal cycle will always lead to more robust estimates of daily totals.

- the IR phone images are only useful if one is interested in spatial variability of ET. Perhaps it can be useful for small-scale investigations of ET variability, e.g. in urban settings.

Investigation spatial variability of ET in specific settings including urban environments is indeed a logical application of the method. However we believe there is a wider potential because of the low-cost nature of the sensors. Ultimately, we see crowdsourcing of ET at larger (i.e. national) scales as a realistic future outlook.

- As I said, no operator is going to stand in the field for 24/7 to take these measurements. Why not buy a cheap weather station, e.g. and Ecowitt one, which is barely more expensive than the WEATHERmeter and supplement it with a cheap surface temperature sensor?

For local applications at smaller scales over which meteorological conditions are not expected to vary significantly, a cheap weather station could indeed provide an alternative to the WEATHERmeter. When supplemented by surface temperature sensor, the same methodology could in principle be applied to the resulting data. However this would still require post-processing on a separate computer. The advantage of using a smartphone as platform is that with a dedicated app, all the flux properties can be calculated, shown, and used on the fly.

- In fact, how well would the model have done with Ta only and not Ts? Do we really need Ts? Can this be tested and discussed?

Ultimately, both sensible and latent heat fluxes are to a large extent driven by the land-atmosphere temperature gradient. This is illustrated by Fig 2a and Fig 4. Fig 2a shows the relation between the temperature difference Ts-Ta and the sensible heat flux. Since most of the net radiation that is not used for sensible heat is used for ET, this already shows that the temperature difference, and not Ta, is the main driving force of ET. Fig 4 further investigates the importance of the different variables. As can be seen, both Ta and Ts carry significant information on ET, so using only Ta would not result in robust estimates. It should be noted that observations were made in humid (non-water-limited) conditions. Under water-limited conditions, the role of Ts will likely be more important. This will be tested in the future.

- This approach is only useful if there are high-quality ET data (e.g. EC-data) available to calibrate the ML model. For most places and users these are not available, and I am not sure how we would get around that. Unless we used cloud free high-resolution ET estimates to calibrate the ML models for specific settings, where lots of hobbyist weather station data are available.

Based on our results, it is not possible to conclude whether the approach will or will not work at other sites without calibration. At Rietholzbach, we were able to test both the validation (i.e. performance on data not used in the training) for the same site, as well as validation for the lysimeter when trained with eddy covariance data and vice versa. In both cases, we got more than satisfactory results which makes us confident that local training will not always be needed. Using satellite ET estimates could be a solution, but this would need to be investigated first. Local-scale validation of such products do not always show good results (e.g. Pardo et al., 2014, Cheng et al., 2021). In the future, we aim to train the algorithm using data from a range of sites with different climate and vegetation cover

- What about emissivity? Do the Ts data not need to be corrected for that? And the fact that the measurements are only measured between 8 and 14 micometer(?).

Good point. Because we are focussing on conditions of partial to full vegetation cover NDVI > 0.5, we assumed emissivity effects to be small with typical values close to 0.99. We will mention this in the revised version.

Specific comments

Line 24: “Traditionally, ET has been measured through the mass-balance principle applied to catchments or lysimeters”. This statement could be expanded a little to help the reader and needs some references.

We will add references to back up this statement, like the reviews by Senay et al. (2011) and Allen et al. (2011).

In fact the whole paragraph between line 24-32 is devoid of any references. This needs fixing.

We will add more references.

LIne 29: Replace “thermal infrared window”, by “thermal infrared atmospheric window of the electromagnetic spectrum”

Thanks for the suggestion, will replace.

Line 58: Say: “model CAT S62 Pro; referred to as S62 from hereon)”

Good suggestion, will adopt.

Line 59/60: Explain better what is mean by saturation here, and what causes this? Why these 2 layers of paper, this sounds rather arbitrary.

A standard light sensor on a smartphone will oversaturate when pointed directly at the sun (i.e. sensor output reaches it maximum possible value). By using a filter such as a small piece of paper, the whole dynamic range of the sensor can be utilized after re-calibration. This procedure is similar to the one proposed by Hukseflux for the Pyranometer App. We will describe this more clearly.

Line 61: Also, what is meant by “phone held straight-up perpendicular to the sun”. You mean that the phone is held vertically? Refer to Fig.1a here? Also, why are you taking a photo of obstacles sitting on the surface, rather than of the actual land surface? I find this confusing. Surely, this angle is only suitable/crucial for the operation of the light sensor, not for the IR image?

Correct, we only do this for the light sensor because a standard smartphone lens leads to angle-dependency of the measured light intensity. This intensity is subsequently corrected for a horizontal surface by using the pitch angle as recorded by the smartphone. We found this procedure to work well (Fig 2a). We will describe this more clearly in the revised version.

Line 64 & 65: use subscripts for Ta and ws (and for Ts in line 62). You use them later in the equation.

Thanks for noticing, will change.

Line 70-71: You say “partitioning of incoming solar radiation into evapotranspiration and sensible heat”. It is net radiation that is partitioned into evapotranspiration and sensible heat, but also into soil heat flux. So, this statement is incorrect. Also “incoming solar radiation” is the same as “global radiation”. Do you want to stick with one term? The latter one is less intuitive.

Good point, this was indeed formulated a bit sloppy. We will use global radiation in a revised version.

Line 71: “Both can be measured by phone’s internal sensors..” What does “Both refer to here?

"Both" refers to the incoming radiation and the surface temperature. We will rephrase the sentence to "Both the incoming radiation and the surface temperature can be measured by the phone's internal sensor."

Line 80-81: Can a little bit more information be given here? “The smartphone and Büel observations are available from Teuling and Lammers (2023)”. How many measurements/ IR images were taken in the field, and of what kind of surface? Only in the footprint of the EC mast or ‘on the lysimeter’?

We will add more information on the measurements and the characteristics of the site.

Line 89: This equation needs a number. Also, the various alphas are not defined properly? Nor is parameter c? Why is pressure not considered? If that is the case ,then take it out of the rest of the paper.

Thanks for the suggestions. We will describe the equation better, and remove pressure from the manuscript.

Line 135-136: What is meant with “.. magnitude of the offset term”. Is this the parameter c in the equation?

Correct. Will describe this more clearly.

Technical corrections

Line 39: It should be “Hukseflux” not “Hukselfux”.
Line 85: It should be “...a lack **of** energy balance closure”
Line 130: ‘it should be **negligible** role’.
Line 132: It should be “these condition**s**”

Thanks for spotting these typos. Will be corrected.

References:

Jiang, L., Zhang, B., Han, S., Chen, H. & Wei, Z. (2021), Upscaling evapotranspiration from the instantaneous to the daily time scale: Assessing six methods including an optimized coefficient based on worldwide eddy covariance flux network. Journal of Hydrology, 596, 126135.

Pardo, N., Sánchez, M.L., Timmermans, J., Su, Z., Perez, I.A. & García, M.A. (2014), SEBS validation in a Spanish rotating crop. *Agricultural and forest meteorology*, **195**, 132–142.

Cheng, M., Jiao, X., Li, B., Yu, X., Shao, M. & Jin, X. (2021), Long time series of daily evapotranspiration in China based on the SEBAL model and multisource images and validation. *Earth Syst. Sci. Data*, **13**, 3995–4017, https://doi.org/10.5194/essd-13-3995-2021.

Senay, G.B., Leake, S., Nagler, P.L., Artan, G., Dickinson, J., Cordova, J.T. & Glenn, E.P. (2011), Estimating basin scale evapotranspiration (ET) by water balance and remote sensing methods. *Hydrol. Process*., **25**, 4037–4049. https://doi.org/10.1002/hyp.8379.

Allen, R.G., Pereira, L.S., Howell, T.A. & Jensen, M.E. (2011), Evapotranspiration information reporting: I. Factors governing measurement accuracy. *Agricultural Water Management*, **98**(6), 899–920.