Response to referee comment Anonymous Referee #1

We appreciate and would like to thank Anonymous Referee #1 for reading our manuscript thoroughly and for raising valuable points of feedback. This will be helpful to improve the manuscript and was used to identify and correct a mistake that we have found in one of the scripts. Please, find below our point-to-point response (comment of the referee in **black**, our response in **blue**).

Review of Jansen et al., Evaporation from a large lowland reservoir—Observed dynamics during a warm summer.

This manuscript is well-organized and generally well-written. The topic is of interest to HESSD readers. The gist of the study is to use eddy covariance sensible and latent heat fluxes from two towers located on a large shallow lake in the Netherlands, to determine which underlying variables drive shallow lake evaporation.

The abstract describes the work pretty well. One thing that caught my attention was the assertion that the Penman equation disagrees with the study results. After reading the rest of the paper, I think the authors are saying that the net radiation Rn and the ground heat flux G terms of Penman are extremely difficult to interpret in the context of a shallow lake. Deriving Rn-G directly from global radiation does not account for the heat capacity within the water column.

Thank you for raising this point. Please find our response to this general point at point 3 of the specific comments.

The experiment is well-designed. The flux footprint analysis is helpful in determining which eddy covariance flux data correspond entirely to lake evaporation and which time periods are contaminated with land fluxes as well. The figures are well-designed. I found Figure 5 (the Venn diagrams) especially informative.

Thank you for this positive feedback on these points!

Specific comments.

1. Page 6 line 18 says that the data were collected during the summer of 2019 and 2020. Section 3.1 describes the large fraction of time that lake fluxes were unavailable or of limited quality. My impression is that these statistics refer to 30-minute averaging periods. If I understand this correctly, it is not clear how daily, monthly, and yearly data were obtained from these data. How were gaps filled? In particular, how would long periods of time with flux footprints falling outside the lake be handled? While the paper is generally clearly written, more attention is needed to specify the time-scales for the various figures, tables, and in-text statistics. Finally on this point,

   We are sincerely thankful for your remark on how we have dealt with the low number of remaining data after filtering (on quality, wind direction). Your remark actually made us go back to the scripts, where we found a small mistake in one of the programming commands which let all daily averages be calculated even if just one hour would be present. This of course was not our intention. We now correctly calculate daily averages only if for at least 66% of the hours (two-third) in a day valid data are available. Of course this leads to less daily data being available and therefore we had to redo some of the analyses, in particular the regression analyses. At a daily timescale the regression model is based on only a few data points (in the order of 12), but it is still found to be statistically significant (p-values < 0.01). The conclusion that in most cases a combination of wind speed and vertical gradient of vapour pressure forms the driving force of LE did not change. Finding the mistake thanks to the comment of the reviewer shows us that the review process is really helpful and needed. A good reminder and learning moment to me and us.

We intentionally did not filled large gaps (>1 hour) in the data. This would create synthetic results interfering with our aim to perform a process-oriented study into the role of evaporation in the surface-atmosphere coupling. We only performed linear interpolation to the data when gaps of at the most one hour occurred. We will add the following section to the manuscript at P7.L14: "*Hourly data was obtained by aggregating the half-hourly flux dataset, where at least one value per hour should be available. Within this hourly dataset large gaps (>1 hour) were intentionally not gap filled, because this would create synthetic results interfering with our aim to perform a process-oriented study. Only linear interpolation was performed to the data when gaps of at the most one hour occurred. Daily averages were calculated from this hourly dataset only if for at least 66% of the hours (two-third) in a day valid data are available*."

In the regression analysis we have only used hourly and daily data. Figures 4 and 9 show monthly averages based on hourly data. The uncertainty bars provide an indication of the ratio between the standard deviation of the hourly observations and the square root of the number of hourly observations taken into account in the calculation of the monthly mean, as a measure of the sampling uncertainty of the estimated mean.

2. When I read that Penman's equation does not work for these data, I was surprised. I thought the authors were claiming that they physics behind the equation were incorrect. But I think the authors agree with the physics of the Penman equation. They simply cannot determine the available energy with any reasonable certainty for this water body. Do I understand this correctly? Read our response at point 3.
3. Maybe this is just a re-wording of the previous comment, but the energy input, heating/cooling of the lake air stability, etc. do determine the fluxes. The temperatures and humidities at various heights and depths adjust according the these principles. The point of the paper is to determine which easily-measured variables give the best access to the fluxes.

   Thank you for your remarks on the disagreement of our results in relation to the Penman equation. In response to your general and specific comments that target the use of Penman's equation we would like to argue the following:
   Lake IJssel is a shallow lake of several metres deep and therefore most probably the heat capacity could indeed not be neglected. Penman developed his theory originally for land surfaces and (infinitely) shallow water surfaces. Therefore, he could make the assumption that the net radiation was divided over the turbulent fluxes and the ground heat flux exactly at the surface-atmosphere interface. This then determines the surface temperature and surface humidity, and eliminates surface temperature from the equation (P3.L2-3). However, in a water body part of the radiation penetrates the water, delivering relatively large amounts of energy below the air-water interface directly. This affects the physical processes

in a way that cannot be described with the PM equation. The large heat capacity of a water body provides the system with a “memory”. As a result, the water temperature at the surface is not directly related to the instantaneous energy balance at the surface, which is how Penman’s equation can be interpreted, but rather it is subject to a delay following the large heat capacity of the water body (P20+P21, L10-11+L1). Using a simple energy balance model of a water layer could help to solve the water temperature for the next time step (Equation 11 from Keijman (1974), and eq. 10 from De Bruin (1981)). However, in our analysis we have not measured *G* based on temperature changes integrated over the volume of the water column. In other words, we have omitted the downward heat flux *G* in the calculation of *Q** to adhere to the original Penman theory (P9, L 17-19). All the results of the Penman model presented in this manuscript are therefore based on the original theory of Penman for land surfaces and (infinitely) shallow water surfaces where *G* is neglected. We think that the Penman model does not present the full story in case of a water body of several meters deep.

We will adjust P20.L10-P21.L1 as follows: “*The large heat capacity of a water body provides the system with a “memory”. As a result, the water temperature at the surface is not directly related to the instantaneous energy balance at the surface, which is how Penman’s equation can be interpreted, but rather it is subject to a delay following the large heat capacity of the water body. We think that the Penman model does not present the full story in case of a water body of several meters deep.* ”

4. In figure 5 in particular, are the diagrams only for the summer months?

   Thank you for this remark. Yes, this diagram is based on only the summer months. In the revised version of the manuscript this will be indicated in the caption of this figure, as well as for other figures where this applies.

5. Page 16 lines 4-6. Typically, if measurements have a restricted range of variability, this results in a smaller R2 value, because random fluctuations are large relative to the observed changes. It looks like the authors are claiming the opposite effect here. Please explain.

   R2 values are determined by the variances of both data series, and the covariance between those series. So we agree that our explanation: ‘higher variance will lead to lower R2 values’, is not covering the whole story. We agree with the reviewer that indeed measurements with a restricted range of variability would result in smaller R2 values given the measurement uncertainties, which are not correlated.

   In an attempt to understand why sometimes the R2 values are higher for the validation period and to check if we are dealing with a ‘bad’ model, we have merged all the data of both summers (2019 and 2020), and subsequently we have split the data randomly in two datasets (each 50% of the data to simulate two summers). The coefficients for the regression model, that has the same form as we originally used, were re-calculated. We now find that in case of hourly data for Stavoren the calibration results in R2=0.8, and validation using the test data results in R2=0.72. Similarly if we study the hourly routine data for Trintelhaven, we find $R2_{cal}$ = 0.48, and $R2_{val}$ = 0.35. This gives us confidence that the original model we have created is not ‘bad’.

6. In Figure 5 d, the central intersection has an R2 value smaller than some of the other intersections. How could adding a variable explain LESS variability than simply not including it?

   This is because the numbers that are depicted in the Venn diagrams indicate the adjusted R2 to study the fit of the model as is mentioned in the caption of the figure. The adjusted R2 takes into account the degrees of freedom and can therefore lead to a decrease in adjusted R2 if the added variable only slightly correlates with the dependent variable.

   We will add the following to P13.L11: "... *the colour is. Adding a variable will not always result in a higher adjusted R2 value, because the adjusted R2 takes into account the degrees of freedom. Therefore, it can lead to a decrease in adjusted R2 if the added variable only slightly correlates with the dependent variable. Venn diagrams (a) and (c) .....*"

7. Throughout the paper I kept wondering why lake water stability was not included. Surely there must be seasonal changes in stability and thus of mixing depth within the lake. Did this have an impact on the data?

   We agree that lake water stability would be interesting to consider since it affects the surface temperature and therefore evaporation rates through the vapour pressure gradient. Evaporation in turn has a cooling effect on the surface temperature (which increases potential mixing). However, we unfortunately did not have the opportunity to measure water temperatures at several depths to study this. A preliminary study performed by one of our master students simulated mixing depths using the model FLake, which showed that in 70% of the time Lake IJssel is fully mixed. During summer it is of course more likely that stable conditions occur and we cannot directly assume fully mixed conditions. However, we considered this phenomenon beyond the scope of the current study.

   We suggest to add the following to P19.L7: "*Another phenomenon that could affect the yearly cycle of evaporation is lake water stability and thus mixing depth within the lake. Seasonal changes in lake water stability affects the surface temperature and therefore evaporation rates through the vapour pressure gradient. Evaporation in turn has a cooling effect on the surface temperature, which increases potential mixing. Supported by a preliminary study where mixing depths were simulated using the model FLake (Voskamp, 2018), we assume that during 70% of the time Lake IJssel is fully mixed. This number is not surprising given the fact that during night time evaporation continues, and with wind speeds that are on average 5.8 m.s^{-1}. In addition, the inflow of the river IJssel into the lake is likely to support mixing as well. During summer it is more likely that stable conditions occur and we cannot directly assume fully mixed conditions. However, we considered this phenomenon beyond the scope of the current study.*"

   Voskamp, T.: The evaporation of Lake IJssel - Comparison of the FLake model with standard methods at multiple timescales for estimating evaporation rates, MSc thesis report, Wageningen University, Wageningen, the Netherlands, 2018.