Referee 1

'This is definitely an important study. I agree that Land Surface Model could be improved in the definition of soil depth. The impact of rooting depth in determining the active soil moisture and the amount of water that can transpire ultimately influence the occurrence and amplification of heat waves, an increasingly pressing issue in an era of climatic changes. The model development proposed in this study could improve climate services that are primary forms of climate adaptation in many sectors.

The manuscript is generally well-written and contains high-quality research. However, I was not familiar with the method used in this manuscript. This is reflected in some questions and comments that the author could consider.'

We would like to thank the referee for the comments. We appreciate the time and effort taken to read our manuscript in detail and to provide us the very useful and interesting thoughts on our research. We will take the comments into account in revising the manuscript.

We have separated the different comments (shown in *italic*) and have written our replies below. Text in the original manuscript is shown in *'italic'* and revised text in '***bold***'. Wherever line numbers are mentioned in our reply they refer to the original manuscript version.

Comment 1.1

'The Memory Method is definitely very interesting. I believe it is reasonable enough to assume that local vegetation adapts the rooting depth according to the drought frequency. However, it's difficult to me to understand the way it is implemented in the model, i.e. varying the total soil depth in the model grid cells. I think it would have been more reasonable to change the rooting distribution Z from the model formulation instead of varying the bottom soil thickness. After all, in reality it is the vegetation adapting the roots, not the soil changing thickness. Why this approach was not followed?'

It is indeed true that in reality vegetation adapts the roots according to drought frequency, and not the soil changing its thickness. The referee suggests an approach to modify the rooting distribution rather than the soil depth.

It should be noted that the soil depth in HTESSEL does not represent the actual soil thickness but instead it represents the hydrologically active depth of the soil as resulting from the actual depth that is reached by the vegetation roots for transpiration. Consistently, the water content in the soil corresponds with the field capacity (i.e. water accessible by vegetation for transpiration), which – excluding wilting point – coincides with root zone storage capacity (S_{r}). We will add a sentence in the model description of HTESSEL (section 2.4) to stress this.

Transpiration is mediated by the amount of water in the soil that can be effectively accessed by vegetation. In the HTESSEL model this depends on three factors: 1) the total amount of water available in the hydrologically active soil (controlled by total soil depth) 2) the relative depth of the individual soil layers and 3) the rooting distribution (i.e. relative density) in each soil layer.

We acknowledge that we could as well have changed the rooting distribution and the individual soil layer depths. However, in this study we did not want to change multiple model parameters at the same time to avoid difficulties in identifying the differences between CTR and MD model output. This study is focused on the effect of the total amount of water in the soil available for vegetation transpiration (as controlled by total soil depth). We will add a sentence explaining this choice better in section 2.5.

In follow-up studies we will consider the effects of changing the rooting distribution and the relative depth of soil layers and we will elaborate on this in the discussion (section 4.3).

We will modify L186 in section 2.4 as follows:

'...with z ***the hydrologically active depth, that corresponds to*** *the combined depth of all soil layers with roots...'*

We will modify L222 in section 2.5 as follows:

'... in Fig. 4. ***Also, the root distribution is not modified in MD, because we aimed for a physical representation of Sr (Eq. 22) and we did not want to change multiple model parameters at the same time.****'*

Comment 1.2

'The modelled approach assumes that the maximum holding capacity should be equal to Sr. If I understand well it should rather a minimum value corresponding to dry years right? For tall vegetation, the root depth is defined through the memory method as a function of the 40 years return period drought and 2 years for low vegetation. With the implementation considered in this study it seems like the soil cannot hold more moisture than the one available in dry years. Maybe this is the reason of the apparent systematic underestimation of Sr by the model.'

If we understand correctly, the referee's perception was that S_{r} represents the maximum available moisture during dry years, and is therefore a lower limit of soil moisture holding capacity. On the contrary, S_{r} is defined by the soil moisture deficit, that maximizes during dry years, and, therefore, represents an upper limit of root zone storage. The apparent underestimation of $S_{\mathrm{r,MM}}$ compared to $S_{\mathrm{r,CTR}}$ is therefore not related to the available moisture in dry years, but related to the moisture deficit in dry years.

We will clarify the relation of soil moisture storage and S_{r} in the methods chapter in lines 127-130 as follows:

' ...is estimated based on catchment hydrometeorological data, according to the methodology described in the studies of De Boer-Euser et al. (2016), Nijzink et al. (2016) and Wang-Erlandsson et al. (2016). $\boldsymbol{S_{r,MM}}$ ***is based on an extreme value analysis of the water storage deficit in the vegetation's root zone (*** $\boldsymbol{S_d}$ ***).*** $\boldsymbol{S_d}$ ***maximizes during dry periods, and, therefore,*** $\boldsymbol{S_r}$ ***represents an upper limit of root zone storage assuming that vegetation has sufficient access to water to overcome these dry periods.*** *The cumulative...'*

Comment 1.3

'Also, the definition of Sr for the observations is a linear superposition of different values that are obtained for high and low vegetation, suggesting the these values are different. Modelled Sr is defined as one value for all vegetation types. So not only the implementation is the model seem to assume that the soil cannot hold more water that than in dry years, but also that is does not depend on the vegetation type. Again, maybea different approach is needed.'

HTESSEL does not implement any sub-grid heterogeneity in the soil discretization and so it does not allow to change soil depths differently for different vegetation types (Fig. 4b)). Accordingly, $S_{\mathrm{r,MM}}$ was defined as one value for all vegetation types (see Eq. 7). The memory method allows us to make a separation in S_{r} for high and low vegetation, but we combined S_{r} values for high and low vegetation to get unique catchment-representative S_{r} and so the MD soil depth consistently with the HTESSEL model formulation. However, we acknowledge that we could get a separation between high and low vegetation S_{r} by modification of the root distribution separately for high and low vegetation. However, as mentioned before, we did not want to change both root distribution and model soil depths at the same time in this study (see also reply to comment 1.1).

We will clarify this in section 2.3 line 155 as follows:

'...maximum storage deficits. ***Theoretically we could treat S_{r} separately for high and low vegetation in HTESSEL, however, this would require changing the root distributions (see section 2.4), which we decided not to do as we did not want to change multiple parameters at the same time.****'*

Comment 1.4

'Finally, I think the modelled Sr could be computed exactly in the same manner as it is done for observations. I understand the author prefer to use a more physical definition, but probably computing it in the same way as in the observation would be a fairer comparison that could be used to better calibrate the modified model.'

Indeed, we did choose a physically based approach (changing the model soil depths) for implementing S_{r} as we aimed to investigate if we could clearly observe effects of the memory method S_{r} estimates on the modelled fluxes in HTESSEL.

However, it would indeed be useful to make a comparison based on soil moisture deficits and the model's effective S_{r} ($S_{\mathrm{r-eff}}$). Thus, following the suggestion of the referee, we explored calculating $S_{\mathrm{r-eff}}$ based on the modelled soil moisture deficits and a similar extreme value calculation as was done in the memory method. We will add a line to the HTESSEL model description (2.4), we will change lines 351-352 in the manuscript and we will include Figure C1 in the supplementary material, together with a short discussion of the results shown in this figure.

We will modify line 189 (section 2.4) as follows:

'is not accessible to roots ... ***It should be noted that we aimed for a physical definition of $S_{r,CTR}$, but that the effective water used by vegetation may be different. We come back to this point more elaborately in the discussion.'***

We will modify lines 351-352 in section 4.3 as follows:

L351-352: *'This formulation represents the theoretical $S_{r,CTR}$, but it might not fully correspond to the soil moisture in the four layers that is actually used by the modeled vegetation.* ***The effective S_r is derived from modelled soil moisture storage deficits ($S_{r,CTR-eff}$) and is smaller than $S_{r,CTR}$ based on depths (Fig C1c). This is likely due to the relatively small root percentage in layer 4 prescribed from look-up tables in this layer for most vegetation types compared to the other layers. However, the $S_{r,MM}$ we implemented in the MD model by changing soil depths is close the $S_{r,MD-eff}$ based on modelled soil moisture deficits in the MD model (Fig. C3d).***

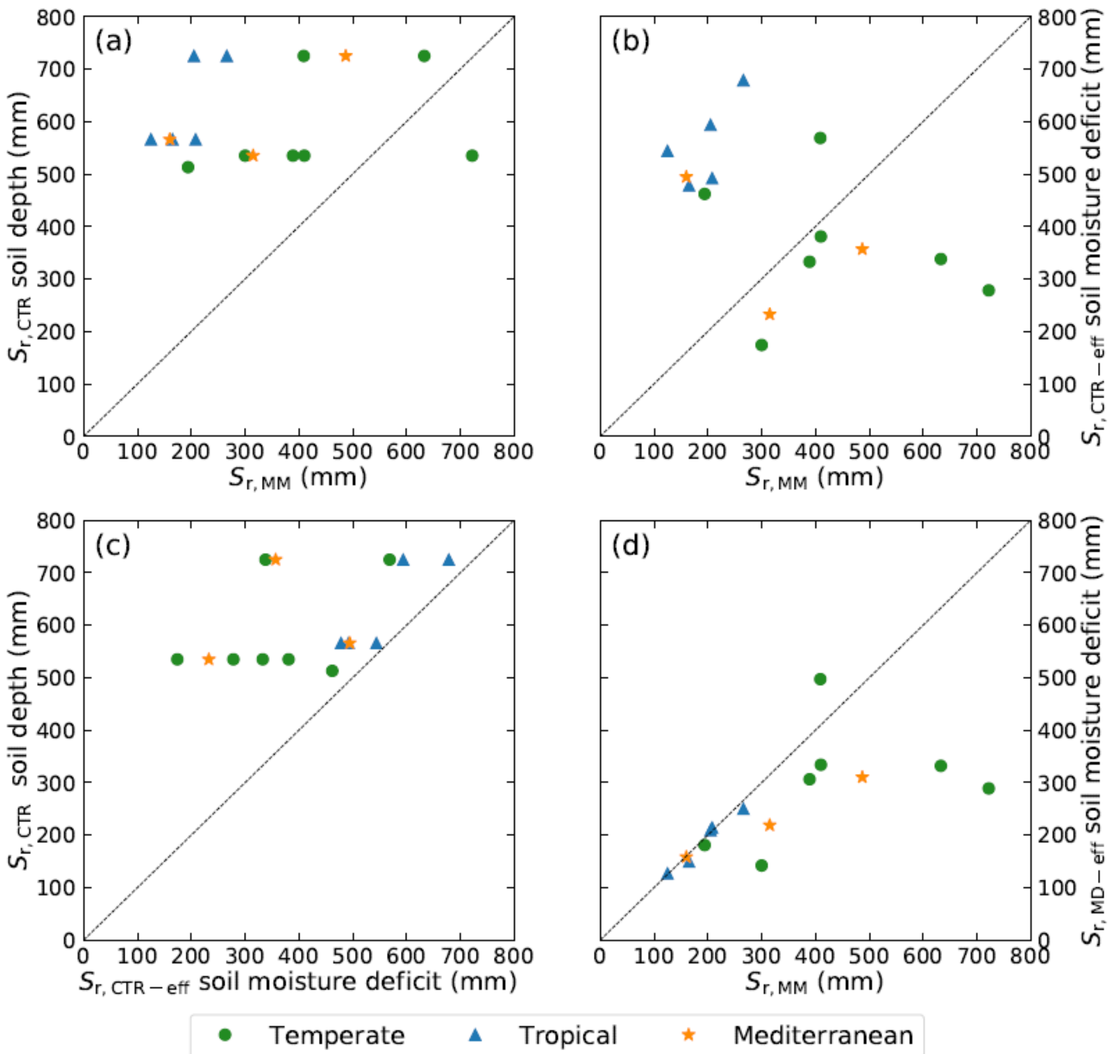

Figure C1 (new Figure S3). Model S_r analysis. (a) $S_{r,MM}$ from the memory method vs. $S_{r,CTR}$ based on HTESSEL soil depth. (b) $S_{r,MM}$ from the memory method vs. $S_{r,CTR-eff}$ based on modelled soil moisture deficits. (c) $S_{r,CTR}$ based on soil depth vs. $S_{r,CTR-eff}$ based on modelled soil moisture deficits. (d) $S_{r,MM}$ from the memory method vs. $S_{r,MD-eff}$ based on modelled soil moisture deficits.

Minor comments

Comment 1.5

'Equation 2 is formally incorrect. The time-dependency S(t) does not match the right-handside where t is the integrating variable that should disappear after the integration is performed. The actual variable that survives should be related to t0 and t1, which are notdefined either than in Appendix A. I think a subscript for the hydrological year should be preferred in that case. About the subscripts, here 'd' is used. 'r' in used elsewhere, why? I think 'd' stands for deficit and 'r' stand for roots. If so, is the right hand side of equation 7 there should be 'd', not 'r'.'

We will change Eq. 2 from:

$$S_{\mathrm{d}}(t) = \max(0, -\int_{t_0}^{t_1}(P_{\mathrm{e}} - E_{\mathrm{t}})\mathrm{d}t)$$

to:

$$\boldsymbol{S_{\mathrm{d}}(t) = \max(0, -\int_{t_0}^{\tau}(P_{\mathrm{e}} - E_{\mathrm{t}})\mathrm{d}t)}$$

with an integration from t_0 that corresponds to the first day in the hydrological year 1973 to τ that corresponds to the daily time steps ending at the last day of the hydrological year 2010.

There is also confusion about the use of the subscripts 'd' and 'r', in which 'd' stands for deficit, and 'r' for root zone. There is an important difference between S_{d} and S_{r}: S_{d} is the storage deficit over time, defined as the cumulative difference between P and E_{t}. On the other hand, S_{r} is the root zone storage capacity that is calculated applying a Gumbel extreme value analysis to annual maximum storage deficits (see lines 154-155). We will clarify this in lines 127-130 as follows:

'$S_{r,MM}$ is estimated based on catchment hydrometeorological data, according to the methodology described in the studies of De Boer-Euser et al. (2016), Nijzink et al. (2016) and Wang-Erlandsson et al. (2016). ***$S_{r,MM}$ is based on an extreme value analysis of the annual maximum water storage deficits in the vegetation's root zone (S_d).'***

Comment 1.6

'I agree that considering constant ratio of actual vs. potential evapotranspiration is a crude approximation, especially in water limited regions. I also agree that the other factors mentioned by the authors (groundwater, irrigation) are important as well. These are difficult to improve in the model in short times. However, the author of implemented an iterative step in their approach to reduce strongly the uncertainty relate to the inter-annual variability of that ratio (Appendix A). Couldn't they use somehow the observed evaporation to further improve the estimation? I think this would improve the estimated Sr, especially regarding the intra-seasonal variability, and eventually improve the modified model calibration.'

The estimation of transpiration in the memory method is solely based on discharge, precipitation and potential evaporation data. This was done because we consider these data reliable for the catchment scale. On the other hand, actual evaporation data (FLUXCOM in this study) is less reliable for the catchment scale, because it is based on point scale estimates of FLUXNET stations that are located far from the study catchments (Fig. 3 and lines 114-117). Therefore, we decided not to use the FLUXCOM intra-seasonal variability for the $S_{\mathrm{r,MM}}$ estimates.

Furthermore, the general finding that $S_{\mathrm{r,MM}}$ is considerably smaller than $S_{\mathrm{r,CTR}}$ does not change when we use FLUXCOM evaporation in the memory method (average S_{r} of 284 mm with FLUXCOM, average S_{r} of 333 mm with scaling E_{p}). As the differences are small, no large effects on the MD model performance are expected.

Comment 1.7

'To the uncertainties, I would add the model drainage rate that, in the current framework,it could as important as the rooting depth. The results obtained by the authors could be due to excessive retention (slow drainage) rather than too deep root zone. This is also supported by the fact that other models are rather augmenting the soil depth adding a groundwater layer instead of reducing it (e.g. CLM).'

We agree that a groundwater layer is important for modelling the base flow and that the lack of a groundwater layer in HTESSEL causes uncertainty in the modelled river discharge. However, in this paper we aimed to improve the representation of the vegetation's root zone, and an analysis of modelled base flow and the potential of an additional groundwater reservoir was out of scope.

We will add this uncertainty to the discussion in 4.3.

Comment 1.8

'Equation 22: where is the equal sign?'

There is confusion about Eq. 22:

$$S_{\mathrm{r,MD}} = S_{\mathrm{r,MM}} = z_{\mathrm{MD}}(\theta_{\mathrm{cap}} - \theta_{\mathrm{pwp}})$$

We will change the equation and line 216 to:

'$\boldsymbol{S_{\mathrm{r,MM}} = z_{\mathrm{MD}}(\theta_{\mathrm{cap}} - \theta_{\mathrm{pwp}})}$

with z_{MD} the total soil depth in the MD model modified to satisfy $\boldsymbol{S_{r,MD} = S_{r,MM}}$.'

Comment 1.9

'Figure 6: Shouldn't Q + E equal P in the long period? If this should be true, the average anomalies of the modified model on the top panel should be equal the average anomalies of the bottom panels as precipitation does not change, but this is not true. E anomalies are much smaller, Why? Am I missing something?'

In the long term the water balance closes and Q + E equals P. In Figure 6 the mean values of Q + E for all the three presented bars (observed, CTR and MD) sum up to P. However, the y-axis scale for Q in Figures 6b and 6c is different than in the other subplots, which apparently confuses the reader. Dr. Guswa (referee 3) suggested to use the same y-axis scale in Figures 6b and 6c, and add zoomed figures of the temperate and Mediterranean results to make the plots readable (comment 3.5). Figure C2 is the new version of Figure 6.

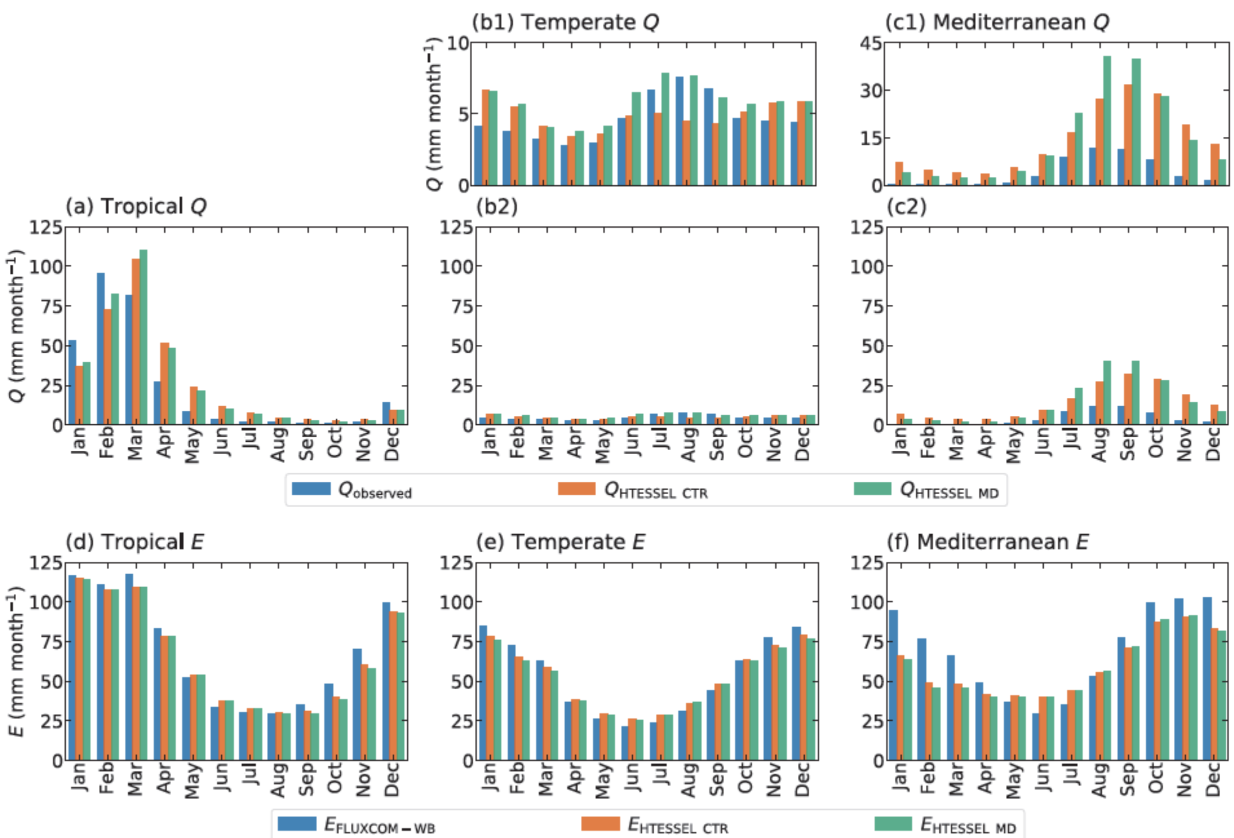

Figure C1(new Figure 6). Monthly seasonal climatology of observed discharge (Q) (top) and FLUXCOM-WB evaporation ($E_{FLUXCOM-WB}$) (bottom) and modeled values in the HTESSEL CTR and MD versions, averaged for the tropical (a, d), temperate (b, e) and Mediterranean (c, f) catchments for the time series 1975-2010. b1 and c1 represent the same data as b2 and c2, but with a different y-axis. Similar figures for the individual catchments are shown in Fig. S1 (Q) and Fig. S2 (E).