Author response to Reviewer 1 for **Improvement of modelling plant responses to low soil moisture in JULESvn4.9 and evaluation against flux tower measurements** by Harper, Williams, et al.

We thank both reviewers for their helpful comments and suggestions. There were a few points raised by both reviewers – we will first address these and then address specific comments from this reviewer.

Clarifying the root parameter d_r

Based on both reviewer's comments, we need to clarify the role of the parameter d_r, which is the e-folding depth of the roots. Figure 2 in the manuscript (included in this response for completeness) shows that roots are present in all common layers, although the fraction of roots may be very small. Therefore, technically the root depth is equal to the soil depth in each experiment. Instead, the $\mathrm{d_r}$ parameter should be interpreted as a weighting factor for the effective root fraction within each soil layer, which will directly shape to the root water extraction and soil moisture stress. The smaller the d_r, the more emphasis is given to shallow layers; while deeper layers are emphasized with a larger d_r. As a specific example: with JULES default soil depth of 3 m, 87% of the root water extraction is from the top 1 m for C3 and C4 grasses (d_r=0.5), compared to 45% in the top 1 m for tropical broadleaf evergreen trees (d_r=3.0). With the 10.8 m-deep soil, 86% of the root water extraction is from the top 1 m for C3 and C4 grasses (d_r=0.5), compared to 29% in the top 1 m for tropical broadleaf evergreen trees (d_r=3.0). Reviewer 2 noted that perhaps grassland and crop ecosystems should not have as deep roots as forest ecosystems. First, we note that according to Canadell et al. (1996), the global average maximum rooting depth for trees is 7±1.2 m, while it is 2.6±0.1 m for herbaceous plants. This supports the reviewers suggestion that grasslands on average have shallower rooting systems than forests. The parameterization of roots in JULES with the 10.8 m soil reflects these observations: for C3 and C4 grasses, when $d_r = 0.5$, 99% of root water extraction comes from the top 2.4 m (this is not the same as the maximum rooting depth reported by Canadell et al. (1996). When d_r is doubled, 99% of root water extraction comes from the top 4.8 m for C3/C4 grasses, which is deeper than the Canadell et al. (1996)

values although that study did find deeper roots for tropical grasslands. In comparison, for the tree PFTs, 95% of root water extraction comes from the top 13 layers (to a depth of 7.8 m) when d_r=3, or 87% when d_r=6.

In the revision, we will clarify that d_r is a root weighting factor throughout the text, not the root depth. We will include this in Section 2: where the parameter is first introduced, and then in the discussion of the different experiments. We hope this also addresses reviewer concerns that we should use site specific or observed root depths, as the d_r parameter incorporates more than just the fraction of roots in each layer but also accounts for other properties and processes such as surface area of roots, conductivity, and hydraulic redistribution.

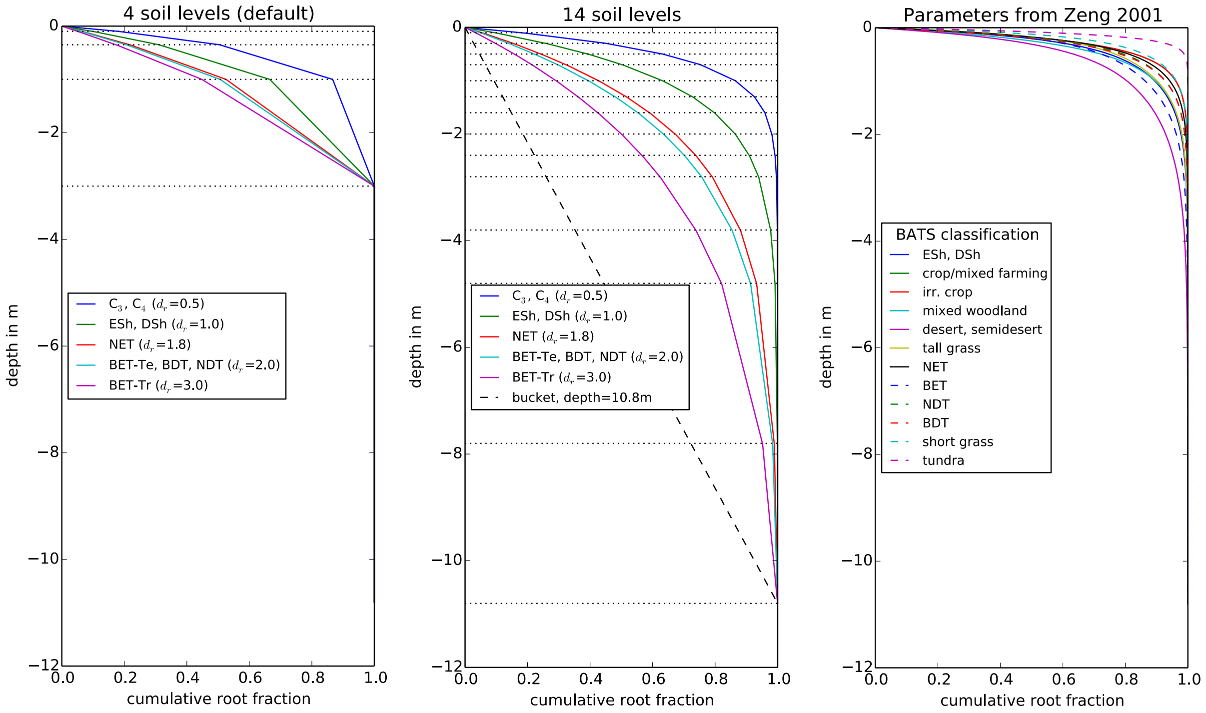

Figure 2: Root water extraction profiles for JULES with the default 4 layer soil (left panel; maximum depth of 3m), with an updated 14 layer soil (middle panel; maximum depth of 10.8m), and compared to root fractions from Zeng (2001) (right panel), where distributions were calculated based on available measurements of root profiles. The parameter d_r in JULES is the e-folding depth for weighting root water extraction and soil moisture stress. The plant functional types are: C3, C4 grasses; evergreen and deciduous shrubs (ESh, DSh); needleleaf evergreen trees (NET), temperate broadleaf evergreen trees (BET-Te), broadleaf deciduous trees (BDT), needleleaf deciduous trees (NDT), tropical broadleaf evergreen trees (BET-Tr).

Above: Figure 2 with new caption.

Justification of the soil moisture stress experiments

Both reviewers have said the modelling choices aren't well motivated. In our revision, we will better justify each experiment in Section 2.3 (in the original manuscript this section was mistakenly labelled as 2.2). The justification for each experiment is alluded to in the Introduction, where the possible need for deeper roots and soils is discussed, along with alternative methods of representing stress (using soil matric potential rather than volumetric water content), and the possibility that stress occurs too early in JULES. However, we can and should make these links clearer as justification for each experiment, and this will be done in the revision.

Clarifying the results section

Both reviewers also found certain parts of the results difficult to follow. To address this, we will update Tables 3 and 4 to have the statistics rather than qualitative assessment. This will enable the numbers to mostly be shown in the table, but for analysis of them to occur in the text.

Section 3.3 evaluates the responses of the model at a subset of sites. The sites are divided into 3 categories, based on evaluation in Section 3.2, where we artificially removed soil moisture stress in the model: sites where soil moisture stress leads to large biases, sites with a Mediterranean climate, and sites with soil moisture stress-related errors plus other biases. At the end of the section, we discuss the average responses across the 11 sites. In the revision, we will set out this logic at the beginning of Section 3.3 and add subsections to reflect the three categories.

We have thought about the best way to visualize the statistics for the 11 sites with the 10 soil moisture stress experiments. The Taylor diagrams are good summaries for selected sites, since they show RMSE, standard deviation (measure of modelled vs observed variability), and correlation for multiple experiments. However, we believe it would be cluttered to show Taylor diagrams for all 11 sites in the main text (these are shown in the SM). As a compromise, and to reduce the need for statistics to be listed in the text, we will add 3 tables to the main text for the average statistics for simulated GPP in the 3 categories of sites examined in Section 3.3. Similar tables for LE will be put in the SM.

Below we respond to Reviewer 1's 3 points, with the reviewer's text in blue and our responses in black.

The manuscript discusses different formulations of the soil moisture plant physiological stress within the model JULES. Multiple formulations are presenting, based on either soil moisture or water potential. Beyond those, the importance of the vertical discretization of the Richards' equation is explored, along with the assumed soil depth and root distribution. For a limited number of sites, where soil moisture and LAI data existed, the authors also used them as prescribed values to the model in order to disentangle the importance of the soil water stress from the remaining model errors related to plant phenology and hydrology. Overall, the manuscript is within the scope of GMD. Even though focused on JULES, almost all state-of-the-art ecosystem models adopt similar formulations, and thus the results are likely important for many similar models. It is also clearly written. However, I have a few concerns that need to be clarified:

1) While it is perfectly reasonable to seek a unified equation to model plant water stress, the same is not true for model parameters, such as the threshold where plants start experience stress, the soil depth, and the root distribution (something also pointed out by the authors in the manuscript's introduction). These would be site-specific parameters, and seeking a model set-up that fits all, would be unrealistic in my opinion. The diversity of plant species found across the globe have been simplified into 9 plant functional types in JULES, which does allow for some variation of parameters between sites (e.g. d_r), depending on the dominant plant types. It is true that some new parameters we introduced should vary by PFT, for example this is particularly true with the ψ_{open} and ψ_{close} parameters (Equation 8). In fact, a

follow-on study is underway which aims to define PFT-specific values of these parameters. The aim of this study was to find recommended parameterizations of stress for global applications. For example, JULES represents the land surface within the UK Earth System Model, and it represents the global terrestrial carbon cycle in the annual Global Carbon Project updates. Therefore, we need a model set-up that produces reasonable results with minimal site-level modifications. In our study, we have struck a balance between including some site-specific parameters to help remove sources of error (e.g. the soil parameters at some sites as specified in SM Table 1) and using a model set-up similar to that used in the UKESM simulations.

The authors need to further explain their rationale regarding the choice of the numerical experiments, and what is the information they wish to extract from each of them. Possibly adding some detailed hypotheses linked to each scenario would help the reader.

Please see our response to both reviewers at the beginning of this document.

As a suggestion, in order to fully evaluate the performance of soil water stress formulations (e.g. moisture-based vs potential based formulations) a small number of sites where the root distribution, the soil depth and the soil hydraulic parameters are known would be very useful.

We already present results from some sites with the observed soil hydraulic parameters, in the revision we will add the information about which sites these are. We aim to find a model set-up that will work for global applications, so we have not adjusted root distributions or soil depths per site, as JULES does not have the capability to vary these parameters spatially in a global simulation.

Actions from comment 1: In a revised manuscript, we will better justify generalised parameters and representations. We will clarify that d_r is a root weighting factor throughout the text, not the root depth. We will revise section 2.3 to better justify and explain the experiments.

2) Increasing the soil layers from 4 to 14 will lead to a more accurate solution of the Richards equation (in terms of numerical accuracy). Solving a highly nonlinear PDE, with a very coarse spatial resolution (e.g. just 4 layers) is expected to lead to biases. Because of that I would suggest the authors to present this choice, not as model improvement, but likely as a warning against using coarse vertical resolutions to gain computational speed.

We agree that the increase in the number of soil layers is desirable for numerical accuracy. We will note the preference for more vertical resolution in the soils when solving Richards' equation (in the discussion when results with the 14 layer soil are discussed).

3) Tables 3 and 4 would be much better if they included the actual numbers (correlation coefficient, RMSE, absolute error) for all sites and simulations (and possibly a rank for each simulation, from best to worst), rather than the classification per biome. I also suggest the authors retain the qualitative classification for the results section itself rather than presenting the numbers there, as this section is currently a bit convoluted and difficult to follow.

This is another good suggestion. In the revision, we will change Tables 3 and 4 to only show the average correlation, normalized absolute error, and variance ratio for each biome. The final column in Table 3 will be removed, and the text in the results section will be revised to include more qualitative discussion. Please see our joint response to reviewers for further modifications in response to this comment.

New references

Canadell J., Jackson R.B., Ehleringer J.B., Mooney H.A., Sala O.E., Schulze E.D.. Maximum rooting depth of vegetation types at the global scale. Oecologia. 1996 Dec;108(4):583-595. doi: 10.1007/BF00329030.

Zeng, X. (2001). Global Vegetation Root Distribution for Land Modeling, *Journal of Hydrometeorology*, *2*(5), 525-530.