Response 2 to Reviewer of Manuscript: Afforestation impacts on terrestrial hydrology insignificant compared to climate change in Great Britain

Comments/Text of the Reviewer is in **black**, our response is in **blue**.

We thank the Reviewer for their time reviewing this article.

The study evaluates the potential hydrological changes under the possible afforestation regime alongside climate changes. Therefore, the results suggested that afforestation has an insignificant impact on the hydrology. The conclusion is not well supported by the modelling results since the overall model performance remains unsatisfactory. Model performance in most of the study sites was not satisfactory (in terms of NSE>0.5, which is more commonly used than KGE), which may require more detailed calibration and validation before future climate scenario is carried out. Further, some hydrography showing the changes might be useful for the presentation since the study evaluates hydrology.

In our response to the other reviewer, you can find a sample of validation studies that have shown the validity of using JULES to explore hydrological and land cover changes. The work here is ambitious in its scope by using a land surface model in which there are a multitude of processes from a range of systems beyond just hydrology. This makes it invaluable for determining complex questions such as that of woodland hydrology. For the ease of the Reviewer, we include a summary here of the key validation studies that illustrate the appropriateness of JULES in this study from our other reviewer response.

Harper et al. (2016, 2021) both validated and improved the plant functional types used in JULES and found that in temperate regions that JULES produced good agreement with latent energy fluxes. The initial implementation of plant hydrology mechanics within JULES can be found in Alton et al. (2009) and shows that global runoff agrees fairly well with this implementation as well as evapotranspiration. Van den Hoof et al. (2013) and Blyth et al. (2019) also illustrate that JULES gets the broad pattern of evapotranspiration changes correct.

Considering streamflow and runoff, the model configuration used is that developed by Martinez-de La Torre et al. (2019) that achieve an NSE score of over 0.8 for the River Thames. Older versions of JULES have also been shown to get the broad structure of hydrological events and extremes correct (e.g. Prudhomme et al. 2011; Harding et al. 2014). More recent version implementations of JULES have produced even more accurate results (e.g. Lewis & Dadson 2021; Mathison et al. 2023).

The main model configuration used in this study can be found in Buechel et al. (2022) as stated on line 152. Buechel et al. (2022) provides a validation of streamflow, soil moisture and evaporation (including at Alice Holt) and these can be found in the methodology section of that study. At the broadleaf woodland sites there was both systematic over and underprediction of soil moisture with an overestimation of evaporation. However, the system responses were broadly correct.

What the Reviewer is touching upon here is the idea of accuracy potentially above faithful representation of processes (Guse et al. 2021). Catchment hydrological models can be calibrated to produce accurate output (such as streamflow) but at the cost of compensatory parameters and an inability to elucidate realistic processes generating changes. This means that it is impossible to deduce potential realistic process responses to afforestation as a more finely calibrated model would alter processes to create an accurate output. We have chosen to use a physically based model that enables exploration of multiple systems and interactions between them.

As shown in the discussion we provide ideas for future improvement of JULES (e.g., lines 582-589) and acknowledge its limitations.

The review of Stratford et al. 2017 illustrates the issue of the current use of hydrological models to determine the impact of afforestation on flooding. Many models, although calibrated and validated to high accuracy, do not replicate the results seen in the observational data. As a community, we are therefore missing crucial processes in understanding the impact of woodland on hydrology within our hydrological models. We therefore need to turn to models of higher complexity that include more processes (especially phenology and stomatal conductance) to understand what the consequences of increased woodland area on hydrology could be. On lines 513-514 we highlight the fact that this model has generated more faithful and realistic results than compared to other models.

In relation to the KGE scores used, we would refer the Reviewer to Knoben et al. 2019 which highlighted that NSE and KGE are not comparable metrics. Although there is indeed a lot of work needed to improve hydrology in land surface models, they are satisfactory when we are more concerned about the multiple processes occurring and not just streamflow output. More validation information can be found in the papers highlighted on lines 150-154 in the main text. The same configuration used by Martinez-de la Torre et al. (2019) achieved NSE scores above 0.8 for the Thames catchment.

Guse, B., Fatichi, S., Gharari, S. and Melsen, L.A., 2021. Advancing process representation in hydrological models: Integrating new concepts, knowledge, and data. Water Resources Research, 57(11), p.e2021WR030661.

Stratford, C., Miller, J., House, A., Old, G., Acreman, M., Duenas-Lopez, M.A., Nisbet, T., Burgess-Gamble, L., Chappell, N., Clarke, S. and Leeson, L., 2017. Do trees in UK-relevant river catchments influence fluvial flood peaks?: a systematic review. https://core.ac.uk/download/pdf/96704761.pdf

Knoben, W.J., Freer, J.E. and Woods, R.A., 2019. Inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores. Hydrology and Earth System Sciences, 23(10), pp.4323-4331.

Martínez-De La Torre, A., Blyth, E.M. and Weedon, G.P. (2019) 'Using observed river flow data to improve the hydrological functioning of the JULES land surface model (vn4.3) used for regional coupled modelling in Great Britain (UKC2)', 925 Geoscientific Model Development, 12(2), pp. 765–784. doi:10.5194/gmd-12-765-2019.

Alton, P., Fisher, R., Los, S., and Williams, M. (2009),Simulations of global evapotranspiration using semiempirical and mechanistic schemes of plant hydrology, _Global Biogeochem. Cycles, 23, GB4023, doi:[10.1029/2009GB003540](https://doi.org/10.1029/2009GB003540

Blyth, E.M., Martinez-de la Torre, A. and Robinson, E.L., 2019. Trends in evapotranspiration and its drivers in Great Britain: 1961 to 2015. Progress in Physical Geography: Earth and Environment, 43(5), pp.666-693.

Buechel, M., Slater, L. and Dadson, S., 2022. Hydrological impact of widespread afforestation in Great Britain using a large ensemble of modelled scenarios. Communications Earth & Environment, 3(1), p.6.

Harding, R.J., Weedon, G.P., van Lanen, H.A. and Clark, D.B., 2014. The future for global water assessment. Journal of Hydrology, 518, pp.186-193.

Harper, A. B., Cox, P. M., Friedlingstein, P., Wiltshire, A. J., Jones, C. D., Sitch, S., Mercado, L. M., Groenendijk, M., Robertson, E., Kattge, J., Bönisch, G., Atkin, O. K., Bahn, M., Cornelissen, J., Niinemets, Ü., Onipchenko, V., Peñuelas, J., Poorter, L., Reich, P. B., Soudzilovskaia, N. A., and Bodegom, P. V.: Improved representation of plant functional types and physiology in the Joint UK Land Environment Simulator (JULES v4.2) using plant trait information, Geosci. Model Dev., 9, 2415–2440, https://doi.org/10.5194/gmd-9-2415-2016, 2016.

Harper, A. B., Williams, K. E., McGuire, P. C., Duran Rojas, M. C., Hemming, D., Verhoef, A., Huntingford, C., Rowland, L., Marthews, T., Breder Eller, C., Mathison, C., Nobrega, R. L. B., Gedney, N., Vidale, P. L., Otu-Larbi, F., Pandey, D., Garrigues, S., Wright, A., Slevin, D., De Kauwe, M. G., Blyth, E., Ardö, J., Black, A., Bonal, D., Buchmann, N., Burban, B., Fuchs, K., de Grandcourt, A., Mammarella, I., Merbold, L., Montagnani, L., Nouvellon, Y., Restrepo-Coupe, N., and Wohlfahrt, G.: Improvement of modeling plant responses to low soil moisture in JULESvn4.9 and evaluation against flux tower measurements, Geosci. Model Dev., 14, 3269–3294, https://doi.org/10.5194/gmd-14-3269-2021, 2021.

Lewis, H.W. and Dadson, S.J., 2021. A regional coupled approach to water cycle prediction during winter 2013/14 in the United Kingdom. *Hydrological Processes*, *35*(12), p.e14438.

Mathison, C., Burke, E., Hartley, A.J., Kelley, D.I., Burton, C., Robertson, E., Gedney, N., Williams, K., Wiltshire, A., Ellis, R.J. and Sellar, A.A., 2023. Description and evaluation of the JULES-ES set-up for ISIMIP2b. Geoscientific Model Development, 16(14), pp.4249-4264.

Prudhomme, C., Parry, S., Hannaford, J., Clark, D.B., Hagemann, S. and Voss, F., 2011. How well do large-scale models reproduce regional hydrological extremes in Europe?. Journal of Hydrometeorology, 12(6), pp.1181-1204.

Van den Hoof, C., Vidale, P.L., Verhoef, A. and Vincke, C., 2013. Improved evaporative flux partitioning and carbon flux in the land surface model JULES: Impact on the simulation of land surface processes in temperate Europe. Agricultural and forest meteorology, 181, pp.108-124.

The conclusion is questionable based on the results of the study. For example, most of the changes in the broadleaf forest area are less than 10 per cent across the study sites. Although the scenario might be realistic under the conditions mentioned in the manuscript, however, not enough to support the conclusion since the hydrological change is not expected to be significant under the level of land cover change. Contradictory results are also referred to in the literature reviews, (L6: Many studies suggest afforestation can reduce overall streamflow). I suggest that results from the relevant studies should be discussed in the manuscript. Further, I believe that per percentage point of afforestation (PPPoA) is not a good indicator to evaluate the hydrological changes. I would suggest the model be simulated under selected paired-experiment sites (at least some examples should be carried out in the manuscript), which better considers the effects of afforestation.

The point being made here is not entirely clear to us. The magnitudes of afforestation are realistic, as stated on lines 132-134. The discussion of contrary findings in previous work provides important context. The choice of PPPoA metric allows for a consistent comparison across all locations. The suggestion to conduct paired-experiments is a good one but would be another study entirely. These points are elaborated below.

The question of this study was to understand the potential hydrological impact of widespread realistic afforestation in the UK compared to climate. As far as we are aware, there is no other study that has a conclusion to this question. Our results suggest that this potential land cover change (which is significant when compared to contemporary land cover changes) is insignificant when compared to potential climate changes.

The contradictory results discussed in the literature review emphasise that there is not a clear answer to the question of widespread afforestation on countrywide hydrology, hence the need for this study to provide further ideas and evidence.

We disagree that percentage point of afforestation is a bad indicator. This enables others to extrapolate the results over multiple spatial scales and see if they match our results, particularly as this work is trying to ascertain the relative impact of afforestation on hydrological processes.

The model used is not appropriate for the small spatial scales of paired-catchment studies (< 10 km^2) in which other more important processes (such as forest management) will have a more significant role on output (and are normally calibrated out and not understood by the modeller). We refer the Reviewer to lines 151-154 in which other studies have evaluated the streamflow, evaporation and soil moisture of this JULES configuration (as well as our response to the other reviewer). In addition, many paired-catchment studies in the UK with enough data to validate this model (as mentioned by the other reviewer) are coniferous and this study does not consider this type of woodland expansion. However, we can add to the manuscript a few lines explaining why a validation study using a paired catchment approach would not be appropriate in this regard.

Various indicators are selected to evaluate the results, but most of them are not well defined in the methods sections or elsewhere in the manuscript (e.g. ANOVA/Kruskal). Terms such as high flow, low flow, summer, and winter should also be defined. The flow duration curve (FDC) is mentioned but rarely explained in the result section.

We thank the Reviewer for highlighting the need for clearer terms. The manuscript can be clarified accordingly. Flow percentile times and seasonal terms can be defined within the manuscript. The flow duration curve has been used in the results section (e.g. line 336 and Figure 6) and is useful for those seeking to understand flow regime change.

Some information about the model setup is missing. For example, it is unclear what temporal resolution is the results based on (hourly, daily or monthly), which is important for model performance evaluation. Also, the period of model simulations is not well described in the method section (e.g. L272 is unclear). The landcover data CAMELS-GB is not mentioned in the manuscript but is shown in the supplementary table.

Thank you for your feedback on this. CHESS-met is given at a daily resolution and is disaggregated by JULES to a 30 minute timestep. The streamflow data calculated and compared to is at a daily resolution as well. Please refer to the comments from the other reviewer in which we have clarified it is daily. The period of model simulations is stated on line 272 (2020-2050).

The supplementary table is meant to provide context for the reader of the various catchments. The data used to derive this data is the same as that used has been used to correct the driving data for JULES and we will clarify these aspects in the revised text.

Specific Comments

L107, L131 & L161 The land cover types are not consistent here, Non-default JULES land cover is referred to in L107 (acid grassland, arable and horticultural areas, heather, heather grassland, improved grassland and neutral grassland), however, how to turn it into 8 JULES land cover types are not mentioned.

The CEH Land Cover 2000 Map is used to identify the maximal possible extent of woodland (which is the same used to create CHESS-land base land cover). Broadleaf woodland is then planted as stated on lines 125-132. The land cover conversion can be found in the supplementary material of Buechel et al. 2022 and the manuscript can be rewritten to alert readers to this fact.

Buechel, M., Slater, L. and Dadson, S., 2022. Hydrological impact of widespread afforestation in Great Britain using a large ensemble of modelled scenarios. *Communications Earth & Environment*, *3*(1), p.6.

L128 I assume the percentage of changes was compared between the CEH 2000 landcover map (the year 2000), and scenarios in the year 2050 (but the selection of dataset was not well described, also the year 2050 is not pointed out clearly).

The manuscript can be rewritten to clarify that the change in hydrological processes for the three parts of this study is between the created afforestation scenarios and the CEH 2000 landcover map.

L159 temporal resolution of CHESS-met is not mentioned in the manuscript, which should be 1 hour.

The Reviewer is incorrect; the temporal resolution of CHESS-met is daily.

L162 Broadleaf tree (and other PFTs) setup: what is the parameter set used in this study? I would believe default parameters are used, however, should be mentioned somewhere in the manuscript.

The configuration is mentioned in the studies referenced on lines 151-154 and can be found in the Rose suite mentioned in the code availability section.

L183 The Kling-Gupta Efficiency Measure (KGE) formula should be described here if it is used for assessment. NSE (also found in the supplementary document) should be a more common indicator for hydrological evaluation.

We have included the formula for KGE in the supplementary text. We respect the Reviewer's opinion that 'NSE … should be a more common indicator' but that is a debated opinion (e.g. Knoben et al. 2019).

Knoben, W.J., Freer, J.E. and Woods, R.A., 2019. Inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores. Hydrology and Earth System Sciences, 23(10), pp.4323-4331.

L186 This should be the results rather than the method. Further, the temporal resolution (hourly, daily or monthly?) of the results is not defined.

We disagree that the validation metrics are results, as the study is exploring the influence of afforestation on streamflow and is not a validation study.

L196 The location of 51 study sites should be pointed out here or in the supplementary document.

The manuscript already refers the reader to Supplementary Material: Figure S1 & Table S4 on Line 196 which is a map and description of the catchments.

L198 Abbreviation as UKBN2 should be explained when it is first referred to in the manuscript.

This can be included in the manuscript.

L202 "that processes simulated are more faithful at JULES' spatial (1 km2) and temporal (30 minutes)" the statement needs more explanation.

A sentence can be added afterwards to explain further: *This is because land surface models are intended to explore processes at countrywide and continental scales.*

L219 Season should be defined here although I believe it follows spring MAM, summer JJA …

Thank you for this recommendation which can be added to the manuscript.

L220 & L227 Theil-Sen slope estimator and Spearman's rank correlation coefficient formula should be explained if it is relevant for evaluating the results.

L224-L232 explains the use of these metrics and further reading regarding Theil-Sen.

L246 ten years (using 2000-2001): should be a typo here.

Data from the period 2000-2001 were used ten times to get the model spun-up. We will clarify this point in a revised manuscript.

L264 To clarify, The CHESS-SCAPE dataset (RCP 8.5) should be referred to in the first place.

In the structure of the manuscript, the third part of the study uses CHESS-SCAPE and thus it is referred to in the last part of the methods.

Table 2 Using percentages to describe the change of FDC might be not that meaningful.

As we are trying to quantify the change in the flow regime, we believe this is appropriate, despite not giving a significant result.

L345 "As the climate changes, land cover is also expected to change." It seems not to change in the three scenarios used in this study.

L345-349 places this comment in context. This study is focusing on the hydrological effects of afforestation in comparison to climate change. We are therefore comparing land cover change, in the form of woodland cover, to potential climate changes when using a land surface model. This work is not proposing overall changes in land cover as it is impossible to predict the exact evolution of land cover change in conjunction with climate and would require a whole separate study. This sentence alone should not be taken as a description of the model experiment, but rather a statement of how the Earth is expected to change. We can clarify the text accordingly.

L348 ANOVA is not explained in the manuscript.

We will clarify that ANOVA stands for Analysis of Variance and provide a reference to the relevant statistical texts to define.

L371 "Rising temperatures appreciably alter many of JULES' hydrological parameterisations." Should affect JULES hydrological processes, not parameterisation.

We can alter the manuscript accordingly.

L383 Kruskal-Wallis (KW) test is not explained in the manuscript.

This statistical test can be described further in the manuscript.

L385 High flow/low flow are not defined in the manuscript.

Thank you for bringing this to our attention and will be clarified in the manuscript.

L437 “This is also seen in observational studies where a 1 % increase in upstream afforestation area does not detectably change streamflow” This could not conclude that” Afforestation impacts on terrestrial hydrology are insignificant.”

We would disagree with the Reviewer on this point. This form of increase in afforestation is realistic in many areas and therefore, objectively, a 1% increase not detectably altering flow is hydrologically insignificant when compared to other factors (such as climate).

L441 “Alternatively, the large epistemic uncertainty within JULES means that highly sensitive hydrological parameters are not included that would lead to diverging regional afforestation responses.” However, this part should be important to evaluate the hydrological regime.

We agree and that is why L440-452 raise these important points to the modelling community.

L560 Figure 8 is missing.

Thank you for noticing this. It is meant to refer to Figure 6 and can be altered accordingly.

L585 “By applying atmospheric changes across the whole country, variations in landcover, topography and soil type are insufficient to substantially alter the hydrological response” This is not real, soil parameters could considerably affect the hydrological regime, and it is not changed in this study before coming to this conclusion.

Soil parameters vary across catchments and regions of the UK. In the land driving data, these parameters were not enough to vary the responses generated by changes in climate and so we disagree with the Reviewer’s point. The following lines of 586-589 put this comment in context.

L666 “Future research should use fully coupled land surface–atmosphere LSMs” I disagree that JULES has already reached its limitation here, there is more room to be improved for better model performance.

We agree that there is further model improvement (as was raised throughout the discussion) to be had. Therefore, we can rephrase that it ‘could’ rather than ‘should’, be potential further research.

We would like to thank the Reviewer for their time and helpful suggestions, which will help us to improve the manuscript.