RC2: "Improving river flow generation over Great Britain in a land surface model required for coupled land-atmosphere interactions" by Alberto Martínez-de la Torre et al.

This paper presents an evaluation of the JULES LSM for 13 catchments in Great Britain, a sensitivity analysis of JULES with respect to runoff generation processes, and a new slope dependent runoff parameterization. The authors demonstrate the sensitivity of the JULES model to a selection of parameters and runoff generation schemes. Insensitive parameters and ill-performing schemes were incrementally removed from the pool of available options for improving the performance of the model. The paper is generally well written and is easy to read (with a few minor exceptions). I have 2 primary concerns with the manuscript.

General Comments

Based on the title and the introduction, it seems the authors are motivated by improving the JULES model for use in coupled land-atmosphere simulations. However, this point is not returned to in detail and the analysis and discussion do not provide the reader with insights as to how this work will impact the coupled behavior of the model. One potential improvement would be to add an analysis of the turbulent fluxes simulated by JULES. More generally, an acknowledgement that improving streamflow, by tuning parameters, may give the "right answers for the wrong reasons." This may be an important point to discuss given the motivation of using this model in coupled simulations.

Authors:

This is true. The coupled behaviour in the UK Environmental Prediction system that motivated this work is being studied and will be presented in the UKC3 (UK Coupled 3) configuration paper following the UKC2 paper published earlier this year (Lewis et al., 2018). However, we are aware of a particular issue with the JULES model when it overestimates evapotranspiration, particularly over winter time, as it was shown recently by Blyth et al. (2018). Another recent study also shows overestimated evapotranspiration at the global level in JULES and other LSMs (Schellekens et al., 2017). Our development has reduced evapotranspiration through an increase of surface runoff over rainy spells (the runoff negative bias has improved) and therefore, even though we do not believe that the runoff bias is the main reason for the model evapotranspiration overestimates, it is a step in the right direction. We have edited the "Discussion" section adding the following at the beginning (page 12, line 2):

"To our knowledge, this is the first study using the CHESS-met dataset (Robinson et al., 2017a; Robinson et al., 2017b) to drive a LSM over a wide region (the 13 selected catchments). This dataset availability opens new possibilities to study land surface hydrology and interactions with the atmosphere using LSMs (that typically require gridded forcing datasets) at the km-scale driven by gridded rainfall derived from gauge stations. A recent study (Blyth et al., 2018) investigates evapotranspiration trends and components in Great Britain over the last 55 years using CHESS-met and the JULES runoff development described in this paper. These authors find that, when comparing to flux tower data, the model overestimates evapotranspiration rates. The new runoff development reduced the negative runoff bias as shown here, mostly from increased surface runoff during the rainy season over mountainous regions. Hence, the evapotranspiration rates in the Blyth et al. (2018) study have been impacted in the right direction by lower soil moisture availability."

Further evidence of this can be seen in the JULES webpage (http://jules.jchmr.org/, evaluation section) where different model configurations (WRR2 configuration include the slope dependency for runoff described here) are compared to global benchmarking observation datasets using the ILAMB system (https://www.ilamb.org/), and the generally positive bias when evaluating against latent heat datasets (FLUXNET-MTE) is reduced, improving all the rest of the ILAMB metrics (see the particular evaluation over the Great Britain region clicking on 'brit' when expanding the tab on top named 'global': http://homepages.abdn.ac.uk/sweat/ilamb_jules/HydrologyCycle/LatentHeat/FLUXNET-MTE//FLUXNET-MTE.html)

The quest for suitable, preferably physically meaningful, model parameters is not new. The authors basically were looking for a method to determine a set of suitable parameters using their 13 catchments and then a method to regionalize those parameters in some meaningful way. In this work, the authors use catchment slope to regionalize the storage term in the runoff generation scheme. Are there other landscape or catchment attributes that hold meaningful information as the best way to distribute these parameters? A substantive discussion on how the method used in this study fits in with the larger context of parameter regionalization and model calibration would be quite instructive. The work by Samaniego et al (2010) and Mizukami et al (2017) provide two poignant examples of how this has been achieved for similar models. Another approach uses catchment similarity concepts for regionalization (e.g. donor catchments, Beck et al 2016).

Authors:

Thanks for the references. We were particularly familiar with the Beck et al. (2016) study. However, we don't think our work fits right in with this effort or Mizukami et al. (2017) as we are trying to avoid previous knowledge of catchment characteristics (similarity to the control catchments in these cases) to assign parameters over data poor regions, but rather use basic grid cell information available to the user. Nevertheless, testing our development over the rest of Great Britain would follow the Beck et al. (2016) approach and that is in a way what we did in the recent publication referenced above (Blyth et al., 2018).

We have edited page 2 (lines 18-21) as follows:

"Nevertheless, there have been significant efforts in the hydrological community to generalise the catchment parametrization for regional scales (Crooks et al., 2014; Wagener and Wheater, 2006) and to estimate parameters over data poor or ungauged regions using catchment similarity concepts (Beck et al., 2016; Mizukami et al., 2017). However, a LSM widely used in the research community like JULES needs physically based parameters that produce sensible results at the regional and global scale, independently of the region studied."

Specific Comments

Page 7, line 23: In the figures showing all 13 catchments, it would be useful to group the catchments in some meaningful way (e.g. geographically).

Authors:

The order we follow through the paper tables and plots is the order of the NRFA station code numbering. It is geographically meaningful as it goes around the island starting in the North coast of Scotland and then following the coast line clockwise in steps of a thousand. Stations for catchments that ultimately end at the same point in the coast have been historically numbered chronologically as they opened (within the same thousand).

Page 7, line 24: The rationale for the range of parameters used in these tests should be explained. It seems, the largest b values give the highest NS performance. This indicates that the authors may not have sampled the complete parameter space.

Authors:

Figure 2 is intended to show how the parameter variability chosen for the PDM tests cover the complete spectrum of possible f_{sat} values as a function of soil water content (i.e. differences tend to diminish as we go towards higher values of b). We stress that it is the S_0

parameter variability that introduced the higher differences, being moreover the novelty of this work as it had a fixed value of zero (not a parameter in practical terms) in the model previous to our development.

Page 8, line 7 and Fig 6: The attempt to characterize the results in the previous sub-sections based on catchment attributes is commended. It may be useful to think more broadly about how to relate multiple attributes to parameters. Slope seems unnecessarily 1-dimensional. The figure itself could use a more descriptive caption or reformat-ting, as the functional relationship between slope and parameter is not made clear.

Authors:

Our initial focus was always to find a simple parametrization of the runoff generation that would not require particular catchment calibration or other landscape properties that were not necessarily available for Great Britain or other domains and that were not already required by a standard JULES user for any hydrological assessment. We did investigate the relationship of our best performing tests with other physical parameters that the model requires (soil properties, topography index), but we decided to base our parametrization on terrain slope as it showed the clearest signal. Furthermore, terrain slope is an accessible parameter that the JULES (or any other LSMs) user can obtain or derive easily from public datasets. We have added the following lines to the "Discussion" section straight after the first paragraph (page 12, second paragraph in the reviewed paper):

"We acknowledge that topographic variability at the grid scale is not new to JULES or other LSMs, as it is considered by the TOPMODEL scheme. However, we have found that for Great Britain regional integrations the surface runoff production by PDM allows for a better characterization of the topographical variability through the S_0 parameter. This finding within the JULES model and Great Britain region framework can have significant impacts over other regions and applied to other models that need to account for subgrid variability in the runoff generation process, using a widely available parameter (from digital elevation model datasets) like the grid cell mean slope as the only input, whereas other physical characteristics might be more difficult to obtain or are simply unavailable."

We have also modified the Fig. 6 caption as follows:

"Figure 6. (a to d): Representation of the b parameter value (x axis) of catchment tests that obtained a better NS metric for a given value of S_0/S_{max} (stated inside each plot), against the mean catchment slope on the y axis. The marker size represent the NS values (larger circles for higher NS values). Tests highlighted with an outer circle indicate the best performance of all tests for a given catchment (so the panel where they are indicates S_0/S_{max} and the x value indicates b). Tests where the mean bias is higher than 30% are not considered. (e): Best PDM parameter tests selected for each catchment following the criterion *mcs* of fixed $b = 2$ and slope dependent value of S_0/S_{max} as follows: 0.0 for mean catchment slopes higher than 5.0° (green background), 0.5 for mean catchment slopes between 3.5° and 5.0° (light green background), and 0.75 for mean catchment slopes lower than 3.5° (white background). For those catchment where *mcs* does not select the test of best NS metric (Tay, Ure, Derwent, Avon), the best performance tests are also represented with a degree of transparency."

Page 9, line 3: The formulation of S0/Smax seems overly subjective. Perhaps a more complete description of how this formulation came about would help. It seems like a more data driven approach would be useful here.

Authors:

We have modified the text in page 9 (lines 3-5) as follows:

"We adopt a simple approach using a linear dependency of S_0/S_{max} on slope for values below a given threshold, representing the PDM parameters in criterion *mcs* presented in Section 3.3 at the model grid cell resolution, as follows:"

Table 2: This table has basically identical information shown in Figure 7.

Authors:

Figure 7 belongs to Section 3.3 and shows results for the tests following the criterion that we called *mcs* (mean catchment slope) in the text.

Table 2 belongs to Section 3.4 and shows the results for the new runs using the slope dependency at the grid cell level (adopting the linear dependency on grid cell slope detailed in the section).

Page 9 Line 13: The acknowledgement that soil characteristics have an important role really highlights the point that the dependence on slope is likely missing other important catchment characteristics. Did the authors consider using a multivariate approach when formulating S0/Smax?

Authors:

We decided to stick to the first order signal given by the slope in regards to the surface runoff parametrization. But yes, we agree that the issue of sub-surface flow still needs to be addressed in the model as highlighted by the poor performance at the Avon catchment. However, our development has been made in the context of the state-of-the-art capability in the JULES model, that either uses the PDM approach (saturation excess surface runoff and free drainage as sub-surface runoff) or the TOPMODEL approach (baseflow using subgrid topographic characteristics as sub-surface runoff). We attempted to discuss this and what we think the way forward should be in the "Discussion" section (second paragraph, will be third after this review).

REFERENCES

Beck, H.E., van Dijk, A.I., de Roo, A., Miralles, D.G., McVicar, T.R., Schellekens, J. and Bruijnzeel, L.A.: Global-scale regionalization of hydrologic model parameters. *Water Resources Research* 52(5), 3599-3622. doi:10.1002/2015WR018247, 2016

Blyth, E.M., Martinez-de la Torre, A. and Robinson, E.L.: Trends in evapotranspiration and its drivers in Great Britain: 1961 to 2015. *Hydrol. Earth Syst. Sci. Discuss.* 2018, 1-47. doi:10.5194/hess-2018-153, 2018

Crooks, S., Kay, A., Davies, H. and Bell, V.: From Catchment to National Scale Rainfall-Runoff Modelling: Demonstration of a Hydrological Modelling Framework. *Hydrology* 1(1), 63. 2014

Lewis, H.W., Castillo Sanchez, J.M., Graham, J., Saulter, A., Bornemann, J., Arnold, A., Fallmann, J., Harris, C., Pearson, D., Ramsdale, S., Martínez-de la Torre, A., Bricheno, L., Blyth, E., Bell, V.A., Davies, H., Marthews, T.R., O'Neill, C., Rumbold, H., O'Dea, E., Brereton, A., Guihou, K., Hines, A., Butenschon, M., Dadson, S.J., Palmer, T., Holt, J., Reynard, N., Best, M., Edwards, J. and Siddorn, J.: The UKC2 regional coupled environmental prediction system. *Geosci. Model Dev.* 11(1), 1-42. doi:10.5194/gmd-11-1-2018, 2018

Mizukami, N., Clark, M.P., Newman, A.J., Wood, A.W., Gutmann, E.D., Nijssen, B., Rakovec, O. and Samaniego, L.: Towards seamless large-domain parameter estimation for hydrologic models. *Water Resources Research* 53(9), 8020-8040. doi:10.1002/2017WR020401, 2017

Schellekens, J., Dutra, E., Martínez-de la Torre, A., Balsamo, G., van Dijk, A., Sperna Weiland, F., Minvielle, M., Calvet, J.C., Decharme, B., Eisner, S., Fink, G., Flörke, M., Peßenteiner, S., van Beek, R., Polcher, J., Beck, H., Orth, R., Calton, B., Burke, S., Dorigo, W. and Weedon, G.P.: A global water resources ensemble of hydrological models: the eartH2Observe Tier-1 dataset. *Earth Syst. Sci. Data* 9(2), 389-413. doi:10.5194/essd-9-389-2017, 2017

Wagener, T. and Wheater, H.S.: Parameter estimation and regionalization for continuous rainfall-runoff models including uncertainty. *Journal of Hydrology* 320(1), 132-154. doi:10.1016/j.jhydrol.2005.07.015, 2006