GENERAL COMMENTS:

This manuscript analyses the effect of varying the values of seven parameters of the JULES model on the validity of simulated runoff and river discharge in three southern African catchments. The manuscript is generally well-written. The results are discussed in an appropriate and balanced way and the figures are of a good quality. I especially enjoyed reading the discussion section which was interesting. However, the description of PDM and TOPMODEL were difficult to understand (I will explain this later). There are also a number of other important issues that require improvement in my view (also explained below). For example; insufficient explanation is given on the method of selection of the 7 parameters that were evaluated in this manuscript. The paper fits within the scope of HESS because the JULES model (which is applied in the paper) is an integrated model (modelling much more than just hydrological processes) designed for global application. However, the audience for this paper is very limited: The paper is probably only useful for JULES modellers specifically interested in the hydrological components of JULES. I do not have very detailed knowledge of JULES, TOPMODEL and JULES' routing model, so I was perhaps not always fully able to judge the validity of statements in the manuscript.

With regard to the concerns about the relevant audience for the paper, we argue that its relevance is not only limited to users of the JULES model, but also to the broader land-surface modelling community. In particular, we feel that the issues we have encountered regarding the surface/subsurface runoff partitioning in the Okavango and Zambezi are not necessarily unique to JULES and could offer insight into the performance of other models in the region. With land-surface models run on regional and global scales, it is not always the case that the performance of the model is evaluated in detail for individual catchments within the domain. For example, in a JULES-TOPMODEL implementation Gedney and Cox (2003) calibrate f according to global mean RO/P ratio and thus do not directly account for catchment-specific sensitivities to this parameter. We wish to highlight the importance of a more detailed and critical evaluation of such parameter sensitivities in LSMs.

SPECIFIC COMMENTS:

Line 12-13, page 11095 "To simulate streamflow in river catchments, runoff routing schemes are also now widely used": It is not clear why it is desirable to have runoff routing in an LSM as streams and rivers probably only account for a very small proportion of the energy, water and trace gas exchanges between land and atmosphere (i.e. it would seem to make sense to represent streams and rivers in a simpler way).

It is true that interaction between river channel processes and regional-scale atmospheric processes is likely to be very small. In fact in JULES there is no representation for direct evaporation or sensible heat exchange from the open river channel. However, there are reasons why such a scheme can be useful in an LSM. It offers a way to evaluate the aggregated model runoff at a catchment scale by directly comparing against observed streamflow. This is very useful when assessing how the water balance of a catchment is simulated by the model. The regional or global scope of an LSM also allows for the simultaneous evaluation of numerous catchments under a common modelling framework. From the perspective of integrated earth system modelling, it is logical to include an online routing scheme rather than directing the grid-point output of the LSM to an offline routing model.

Lines 6-7, page 11096, "relationships to measured physical properties": Could these not also have the same equifinality problem as mentioned previously (Presumably not because they are probably not very complex, but this is not clear from the text)?

As the reviewer suggests, a simple statistical model is less likely to be subject to equfinality. What is probably a more relevant issue here is choice of model and predictor variables used to derive such a relationship and whether this model is conceptually an appropriate way to estimate such parameters.

Line 2, page 11097, "at monthly timescales": JULES should probably work well for 6-hour time steps because it was designed to be coupled with a GCM which probably has 6-hour time steps. Therefore it surprises me that a monthly time step was used for this paper. Could the authors explain why this was done?

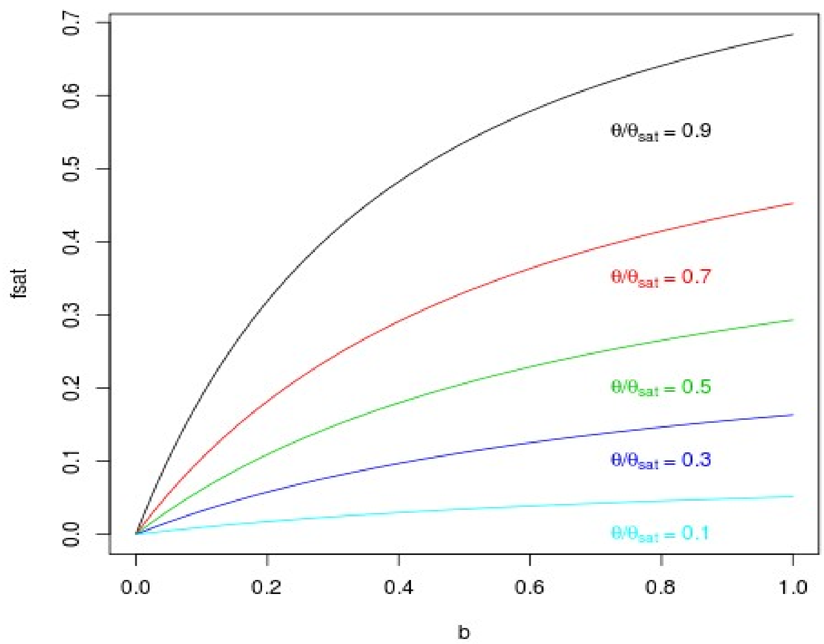
The time step of the meteorological forcing data is in fact 3 hours and the model time step is 1 hour. Forcing data are temporally interpolated to match the model time step. The monthly time scale that is referred to is simply how we have aggregated the model output for evaluation, which is consistent with the context of the intended use of the model (ie. long-term climate-hydrology studies). We can insert the following sentence into section 2.3 (Line 22, page 11101) to clarify this: "WFD input fields are supplied at a 3-hourly time resolution and are interpolated within JULES to match the model time step of 1 hour."

Lines 8-9, page 11097, "evaluate whether one offers a consistent advantage over the other": I think that the number of test catchments should be larger than 3 for such an evaluation. Could the authors comment on this?

We agree with the reviewer here. Although the 3 catchments provide interesting case studies for comparison, we agree that a fully-representative evaluation is not possible. We can remove this sentence from the revised manuscript.

Line 11-13, page 11098, "A higher value of b results in fewer high capacity ("deep") stores relative to low capacity ("shallow") stores and so will result in a more rapid production of surface runoff": This is incorrect, because a higher value of b in Eq. 1 results in a lower value of fsat, and thus less runoff production.

We maintain that our description is correct. The figure below plots fsat as a function of b for different values of θ/θ_{sat}. A higher b results in higher fsat and hence more runoff as the fraction of the grid cell surface area that is saturated is greater.



Line 16, page 11098, "per unit contour length": This is a bit confusing in the context of a grid-based approach. Do the authors not mean "per unit of grid cell area"?

This is a bit more complicated than per unit of grid cell area, but is rather defined as specific catchment area (in m, not m^2), which is the upstream area contributing to drainage at a pixel divided by the length of the outflow contour within the pixel. It is probably not necessary to go into a more detailed description in this paper, but we can include the following two methodological references should the reader wish to know more:

Quinn, P.F, Beven, K.J., Chevallier, P. and Planchon, O: The prediction of hillslope flow paths for distributed hydrological modelling using digital terrain models, Hydrological Processes, 5, 59-79, 1991.

Quinn, P.F., Beven, K.J. and Lamb, R: The ln(a/tanβ) index: how to calculate it and how to use it within the TOPMODEL framework, Hydrological Processes, 9, 161-182, 1995.

Line 16-18, page 11098, "the representation of soil stores in PDM, which is not related to any physical characteristics of the catchment": This might be a little overstated because a modeller may choose to base values of PDM parameters on empirical relations with physical catchment characteristics that are commonly mapped (e.g. soil type or elevation).

Since PDM is a conceptual model it is not designed to represent explicitly any physical characteristics, but yes, the parameters can be empirically related to some physical properties. We can remove this sentence from the text as we feel it is not really necessary.

Line 19, page 11098, "Sub-grid variation in lambda is modelled using a gamma distribution": Presumably the authors have switched here from describing TOPMODEL to describing the specific way in which TOPMODEL is implemented in JULES. If so, then please make this clearer in the text. Also, it is difficult for me (and other readers of this paper) to judge the quality of this modelled sub-grid variation because section 2.2 ("Data sets and catchments") does not mention the observations that were used to fit this gamma distribution.

Yes, this refers to the implementation of TOPMODEL in JULES following Gedney and Cox (2003). We obtained the topographical data from the UK Met Office and these are based on the 30 arc-second DEM of Verdin, K. L., and S. Jensen: Development of continental scale DEMs and extraction of hydrographic features, Third International Conference Workshop on Integrating GIS and Environmental Modeling. Santa Fe, New Mexico, 1996.

Lines 23-24, page 11098, "an additional storage layer beneath the standard 4-layer, 3m deep soil column" For those that do not already know JULES and TOPMODEL, it is not clear whether this ’standard 4-layer, 3m deep soil column’ is standard to JULES, or standard to TOPMODEL (presumably the former).

This is the standard configuration for JULES.

Line 17, page 11098, Equation 2: This equation does not convey any information that was not already conveyed by the preceding sentence, so it seems redundant to me. In addition, this equation is only about the fifth layer whereas subsurface runoff is generated from any layer below or containing the top of the water table. Thus Equation 2 seems to govern only a small part of the generation of subsurface runoff. Why is subsurface-runoff generation from other layers not explained in the paper?

We feel this is relevant because of the sensitivity of total runoff to changes in the f parameter, as shown in the results. It is perhaps surprising that this equation does indeed have a marked impact on total runoff as we show in the results. In the 4 upper layers Ksat is uniform (we will add this to the text) and, as already noted, subsurface runoff occurs from and layer intersecting with or below the water table.

Line 3, page 11101, "Of particular interest for the selected catchments is the contrasting geological environments represented.": Please also indicate why the geological environments of the Okavango and the Zambezi catchment are contrasting (it is not mentioned anywhere in the paper).

The headwaters of the Okavango and Zambezi do in fact share similar geological characteristics in that large parts of the upper Zambezi consist of extensive floodplains underlain by deep Kalahari sand deposits. So the contrast is really between the Orange on one hand and Okavango and Zambezi on the other. We will clarify this in the text.

Line 21-22, page 11101, "The grid resolution is 0.5": This is much finer than the resolution of GCM's (to which JULES was designed to be coupled). Is it reasonable to assume that all JULES parameters are still valid at this finer resolution? I think it is necessary to touch upon this scaling issue somewhere in the paper.

The applicability of JULES ranges from point scale (ie. a single observation site) to GCM grid scale and so it is not outside the model's design scope to run at a 0.5 degree resolution. There will nevertheless be scale dependencies when it comes to estimating parameters. For globally-set parameters (eg. PDM and TOPMODEL exponents) it is reasonable to argue that by calibrating for a specific domain there is some scale-related dependency that is implicitly accounted for by performing this adjustment. This could be tested by running the model at different resolutions and seeing whether the same parameters yield different results. In this paper we explore only parameters related to runoff, but we cannot exclude similar scale-related sensitivities in other model components. In the case of spatially variable parameters (eg. soil texture, topographical index, soil hydraulic properties) these are interpolated from various gridded datasets and so will have a spatial resolution matching the model grid, however the degree to which each grid value adequately represents that area may be problematic. This is part of the motivation behind incorporating sub-grid scale runoff parameterisations to the LSM. However, this does not entirely overcome the scale issue as Beven (2001) notes that the scale of the DEM from which a topographical index is calculated can affect the values of the parameters and hence the sub-grid runoff variation. Gedney and Cox (2003) also note that the 30 arc-second (~1km) DEM resolution from which they produce the distribution of sub-grid topography may be too coarse to accurately resolve hills and valleys.

Line 26-28, page 11101, "Initial experiments identified that river discharge simulated by the PDM scheme is highly sensitive to the b shape parameter and that TOPMODEL is most sensitive to the f exponent (see Sect. 2.1).": Please mention the other parameters and configurations that were tested, and how this was done.

The sensitivity of the JULES-TOPMODEL configuration was tested for sensitivity to changes in the topographical index (TI) mean and standard deviation. These values were altered within the range of values observed across the domain in the ancillary field. It was found that although the results were sensitive to changes in TI parameters, these were considerably smaller than for f and so these experiments were omitted from the analysis. Furthermore, since the TI parameters have been derived through an objective method it wouldn't really make sense to adjust these arbitrarily. This test was therefore only done to ascertain the relative importance of uncertainties in the estimation of the TI compared to changes in f. The conclusion from this is that the inclusion of an improved spatial map of TI parameters is not likely to greatly improve the simulation results. Another parameter that was tested with little effect were the maximum water table depth in TOPMODEL.

Lines 6-7, page 11102, "mean and standard deviation of the topographical index (TI)": Presumably, this is the data that was used to fit the gamma distribution mentioned in section 2.1. If so; then this dataset should be mentioned in section 2.2. It is difficult for me (and probably for other readers of this paper) to judge the quality of this modelled sub-grid variation without having an indication of the density and type of observations underlying this "spatially-varying ancillary field".

This is the same issue that we have addressed in our reply to the previous comment concerning Line 19, page 11098.

Line 5, page 11102: "within the range of previous regional implementations of the respective schemes" Please clarify whether this means implementations of these schemes within JULES, or all implementations of these schemes.

This refers to implementations both within JULES and other modelling structures. We will clarify this in the revised text.

Line 9-14, page 11102, "There ... (TOP1.0cr).": Please explain how those 5 routing model parameters were selected.

The initial wave speed values were taken from a previous implementation of JULES and because these resulted in a flood peak that occurred to early in the year, these were manually decreased until the timing of the flood peak more closely matched observations. Likewise for retr (return flow from subsurface to surface river channel), the initial value of 0.00068 was take as a “default” and we varied this parameter between 0 and 1. For the Okavango and Zambezi, a value of 0.01 yielded a contribution to dry-season discharge similar to that given by 0.00068, but also an undesirable increase in peak discharge. Increasing retr beyond 0.01 did not yield any further contribution to surface river channel discharge.

Line 20-21, page 11102, "Three performance metrics are used to assess the efficacy of the model simulated monthly mean river discharge.": Can parameter criver be expected to have any sensitivity at a monthly time step given the river flow velocities and catchment areas of these three catchments?

As pointed out in our response to the previous comment regarding the model time step, the time step is not 1 month, but rather 1 hour. So, yes, changing the criver parameter does induce sensitivities at the model time step.

TECHNICAL CORRECTIONS:

Line 16-17, page 11096, "intra-annual variability in mean annual runoff": This is a word combination contradicting itself.

The sentence will be modified to read: “a pronounced seasonal runoff cycle and high inter-annual variability in mean annual runoff”.

Line 18, page 11097: "in" should be changed to "as".

This will be corrected.

Lines 20-21, page 11098, "which is calculated from moisture conditions in the soil profile": I would leave this out or modify because it seems rather obvious that the saturated grid-box fraction must be a function of (assumed?/calculated?) soil moisture conditions. Presumably, these "moisture conditions" are a modelled grid-cell average condition. If so, then please state this explicitly.

This will be changed to “grid box average moisture conditions in the soil profile”

Lines 1-2, page 11099: "an exponent" should be replaced with "a parameter".

This will be corrected.

Lines 4-5, page 11099, "When the water table intersects with the land surface, saturation excess overland flow is produced.": This sentence seems redundant because lines 21-22 on page 11098 give the same information.

We agree this is redundant, but see no harm in leaving it here.

Line 1, page 11101, "discharge": I would either use the word "discharge" or "flow" (not both).

“Flow” will be replaced by “discharge” where appropriate.

Line 18, page 11104, "TOP0.1": This should be changed to ’PDM0.1’.

This will be corrected.

Lines 20-21, page 11104, "This dry-season flow is absent in the PDM simulations.": This sentence might confuse some readers. Presumably, this sentence means that the simulated dry-season flow has negligible values. However, one might think that it means that the PDM model does not include dry-season flow.

The sentence will be changed to: “The PDM simulations produce negligible values for this dry-season discharge”.

Line 1, page 11106, "along with a": These words should be removed.

This will be corrected.