The paper presents a potentially interesting evaluation of the hydrological model part of a coupled land surface model. However, there are a number of problems with the paper.

The introduction does not take into account recent literature on hydrological modelling in general (PUB, uncertainty, models for hypothesis testing, sensitivity analysis, etc.) of which there is a lot, nor does it take into account the hydrological modelling literature on the specific region and catchments included in the paper. This is a serious issue given the generally poor results obtained from this study compared to much better results for other models.

We find that the volume of relevant modelling literature for the region is relatively small, but we agree that there are some studies that we have overlooked that provide a basis against which we can compare our results. For the Okavango, Wilk et al. (2006) noted that at the time of their study there had been much work done within the Delta, but little on the overall basin dynamics. They suggest that this is partly due to lack of rainfall and discharge observations in Angola. Since then Hughes et al. (2006, 2011) have implemented the Pitman rainfall-runoff model over the Okavango catchment. Hughes et al. (2006) calibrate the model with a rainfall product derived from the limited available observations from within the catchment and satellite estimates (Wilks et al., 2006) and find the model does an adequate job of replicating observed discharge. Their values for Nash-Sutcliffe efficiency (E) are generally above 0.6 and absolute error < 5 % for 16 sub-catchments. Interestingly, they found that performance was generally lower (-1.115 <= E <= 0.726) for the sub-catchments in the eastern headwaters, which are underlain by thick Kalahari sands and characterised by low drainage densities, high baseflow and low seasonal variability. This suggests that their model has difficulty representing baseflow-dominated catchments. In Hughes et al. (2011) the same model is calibrated with input from the University of Delaware (UDEL) rainfall product and they find that the model does not represent well the seasonal distribution of runoff in the catchment, with the magnitude of peak discharge underestimated and the timing of the peak shifted one month too early. They attribute this weakness to the unreliability of the UDEL data post the early 1970s period where rainfall records in Angola cease. Uncertainty in simulated output resulting from uncertainties in rainfall input for the Okavango catchment are demonstrated by Milzow, et al. (2011) who show large differences in simulated discharge when using different rainfall products to force the Soil and Water Assessment Tool (SWAT) model. Their model also simulates the seasonal discharge peak occurring too early in the year.

In the upper Zambezi catchment, Winsemius et al. (2006) compare a gridded GIS-based model (STREAM) with a semi-distributed conceptual model (Lumped Elementary Watershed; LEW) and found the latter to represent better the actual processes in the catchment.

Regarding the general modelling topics that the reviewer mentions, this is a rather broad topic. Since the paper goes to a hydrological journal, we believe the audience should be familiar with recent advances in general hydrological modelling concepts. Additionally, we are not introducing new concepts, but are rather assessing the usefulness of a particular type of a model in a particular context.

On page 5 starting at line 8 reference is made to the processes explicitly simulated but does not mention subsurface runoff or GW recharge (these are only mentioned later). There are also many other aspects of the models that could be inconsistent with the hydrological processes that are dominant in some of the catchments used. Deep groundwater conditions do not seem to be modelled and the models rely on intersection of the GW with the soils. The models do not seem to account for surface runoff that is driven by rainfall amounts only (rather than wetted soils through saturation excess). The latter will be a problem for the Orange River, as the authors appear to have discovered.

The sentence in question is simply referring to the surface processes, whereas the subsurface are, as the reviewer points out, introduced later. We see no reason why this is not appropriate. JULES does in fact represent surface runoff due to infiltration excess at the grid box scale. We will clarify this in the revised text.

I found that the model descriptions were generally inadequate given that the model is being evaluated in the context of likely reall hydrological responses.

It is not clear from this comment how the model descriptions can be improved.

P8, L14: There appears to be an implication here and later (in the discussion) of the results that the MOD16 data will not be biased and therefore the model outputs can be assessed relative to the MOD16 data. I do not think that there is any evidence that can be presented to confirm this and in fact the MOD16 data might be totally wrong (as we have found in other parts of southern Africa). The other point is that if the JULES model gets the streamflow volume correct for the correct rainfall, then how can the evaporation be a long way off (i.e. the Jules model does close the water balance). Perhaps MOD16 is OK for Okavango and Zambezi but not for the Orange?

This is not what we meant to imply as we are aware that the MOD16 product may be unreliable in some regions. We do point out that this is a comparison between two model outputs, rather than model vs. observation. A recent assessment of various ET products over Africa was done by Trambauer et al. 2013. From this it appears that MOD16 strongly underestimates ET over the upper Orange, and also underestimates in the Okavango and Zambezi, but not by as much. We will include this reference in the revised text.

Regarding the second point, we are comparing ET from JULES model simulations with ET from MOD16. The somewhat implicit assumption is that our best models have the water balance right. If so, the discrepancy between model ET and MOD16 is either because MOD16 is wrong, or because we got our water balances right for the wrong reason. We do discuss these issues in p. 11110.

Some of the results for the Okavango seem to ignore some of the processes that are possibly active in large catchments with extensive floodplains. Why is there no mention of the extensive amount of previous hydrological modelling work that has been done (and published) on the Okavango? P15 refers to the lack of water balance studies in the Okavango headwaters - some of these have been done as aprt of previous modelling studies.

It is not clear which processes the reviewer is specifically referring to here. One potentially major limitation of the routing scheme in JULES is that it does not account for channel transmission losses, which are important in the lower semi-arid reaches of the Okavango (Hughes et al., 2006). JULES does include a scheme to simulate floodplain inundation whereby water is removed from the river channel to flood low-lying topography, from where it is allowed to evaporate. We did test this scheme for our domain, but found it had a negligible impact on simulated discharge. This is, however, something worth exploring further. We note that Wilk et al. (2006) have produced water balance estimates for the Okavango that we will refer to in the revised text. Regarding reference to other modelling work in the Okavango, we note this above in our reply to the first comment.

I would be interested to know where the authors expect to get reliable soils data from and how this will be integrated at the very large grid scales that they are using.

The soil data is provided by the FAO global soil type map and the dominant type for each JULES grid box is used.

I get the impression that the authors have applied a very limited conceptual view to the hydrological processes that are likely to be dominant in these catchments and that view has been cloded by the limitations of the two models selected for use. I would argue that this is one of the problems with adopting a very coarse grid scale type of modelling approach. If one looks at the structure of the TOPMODEL it is quite clear that it cannot work form amny catchments in the drier and more mountainous areas of southern Africa simply because the dominant processes are not represented. Models that do represent deeper and slowly draining groundwater as well a infiltration excess type surface runoff (even simulated implicitly) have been frequently deomonstrated to be at least reasonably successful. Unfortunately, the authors appear to have decided to ignore the 30+ years of modellign experience in the region.

The aim of this study is not develop models for specific catchments, but to evaluate the performance of an existing land surface modelling system (LSM) for individual catchments. We are aware that better models exist for representing individual catchments and can be calibrated to closely match observed discharge, but our interest is in the performance of a physically-based, minimally calibrated LSM that is applied over a large area covering multiple catchments. This is important in the context of how LSMs are typically applied, ie. as coupled components of regional or global climate models, where resolution is coarse and intensive calibration procedures are generally not applied. Through an evaluation such as the one we present, we hope to understand better the limitations of the LSM at simulating hydrological processes in the context of long-term climatic change and variability.

Specific comments:

P4, L12: What is meant by an ’unstable’ regime? Do the authors mean ’variable’?

“Unstable” refers to the non-linear response of runoff to a change in rainfall demonstrated by de Wit and Stankiewicz (2006) whereby changes in runoff caused by a specified decrease in rainfall are progressively greater as mean annual rainfall decreases.

P4, L14: I am not convinced about the existence of thresholds and the authors do not support this statement with any literature or other evidence.

This statement refers to the work of de Wit and Stankiewicz (2006), which is cited at the beginning of the sentence.

P4, L25: It wuld be a good idea if the authors look at recent literature (by Beven and others)on the use of models for hypothesis testing.

Since the methodology doesn't include formal a hypothesis testing or limits-of-acceptability approach, it is probably not best to include a discussion on this when outlining our objectives. We could, however, include some discussion on this and how is relates to our results in Section 4.

P7, L9: 10% of what?

This refers to differences in mean monthly rainfall between the WFD and WFD-ERA-Interim products.

P8, L10: Please specify if this is actual evapotranspiration.

Yes, this is actual ET.

P10, L3: What is meant by mean climatolgies when the graphs show rainfall and streamflow (these are not climatologies).

We use the word in the sense that the graphs show long-term monthly mean values, which is typically referred to as a climatology when presenting an atmospheric variable. If this is not appropriate terminology when referring to river discharge, then we can change this to "mean monthly values".

P12, L5: Grammatical error - repetition.

This will be corrected in the revised text.

References: There are some page numbers missing from the references and these are not replaced with doi numbers.

These were pointed out by the copy editor and were corrected for the online discussion paper.

References not already listed in the manuscript:

Hughes, D.A., Andersson, L., Wilk, J. and Savenije, H.H.G.: Regional calibration of the Pitman

model for the Okavango River, Journal of Hydrology, 331, 30–42, 2006.
Hughes, D.A., Kingston, D.G. and Todd, M.C.: Uncertainty in water resources availability in the Okavango River basin as a result of climate change, Hydrology and Earth System Sciences, 15, 931–941, 2011.
Milzow, C., Krogh, P.E. and Bauer-Gottwein, P.: Combining satellite radar altimetry, SAR surface soil moisture and GRACE total storage changes for hydrological model calibration in a large poorly gauged catchment, Hydrology and Earth System Sciences, 15, 1729–1743, 2011.
Trambauer, P., Dutra, E., Maskey, S., Werner, M., Pappenberger, F., van Beek, L.P.H., Uhlenbrook, S.: Comparison of different evaporation estimates over the African continent, Hydrology and Earth System Sciences Discussions, 10, 8421–8465, 2013.
Wilk, J., Kniveton, D., Andersson, L., Layberry, R., Todd, M.C., Hughes, D., Ringrose, S. and Vanderpost, C.: Estimating rainfall and water balance over the Okavango River Basin for hydrological applications, Journal of Hydrology, 331, 18–29, 2006.
Winsemius, H.C., Savenije, H.H.G., Gerrits, A.M.J., Zapreeva, E.A. and Klees, R.: Comparison of two model approaches in the Zambezi river basin with regard to model reliability and identifiability, Hydrology and Earth System Sciences, 10, 339–352, 2006.