We thank the reviewer for their comments. Each comment is addressed below with the original review in italics and our responses in normal font.

General comments:
This paper try to analyze the impact of changing the MOSES land surface scheme by the CABLE land surface model in the ACCESS climate model. The introduction is very short and any references/discussion about land/atmosphere coupling processes (c.f. Betts et al. 1996; Betts 2009) were done. Only 2 references on previous works analyzing land surface-atmosphere interactions were done without any discussion.

We will provided extra detail in the introduction especially around the Betts 2009 paper and recent work by Hirsch et al.

In the section 2, differences between the both land surface modules are relatively well described even if some details are missing.

As requested by reviewer 2 we now include a table of differences including some further details about the snow scheme.

The section 3 presents the experimental design of the study. There is only two sentences that present the observations used to evaluate the model, that underline the poor scientific quality of this study. ERA Interim reanalysis alone cannot be used as observations. There is many "real" observations available to evaluate the model.

The main purpose of this paper was a comparative study of simulations with different land surface schemes. For this reason most of the results presented are model-model comparisons (and a process understanding of these model differences) rather than model-obs comparisons. Where we show zonal mean cloud and precipitation we will add an observational dataset and will now show land only zonal means to be compatible with the GPCC dataset. We understand the limitation of only using ERA interim reanalysis but have now checked our model simulations against CRU temperatures as well. This will be noted in our comments on the relative bias of each model temperature compared to ERA Interim/obs.

The section 4 presents the main results. The text is very descriptive and generally boring even if the fact to use off-line runs to explain some in-line behaviours is a very good idea.

We will revise section 4 to remove some unnecessary detail and to try to make it easier to read.

Any tests of significance is done for all differences model versus observations and model versus model shown in this manuscript.

See reply below about significance testing.

The conclusion is well written and brings into focus the qualities and the defaults of this study. After having hesitated for a long time between rejected this paper or reconsidered it after major revisions, I think that this paper must be largely improved and it deserves a chance.

We will endeavour to address both reviewer comments, noting that at times they had substantially different views on the relative importance of different parts of the paper and how it might be improved. We believe that our changes will deal with all critical issues and will result in an improved manuscript.

Specific comments:
P.1-2: The introduction did not give the readers a sense of the state of the art, e.g., what are the land/atmosphere coupling processes, and which ones are important or not? Did anyone else in the community attempt to analyze land/atmosphere coupling for global climate simulations? Only one sentence to sum up the work of Koster (2004) or Seneviratne et al. (2010) is not enough.

As our response above we will extend the introduction to give more information about the cited work and to add some other relevant literature.

P.1, l.16: I am not sure that LSM is a "key" component of a climate model, it is an important component but the key component is the atmospheric model (or the oceanic model).

Sentence will be modified to note that LSM is one of the key components.

P.3, l.25: What is the value for Cs or how it is computed?

Cs is a volumetric heat capacity calculated as the weighted sum of the heat capacity of dry soil, liquid and ice ($JK^{-1}m^{-2}$) and this will be added to the manuscript.

P.3, l.27: What is the value for fr or how it is computed?

Fr is $1\text{-}e^{LAI/2}$ and this will be added to the manuscript.

P.4, l.19: What is the value for c in the exponential or how it is computed?

C is an extinction coefficent for beam radiation and black leaves and this will be added to the manuscript.

P.4, l.20 and l.30: Are you sure that CABLE "has a more complex representation" of canopy or displacement height than many LSMs? If so, prove it and explain the main differences with "many other LSMs".

Most other LSM use conventional rough wall boundary layer theory and canopy parameters, e.g. displacement height, that are a constant fraction of canopy height. CABLE more complex representation is based on Raupach 1989, 1994 which are referenced in the paper. We will note in the text the simpler methods used by most other models but do not feel it is necessary to repeat all the detail on the canopy turbulent transport described in the Raupach references.

P.5, l.17: "MOSES does account for this heat exchange." These sentences are not clear to me.

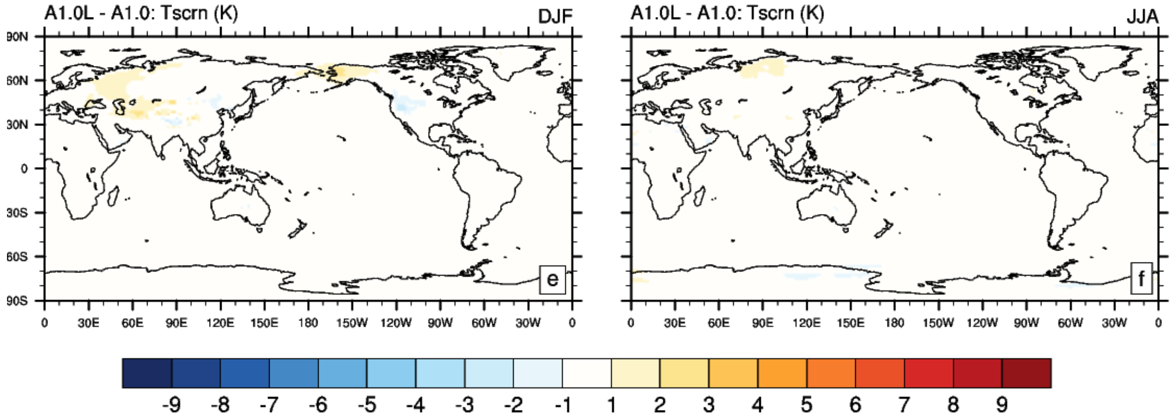
This will be clarified - Moses calculates the advection of heat by moisture fluxes within the soil column.

P.5, l.31 : The description of snow modules are very short while in the paper you access that "Warmer winter temperatures simulated by ACCESS1.1 over the snow covered areas of mid and high latitudes are attributed to differences in the snow parameterization in CABLE compared with MOSES." Please provide more details. Does CABLE "has a more complex representation" of snow processes than many other LSMs?

Differences in the snow parameterisation will now be captured in the new table of CABLE-MOSES model differences which was requested by reviewer 2.

P.6, l.27: LAI data are not the same for MOSES and CABLE. How you can be sure that most of the impacts in summer are not due directly to this difference rather than in the differences in land surface physics? This fact must be addressed in order to improve the scientific quality of this paper. Normally, in this paper, the same LAI should be used in both models, perhaps by introducing an intermediate simulation (MOSES with LAI from CABLE or inversely).

We have now performed a simulation in which ACCESS1.0 (MOSES) is run with ACCESS1.1 (CABLE's) LAI. The change in LAI has very little impact on the simulation. The figure below shows the difference between the sensitivity test and the original A1.0 and should be compared with Fig 4 e/f from the paper. The very small differences shown do not explain the A1.1-A1.0 differences from Fig 4. Appropriate comments will be added to the manuscript when explaining the temperature differences found for the Northern continent in summer and winter.



P.7, l.: This is the most important negative point of this work. Any "real" observation was used. For continental precipitation, there is product: GPCC (certainly the best), GPCP, TRIMM (over tropics), etc. Try to use these products in your comparison. The same is true for temperature or cloud: CRU data gives Tasmin/Tasmax/Tasmean that allow to evaluate the diurnal cycle; CERES, MODIS, etc. for cloud (see Pincus et al. 2012)

As noted above we will add some extra comparisons with other observational products where the paper currently focusses on model differences from observations (Fig 3 & 4 a-d). The rest of the paper focusses on model-model differences and a process understanding of these differences and hence additional comparisons to observations are less relevant for that analysis.

P.8 and after: The result and figure part are too descriptive and not enough scientific.
Tests of significance (for example T-Test) must be done on the differences model vs obs as well as for model vs model. On figure, generally, the pattern where the differences are significant is shown with dotted panel. Without that, this article cannot be accepted. Please only comment where the differences are significant statistically.

In general significance testing shows that model-model differences are significant almost everywhere (see figure below for seasonal mean temperature) while the significance of model-obs differences is both seasonally and model dependent. Model-obs significance testing is consequently useful for determining if one model better simulates observations than the other model. Testing against both CRU and ERA Interim temperatures suggests that the models can't be differentiated in DJF but that A1.1 (CABLE) produces the better simulation of temperature in JJA. T-test stippling will be added to Fig 4 a-d but we will not add stippling to the other figures of model-model differences since it would be required almost everywhere and tends to obscure important features of these figures.