**General comments:**

The premise of this study, which is "to systematically evaluate the ability of land surface models to simulate [biological and physical] processes [and ecosystem dynamics] during soil moisture deficits" (lines 7-9, pg 10791) is a critical scientific objective for hydrological, ecosystem and climate change research and is a prerequisite for making predictions about how the function and services of different vegetation types will be altered by anthropocentric forcings in the coming century. The subject matter and scope of this study is appropriate for HESS. This study has potential to become a high impact and well cited paper. However, in this current version, this manuscript falls short in making significant advances in model understanding and in convincing me that they have appropriately interpreted their statistics. Therefore, this manuscript needs significant revisions before it should be considered for publication in HESS.

First, it strikes me as strange to use the variation that soil texture produces in model output as a way to evaluate the skill of the model to capture Qe of particular flux towers with a known soil texture. We would expect different soil textures to produce different magnitudes of Qe for different vegetation types. We would also expect different soil textures to produce different patterns of Qe during drying periods for different representations of the soil physics. These two results are indeed shown by the red curves in Figures 6 and 7. But what is not clear to me is how the observations relate to the variation produced by these two alternative formulations and by the contrasting soil type. Without observations of the same vegetation under the same climate, but growing on different soil types, there is no way to tell which model formula is correct. In other words, how would the observations change if the vegetation were growing on a different soil type? The default formula predicts that Qe would behave one way while the new formulation predicts that Qe would behave another way—but which one is correct, you cannot tell from the information presented in this study.

The same case as above can also be made for LAI since the authors prescribed it rather than evaluate the model's ability to predict it. Therefore, the authors should use the known LAI as a constraint on Qe in order to understand other aspects of the model that are poorly constrained. In contrast, how gs is regulated is not known and therefore, this type of comparison to observations does make sense.

Second, it is not clear how to interpret Figures 3, 4, and 8. We would expect NME and MBE to get increasingly large when the model is configured for a soil type that is not consistent with the known soil type for the flux tower. But, that doesn't mean the model is performing poorly ("performance" ln13 pg 10799). Indeed the model may be performing correctly for the vegetation that its soil type is configured for. In fact, I would be alarmed if NME = 0 even though soil texture was changed. Therefore, it should be argued that NME values near 0 or 1 indicate that the model is performing poorly when the soil type in the model does not match the observed soil type. But this distinction is not clear in the manuscript. More importantly, what can we really learn from the reported NME and MBE values when all values are groups together even though there is a mismatch between LAI and soil texture for some of the $M_i$ and $O_i$ values, but not others? How do they inform us in terms of model development when some of the values being put into $M_i$ and $O_i$ are not the same thing?

Third, the scope of the study is of limited appeal as it is presented. The manuscript is written as if it were speaking mainly to those interested in modeling the soil boundary condition for a land surface model. The model is just a tool for gaining more detailed understanding (or making predicitons) about the system of interest. The study would be appealing to a much broader audience if the authors described what the predictions of the competing hypotheses (i.e. parameterizations) mean in terms how we understand ecology, physiology, and hydrology in a world with a changing climate, and not make the central focus of their discussion simply about model errors. As one example, the authors used two alternative formulas for gs, each representing very different hypotheses about stomatal regulation. Interestingly, the models predicted that the mode for stomatal regulation has very little effect on Qe during periods of water stress for all sites except Howard Springs (Figs. 6 & 7). This is a remarkable result with significant ecological, hydrological and climatological implications that needs to be expanded upon in the Discussion. There are many other example as well. After reading this paper, I did not come away with a clear sense about new hypotheses to test, observations and experiments to make, and model formulas to develop. (See also Specific comments 1c and 1d).

Third, there is a considerable amount of information contained in the figures that should be flushed out in order to give greater clarity about the relative contribution each parameterization contributes to the variability in Qe. Take Figure 5 for example (but this comment pertains to all the figures), all of the "alternative LAI, gs, and soil parameterizations" (Fig.5 caption) are all mixed together to show the variation of the time series of Qe. Does one particular parameterization account for most of the variation on either the high end or the low end? If not, say so in the discussion. If so, what does the sensitivity (or lack thereof) to a particular parameterization mean in terms of the ecology, hydrology, physiology, and climatology of the different systems? What are the implications of the predictions of the different parameterizations? Constructing the analysis and discussion in this manner will give much clearer guidance to modelers and empiricists about modeling, experimental and observational needs.

**Specific comments:**

1. The message of this paper needs to be tighten-up considerably throughout the existing text and expanded upon in the Methods and Discussion. For example:
   a. The Introduction is not particularly focused. It would be helpful if the Introduction were organized around a Problem Statement that is explicitly articulated at the beginning. The Problem Statement should address the culminating result of the study (i.e. lines 4-7, pg. 10804). Unfortunately, the reader has to get all the way to lines 7-9, pg 10796 before they encounter the actual Problem Statement that this analysis attempts to resolve.
   b. Lines 19-25, pg 10791. Why do these models get these results and how do these results relate to the Problem Statement? In other words, what is the rationale for focusing on soil physics instead of biological processes? There is a huge body of literature that suggests we need to emphasize improving our understanding and representation of biological processes such as phenology or plant water-transport, rather than focusing on improving the soil boundary conditions.
   c. The Methods need to include equations for all of the alternative parameterizations examined. The Methods also need to include a Table of parameters and parameter values to maximize the transparency and reproducibility of this study.
   d. The Discussion needs to map out how the equations and parameters (i.e. from 1b above) explicitly link to the different Results illustrated in the Figures. Without doing 1b and 1c, the model remains a bit of a black box, and therefore, it is difficult for modelers to know how to

improve the existing formula and what specific parameters are controlling the output. Making these linkages is also important for informing empiricists on which field measurements should be prioritized.

e. The influence of the "slope parameter" seems to be a key finding, yet it is given very little attention at the end of the Results and there is no mention of it in the Discussion. The authors state: "The slope appears more critical for simulation of Qe than the other parameterizations investigated here and has strong effect on the magnitude of the **fluxes primarily during dry-down**" (lines 19-21, pg 10803). The authors also state "our goal was to determine whether CABLE can **capture dry-down** associated with rainfall deficits as the components of the model are varied [among which is the hydrology scheme and slope parameter], or whether the model lacks the mechanisms to simulate this phenomenon" (lines 9-11, pg 10808). [Bold type face is the Reviewer's emphasis.] The authors fall short on meeting this goal when they fail to mention the role of one of the most "critical" parameters in the Discussion.

f. Many statements throughout the Discussion need to clearly reference a figure (a few examples are given below). Also, each figure published in the Results section needs to be referenced and discussed in the Discussion section. Otherwise, any figure that is not discussed in the Discussion section should be moved to the Supplement because it is clearly not central to the main message of the study; rather, it is just supporting information.

g.

2. Lines 23 & 28, pg 10800. "Likely due to" This is speculative in both cases. The beauty of using a model is that you can know these two things. By not exploring the output and knowing these for sure, statements like these are not very useful for either modelers or empiricists because they do not unequivocally tell us where to concentrate our efforts (or even worse—speculative statements can lead us down the wrong road). Also, "drying soil" and "compensating errors" both need to be quantified and demonstrated.
3. Lines 21-22, pg 10806 "high soil evaporation **may** result from…" This is speculative. The authors can know this with closer inspection of the canopy turbulence output of their model.
4. Lines 3-4, pg 10807. "seasonal droughts". Do you mean dry season? I am not sure what a "seasonal" drought is. Droughts by definition are some type of water-deficit anomaly--be it measured in terms of rainfall, soil moisture, streamflow, etc—and anomalies are not seasonal, they are atypical. This is an important distinction to make because vegetation in areas with dry seasons are adapted for those dry seasons. However, depending on its severity, the plants may not be adapted for a drought that is layered on top of a dry season, which could be an important ecological filter for certain species as climate changes.

**Technical comments:**

Lines 24-26, pg 10790. Awkward sentence. Reorganize as: "LSMs form an integral part of global climate models by controlling how net radiation is partitioned…"

Lines 22 & 23. "87%" and "66%" These do not match Table S2.

Line 2, 10798. "empirical approach" What is this? Eloborate.

Lines 16-17, pg 10800. “Overall, both hydrological…” This sentence is not really true for all sites. E.g. see Harvard Forest or Umich.

Lines 9-10, pg 10802. “due to compensating biases” What are these?

Line 3, pg 10804. “Have shown” Needs to reference a figure.

Line 4, pg 10804. “have also shown” Needs to reference a figure.

Line 19, pg 10804. “showed” Needs to reference a figure.

Line 23, pg 10804. “the contribution of LAI (Fig. Xa), gs (Fig. Xb), and soil parameterisations (Fig. Xc)” Each parameterization needs to reference their respective figures.

Line 2, pg 10805. “We identified” Needs to reference a figure.

Lines 6-8, pg 10805. Last sentence of the paragraph is not true for all sites. This sentence needs to include a qualifier at the end (before (Fig. S7)). For example, insert “for most sites”.

Lines 17-19, pg 10805. Which sites in Figure S7. Clearly reference the figure at the end of the sentence e.g. (Fig. S7a,b,d,f).

Lines 9-10, pg 10807 and elsewhere in the text. “monthly climatology”. LAI is a vegetation property, not a property of the climate. Therefore, it strikes me as confusing when LAI is referred to as being part of the climatology.

All Figures. The font size is way too small.

Figure 8. Legend needs labels. And, caption needs to state what colors go with each of the parameters.

Table S2. Check numbers on “medium soil”. Should be decimals?