Dear authors,

The subject of the manuscript is for sure very relevant for the climate modelling community dealing with land-surface processes and their interactions with the atmosphere. However, I’m sorry to conclude that this manuscript does not fulfil the expectations I have on a scientific documentation of theory, experiments and results. The reason is partly that the English language is now at such a level where it becomes difficult to actually understand what the authors wish to describe in some background and results. The reasons are also that the structure, motivation and balance of the text and results are not satisfactory. For example, the Introduction section lacks clarity (see below), Section 2.5 “Statistical analysis” refers to details that are described in later sections. It is not clear what is the objective with Section 4 “Results and discussion”. From all presented experiments and results I expect to find some indication in the end on how these experiments are ranked with respect to performance but no such ranking is reached, only a conclusion that experiments indicate in general improved performance compared to the reference experiment. Thus, there is no balance between all presented details including the statistical analysis and the overall outcome of the results. Therefore, as an overall judgement I must recommend major revision. Both language level and structure, motivation and balance need major improvements in my opinion.

The background and motivation for this manuscript as given in the Introduction is not clear enough as it is written now. For example, in lines 40-42 you state that “the evapotranspiration simulated by ... TERRA-ML ... was found to be systematically underestimated from April to October during the growing season.” But you give no reference to this statement and it is not clear over which region or regions this conclusion refers to. Is it Europe only or also other regions? Are there no publications available where this underestimation is shown? You refer to published evaporation and transpiration fractions but I miss any comment on how these fractions are estimated by TERRA-ML. Later on (lines 51-52) you state that “plant transpiration is calculated in current version of TERRA-ML with errors (Stockle, 2001)”. Here it would be good to also say what kind of errors you mean. I would also say that even if TERRA-ML, now based on empirical stomatal conductance parametrisation, would have given good evapotranspiration in
validations of hindcast simulations it can still be motivated to introduce a more advanced stomatal conductance formulation since an empirical formulation may not be valid in changing climate conditions including rising CO2 levels. But if you wish to motivate your work based on bad performance you need to show this bad performance more clearly.

Overall, the Introduction section now gives a bit jumpy feeling between very overview style paragraphs (lines 38-38, 53-63, 70-74) and on the other hand very TERRA specific comments (lines 40-42, 48-53, 63-68). Also, all the version details in lines 75-83 do not clarify much. I would recommend that you revise the Introduction to find a better balance between background, TERRA details and motivation for your work.

Detailed comments:

Line 30: In my mind, for such a very general statement like “The land surface processes significantly affect the conditions in the low-level atmosphere” it is more appropriate and respectful to refer to well recognized reviews in the area like e.g. Betts et al. 1996 than to one’s own very recent paper.

Line 49: What do you mean by “not sufficiently represented”? Please be more specific.

Line 52: “Stockle” should be “Savabi and Stockle”.

Lines 73-74: I find the sentence and statement “However, these schemes have not been implemented into production (exploitation) at convection-permitting scale” a bit strange. Okay, so you mean that dynamic vegetation should be implemented just because it is missing or for some other reason? Please be more specific.

Line 85: It is not clear now what “these limitations” exactly refer to. Please be more precise.

Line 115: The formulation “atmospheric parameters under the soil” is probably not correct I assume.

Line 123: Hmhm, just wonder if the factor Ld, representing Leaf Area Index, in Eq 58 in Dickinson et al. (1993) is missing here or it is just a different definition of transpiration?
Line 124: In Eq 1 it says “Tr” but here “Trk”. Please make it consistent.

Lines 169-171: Although it is very precise to divide the text in “Current” and “New” formulation subsections it is from a stylistic perspective a bit awkward when the “Current” section is represented by only one sentence. Therefore, I would recommend to remove the subsections here. The same comment is valid for Section 2.4.


Line 196, Eq 8: Please replace “sun” with “sha”.

Lines 287-292: Very complicated paragraph where I assume the main message is simply “Grided observational data sets (E-OBS, HYRAS, GLEAM) were interpolated to the COSMO-CLM grid for comparison.”, right?

Lines 316-349: I don’t see the point to spend a considerable part of the discussion on how values look for the inactive vegetation periods (wintertime and night-time). In my mind the most interesting part is how they differ during summer daytime. But this part cannot be analysed by these figures since one cannot distinguish any differences due to the y-axis scale. I would recommend to focus your analysis more on the summer daytime part.

Lines 351-371: You start the paragraph by concluding that “stomatal resistance ... is a highly intermittent phenomenon, extremely localized on the leaf level, and varies with leaf positioning on a plant and from leaf to leaf and from plant to plant” but then you compare your model results with observations from literature based on “located in the North America region” with no further comments on if these observations can at all be considered to be representative for your model results. Thus, this first sentence and your final comparison does not make sense to me.

Lines 374-421 (Section 4.2 and Figures 4-5): The comparison between model results and GLEAM datasets in Figure 4 shows that the difference between the GLEAM datasets are often as big or bigger than the differences between the model versions, especially for AEVAP. Thus, in my mind it is difficult to draw any further conclusions from this comparison other than perhaps that ZVERBO for the new model versions is better than CCLMref. The statistical analysis with all numbers presented is not necessary to reach this conclusion I would say. And the analysis gives no indication on which of the new model versions are better or worse, right?

Lines 423-437 (Section 4.3 and Figure 6): As for the section on “Evapotranspiration and evaporation” the statistical analysis with all detailed numbers of sensible and latent heat
fluxes is not needed to reach your conclusion (visible from the figure) that “experiment results are similar to the CCLMref data”. Thus, in my mind unnecessary long details for this conclusion.