Reviewer's comment in bold, author's response in non-bold blue text.

This study uses the CLM5 model over Africa with different atmospheric forcings and soil texture inputs and explores their role in estimates of ET, runoff and soil moisture. Overall, there is a larger effect from the forcing dataset than from the soil texture inputs.

General comments

In my view, the paper in its current form requires major revisions before it can be considered for publication. The authors should specifically address the following limitations:

- **Since the study period is relatively short (two years): how can we know the results are generalizable to other years? Would it be an option to extend the study period?**

  This work investigates the interaction between uncertain soil properties and the temporal resolution of atmospheric forcings at the continental, regional and local scale with a focus on the rainfall intensity representation in the land surface model input. It therefore suggests that as long as the CLM5 model does not handle precipitation differently, extending the study period may not change the current outcome.

- **The study misses an evaluation with reference data of ET, soil moisture, and runoff to better appreciate the impact of the work: which forcing and soil texture input combination do the authors recommend the community to use in future work? Section 3.3 (local results) could be dropped in favour of this.**

  Performing an evaluation with measurement data is beyond the scope of this paper as stated in lines 60-61. The aim of the paper is not to select the best model settings over Africa or extensively evaluate the model performance, but to investigate the impact of certain model inputs (different soil input data versus different atmospheric forcings and role of temporal resolution) in a more generic sense. However, we will present a comparison of our ET, runoff and soil moisture estimates with widely available gridded datasets.

- **Often small impacts are extensively discussed, but it's not clear whether they are significant or not. Statistical tests should be performed to verify this.**

  We will reformulate the text to take this comment into account and provide information on statistical significance if possible.

- **I have methodological concerns about using the "random selection" upscaling technique for a single grid cell and then comparing the results with**

other upscaling techniques. Such results are not reproducible, so how should they be interpreted?

Random results are reproducible using "seeding" in programming. Seeding is used to initialize the random number generator (RNG) such that it produces a reproducible sequence of numbers. We specified seeds before performing upscaling using the Random method. If we therefore perform upscaling using the Random method multiple times, we will have the same outcome.

- **The Methodology is too detailed in terms of equations that are used by the LSM. These equations are only useful if they are referenced to in the text to explain or discuss the results, which is currently not the case.**
- We will link the discussion in the paper better to the method section and might also move some of the material presented in Methodology to an appendix,
- **The Discussion is too descriptive and fails to convincingly explain some of the results.**

  We will deepen the discussion and discuss the results more by making more reference to equations in the revised version.

Specific comments

1. **The title hints an impact of "temporal model resolution", but this should be the resolution of the atmospheric forcing dataset (input for the model). Temporal model resolution is always 30 minutes, and its impact was not examined.**

   Yes, Land Surface Model temporal resolution is usually 30 minutes. We will revise the title in the revised manuscript version for consistency. We are considering "High Resolution Land Surface Modelling over Africa: the role of uncertain soil properties in combination with forcing temporal resolution" in the revised version.

2. **L62-69: Part of this paragraph should go to the Methodology, for example, how the dataset was upscaled to the model resolution.**

   Lines 67-69 where upscaling is described will be moved to methodology in the revised version.

3. **L70: many new acronyms which should be written in full the first time they are introduced.**

   CRUNCEP (Climatic Research Unit (CRU) and National Centres for Environmental Prediction (NCEP)), GSWP (Global Soil Wetness Project) and WFDE5 (Watch Forcing Data methodology on ERA 5) will be written in full in the revised version.

4. **L83: "the novelty of the work lies in ..."; this is more or less a repetition of line 58.**

   We will merge both sentences into lines 83-85 in the revised version.

5. **L88: the Introduction is finished with raising two main research questions on which the authors want to find the answer. I expect to explicitly find an answer to question 1 and question 2 somewhere in the Discussion or (preferably) the Conclusions.**

   Answers to the research questions as shown by the work will be explicitly stated in the conclusions of the revised version of the manuscript.

6. **L94: Add a citation to CLM5.0. Additionally, be consistent in the use of CLM5.0 and CLM5: both are currently found throughout the text.**

   CLM5.0 (Community Land Model version 5.0) will be written in full in L94 and thereafter referred to as CLM5. All other uses of "CLM5.0" within the text will be changed to CLM5 in the revised version.

7. **L101: "CLM5.0 therefore has features of great interest for land surface modelling over Africa at a high resolution". Why specifically Africa?**

   We did not want to place the focus on Africa, but on the high spatial resolution. The sentence will be changed to clarify this.

8. **L106: Acronym for CLM5 should be written out in full the first time it is used (in the Introduction).**

   This has been addressed in our response to comment 6.

9. **L107: Change "The total porosity is given by" to "The total porosity for level i is given by".**

   This will be changed as you suggest in the revised version.

10. **L117: Citation for the Brooks and Corey model? What is the physical interpretation of "the exponent B"?**

    The B in the Brooks and Corey model represents the pore size distribution. The text "which represents the pore size distribution" will be added to line 117 in the revised version.

11. **Overall, the introduction has too many equations which are not referenced anywhere in the text. So it is not clear what their added value is. Either they should be used in the Discussion to support in explaining the observed results, or they should be replaced by a citation of the CLM5 paper in which they can probably also be found?**

We will link the discussion in the paper better to the method section and might also move some of the material presented in Methodology to an appendix.

12. **L162: Add that the time step Δt is expressed in seconds.**

    The units for Δt (seconds) will be added to L162 in the revised edition.

13. **L167: What is the native resolution of IGBP-DIS? Is it 3km? If not, why is only SoilGrids250m upscaled?**

    The IGBP-DIS soil texture information is 8km resolution. This was internally downscaled by CLM5 using the nearest neighbour algorithm.

14. **L170: "The 10 upper CLM5 soil layers": how many are there in total? This type of information on the model should be in section 2.1, rather than the equations which are currently there.**

    There are 20 active soil layers in CLM5 (8.6m deep). In this work we are concerned with the top 2m. L170 will then be changed to “The soil texture dataset provides information for the top 10 CLM5 soil layers: at 0.0175, 0.0451, 0.0906, 0.1656, 0.2892, 0.493, 0.829, 1.3829, 2.2962 and 3.4332 meter depth.” From “The soil texture dataset provides information for the 10 upper CLM5 soil layers: at 0.0175, 0.0451, 0.0906, 0.1656, 0.2892, 0.493, 0.829, 1.3829, 2.2962 and 3.4332 meter depth.”

15. **L181-182: Too detailed and not relevant for this study.**

    We will delete this sentence in the revised version of the manuscript.

16. **L200: Is method (iii) the novelty of this paper? If yes, it should be mentioned explicitly and motivated (why is a third new method necessary if two other already exist). If no, a citation to another study using this approach is missing.**

    Yes, the “Random upscaling” method is a novelty of this paper as it is yet to be documented. While the Dominant upscaling method introduces bias towards the prevalent soil types and the Averaging method presents a conservative estimate that misses extremes, the random upscaling method, creates an opportunity for outliers to be the effective soil parameter. This detail will be included in the revised version.

17. **Table 1: the longitudinal extent is not relevant for this study: all three are global and thus cover the study domain.**

    The longitudinal extent will be removed in the revised version.

18. **L256: "A spatially varying soil thickness dataset": which one?**

The spatially varying Soil Thickness dataset is new in CLM5. It replaces an assumed soil thickness of 2m present in earlier versions of the CLM. The reference to the dataset has been provided in L257.

19. **L270: "the hourly WFDE5 forcings were also aggregated to 3 hours and used in new simulations". Perhaps this should go to section 2.4. Why not 6h, so all three forcings are comparable?**

    Aggregation of WFDE5 into 3 hourly timesteps is an additional model setup performed after initial results already confirmed the effects of the higher temporal resolution of the WFDE5 data. We reasoned that making WFDE5 similar to GSWP (3 hourly) was sufficient to validate our claim.

20. **L276: "A simulated variable for a certain atmospheric forcings- soil texture map combination at a given time step is denoted by M1(t), M2(t), M3(t) and M4(t)." Don't you have 12 different of such combinations? Or do you mean that you have 4 such combinations for each atmospheric forcing? Please be more specific in the definition. Also, define t as the time step.**

    Yes, we have a combination of the 4 soil texture map outputs for each atmospheric forcing. The text on L276 stating "A simulated variable for a certain atmospheric forcings- soil texture map combination at a given time step is denoted by M1(t), M2(t), M3(t) and M4(t)." will be changed to "The average margin for a simulated variable for a certain atmospheric forcing/soil texture map combination at a given time step is denoted by M1(t), M2(t), M3(t) and M4(t)."

    In L283, "where T represents the total number of time steps in the time series." will be changed to "where T represents the total number of time steps in the time series and t denotes time step."

21. **L290: ET is evapotranspiration, not evaporation.**

    L290 will be changed to Evapotranspiration in the revised version.

22. **L298: "marginally" implies that you performed a statistical test. One should actually be performed to assess the significance of the differences in mean/max ET (and other variables), depending on the texture, forcing and their interaction.**

    The word "marginally" used in L298 will be changed to "slightly" in the revised version.

23. **L330-332: again, the statements made here require hypothesis testing.**

    The word "significantly" in the sentence will be changed to "substantially" in the revised version. We will provide information on statistical significance if possible.

24. **L335: WFDE5 are not hourly because they have been aggregated to 3h (L270)?**

At this stage, WFDE5 is still hourly. We will clarify this in the text.

25. **L337: "The increased surface runoff in the WFDE5 forced simulations reduces the availability of water for ET processes"; can this claim be made with certainty if it also rains more, so more water is coming in (cfr. Fig. 2)?**

    The statement "The increased surface runoff in the WFDE5 forced simulations reduces the availability of water for ET processes" will be changed to "The increased surface runoff in the WFDE5 forced simulations reduces the availability of water for ET processes especially during sheet flow."

26. **L339: "A higher temporal resolution of the atmospheric forcings as for WFDE5 will result in higher peaks of precipitation intensity, whereas a coarser resolution of 6 hours like for CRUNCEP will average out intensive precipitation over longer time periods with less high peaks in precipitation intensity"; but wasn't this effect to be mitigated by aggregating the hourly dataset (L270)?**

    This effect was mitigated in the validation set-up where WFDE5 was aggregated to 3 hourly. But in the initial run, which first indicated the finding, the statement holds. We will clarify the text accordingly.

27. **L353: "significantly influenced": be careful using this word.**

    The statement “The estimation of subsurface runoff is more significantly influenced by soil texture variations and the upscaling of soil texture properties compared to ET and surface runoff simulations.” Will be changed to “The estimation of subsurface runoff is more substantially influenced by soil texture variations and the upscaling of soil texture properties compared to ET and surface runoff simulations.” In the revised version.

28. **L355: "The most substantial differences in simulated subsurface runoff are observed between the FAO soil map the SoilGrids250m maps"; actually the forcings still result in the largest differences.**

    We will change this sentence to “The most substantial differences in simulated subsurface runoff among the soil texture maps are observed between the FAO soil map and the SoilGrids250m maps, while the disparities among the upscaled SoilGrids250m maps are smaller.”

29. **L364: Fig. 4 should be referenced to at the start of the section.**

    We will change this as suggested by the reviewer.

30. **L371: It's confusing to talk about ET in the section on runoff, save this for the Discussion.**

    We will remove this sentence.

31. **L382: "weighted average": how are the weights defined? Perhaps this is something to be described in the Methodology.**

    The weights are defined according to the thickness of each soil layer. In L267, the sentence “For continental scale results, yearly sum of evapotranspiration (ET), surface runoff, and subsurface runoff were computed as well as seasonal mean soil moisture contents.” will be changed to “For continental scale results, yearly sum of evapotranspiration (ET), surface runoff, and subsurface runoff were computed as well as the seasonal mean of the weighted average of the top 2 meters soil moisture content. The weights were defined according to the thickness of each soil layer in CLM5”

32. **L394: This statement doesn't hold for the mean: similar effect when varying the forcing and the texture.**

    L394 will be changed to “Like for ET and surface runoff, varying the atmospheric forcing impacted continental maximum of soil moisture more than variations in soil texture input.”

33. **Figure 5:**

    - **Would it make sense to also look at min soil moisture, next to mean and max?**

      Minimum soil moisture will be displayed in the revised version.

    - **Units of the colorbar are mm^3/mm^3, while elsewhere in the text cm^3/cm^3 is used. Be consistent. Also, consider using more sensible intervals for the colorbar ticks.**

      Colorbar units for soil moisture will be harmonised and intervals made more relatable in the revised version.

    - **Perhaps one could keep the first column as is, and the other columns could show the difference w.r.t. the first column? Now all 12 maps look identical and it's hard to see where the differences are.**

      This will be done for soil moisture as a supplementary material and referred to within the text in the revised version.

34. **Section 3.2: Consider moving the definitions of the regions to the Methodology.**

    The definition of regions will be moved to methodology in the revised version.

35. **L410: Motivate the choice for Sahara and Central Africa (dry and wet region?).**

    Yes. The choice of the Sahara and Central Africa are based on their moisture availability. The sentence “Two regions are discussed while more information for

other regions is available as supplementary material." In L410 will be changed to "Two regions are discussed while more information for other regions is available as supplementary material. These two regions are shown here while they have strongly contrasting moisture availability."

36. **L417: This statement is true on average (it may well exceed this in certain pixels).**

    The statement "The Sahara region is the most moisture deficient region in Africa…" in L417 will be changed to "The Sahara region is generally the most moisture deficient region in Africa…" in the revised version.

37. **Figure 7:**

    - **"Soil water content", "soil moisture content", "soil moisture" are all used interchangeably. Be consistent.**

      We will adhere to using soil moisture content in the revised version.

    - **Dominant and Random are very similar, as can be expected. The first one has already been used in literature. Please motivate (in the Methodology) the advantage of Random to justify it as a new technique (e.g., are there computational gains?). Also: which one is best? Why was no evaluation performed? There are many global grid-based ET products available that can serve as reference.**

      While the Dominant upscaling method introduces bias towards the prevalent soil types, the random upscaling method, creates an equal opportunity for all grid cells, especially for outliers to be the effective soil property. This detail will be included in the revised version.

38. **L430: Why is the long name for ET suddenly introduced here?**

    This will be changed in the revised version.

39. **L434: "Subsurface runoff shows a decreasing trend, which is attributed to initially higher groundwater levels"; was it already decreasing during the spinup? In this case, the spinup time of 1 year may not have been sufficient. What spinup times are used in similar studies?**

    To avoid confusion, the sentence "Subsurface runoff shows a decreasing trend, which is attributed to initially higher groundwater levels "will be removed in the revised version. The observed decline in subsurface runoff over the Sahara is not a spin-up effect, as such systematic decline is not observed over other locations. Furthermore, spinup was increased to 2 years in the validation runs. Nevertheless, we will consider performing tests with a longer spinup to study the impact on the results and in case larger impacts are found, redo the simulations.

40. **L443: Again, I am confused that WFDE5 would have an hourly temporal resolution while it is states in the Methodology that it was aggregated to 3h.**

    WFDE5 has hourly temporal resolution in the initial run but 3 hourly only in the validation run.

41. **Figure 8: can we conclude from this that one shouldn't use mean soil texture properties, as it is the only one deviating from all others? Of course, this claim would be stronger if an external product were used for validation.**

    A definitive statement will be made in the revised version of the manuscript after comparison with a reference dataset.

42. **L465: "... variations in ET values across different soil texture maps", not really: only a single one shows a variation (and this is probably also the limitation of your average margin metric, which isn't able to identify this "outlier"). The following paragraph fails to mention or discuss this fact.**

    The statement will be changed to "... exhibits a variation in ET values across different soil texture maps", in the revised version.

43. **L481: "We selected one location for each of the eight climate regions"; only two are in the main text.**

    The text will be changed to "We selected one location for each of the eight climate regions. Two are presented here while the reader is referred to the supplementary materials for others."

44. **L490-492: Add a paragraph on the crop model/irrigation model in CLM5 to the Methodology.**

    A brief description of the irrigation model of CLM5 will be added to section 2.1 under CLM5.0

45. **L494-495: Not clear from the Figure.**

    The differences are not expressed in absolute terms (mm) but in terms of the percentage of ET for each forcing.

46. **L499: "While slight variations in surface runoff are found ..."; again, be more specific which soil texture maps differ, since they don't all differ from one another!**

    The statement “while slight variations in surface runoff are found for WFDE5” will be changed to reflect a quantity in the revised version.

47. **Figure 9:**

    - **Panel "S.Runoff-WFDE5": is the Mean on top of the other textures, or the Rand?**

Other textures are 0 while the mean stands out. This is further clarified in Figure S31 and Table T10 with average margin calculation.

- **What is the FAO/Dom soil texture type for this grid cell?**

  Sandy loam. This will be clarified.

- **The Rand soil texture may be very different when the experiment is performed again, no? Since this is only a single pixel. Hence, the study is actually not reproducible in this sense. Also see Fig. 10: here Rand just happens to coincide with Dom in terms of soil texture class, so the lines overlap. But run the experiment again and the results may be completely different. I suggest performing "Rand" several times and reporting an ensemble average to resolve this issue.**

  Running these experiments multiple times in a Monte Carlo fashion is not feasible given the required compute time for the high spatial resolution model runs over Africa. Instead, the focus of this work is to identify and not to quantify the sensitivity of uncertain soil properties to atmospheric forcing's high temporal resolution.

48. **L519-529: Same remark as earlier: the difference in soil texture maps is always just cause by Mean vs. all the rest.**

    A definite claim will be made after comparison with other datasets.

49. **Table 3: Not clear what is presented here: difference between which two things? Units?**

    After aggregating WFDE5 to 3 hourly, average margin was calculated for the ET, surface runoff, subsurface runoff and soil moisture. The table shows that WFDE5 which was aggregated to 3 hours and GWSP (3 hourly) had similar values for the margins, while WFDE5 (hourly) had larger margins. All values are in mm/month.

50. **Table 5:**

    - **How come the SGMean column has a specific soil texture class? I thought this approach created a "new" class with averaged soil metrics.**

      The averaging was performed at the level of the clay and sand fractions which are in percentages. The average clay, sand and silt percentages were used to determine the soil texture class according to the USDA triangle.

    - **Is referenced nowhere in the text. Please use it to discuss for example the bottom row of Fig. 9. Overall, the Discussion describes the figures**

and results too much, rather than explaining what we see in them (and why!).

This will be considered in the revised version of the manuscript.

51. **Conclusion: misses outlook discussing how future research can benefit from the results of the study.**

    We will discuss how future research can benefit from the result of this study in the revised version of the manuscript.

Technical corrections

- **L55: "0.5o" instead of "0.5°"**
- **L162: "kg/m^{2}s" instead of "kg/(m^{2}s)"**

  Both adjustments will be made in the revised version of the manuscript.