1. Zachary M. Subin and co-authors
2. October 15, 2014
3. Response to Reviewers for Hydrology and Earth System Sciences Manuscript Titled "Resolving terrestrial ecosystem processes along a subgrid topographic gradient for an earth-system model."

I. Summary:

4. We appreciate that both reviewers recognized the importance of our goals and believed that our approach was fundamentally viable and a potentially worthwhile contribution to the literature. Both reviewers raised issues of (a) clarity of presentation, and (b) the importance of further evaluation for confidence in the validity of the modeling approach and its global application. We should be able to address much of the clarity issues (a) in the revision, and we answer some of the specific questions raised by the reviewers below. Regarding further evaluation (b), in particular detailed evaluation of the model at an ecosystem- or hillslope-scale with known characteristics and responses to atmospheric forcing, we contend that it would be worthy scientifically but is beyond the scope of this paper. Moreover, this concern has been implicitly addressed by previous papers this work relies upon, such as Paniconi et al. (2003), expanded upon below. Our formulation is an extension of the previous formulation in Milly et al. (2014), and our formulation and global evaluation approach is consistent with other published analyses. Consequently, while site-level validation of the model would be a worthy exercise, we contend that it is not necessary for the initial presentation of the model and global evaluation here, nor is it necessary for our primary conclusions that (a) this novel method is worthy of further research, (b) additional observations are needed to parameterize subgrid / finely-resolved hydrological models, and (c) existing Earth System Model (ESM) predictions of future soil carbon loss may be biased low by failing to adequately represent the accumulation of carbon in wet soils, due to the nonlinear coupling of hydrology and biogeochemistry poorly represented by the gridcell-mean state.

II. Relationship to Previous Work:

5. The use of a 2-dimensional (2D) hillslope conceptual unit (with ~1 km length and 200 m depth to bedrock) to estimate saturated fraction and runoff in each land model gridcell was presented in Milly et al. (2014). The main innovation of this work is to discretize the hillslope into "tiles" and explicitly solve for the 2D flow field, with separate vegetation, soil thermal, and soil biogeochemical states tracked in each tile. In Section 3.5 we showed that the discretization of the hillslope caused little change in simulated surface energy and water fluxes (and thus the partitioning between evapotranspiration [ET] and runoff), compared with the model presented in Milly et al. (2014). The main differences appear in being able to directly define a "wetland area" based on the fraction of tiles with a near-surface water table, and in the biogeochemical predictions (including leaf area and soil carbon). For the latter, we do not contend here that the new formulation is

suitably parameterized to represent global biogeochemical dynamics with high fidelity (noting that many of the state-of-the-art ESMs used in the Climate Model Intercomparison Project Phase 5 [CMIP5] also lacked evidence of this fidelity): we merely qualitatively illustrate the sensitivity of the simulated biogeochemistry to the omission of subgrid hydrology, due to the nonlinear coupling of hydrology and biogeochemistry. As we noted in Section 4 of the submitted manuscript, we agree with the reviewers that quantitative evaluation of the representation of vegetation and soil carbon would improve confidence in predictions.

6. Furthermore, the work in Milly et al. (2014) built on the work of Paniconi et al. (2003), cited in our Section 1, who showed that solving an equation for the water-table depth in a 2D hillslope (the Hillslope Boussinesq Equation) matched the predictions of a 3D Richards equation model of the kind routinely used in hydrology studies. Indeed, our formulation is potentially more general than the one used in Paniconi et al. (2003), as it does not rely on the assumptions underlying the Hillslope Boussinesq Equation (i.e., hydrostatic conditions in each vertical slice of the hillslope) but directly solves Darcy's Law to determine the 2D flow field. Our approach includes some of the same limitations as, but is less simplified than, current treatments of groundwater in ESMs [i.e., Krakauer et al. (2013), Lei et al. (2014), Niu et al. (2007)].

III. Detailed Responses:

7. We include the relevant Reviewers' Comments *in italics* and our responses below.

IV. Reviewer 1:

8. *The results are, however, mixed with relatively little evidence of improved wetland simulations, although much of the deficiencies are blamed on an inadequate ability to parameterize the model effectively due to lack of observational datasets…. How plausible is it to realistically parameterize the model using observations? I think this is a reasonable question given how much uncertainty there seems to be in how to parameterize the model globally. Is this something that is ever going to be possible? Would it just take a concerted effort to derive from presently available datasets? New observations required? From my reading, it seems appropriately parameterizing the model would be a monumental task.*

9. Observations of hillslope geometry required to parameterize TiHy are in principle derivable from digital elevation models (DEMs), and work is under way to develop such datasets (J. Pelletier, Personal Communication, 2013). Observations of subsurface properties like depth-to-bedrock and bedrock permeability will always inherently be globally sparse, and this is a limitation. However, simple assumptions about these subsurface properties are required for any treatment of groundwater in ESMs (see #6 above). Evaluating the reliability of these assumptions and their effects on the hydrological simulations would be a valuable additional exercise (see #4 - #6 above). Were additional datasets made available providing high-resolution subsurface characteristics (i.e., depth to bedrock and

bedrock permeability), our model formulation would be capable of incorporating these directly, as we include these parameters explicitly for each hillslope tile, while previous efforts could only parameterize their effects at a large scale.

10. *I feel that the authors should provide more site-level analysis demonstrating that the model formulations are valid, before attempting these global-scale simulations. If the authors can take their model and run it for sites where the model parameters can be appropriately set and the model outputs evaluated, it would provide much more confidence that the model formulations are appropriate. The authors do show three example regions with different water table regimes (Fig 3), but these are just taken from the global simulations and have no observations for validation against. I do understand that water table information is sparse, but with no effective evaluation of the model outputs, it is hard to accept the global simulation results. The authors propose something similar on p. 8471 line 24 and I would suggest that, as they say, evaluation at sites or regions where the topography, substrate properties, forcing climate and resulting hydrology are well characterized should improve confidence in the model. I feel that this has to occur before the underlying model structure behind these global scale simulations can be assumed to be correct. I think these proposed simulations could form a first paper, with this manuscript reviewed here as a follow-up paper after having laid the groundwork and demonstrated the appropriateness of the model structure in the first regional and site-level paper.*

11. See #4 - #6 and #9 above. Hillslope-scale evaluation is a task for another study, and it is not inappropriate to evaluate the model at the same spatial scale at which it will be applied (coarse gridcells across a global domain), as we have done here.

12. *While their results demonstrate the LM3-TiHy, as this present stage, does not represent a significant advancement for modelling wetlands, it could provide a means for better capturing vegetation producivity differences between uplands and lowlands. The authors do look at LAI in Fig 7 and also evapotranspiration, but no attempt is made to compare to observations. The authors also themselves suggest comparing their results to observations (p. 8469 line 6) and I encourage them to do just that.*

13. We reiterate the value of this comparison, but it would not fit within this paper, which includes several observationally-based datasets for globally evaluating the hydrology. Further developments are underway to improve the biogeochemistry in LM3 that would deprecate such a comparison; as noted in #5, we included results for LAI and soil carbon to qualitatively illustrate these sensitivities, not because we are confident in the quantitative predictions of this model version. We will make sure this is clear in the revised manuscript.

14. *The authors present some results derived using the unpublished CORPSE model. Since this model is not published and no description of the model is provided, I have no way to evaluate if these results are in any way reasonable. I request the authors to remove the CORPSE results from the paper since it is not possible to adequately evaluate these results.*

15. The CORPSE model is presented in a manuscript that is now accepted and can be cited as such in our revised manuscript. The relevant features of CORPSE are that (a) it includes vertically-resolved soil carbon and (b) a large reduction of soil decomposition at saturation. While the quantitative predictions of Section 3.6 may depend on the details of CORPSE, we expect that the qualitative results would hold true with any model with these features. We are also happy to provide a more detailed description of CORPSE in the Supplement if desired.

16. *The authors suggest that the LM3-TiHy model can provide a new approach to investigating the vulnerability of boreal peatland carbon to climate change. I am not so sure of this as peatlands are generally areas of low-relief and the West Siberian Lowlands and Hudson's Bay Lowlands were both poorly simulated by LM3-TiHy. This is important as the WSL and HBL are the predominant regions of boreal peatland carbon.*

17. This is a valid concern that should be addressed by future work with TiHy, and we will highlight this in the revision and reword the conclusion appropriately. Simulating these regions well is a challenge even for peat-specific ecosystem models which are often parameterized for site-specific applications (Frolking et al., 2010; Morris et al., 2012); peatland scientists continue to explore what factors coincide to explain the initiation of peatlands or their persistence in a subset of areas of low-relief terrain (N. Roulet and L. Belyea, Personal Communications, 2014).

18. *p. 8449 line 12 - Expand on what is meant by 'simple plant functional type transition dynamics'. This sentence is too vague as is.*

19. We mean "dynamic vegetation" as the term is commonly used in the ESM community: the plant functional types can change as climate changes (for instance, drying could convert a forest to a grassland). We refer to the referenced Shevliakova et al. (2009) for further details.

20. *p. 8453 line 3 - The surface runoff is assumed to flow directly downstream, bypassing the intervening tiles. Any idea how important this simplification is?*

21. This simplification could be important in areas where rapid infiltration into the downslope tiles could occur, or in areas where microtopography would allow the accumulation of surface water storage before runoff could continue. We highlighted the further development of surface-runoff and inundation processes as a priority for future work. However, we note that in most cases, downslope tiles will be at least as wet as the tile being considered, so infiltration somewhere downslope is not likely.

22. *p. 8459 line 1-4 - I am not sure I understood this. By only considering the water table to a depth of 0.1 m, the authors are not looking at where water table is most important in wetlands, i.e. shallower than 0.1 m. Perhaps expand on this.*

23. As the model does not include inundation, and the soil moisture in the top 0.1 m of a soil tile where the water table is at 0.1 m depth is likely to be in the capillary fringe and therefore close to saturation, there will be little difference in behavior when the water table is at the surface vs. 0.1 m depth. Consequently, as we

wished to emphasize the shallow water-table-depth variability that is important in wetlands (i.e., 0.1 m to several m) and for distinguishing wetlands from non-wetlands, we evaluated the log of water-table depth, and bounded the depth by 0.1 m to prevent the log from decreasing without bound as the water-table depth approached zero.

24. *p. 8459 line 12 - I assume I am being thick here, but the description of the sensitivity and precision read as though they would give the same result.*
25. The sensitivity is the proportion of *observed* wetland that is identified as wetland by the model, while the precision is the proportion of *modeled* wetland that is actually observed as such. The distinction is between one of Type 2 and Type 1 errors, respectively, as commonly used in statistics.
26. *p. 8461 line 10 - This suggests to me that the perched water tables in permafrost zones might correspond to a separate regime from the three identified by the authors as the underlying processes would differ from other regions, i.e. the regions could shift regimes as the active layer deepens.*
27. The perched water table in permafrost zones could be considered a fourth regime, and we will revise this section to be consistent with that definition.
28. *p. 8470 line 20 - The authors suggest that observations of inundated area are better than those of wetland ecosystem extent. I don't agree considering the problems with inundated area observations in areas with high canopy cover and flooded agricultural fields (see discussion of GIEMS in Melton et al. 2013). I ask the authors to provide more support for the statement.*
29. We acknowledge that there are deficiencies with inundated area observations. However, inundated area observations are constructed with a consistent global algorithm from satellite-retrieved properties, in contrast with wetland datasets which include geographic databases that have inconsistent coverage and conventions across culturally defined regions. This is what we intended to emphasize in our statement and will revise to make this more clear.
30. *p. 8471 line 13 -17 - Yes, including ground water pumping, irrigation, and drainage should improve wetland simulations. However the LM3-TiHy model seems to perform most poorly in regions that are relatively unaffected by these processes. What about floodplain processes or improving processes in areas with low-relief?*
31. We agree that improving floodplain processes and model behavior in areas of low relief should be a priority for future work. We already highlighted the former in the Discussion section, and we will revise to highlight the latter.

V. Reviewer #2:

32. *This paper stems from a brilliant idea to describe sub-grid soil moisture variability in land surface models, based on a tiling approach allowing to account for water table depth (WTD) gradient along hillslopes. Yet, the paper struck me more by what it lacks (a good structure; a comprehensive description of the*

model, simulations, forcing and validation data; a critical look at some results and the work's limitations) than by what it brought me as an interested reader. Basically, I felt the paper wanted to say too much, and couldn't say it well in the standard length of a scientific paper. I thus recommend major revisions, and hope the comments below can help.

33. We appreciate the reviewer's recognition of the value of our approach. We regret that the reviewer found the presentation to be unclear. We carefully considered the pros and cons of various structures for the text, and eventually converged on this one. We still think it will be the clearest to most readers. We agree that the paper does attempt to cover a lot of ground, and it would be difficult to add more. We are not sure if there is a good section to remove for shortening it.
34. We address the detailed comments below: some of these suggestions would increase the clarity of the manuscript.
35. *1) Introduction: The paper claims to pioneer the implementation of below-ground, hydrologically-based tiling, but proper credit should be given to Koster et al., 2000, who proposed the Catchment model with the same purpose 15 years ago, although with tiles of dynamical extent, based on TopModel's analysis of the relationships between topography and WTD. Note also that the underprediction of wetlands in flat areas (p8447, L10) is not an intrinsic failure of TopModel, but is restricted to its simplistic implementation in most LSMs, following Gedney & Cox, 2003.*
36. We were intending to highlight the first implementation of belowground heterogeneity in natural ecosystems in a land model capable of being coupled into an ESM, of the class of models included in CMIP5. We regret not including Koster et al. (2000), which was cited in Milly et al. (2014), and Ducharne et al. (2000), and will recognize this work in the revision. We will cite the Gedney and Cox (2003) in the revision along with the other models using TopModel to diagnose saturated fraction. The limiting assumption they addressed in their study to which the reviewer may be referring is that the hydraulic conductivity follows a fixed exponential decay profile with depth in some other models but not their formulation. However, TopModel still assumes steady-state recharge at all times, water-table slope parallel to surface topography, and hydraulic conductivity decaying to zero at depth. TiHy assumes none of those.
37. *2.1) LM3: p8451, L10: if several representative hillslopes are used in one grid-cell, are they hydrologically independent (i.e. no exchanged water)?*
38. Yes, they are hydrologically independent. Future work could include backflow from the stream to the land, which would provide some interaction. However, we do make a simplification by only including one order of hillslopes, and not considering the baseflow from first-order hillslopes into larger-scale hillslopes.
39. *2.2) LM3-TiHy: The approach is far from being straightforward, and the paper is not self-consistent to explain it, as it is necessary to search for many important information in Milly et al., 2014, and in the technical note provided as Supplementary material. I recommend to present in the paper all the equations*

and parameters that are further used to discuss the results. If the authors feel some equations need to be annexed, then restrict the technical note to what is needed for this paper, and make sure the notations are consistent with the ones of section 2.1. It would be nice to illustrate the variables of Eq. 4 in Fig 2, including the relative positions of the tiles (j-1), j, (j+1). I suggest that the beginning of Section 2.3 is used in 2.2 to explain Ln (adding the definition of Z(x)). I didn't understand the rationale of Eq.5 and how it is used to define Kl in Eq. 4. Finally, it is said several times throughout the paper that inundation processes are not fully described in LM3-TiHy, but I couldn't find a clear description of what is implemented and what is lacking. It should be added since the authors justify many of the poor performances of the model by these shortcomings.

40. We believe that we included most of the equations in the paper that are necessary to understand the results. One exception may be the modification of Eq. 4 for the case of the simulation with the converging hillslope, and we can include this. We don't see which additional equations from Milly et al. (2014) are being suggested for inclusion. We depend on the vertical solution of Richards equation in that work, but it is not essential for interpreting the results we present.

41. The technical note is intended to mirror the code itself and be sufficient to reconstruct the model. By necessity it includes features that are implemented in the code but not exercised in this paper (i.e., multiple tiles at the same elevation class in each hillslope due to, e.g., land cover change) and details that are critical for implementation but extraneous to the paper (i.e., energy conservation as heat is advected horizontally by water flows between tiles). Because the technical note is more detailed and general, it is difficult to maintain the same notations while keeping clarity. We can illustrate the variables of Eq. 4 in Fig. 2 in the revision.

42. The beginning of Section 2.3 includes features that are not intrinsic to TiHy but are used in the implementation here (such as the power-law hillslope shape), while Section 2.2 is intended to be more general. We can include an additional equation to clarify the relationship between Eq. 4 and Eq. 5.

43. We will clarify the inundation processes that are missing. The key simplification is instantaneous surface runoff from each tile directly to the stream. In reality, the runoff speed is finite, water may evaporate from or infiltrate into downslope areas before reaching the bounding stream, and water may pool in surface depressions due to microtopography, leading to a surface water storage state that is not currently represented in the model.

44. *3) Sections 2.3 to 2.5 of the Methods: I found this part really messy, as it requires lots of back and forth reading, and I never had the information I wanted when I wanted it.*

45. *3.1) An alternative structure could be to separate the input data (slope, permeability), the performed simulations (including the way prescribed parameters are prescribed, and ending with the spin-up procedures), and the validation data, which lack important details as the spatial and temporal resolution and covered period. The interpolation procedure for GLWD (p8460) was not crystal clear to me, while it is called as a potential explanation of the*

model's poor performances (p8465). By the way, if the authors really believe this, why not improving this interpolation procedure before further analyses?

46. See #33 above. We will include the spatial and temporal resolution and covered period of each validation dataset in the revision. If computational resources are available, we can attempt to re-do the interpolation to see if this affects the results.

47. *3.2) More importantly, I had problems really understanding the assumptions behind the simulations and the differences between them: do all the tiles of a grid-cell have the same characteristic slope (zeta) ? Does zeta take only three possible values using the FAO data? Are Untiled and CORPSE-Untiled defined by having only one tile per grid-cell? How do they differ from the LM3 simulation in Milly et al., 2014? Why is the hillslope length L=1km ? How can you describe 2_x2.5_ grid-cells using one 1- km hillslope? Why do you assume a depth to bedrock B=200 m and what are the implications regarding the results? Why don't you build the Concave and Converging simulations from the same simulation? HSWDSlope Bimodal is not clear to me, does it assume a rectangular shape with two different slopes? If so, how do they alternate along x (zeta1, zeta2, zeta1, zeta2, etc...; or 5 times zeta1, then five times zeta2)? I also suggest to add some information in Table 1 to clarify the differences between the simulations, and to explain CORPSE only once, when presenting the simulations.*

48. All the tiles have the same characteristic slope if a linear hillslope is used (i.e., $\beta = 1$ in Eq. 6) but not otherwise. Although the fine-scale FAO data have only 3 possible slope values, the average at the 2° scale can vary more continuously. *Untiled* and *CORPSE-Untiled* do have one tile per gridcell, but the calculation of runoff still assumes the same hillslope concept as here, and the relationship between gridcell-mean water-table depth and runoff is calculated offline: see Milly et al. (2014). They only differ from the LM3.1 simulation used in Milly et al. (2014) based on the assumed depth to bedrock, the macroporosity, and the xylem resistances, as described in the text. Describing a 2° gridcell with one hillslope is certainly a vast simplification, but existing ESMs tend to describe it with only one *tile*, which is even more of a simplification. The bedrock depth of 200 m is as in Milly et al. (2014), and, as noted in the text, the simulations are not highly sensitive to this choice as the hydraulic conductivity of bedrock varies over many orders of magnitude, so that is the primary bedrock control on water-table regime.

49. We made an error in Table 1: the *Converging* and *Concave* simulations were actually branched from the same simulation using the FAO slope dataset, and we will correct this error. *HWSDSlopeBimodal* assumes two non-interacting hillslopes, not one hillslope with multiple slopes. It is a matter of preference whether to include all of the methods together in Section 2 or present some of them together with the results in Section 3; we chose to do the former.

50. *3.3) Regarding the details about the specific analyses (how to define WTD, wetland and inundated fractions, Budyko's index, the WTR), I would have preferred reading them directly in the Results section, when these specific features are analyzed. Regarding the diagnosed inundation fraction, the authors*

write p8458, L16, that they didn't attempt to to tune topographic parameters to match the observed inundation, but it is a bit abusive, as they did tune zm, which is not very different from correcting topographic indices from their known dependence onto DEM resolution. Regarding WTR, some variables in Eq. 10 do not have the same meaning as in Gleeson et al., 2011a (B replaces maximum terrain rise, total runoff replaces recharge): please discuss the consequences.

51. As above, the placement of the analysis details in the Methods or the Results section is a matter of preference. Regarding the tuning to yield observed inundation: the distinction is that we tuned *one* global parameter, whereas previous analyses cited tuned geographically varying parameters to match observed inundation in a spatially explicit manner. *B* does not replace the maximum terrain rise but rather the thickness of the aquifer, denoted as "the average vertical extent of the groundwater flow system" in Gleeson et al. (2011a). We collapsed a factor of their L/d (the distance between surface water bodies divided by the maximum terrain rise) into our $1/\zeta$. We will clarify the adaptation of their Equation 1 in the revision. We chose to use total runoff rather than recharge as it is more independent of the belowground hydrology we wish to evaluate.

52. *4.1) All the maps require a higher resolution, as we need to zoom in a lot to see anything.*

53. We provided a PDF with high-resolution images to the publisher, and the resolution must have been coarsened during the production of this proof. The final version would certainly have higher resolution and we would be happy to provide the PDF so that reviewers can access it directly.

54. *4.2) In 3.1, I would present Fig 4, with the main forcings of simulation Base, before the typical WTD regimes (Fig3). More importantly, I was intrigued by the selected hillslopes shapes, with 5 to 10% slopes over zones supposed to extend 1km from the streams: I suspect this is quite rare, and that river beds usually have flatter cross sections; and what happens further from the streams, since it can take more than 1 km for overland flow to reach the closest stream? Can you discuss these strong assumptions in the paper?*

55. I am not sure what is meant by the "main forcings": there are 7 different atmospheric variables alone in the forcing, some of which can be directly examined in the cited Sheffield et al. (2006). We intended to present the typical water-table regimes and their controls before evaluating the global water-table depth. The constant slope from the stream to the hilltop is a simplification in the *Base* simulation; the *Concave* simulation tests the sensitivity to this assumption by having a flatter cross section near the stream. We did not include a sensitivity to the hillslope length of 1 km here for brevity, but we could include this if desired. As the discharge timescale is proportional to the inverse square of the hillslope length, we would expect longer spinup times and shallower water tables / more wetland area with a larger hillslope length, which would not be likely to improve the results.

56. *4.3) I also regret that no attempt is made to validate the simulated results: Fig 4a could be compared to the map of Gleeson et al, 2011a, over the USA (with significant differences over Florida for instance); couldn't the authors find any monitored transect showing the kinds of WTD gradient they highlight? One could think of the Sleepers River catchment for the first regime. Refer to above comment regarding the second regime illustrated by a Florida grid-cell. And regarding the third one, with groundwater level below stream level, there should not be any baseflow in such a case, so streamflow could become ephemeral: is it consistent with typical hydrological regimes in the Rocky mountains?*
57. Fig. 4a can indeed be compared to the Gleeson et al. (2011a) if this would fit in the appropriate length of the paper. We provided the information in Fig. 4b-d showing what datasets [i.e., the permeability from Gleeson et al. (2011b)] and simulated results [i.e., the runoff, which was comparable to LM3.1 in Milly et al. (2014)] which determined the features of this comparison. Regarding comparing to a monitored transect, see above.
58. Streamflow would become ephemeral when the water table is below the stream level consistently. This could be reinterpreted as suggesting a larger hillslope length (and upland elevation over the stream) in this region, but the results when averaged to the land model gridcell would be nearly unchanged, as the water table would still be well below the surface over nearly all of the hillslope.
59. *4.4) Upland vs Lowland: this analysis is split between Sect 3.2 for WTD, and Sect 3.3 for the surface properties (ET, Ground temperature T, LAI). (i) I didn't find a clear definition of the upland and lowland tiles, even though I supposed they were the two extreme tiles along the simulated hillslopes. (ii) I would have liked seeing the difference in WTD between up and low to better understand the corresponding differences in ET, T, and LAI. (iii) I was very surprised that the mean values of these surface properties (Fig 7, left) are not at all assessed against a reference, whether from LM3 or from observations. (iv) An important result is that the LAI increase in lowlands compared to uplands may be excessive, but it is only said in the Conclusion (p8469, L7-9); then, is it related to the hydrology or to the vegetation parameterizations?*
60. This definition is found in Section 2.2. We did show the water-table depths for both the upland and the lowland in Fig. 6. The mean values of the surface properties are compared to the LM3.1 configuration of Milly et al. (2014) in Fig. 9. As they were largely unchanged for temperature and evapotranspiration, we defer to the referenced paper for evaluation. We highlighted that the LAI increase may be excessive; we think this is at least partly due to the vegetation parameterization based on Shevliakova et al. (2009), as there is unrealistically high LAI (~8) in the Boreal forest. See #13 above regarding additional developments. We will make sure this is clear in the revision.
61. *4.5) I have serious doubts on the comparison between CORPSE, Base, and Untiled (Sect. 3.6, p8467): if I understood well, the main result here is that CORPSE simulates a much larger accumulation of soil carbon than CORPSE-Untiled, despite the two simulations having the same WTD. This accumulation is*

attributed to larger areas with wetland fractions (Fig. 8b). Yet, Fig 8b is not from CORPSE but from Base, and Table 2 shows problematic inconsistencies between Base, Untiled, CORPSE and CORPSEUntiled. On a global mean, Base and CORPSE have the similar WTD, wetland and inundated fractions, but Untiled and CORPSE-Untiled don't; Untiled has lower WTD, wetland and inundated fractions than Base, which makes sense, but Untiled-CORPSE has the same WTD than CORPSE, yet much lower wetland and inundated fractions (by an order of magnitude); as a result, Untiled-CORPSE has a larger WTD than Untiled, but much lower wetland and inundated fractions. There seems to be a problem here. Does it come from the CORPSE simulations (spin-up ? feedback of soil carbon onto hydrology?) Does it come from the Untiled simulations? All this has to be checked and elucidated before any conclusion can be drawn.

62. We regret this was unclear. Untiled and CORPSE-Untiled indeed have different global water-table depth. As the solution for water-table depth is different in LM3 and in LM3-TiHy (i.e., different stream boundary condition), it is not surprising that *Base* and *Untiled* have different water-table depths. As noted in Section 2.4, we adjusted the macroporosity in *CORPSE-Untiled* as compared with *Untiled* so that *CORPSE* and *CORPSE-Untiled* would have similar geometric mean water-table depths; otherwise the comparison of soil carbon accumulations in Fig. 10 would have been inappropriate. We wanted to examine the effect of the soil-moisture heterogeneity on soil carbon accumulation, not the effect of a slightly different water-table-depth solution method. *Untiled-CORPSE* has a deeper water-table depth than *Untiled*, *therefore* it has much lower wetland and inundated fractions. *Untiled-CORPSE* has much lower wetland and inundated fractions than *Untiled* because we have defined these fractions by the proportion of soil area with near-surface water tables (see Section 2.5), and *Untiled* is largely incapable of resolving these areas (there is a diagnosed saturated fraction, but it is not associated with a distinct soil state): this is one of the main innovations of LM3-TiHy. We do not see the inconsistency to which the reviewer is referring.

63. *4.6) Sensivity analyses (Sect. 3.7): I suggest not to discuss ConstGeo before this section (L14-20 p 8463 is rather distracting there, especially since no explanation is given to the defects of this simulation. In particular, I didn't understand why ConstGeo lowers WTD so much. The same applies to Untiled with a mean WTD which is 5 m lower than Base (Sect 3.5). Some explanation should be proposed to explain these important differences and help understand LM3-TiHy. I also found problematic that the two experiments showing the effect of hillslope shape (Concave and Converging) did not proceed from the same reference simulation. Finally, what conclusions can be drawn from the sensitivity to the slope input (Base vs HWSD and HWSDSlopeBimodal)? Can you conclude that one slope dataset is better? Can this be separated from interpolation procedures? From the influence of hillslope shape? And how do all these simulations compare to Untiled (Sect. 3.5) and why?*

64. We can move the discussion of *ConstGeo* to Section 3.7 if desired. *ConstGeo* had an arbitrarily chosen constant set of hillslope and soil texture properties, so it is unsurprising that it had a different mean water-table depth. We were not

interested in the change in the mean but the change in the correlations with observations. We discussed the difference in mean water-table depth between *Untiled* and *Base* in #62.

65. We regret the error in our Table 1 regarding *Converging* and *Concave* (see #49). We discussed that neither topographic dataset performed uniformly better than the other, and that although HWSD is a newer dataset south of 60°, using it globally requires filling in north of 60°, leading to artifacts. The interpolation procedures were the same for FAO and HWSD. We chose *Base* to be as similar to *Untiled* as possible, which were both similar to the LM3.1 configuration in Milly et al. (2014) except for the above-mentioned differences.

66. *4.7) Overall, I found the "Results" section too descriptive, and not enough focused onto giving useful insights on the pros/cons of the model, and the difficulties to parameterize it. An example of what could be discussed is the respective contribution of the studied parameters and recharge to control the WTD patterns, the effect of the latter being very overlooked in the paper. The differences between the different simulations, or the upland and lowland tiles, would be more convincing if they were compared to the mean values and to variability metrics, as often realized owing to statistical significance analysis.*

67. Such an analysis was intended in Section 3.1 and Fig. 3-4. The interannual variability of the water-table depth (typically < 1 m) is an order of magnitude smaller than the typical differences between the upland and lowland tiles (> 10 m), so these differences are trivially statistically significant. The sensitivity experiments are illustrative and not intended to be comprehensive, and doing statistical analysis on them would add little value to the paper.

VI. References:

68. Ducharne, A., Koster, R. D., Suarez, M. J., Stieglitz, M., and Kumar, P.: A catchment-based approach to modeling land surface processes in a general circulation model 2. Parameter estimation and model demonstration, Journal of Geophysical Research-Atmospheres, 105, 24823-24838, 10.1029/2000jd900328, 2000.

69. Frolking, S., Roulet, N. T., Tuittila, E., Bubier, J. L., Quillet, A., Talbot, J., and Richard, P. J. H.: A new model of Holocene peatland net primary production, decomposition, water balance, and peat accumulation, Earth Syst. Dynam., 1, 1-21, 10.5194/esd-1-1-2010, 2010.

70. Gedney, N., and Cox, P. M.: The sensitivity of global climate model simulations to the representation of soil moisture heterogeneity, Journal of Hydrometeorology, 4, 1265-1275, 10.1175/1525-7541(2003)004<1265:tsogcm>2.0.co;2, 2003.

71. Gleeson, T., Marklund, L., Smith, L., and Manning, A. H.: Classifying the water table at regional to continental scales, Geophysical Research Letters, 38, 10.1029/2010gl046427, 2011a.

72. Gleeson, T., Smith, L., Moosdorf, N., Hartmann, J., Durr, H. H., Manning, A. H., van Beek, L. P. H., and Jellinek, A. M.: Mapping permeability over the surface of the Earth, Geophysical Research Letters, 38, 6, 10.1029/2010gl045565, 2011b.

73. Koster, R. D., Suarez, M. J., Ducharne, A., Stieglitz, M., and Kumar, P.: A catchment-based approach to modeling land surface processes in a general circulation model 1. Model structure, Journal of Geophysical Research-Atmospheres, 105, 24809-24822, 10.1029/2000jd900327, 2000.

74. Krakauer, N. Y., Puma, M. J., and Cook, B. I.: Impacts of soil–aquifer heat and water fluxes on simulated global climate, Hydrol. Earth Syst. Sci., 17, 1963-1974, 10.5194/hess-17-1963-2013, 2013.

75. Lei, H., Huang, M., Leung, L. R., Yang, D., Shi, X., Mao, J., Hayes, D. J., Schwalm, C. R., Wei, Y., and Liu, S.: Sensitivity of global terrestrial gross primary production to hydrologic states simulated by the Community Land Model using two runoff parameterizations, Journal of Advances in Modeling Earth Systems, 06, 10.1002/2013MS000252, 2014.

76. Milly, P. C. D., Malyshev, S. L., Shevliakova, E., Dunne, K. A., Findell, K. L., Gleeson, T., Liang, Z., Phillips, P., Stouffer, R. J., and Swenson, S.: An enhanced model of land water and energy for global hydrologic and earth-system studies, Journal of Hydrometeorology, 10.1175/JHM-D-13-0162.1, 2014.

77. Morris, P. J., Baird, A. J., and Belyea, L. R.: The DigiBog peatland development model 2: ecohydrological simulations in 2D, Ecohydrology, 5, 256-268, 10.1002/eco.229, 2012.

78. Niu, G. Y., Yang, Z. L., Dickinson, R. E., Gulden, L. E., and Su, H.: Development of a simple groundwater model for use in climate models and evaluation with Gravity Recovery and Climate Experiment data, Journal of Geophysical Research-Atmospheres, 112, 14, 10.1029/2006jd007522, 2007.

79. Paniconi, C., Troch, P. A., van Loon, E. E., and Hilberts, A. G. J.: Hillslope-storage Boussinesq model for subsurface flow and variable source areas along complex hillslopes: 2. Intercomparison with a three-dimensional Richards equation model, Water Resources Research, 39, 17, 10.1029/2002wr001730, 2003.

80. Sheffield, J., Goteti, G., and Wood, E. F.: Development of a 50-year high-resolution global dataset of meteorological forcings for land surface modeling, Journal of Climate, 19, 3088-3111, 10.1175/jcli3790.1, 2006.

81. Shevliakova, E., Pacala, S. W., Malyshev, S., Hurtt, G. C., Milly, P. C. D., Caspersen, J. P., Sentman, L. T., Fisk, J. P., Wirth, C., and Crevoisier, C.: Carbon cycling under 300 years of land use change: Importance of the secondary vegetation sink, Global Biogeochemical Cycles, 23, Gb2022, 10.1029/2007gb003176, 2009.

82.

83.