The Water Table Model (WTM) v2.0.1: Coupled groundwater and dynamic lake modelling

Kerry L Callaghan[1,2,3], Andrew D Wickert[2,4,5], Richard Barnes[6], and Jacqueline Austermann[3]

[1]Department of Earth & Environmental Sciences, University of Illinois, Chicago, USA
[2]Department of Earth & Environmental Sciences, University of Minnesota, Minneapolis, USA
[3]Lamont–Doherty Earth Observatory, Columbia University, New York, USA
[4]Saint Anthony Falls Laboratory, University of Minnesota, Minneapolis, USA
[5]Deutsches GeoForschungsZentrum (GFZ), Potsdam, Germany
[6]Lawrence Berkeley National Lab, Berkeley, USA

Correspondence: Kerry Callaghan (kerryc@uic.edu)

We would like to thank Reed Maxwell and the two anonymous referees for their constructive comments on our work. We have worked to implement changes into our submission and believe that this has strengthened our manuscript.

We will address firstly the comments of the first, anonymous referee.

Comment: It would be better to discuss the impact of Dupuit–Forchheimer approximation on the accuracy of water table depth simulation, such as the possible range of errors in this study. Because this study does not include vertical hydraulic infiltration, the critical infiltration process. Relevant studies can be cited to strengthen the explanation of the applicability of this assumption. For example, considering the increased computational efficiency due to this assumption, the error is not a major issue, etc.

Response: Thank you for noting this. An important assumption of the Dupuit-Forchheimer approximation is that the slope of the water table should be gentle. Quoting Freeze and Cherry (1979), "Calculations based on the Dupuit assumptions compare favourably with those based on more rigorous methods when the slope of the free surface is small and when the depth of the unconfined flow field is shallow." To be more specific, Smith and Wheatcraft (1993) state that the assumptions do not introduce significant errors when water table slopes are less than 0.18. We have added text noting the assumption of a gentle slope in the water table.

In the case of the climate-influenced present-day simulation that we include in this submission, more than 99% of cells have a slope below the threshold of 0.18. In addition, as the reviewer states, computational efficiency is much improved when using these assumptions.

Other model users should only use this model when valid for their conditions and, as such, may elect not to use a model making the Dupuit-Forchheimer approximations in cases where vertical flux is of great importance, or when water table is known to be steeply sloping. Steeper slopes in water table may be observed more frequently at higher spatial resolutions, so that users should be cautious about the spatial resolution of their WTM simulations.

Comment: If I understand correctly, only Figures 4 and 5 compared the model with the observations, but the time used for the comparison is unclear. Therefore, the overall accuracy of the model remains vague.

Response: Thank you for pointing this out - we would like to have our figure captions be as clear and informative as possible. The simulation data shown in Figures 4 and 5 is from our present-day climate-driven simulation (this was stated in the caption of Figure 4 and has been added to the caption of Figure 5), representative of the late 20th/early 21st century. The observational data covers a similar time period. Groundwater well data from Fan et al. (2013a) included readings from 1927-2009. Dates covered by the Kourzeneva et al. (2012) lake dataset are not entirely clear, but seem to predominantly be from the 1990s and early 2000s. Wetland data Zhang et al. (2023) was based on remotely sensed imagery from 2020. Although it is not possible to find all data for the same year, we believe that all of the data sources used are applicable for the scale of work shared here. We have added text to the figure captions indicating that observations correspond to years 1927-2020.

Comment: The three data sources for the observations of water table level are groundwater-table depth, lake extents and depths, and wetland depth (assumed to be zero). Different data sources may have measurement discrepancies and issues with the validity of assumptions. Ultimately, all three data sources are treated as equally valid and compared with the model values. Providing some evidence for this approach would enhance the credibility of the paper. The basic attributes of the observational data should also be listed in the appendix or the main text. It is uncommon to only cite the paper without any data description.

Response: One reason that these three data sources (groundwater-table depth, lake level, wetland locations) are treated the same way is that a part of the philosophy of our model design is that groundwater, lake, and wetland levels are all represented by the same water table. Whether above or below the land surface, the water table is a single continuous plane. As such, any data that defines the location of the water table can be used for validation. With that said, it is true that each data source may have its own associated discrepancies and assumptions. One notable difference between these is the type and resolution of the data. Both lake and wetland datasets define entire cells as ‘lake’ or as ‘wetland’, and we are able to match these with the cells in our simulation. The lake dataset, like our simulation, has a spatial resolution of 30 arcseconds, though temporal differences in lake level still cause some mismatches. The wetland dataset has a spatial resolution of 30 m, and so we define a 30-arcsecond cell as ‘wetlands’ if it contains more than 50% of these 30 m wetland cells. The groundwater well data, however, does not define depth-to-groundwater across an entire cell: it is point data. Given the changing topography within a single 30-arcsecond cell, a point at one end of the cell may look significantly different from a point at the other end, while our simulation should represent something close to the cell mean. A brief discussion of this differentiation has been added to the manuscript, and a non-exhaustive discussion of the data-specific discrepancies and assumptions is given in section 6.1.1 of the paper.

Comment: Regarding the schematic diagram Figure A1 of the model, it appears that the soil and channel are unidirectionally coupled, meaning that under saturated soil moisture conditions, water flows from the soil to the channel/lake, but the channel/lake is impermeable, with no water returning to the surrounding soil. This situation could introduce errors. If so, could the author discuss the potential impact on this study?

Response: In lakes, water can and does return to the surrounding soil: any cell in the domain that contains lakewater will first allow this water to seep into the subsurface to saturate the subsurface, if needed. In addition, during the groundwater flow step, it is possible for lakewater to horizontally flow to a neighbouring cell with a higher elevation and pass to the subsurface in this way. Finally, if the infiltration option is turned on by the model user, then some surface water will infiltrate into the subsurface as it moves across the surface during the 'Move surface water' step. Our simulations shown in the paper did not have the infiltration option turned on, but it is available. If it was unclear what we meant by 'move groundwater' and 'move surface water' on Figure 1A (reproduced below as Figure 1), we have updated these boxes to read 'Move groundwater: lateral; possible exfiltration' and 'Move surface water: lateral; possible seepage'.

Comment: Figure 1: The elements in Figure 1 are relatively simple, it is recommended to use a textual description instead. Compared with Figure 1, Figure 2 sufficiently demonstrates the relationship between groundwater levels and soil moisture infiltration. Alternatively, the author may provide more information for the water level simulation in Figure 1. The author can decide which option is more suitable.

Response: Figure 1 is more of a conceptual figure than a model process figure. As such, it is not intended to provide information about the actual water table simulation, but rather to show the context for the importance of the water table. This is particularly important for GMD given that it is an interdisciplinary journal. We would like to ensure that readers can understand the broader context and importance of our work. Anecdotally, several non-hydrologists who have seen this figure have expressed that it has helped them to contextualise our work.

Comment: Figure 2: It lacks soil moisture contours that would indicate the gradient of total hydraulic head in Darcy's law or soil pressure head ψ. In soil moisture-related studies, displaying such schematic diagrams is a common practice. Showing the gradient helps illustrate the process of the system gradually approaching equilibrium from its initial state. In this study, in other word, it would be better to draw a schematic diagram of the hydraulic gradient, which is equal to the slope of the water table and does not vary with depth below the water table.

Response: The reviewer is correct that we do not include soil moisture contours in Figure 2. In our model, P-ET is directly applied to the groundwater table, after subtracting runoff based on a user-supplied runoff coefficient. This is discussed in Section 3.3 of the paper. The WTM neglects unsaturated-zone processes, i.e., soil moisture is not explicitly considered. Our reasons for this choice are (1) to maintain simplicity in the model; (2) in longer-term simulations such as those included here,

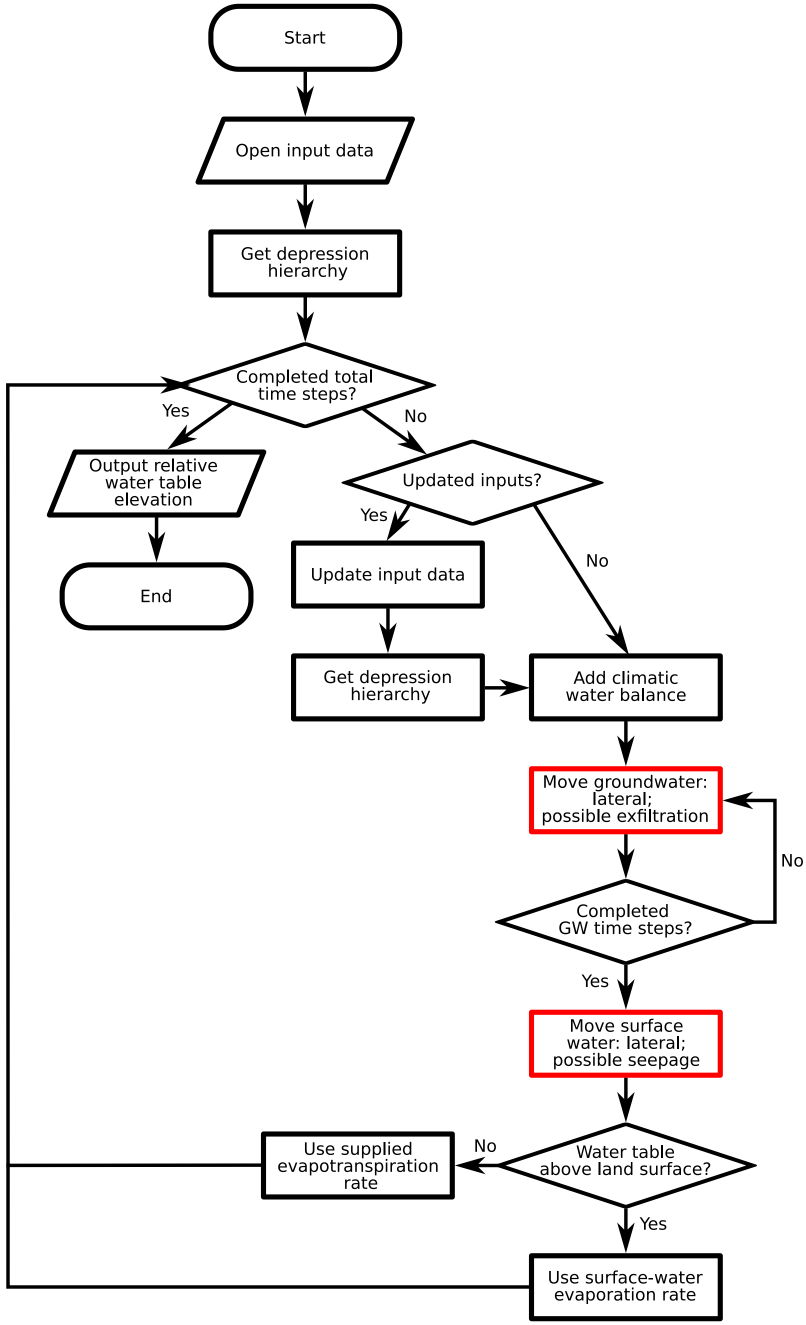

Figure 1. Steps taken by the WTM. The two red boxes indicate the components used to couple groundwater and surface water.

unsaturated zone processes become less important because of the time scales, allowing us to assume that water has percolated through the unsaturated zone and recharged the water table; and (3) to prevent the model from becoming too computationally expensive to be able to perform the long-term, continental-scale simulations that are our focus.

Comment: Figure 3: Could the author include the simulation time period in the figure caption? The same issue exists with the other figures as well. The Figure 6 simulation period is also unknown.

Response: I have added both the date that the water table is relevant for, and the duration of model spin-up to the captions for Figures 3 and 6.

Comment: Figure 4: The y-axis values in Figure 4 are very small. Please confirm that the area enclosed by the curve and the x-axis is 1. If the sum of the probabilities is not 1, please explain the reason beside the x-axis limits. Figure b shows good results. There are a few issues. The missing observations are replaced with assumed values; could the authors exclude comparisons for the missing observations from the histogram? Observations cover 11.3% of the cells within the research domain. Could the authors display the location of these 11.3% in a raster figure in the appendix? This would allow readers to intuitively understand the range shown by the histogram.

Response: The axis labels appear to be similar in size to other text in the document, although of course these can be modified if needed. These are probability density plots, so that if all values were included, the area under the curve would be equal to 1. For ease of visual processing, we have chosen to cut off the x axis at values that included most of the data while also allowing us to show the data at a larger scale and not only focus on outliers. In the updated plots, Figure 4(a) includes 88% of the total data, and Figure 4(b) includes 96.5% of the total data. Our figure caption does state that a small proportion of data lies outside the axis limits.

To combat some of the issues seen with observational values, we have made changes to the dataset as detailed in the next response. As suggested by the reviewer, we have added a figure (reproduced below as Figure 2) showing the location of cells used in the data comparison in Appendix F.

Comment: Figure 5: In Figure 5(a), there is a noticeable distinction in the number of observation points in the regions where both x and y axes are negative and where x is positive and y is negative. What causes this? This difference appears to be caused by interpolation. The model has a feature where many simulated data points have the same observed value. In Figure 5(b), similarly, many model simulation values have the same observed values. Please ensure that this figure has accurately excluded meaningless simulation results and explain the reasons for such discrepancies/phenome. In line #329, the discrepancies include seasonal variations in observed data, one solution would be normalized data for each season.

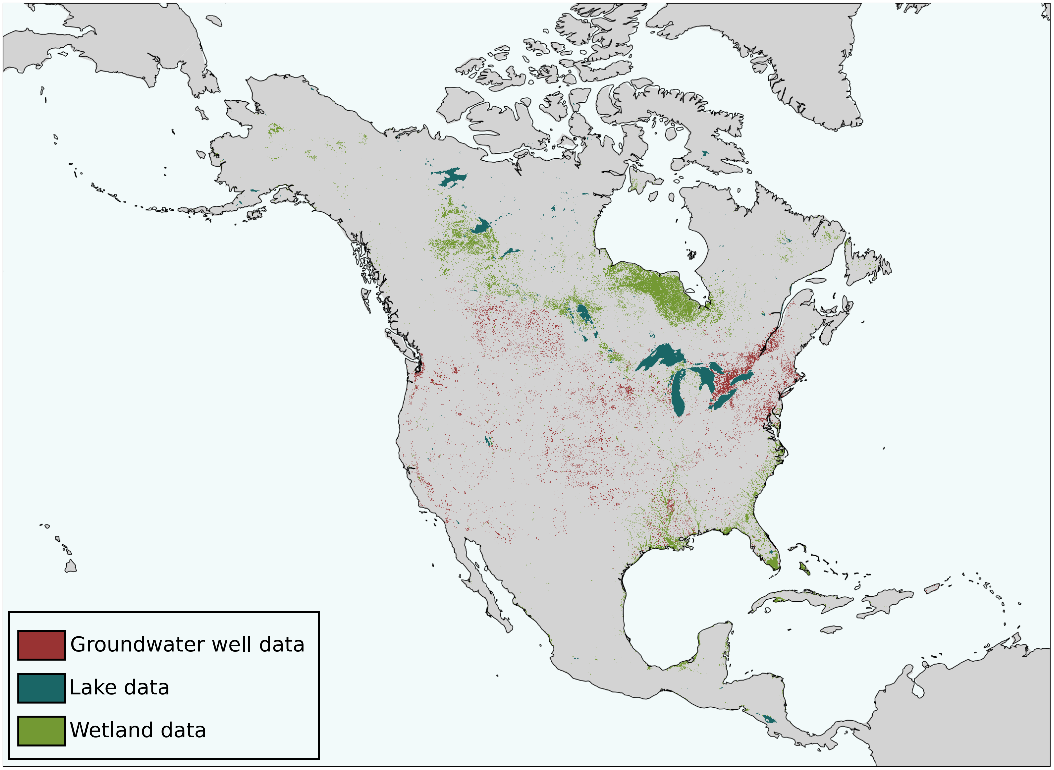

Figure 2. Location and spatial extent of each of the three data sources used in validation of depth to water table.

Response: Based on this and other comments regarding our validation, we made some changes to the observation dataset used for validation. These changes were as follows: (1) we were more rigorous in the selection of wetland cells. The wetland dataset has a spatial resolution of 30 m. Previously, we mistakenly resampled this to our coarser 30 arcsecond resolution using a nearest neighbour approach, which was not the best approach. Now, we instead evaluated which 30-arcsecond cells would contain >50% wetlands based on the finer-resolution input data, and only evaluated these as ’wetland’ cells. (2) we noted that many of the data-model mismatches that were associated with horizontal lines on the originally included scatter plot were related to a mismatch between the lake dataset and lake levels captured in the topography dataset. Some lake extents in the lake dataset were greater than lake extents represented by the topography. This could be because lake levels were actually higher when these extents were captured, because the lake dataset purposely assumed a high water level, or because of errors in either dataset. In lakes that were represented by only a ‘mean depth’ in the dataset, the result was that single, mean depth values for a lake were being compared to variable values rising up the slope alongside a lake. To combat this, we ‘shrank’ each lake by 5 cells. Although this would have removed some good data, bad data was successfully removed and spurious lines on the scatter plots dramatically reduced. The updated scatter plots, included with this submission, show better matches with significantly fewer of the linear features described by the reviewer. The updated plots are shown below (Figures 4 and 5).

Regarding the discrepancy between where x and y axes are both negative and where x is positive and y is negative: all of these points are comparisons to groundwater well data (hence the observation is negative, below the land surface). In most cases, the simulation is also negative, leading to many points where both is negative. In some cases, the simulation is instead positive. This could be because the size of a cell could accommodate both a lake (positive result in the simulation) and a well with a negative (below the land surface) water table within the same 30-arcsecond grid cell.

Comment: Figure 6: G3M treats surface water as a static boundary condition, all water above the land surface would either evaporate or run off. Therefore, the G3M's poor performance within +/-10m is obvious. Why choose to compare with G3M simulation results that have obvious assumption biases?

Response: G3M and Fan et al. (2013a) are two prominent models that are being used to work on the integration of terrestrial water into Earth system models. One reason to compare our results with these models because of their similar treatment of groundwater, which makes them an appropriate for comparisons of groundwater results. In addition, these models are pushing science in the same direction that we are attempting to, with one major aim being the improvement of large-scale representations of the water table. Here, we add a step improvement to this larger purpose through our inclusion of the dynamic lake component. Our comparison emphasises the difference in treatment of surface water: both G3M and the Fan et al model treat surface water as static boundary conditions, in different ways. The WTM does not, which is one of the primary rationales behind the work. The inclusion of this temporally dynamic surface water component allows simulation of changing lake depths through time.

Comment: Figure 7: The raster figure located in the Canada region does not have a corresponding color scale. Another way is adding a note in the caption to show thickness value. Figure 3 has the same issue.

Response: I think the reviewer is referring to the raster displaying ice thickness. If so, the caption already states: 'Continental ice thickness from ICE-6G (Peltier et al., 2015) varies from blue-grey (thin) to white (thick)'.

Comment: Figure 8: If I understand correctly, it is not very clear how accurately the model's ice cover section compares with the observations. So it is unclear of the water table change from LGM to present day, the Figure (a) shows slight increases in northern part of Northern America. In the context of overall climate change, additional evidence may be needed to support this result.

Response: The ice cover used in this simulation was based on the widely-used ICE-6G model, which was made to match mapped ice outlines from Dyke et al. (2002). For our simulations, ICE-6G provides ice thickness, which is indeed a source of uncertainty for Glacial Isostatic Adjustment (GIA) (Simon and Riva, 2020; Melini and Spada, 2019). GIA modelling is a very active area of research and we anticipate that it will see continued improvement in the coming years and decades. With regards to the shared need of GIA, past ice extent, and past climate to have more knowledge about the past in order to create better simulations, the WTM also has the potential to aid in the connection of these simulations to geological data. For example, water tables from the WTM can be matched to palaeo lake shorelines as in independent check of lake water levels. We plan to evaluate these connections further in future work.

Bringing the focus back to our results shown in Figure 8, this figure shows that there is there is a strong correlation between differences in the water table and differences in P-ET during the two time periods simulated. Increasing P-ET in Northern North America (Figure 8b) is the likely driver behind increases in water table in this region.

Comment: Figure 10: During the period of 17 - 16 ka BP, groundwater shrank rapidly, while surface water remained relatively stable, with only a slight decrease. This result is quite unexpected; could the authors explain the reason?

Response: Figure 9(b), reproduced below as Figure 3, shows the total change in water table from 21 ka to 16 ka, much of which took place in the time period from 17-16 ka. As can be seen on this figure, the change was concentrated in the more Southern portions of the continent (here we add a thick black outline to this region), where fewer lakes exist based on the region's topography. Most lakes were located in regions of the continent where the total water table was more stable. This means that lakes saw relatively little change in water storage volume because they existed in regions where P-ET was relatively stable; while groundwater saw more dramatic changes in water storage volume as a result of climatic change over some of its area.

Based on other comments regarding the scope of our work, this section has been removed from the current paper, with the aim of providing these interesting results with enough space to be thoroughly reported elsewhere.

Comment: In the conclusion section, the results of other studies related to the significant findings of this research could be discussed here, such as the main finding of reduction trend in lake volume. For example, this article might mention similar conclusions 1. The conclusion section also needs to be kept complete.

Response: Thank you for sharing this fascinating paper. This paper focusses mostly on changes between 21 ka, 6 ka, and the present day. Part of our plans for future work include a transient simulation of water table from 21 ka to the present day, and we look forward to comparing our results with Zhang and Li (2022) and citing their paper in this future work, once our results include simulations over their time period. We are not aware of other, similar studies to discuss in this portion of the work. Because the focus of the paper is on the model itself and simulations included herein are only intended as proof-of-concept, we focus on the importance of the model construction in our conclusion.

Comment: If climate models with future global warming/radiation/aerosol scenarios are used as model input, it is possible to show future changes in groundwater levels and lake levels. If there are some eye-catching conclusions, this could potentially

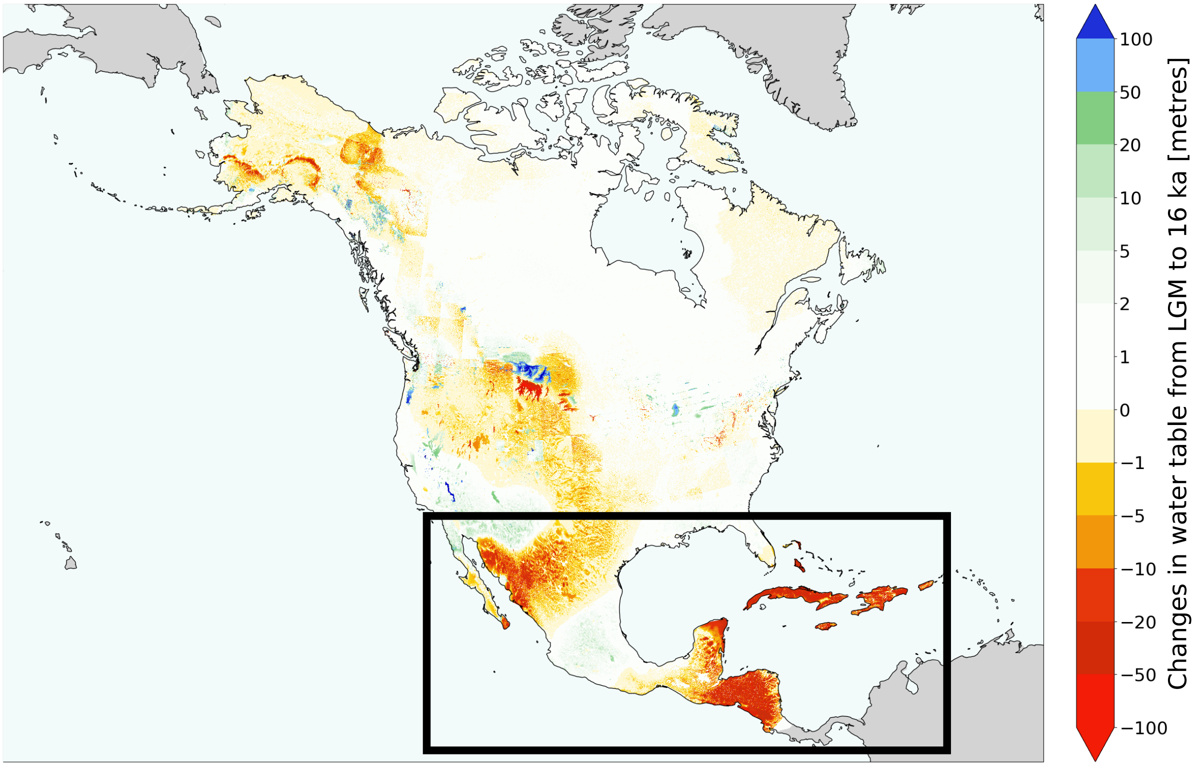

Figure 3. Change in water table depth from LGM to 16 ka. The thick black line indicates the region where the greatest change in volume of stored water was seen. This region includes the area that saw the greatest change in P-ET over the simulated time period.

increase the citation rate of the paper.

Response: This is one of the ideal use-cases of the model, and will be addressed in future research. It is beyond the scope of this paper.

Comment: Line #452, “wherever the water table”, “the” is the typo.

Response: Thank you, we have corrected this typo.

Next, we will address comments from the second reviewer, Reed Maxwell:

Comment: The scope of this work is a bit broader than it might otherwise need to be. I find a few general themes / messages in the work:

A combination of a large scale groundwater parameterization similar to Fan et al and a lake model similar to Barnes et al (2021; fill-spill-merge) A robust numerical implementation of the otherwise simple (explicit FD) formulation of Fan et al using the PETSC numerical library that is also documented and open source, etc. A series of large scale simulations, including a present day north american simulation and a reconstruction or evolution of the water table since the last glacial maximum. A lengthy series of appendices that detail model inputs, assumptions and a range of equations presenting infiltration and long-term ET

While most of these themes are interesting and this clearly represents a lot of work, it is hard to tell what the central focus of the manuscript is. Is this a possible climate application? A model to represent present-day conditions? (i.e. is the science / results the focus); is this a model description paper that comprehensively talks about these interactions? (i.e. is the model formulation GW-lake or GW-lake-land surface the focus); is this a numerical applications paper that talks about performance and the need for putting this framework on something like PETSC? There are multiple GMD paper-worthy topics in here (in my opinion) and having them all in one manuscript so prominantly makes it a bit hard for the reader to follow the main themes and take-home messages and weakens the work overall. I'm not suggesting a complete re-write of the manuscript, but some careful discussion among the authors to help focus and tighten this work around a central theme.

Response: Thank you for your clear and constructive thoughts here. We agree that the paper as submitted here became overly bloated and lacks focus. Our intended focus is on the GW-lake-land surface model formulation, with proof-of-concept simulations. To make this clearer, we have reduced the length of Section 6 (the simulations) and worked to emphasise further that these are intended as a proof-of-concept, leaving the actual model design as the focus of the paper. We have removed the transient simulation and been more succinct in our descriptions of the present-day and LGM simulations. Our intention is to keep the main focus on the description of the model itself.

Comment: Introduction: The intro starts with a nice motivation on the changes of terrestrial water storage over geologic time scales. It defines the need for a GW+lake model that can run over much longer timescales than the current literature provides. My feeling is that in the paragraph around line 65 the authors might keep this focus instead of diving into anthropogenic impacts, etc. I recommend expanding this out (if this is the focus the authors decide to go with), and focusing on the details of the work that will enable these types of simulations. In other words, I feel this paragraph is where the manuscript starts to split into a few different directions losing the nice focus it had in the rest of the intro. Additionally, the authors might also consider the conceptual framework of Condon et al WRR 2021 (figure 4) as a way to further classify the space their work sits within.

Response: Again, thank you for your clear and constructive thoughts here. We have made modifications to the introduction to try to make our focus clearer.

Comment: There are a number of really short paragraphs that are distracting to the reader. Examples include lines 81-83 (perhaps leave this to the model / data availability statement or put in a footnote); lines 94-97; lines 266-268; lines 269-271

Response:We have edited to remove the short paragraphs that you point out, as well as several others that we noticed.

Comment: Comparison to observations for the present-day simulation: The model comparisons to water table observations as currently written are very hard to follow and in my opinion do not support the message of the manuscript. Figure 5 is confusing and looks like it might actually contain plotting errors or artifacts. In (A) he pattern and striped lines (some horizontal and vertical, some at 45-deg angles) look like there might be an aliasing issue or even something numerical going on. The values, even when plotted as head (incuding topography, in B) in which some of the areas that should have heads over 1km are below sea level. These graphs plus the associated analysis need to be carefully reviewed and possibly reconsidered. I understand the need to compare models to observations but this section as presented does not instill confidence in the model formulation. Perhaps other benchmarks could be used with the model? Are there analytical solutions to which the model could be compared that might also serve as unit tests for the code framework?

Response: Based on this and other comments regarding our validation, we made some changes to the observation dataset used for validation. These changes were as follows: (1) we were more rigorous in the selection of wetland cells. The wetland dataset has a spatial resolution of 30 m. Previously, we mistakenly resampled this to our coarser 30 arcsecond resolution using a nearest neighbour approach, which was not the best approach. Now, we instead evaluated which 30-arcsecond cells would contain >50% wetlands based on the finer-resolution input data, and only evaluated these as 'wetland' cells. (2) we noted that many of the data-model mismatches that were associated with horizontal lines on the originally included scatter plot were related to a mismatch between the lake dataset and lake levels captured in the topography dataset. Some lake extents in the lake dataset were greater than lake extents represented by the topography. This could be because lake levels were actually higher when these extents were captured, because the lake dataset purposely assumed a high water level, or because of errors in either dataset. In lakes that were represented by only a 'mean depth' in the dataset, the result was that single, mean depth values for a lake were being compared to variable values rising up the slope alongside a lake. To combat this, we 'shrank' each lake by 5 cells. Although this would have removed some good data, bad data was successfully removed and spurious lines on the scatter plots dramatically reduced. These changes also improved the discrepancies previously seen on the head scatter plot. The updated plots are shown below (Figures 4 and 5).

You will still notice the 45-degree angled lines on the scatter plot. These are associated with several northern-latitude lakes in which our simulation has underfilled the lakes. The result is that observations have higher values than simulations, and both increase with depth in the lakes. The vertical line at simulation == 0 is also associated with this (simulated lakes having a water table at the land surface rather than containing surface water). Given that this is only happening in northern-latitude lakes, we believe that this is because our dataset for open-water evaporation did not take the influence of lake ice on evaporation into account. This is external to the actual WTM simulation, which takes open-water evaporation as a data input. In future work, this issue could be remedied by users incorporating lake ice into this input data, or by an add-on module to the WTM.

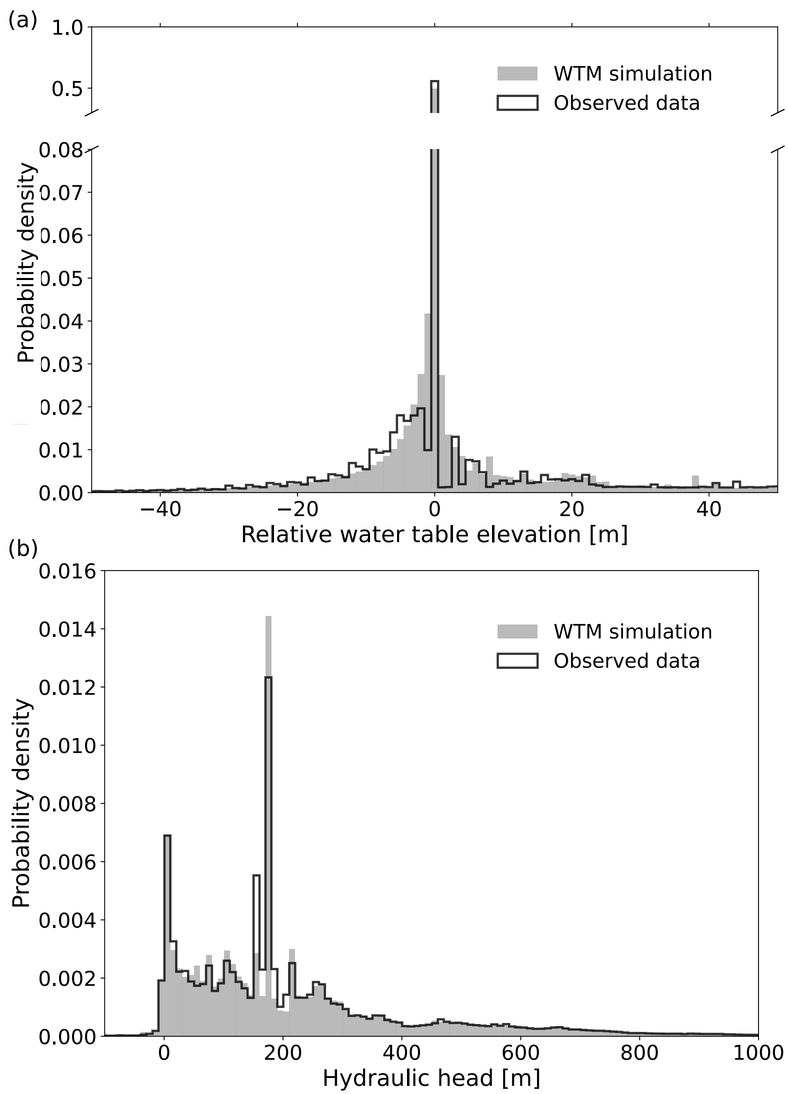

Figure 4. Simulated versus observed present day climate-driven water table in North America. (a) Relative water-table elevation; **(b)** hydraulic head. Observations include lake, wetland, and groundwater-well data from Kourzeneva et al. (2012), Zhang et al. (2023), and Fan et al. (2013a), respectively. The dates represented by these data range from 1927 (for some of the wells) to 2020. A small proportion of both observations and simulated relative water-table elevations and heads lie outside the x-axis limits.

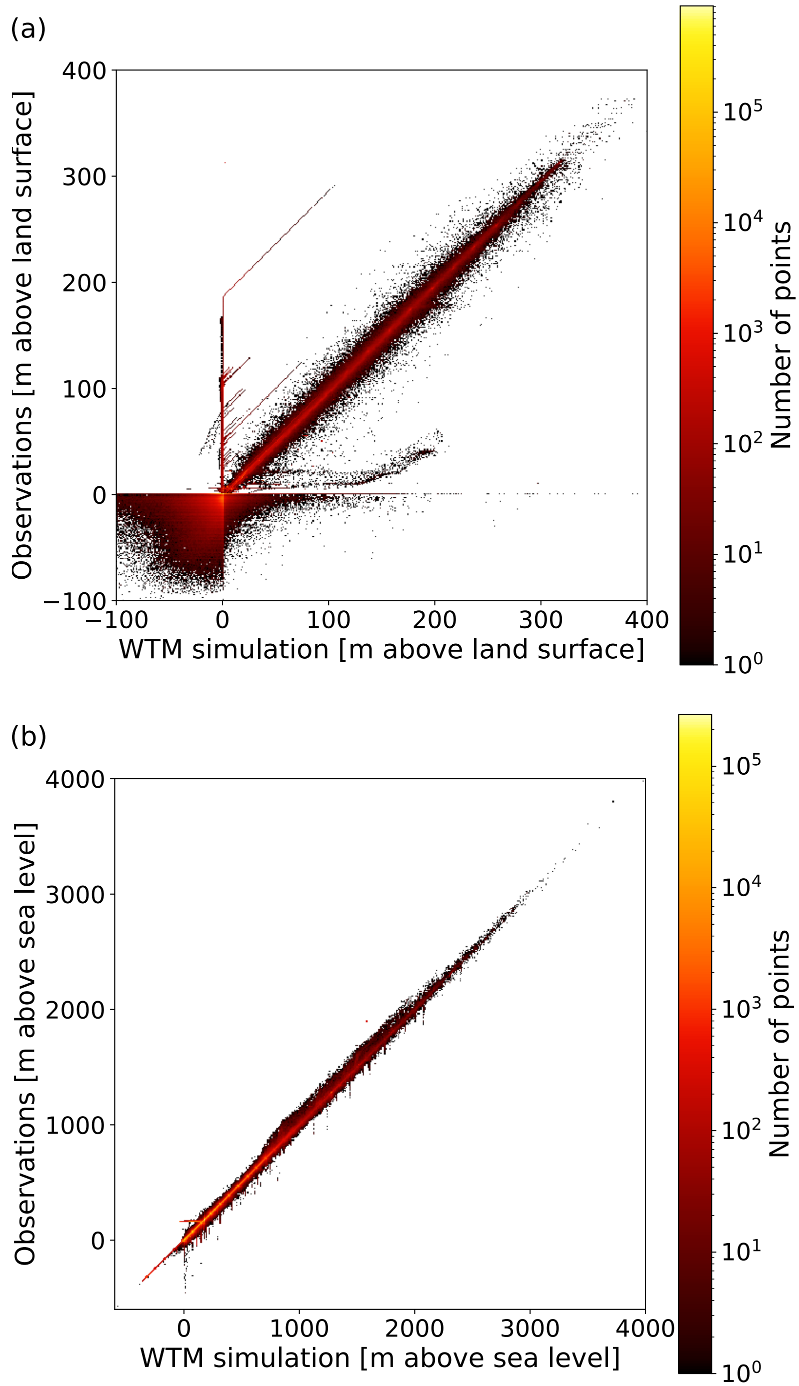

Figure 5. Observations vs. WTM present day climate-driven simulation results. (a) Relative water table elevation; **(b)** Hydraulic head. Observations include lake, wetland, and groundwater-well data from Kourzeneva et al. (2012), Zhang et al. (2023), and Fan et al. (2013a), respectively. The dates represented by these data range from 1927 (for some of the wells) to 2020. These comparisons include only those model cells that contain observations.

Comment: What is the impact of including lakes in the simulation? This is discussed in text (lines 343-352) but an approach might be simulate with and without the lake model (or with a simpler approach the removes water) to show the difference spatially and using something like histograms.

Response: We had time to run a simulation without lakes for ∼12,000 years for this response. We include here some figures showing how the results differ. First, we reproduce the data-model comparison figures (Figures 4 and 5 from the paper) using the no-lake result. These are shown below as Figures 6 and 7 of the response. It may be unfair to compare a no-lake simulation to data that includes lake levels, but nevertheless, we do so. Figure 6 shows, of course, that the data includes no surface water in panel (a), but also that the data does not reproduce the peaks seen a little below 200 m head in panel (b). These, while not exactly matched with the full simulation, are much more closely matched there. Figure 7 (a) again somewhat unfairly shows lake data against a no-lake simulation: surface water has a strong correlation with observations in the full simulation, but, obviously, not here. Note the new horizontally-oriented blips seen on the head scatter plot (panel b): these appear to be related to large, deep lakes (such as the great lakes) which are now empty in the simulation.

Finally, we provide here Figure 8 to show the spatial difference between these two simulations (with-lake vs no-lake). Observe: (1) lake levels are higher when lakes are simulated (obviously), most visually notable in the USA's great lakes. (2) there are several areas where groundwater is regionally higher when lakes are simulated, the largest of which is located in Mexico. There is little surface water in these regions, so that the reason for the discrepancy is not immediately obvious. One possibility is that in the case where lakes are simulated, surface water runs off into depressions and there seeps to the water table, raising regional water table. In the simulation not including lakes, this water was removed from the system before having a chance to seep to the water table (because the distribution of water into lake locations was not simulated; water was simply removed from the land surface). (3) There are a few areas, e.g. Florida, where the simulation with no lakes actually has a higher water table. This is likely because this simulation was not run for as long (∼12000 years vs ∼20000 years) simply because of time constraints in completing this response. The water table was not yet equilibrated in these regions in the no-lake simulation.

A comparison of the two simulations shows that the lake-enabled simulation contained ∼17.1 cm sea-level equivalent (SLE) more water than the lake-free simulation, of which ∼13.2 cm SLE was lakewater and the remainder was groundwater.

Unfortunately, although it is clear that the lake-enabled simulation performs better in areas that contain lakes, it remains unclear whether either of these simulations performs better where groundwater is concerned.

Comment: LGM simulation. This section is in my opinion pretty novel. Very few efforts have tried to simulate dynamics across these spatial temporal scales (e.g. Lemieux et al) and this section ties well into the motivation for the work.

Response: Thank you: this section is intended as a proof-of-concept to illustrate what we believe to be one of the most appropriate use-cases for this model, and we are excited to continue simulating long-term temporal dynamics of the water table in future work.

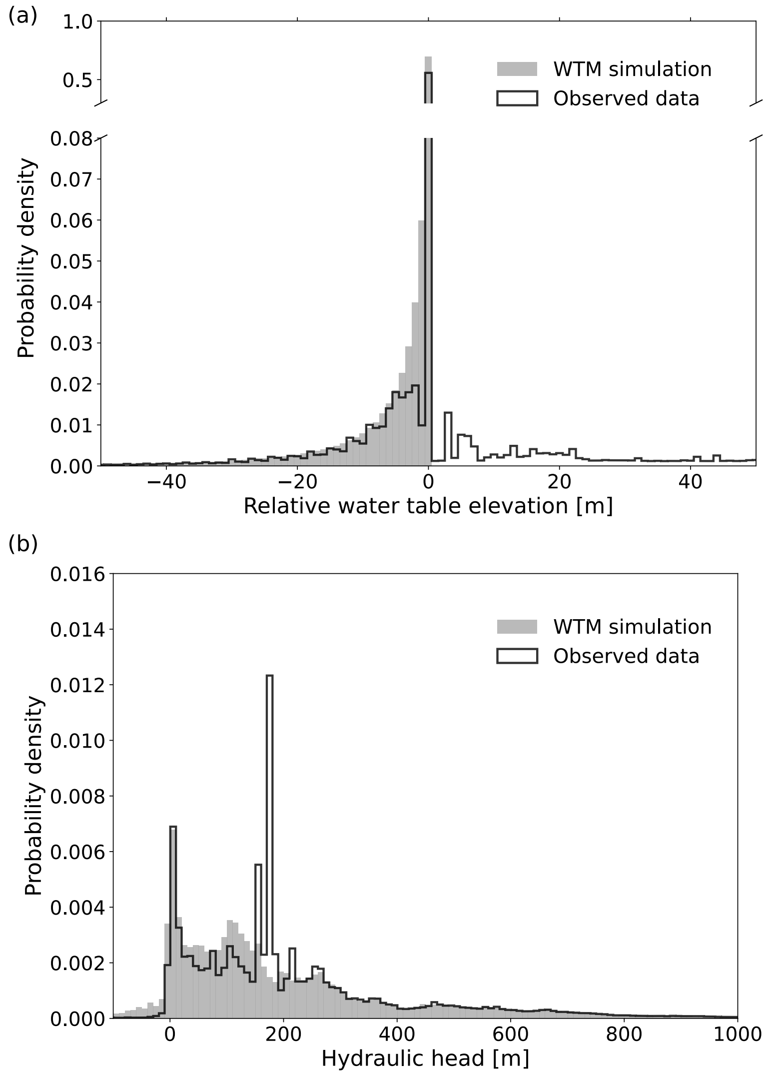

Figure 6. Simulated versus observed present day climate-driven water table in North America, with the simulation removing all surface water. (a) Relative water-table elevation; **(b)** hydraulic head. Observations include lake, wetland, and groundwater-well data from Kourzeneva et al. (2012), Zhang et al. (2023), and Fan et al. (2013a), respectively. A small proportion of both observations and simulated relative water-table elevations and heads lie outside the x-axis limits.

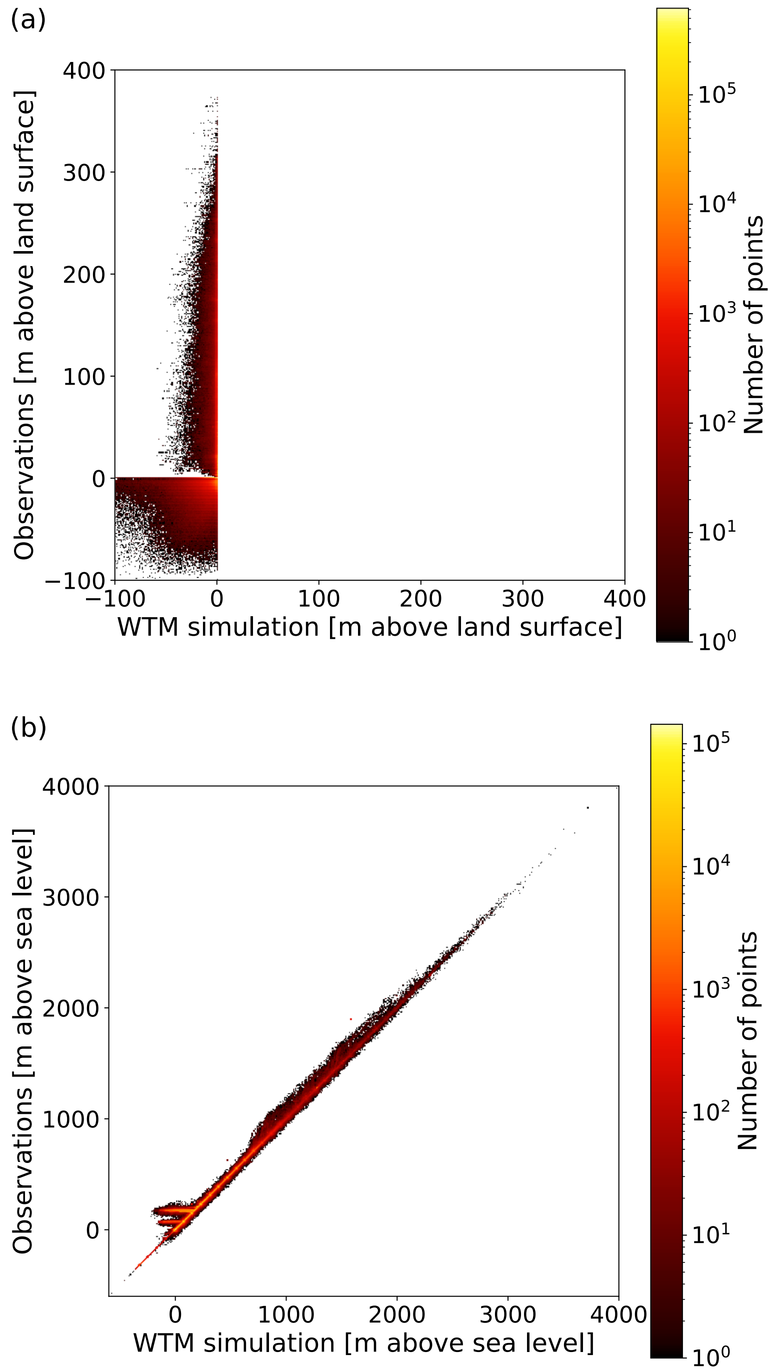

Figure 7. Observations vs. WTM simulation results, with the simulation removing all surface water. (a) Relative water table elevation; **(b)** Hydraulic head. These comparisons include only those model cells that contain observations.

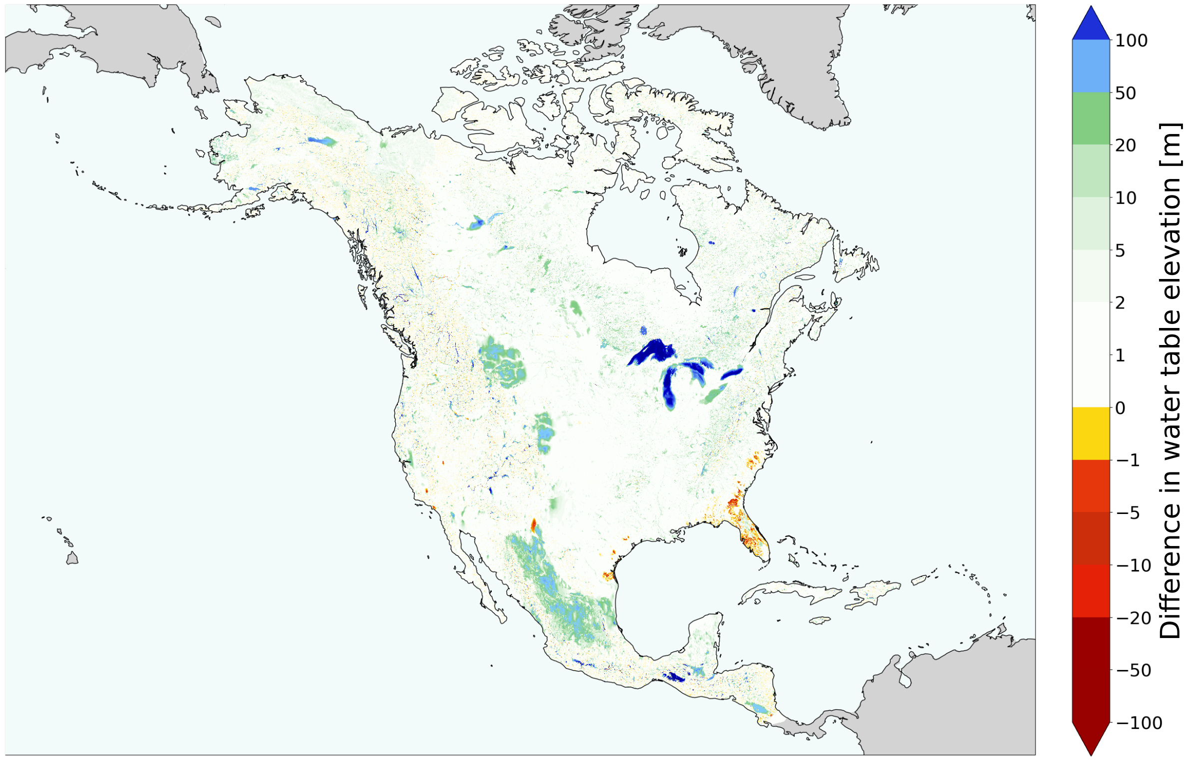

Figure 8. Simulated water table including lake surfaces versus removing all surface water. Blue and green colours indicate locations where the simulation contained more water when lakes were included.

Comment: Uncertainty. Parameter uncertainty is a primary challenge in these types of modeling studies. It seems that one advantage of a simplified / reduced physics model is the ability to run very rapid simuations that include parameter sensitivity and uncertainty. This would greatly enhance the work to provide understanding of what parameters might be driving outputs and to address the substantial uncertainty in the assumptions made around the geologic framework. If this is outside the scope of the work, it would be very helpful to provide discussion of uncertainty and how it might impact the results and conclusions.

Response: Thank you for another thoughtful comment. Yes, a sensitivity analysis is a possibility with this model, and one that we hope to pursue in future work, in particular when we present work with more of a focus on simulation results (versus our focus on model development in this paper). With respect to input data to the model, qualitatively, we believe that $P - ET$ is a primary driver of change in water table in our simulations. A study of uncertainty in the CCSM3 climate model (which is the basis of the Trace-21K simulation used as input data for our LGM run) (Deser et al., 2012) found that weather noise was responsible for much of the internal variability. In our simulations, we attempt to reduce weather noise from the input data by averaging inputs over multiple years, so that our results are better able to represent steady-state conditions for a certain point in the longer-term climate evolution of the system, rather than for a single year which may represent its own set of weather conditions. Hopefully, this processing choice has reduced the impacts of climate uncertainty on our simulation results. The CESM Large Ensemble project (Kay et al., 2015) notes that internal climate variability is of particular importance at the regional scale; it will therefore be important to consider multiple realisations of climate in future work.

With respect to parameters set in the model's configuration file, most of these involve no uncertainty and are simply facts about the domain (such resolution) and choices about how to run the model (such as the number of time steps to complete). The only parameters here which include some uncertainty are those pertaining to the e-folding depth for decay of hydraulic conductivity; we based our choice on calibration done by Fan et al. (2013b). We choose to use the same values they proposed both because of the similarity of our approaches and because we want to make our simulation more comparable to the Fan et al. (2013a) work for benchmarking purposes. Fan et al. (2013b) note that these parameters are dependent on grid cell size; we use their recommendations for 30-arcsecond cells, since this is the spatial resolution of our simulations.

Finally, we address the third, anonymous reviewer.

Comment: Uncertainty in evapotranspiration (ET) estimation: One of the key inputs for the WTM is evapotranspiration (ET), which is widely recognized as a challenging and uncertain variable in hydrological modeling. Given the significant impact of ET on water table dynamics, it is crucial to discuss the extent to which ET uncertainty may affect the water table estimation. I recommend including a sensitivity analysis or discussion on how variations in ET estimations could influence the water table simulations.

Response: Uncertainties in the input data can certainly propagate into our simulation, and unfortunately, these can be hard to quantify. As shown in Figure 10 of the original submission, groundwater storage volume closely follows P-ET, so it seems likely that uncertainties in these input data would be translated directly into our results. For the present-day simulation, I refer the reviewer to Figure 5 of Abatzoglou et al. (2018), which provides information on the spatial coverage of the data underlying the Terraclimate dataset. As seen on this figure, there is less coverage in the northern part of North America for both precipitation and vapour pressure data, which may translate into higher uncertainty in this region.

Although a sensitivity analysis of the WTM would be a fascinating exercise – and one we intend to pursue in the future – it is beyond the scope of this (already long) work. In the future, we intend to take several approaches to evaluating the relationships between P-ET inputs and the water table depth simulated by our model. This will include use of multiple ET produces to evaluate which provides the best match to water table data when used in our simulations. However, this is a model description paper, and such an exercise is beyond the current scope.

Comment: Effect of Water Depth on Heat Storage and Evaporation: The paper does not clarify whether the effect of water depth on heat storage within water bodies has been considered in the evaporation computation. Heat storage significantly impacts evaporation rates, especially across different seasons. Shallow water bodies, with less thermal inertia, experience more rapid temperature changes, while deeper bodies retain heat longer. These variations affect seasonal evaporation rates. The authors should address how water depth and thermal storage are factored into the model or discuss the potential integration of these elements to improve the model's accuracy.

Response: The effect of water depth on heat storage within water bodies has not been considered in this work. This is mostly to preserve the simplicity of our formulation, but we also think that the simplification is not unreasonable at our scales: in a study focussed on understanding the impacts of including the effects of heat storage on evaporation, Gan and Liu (2020) conclude that although the heat storage effects are not negligible, lake evaporation is mainly governed by the available energy and latent heat flux, which are covered by our formulation of the Penman equation.

Currently, open-water evaporation is given as an input dataset to the model. Although we did not consider water depth and heat storage in our formulation of the open-water evaporation layers, a model user would certainly be free to do so, or it could be considered as an add-on module to the WTM in future work.

Comment: Inclusion of human water usage: The current WTM does not account for human water usage, including agricultural irrigation, industrial demand, and urban development, all of which heavily rely on both surface water and groundwater resources. This omission could significantly limit the model's applicability in real-world scenarios where human activities play a critical role in water table dynamics. It is recommended that the authors consider integrating a component to represent human water usage or at least provide a discussion on how such factors could be incorporated in future model developments.

Response: The current capability of the model focuses on climate- and topography-induced changes in the water table. It is often useful to evaluate only certain aspects of a system and here we choose to focus on these. Note that the model is intended to work on continental and long-term scales. During many of the times that can be simulated with this model – such as the LGM simulation included herein – human water usage is a non-issue, while changing climate is. With that said, while the model does not currently include human water usage, its modular nature means that modules representing various aspects of human water usage could certainly be constructed in future work. The code is open-source, so other researchers would also be able to make their own changes to the model, if desired. We have plans to incorporate some results from the WaterGAP (Alcamo et al., 2003) Global Water Use model as input data into a human impact module of WTM in future work, but that is beyond the scope of the current work.

Comment: Initialization of the water table: The manuscript lacks clarity on how the initial water table is set in the WTM simulations. It would be beneficial to discuss whether surface water body products from satellite remote sensing (e.g., JRC-GSW, https://global-surface-water.appspot.com) could be utilized for initialization purposes. Additionally, for simulations extending over thousands of years, it is essential to clarify the approach used for initializing the water table and how this might impact long-term model outcomes.

Response: The water table may be initialised by the model user in cases where the initial water table depth is known (note that this is a requirement for transient simulations). For equilibrium simulations, when the initial water table is not known, WTM will initialise it at 0, i.e., the land surface. Users may still opt to initialise the water table at some other value if it is known. In the case of these equilibrium runs, a simulation that lasts long enough should ultimately converge on the same result; because hydraulic conductivity decreases with depth, we recommend against initialising the water table at great depths unless data supports this approach, since the slow movement of groundwater at depth could result in long runtimes. The initialisation of the water table is discussed in Appendix A, along with the other data requirements.

In the specific examples shown in section 6 of the original submission, both equilibrium simulations (present-day and LGM) were automatically initialised with water table depth equal to 0. The result from the LGM simulation was then used to initialise the transient simulation through 16 ka. Naturally, this means that the results of the transient simulation are all predicated on the assumption that the water table was at equilibrium at the LGM. We choose to make this assumption based on the observation that climate had been relatively steady in the preceding several thousand years. Note that, on the basis of other comments, we no longer include the transient simulation in this submission; we hope to give this work enough space to be thoroughly explored in another submission.

Comment: Simulation of Flooding and Inundation Processes: The manuscript currently does not provide sufficient detail on whether the WTM is capable of simulating flooding and inundation processes under transient state conditions. The ability to model these processes would significantly enhance the utility and applicability of the model, particularly given the increasing frequency and severity of flood events due to climate change and anthropogenic activities. This represents a highly promising area of application, especially for regions prone to extreme hydrological events. A detailed discussion in this regard could greatly expand the potential impact and use cases of the model.

Response: Currently, the WTM would not be an appropriate choice for simulation of flooding or inundation during short-term events if this flooding does not occur in depressions (for example, if it is a result of unusually high river discharge). This is because its handling of surface water assumes that surface water moves into depressions or off the landscape during each time step. The reason for this is that it allows us to deal with the surface water very quickly, therefore making simulations over thousands of years possible. Again, the modular nature of the WTM means that a representation of surface water that simulates flooding could certainly be added in future work, and used in the place of FSM for shorter-term studies where surface water that is not in a depression is more important.

Temporary flooding of depressions can certainly be simulated using the WTM or, depending on the parameters of the study, FSM (our dynamic lake component) alone. For example, Preisser et al. (2022) use FSM to estimate flooding in Austin, Texas, using FSM.

Comment: Parameterization and uncertainty analysis: The WTM involves substantial parameterization and a large number of model parameters. It is critical to assess how parameter uncertainty might influence the model results. A more detailed uncertainty analysis or discussion is recommended to understand the robustness of the model outcomes and provide guidance on parameter calibration and sensitivity.

Response: It is unclear whether the reviewer refers to input data, or to parameters set in the configuration file. If the reviewer refers to configuration parameters, then most of these involve no uncertainty and are simply facts about the domain (such resolution) and choices about how to run the model (such as the number of time steps to complete). The only parameters here which include some uncertainty are those pertaining to the e-folding depth for decay of hydraulic conductivity; we based our choice on calibration done by Fan et al. (2013b). We choose to use the same values they proposed both because of the similarity of our approaches and because we want to make our simulation more comparable to the Fan et al. (2013a) work for benchmarking purposes. Fan et al. (2013b) note that these parameters are dependent on grid cell size; we use their recommendations for 30-arcsecond cells, since this is the spatial resolution of our simulations. If the reviewer refers to input data, then as discussed above, we agree that our results are highly dependent on the input data and that uncertainty may be propagated to our results. We have added some text to this effect.

References

Abatzoglou, J. T., Dobrowski, S. Z., Parks, S. A., and Hegewisch, K. C.: TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958-2015, Scientific Data, 5, 1–12, https://doi.org/10.1038/sdata.2017.191, 2018.

Alcamo, J., Döll, P., Henrichs, T., Kaspar, F., Lehner, B., Rösch, T., and Siebert, S.: Development and testing of the WaterGAP 2 global model of water use and availability, Hydrological Sciences Journal, 48, 317–337, https://doi.org/10.1623/hysj.48.3.317.45290, 2003.

Deser, C., Phillips, A., Bourdette, V., and Teng, H.: Uncertainty in climate change projections: The role of internal variability, Climate Dynamics, 38, 527–546, https://doi.org/10.1007/s00382-010-0977-x, 2012.

Dyke, A. S., Andrews, J. T., Clark, P. U., England, J. H., Miller, G. H., Shaw, J., and Veillette, J. J.: The Laurentide and Innuitian ice sheets during the Last Glacial Maximum, Quaternary Science Reviews, 21, 9–31, https://doi.org/10.1016/S0277-3791(01)00095-6, 2002.

Fan, Y., Li, H., and Miguez-Macho, G.: Global patterns of groundwater table depth, Science, 339, 940–943, https://doi.org/10.1126/science.1229881, 2013a.

Fan, Y., Li, H., and Miguez-Macho, G.: Global patterns of groundwater table depth, Science, 339, 940–943, https://doi.org/10.1126/science.1229881, 2013b.

Freeze, R. A. and Cherry, J. A.: Groundwater, Prentice-Hall, Englewood Cliffs, New Jersey, 1979.

Gan, G. and Liu, Y.: Heat Storage Effect on Evaporation Estimates of China's Largest Freshwater Lake, Journal of Geophysical Research: Atmospheres, 125, https://doi.org/10.1029/2019JD032334, 2020.

Kay, J. E., Deser, C., Phillips, A., Mai, A., Hannay, C., Strand, G., Arblaster, J. M., Bates, S. C., Danabasoglu, G., Edwards, J., Holland, M., Kushner, P., Lamarque, J. F., Lawrence, D., Lindsay, K., Middleton, A., Munoz, E., Neale, R., Oleson, K., Polvani, L., and Vertenstein, M.: The community earth system model (CESM) large ensemble project : A community resource for studying climate change in the presence of internal climate variability, Bulletin of the American Meteorological Society, 96, 1333–1349, https://doi.org/10.1175/BAMS-D-13-00255.1, 2015.

Kourzeneva, E., Asensio, H., Martin, E., and Faroux, S.: Global gridded dataset of lake coverage and lake depth for use in numerical weather prediction and climate modelling, Tellus, Series A: Dynamic Meteorology and Oceanography, 64, https://doi.org/10.3402/tellusa.v64i0.15640, 2012.

Melini, D. and Spada, G.: Some remarks on glacial isostatic adjustment modelling uncertainties, Geophysical Journal International, 218, 401–413, https://doi.org/10.1093/gji/ggz158, 2019.

Preisser, M., Passalacqua, P., Bixler, R. P., and Hofmann, J.: Intersecting near-real time fluvial and pluvial inundation estimates with sociodemographic vulnerability to quantify a household flood impact index, Hydrology and Earth System Sciences, 26, 3941–3964, https://doi.org/10.5194/hess-26-3941-2022, 2022.

Simon, K. M. and Riva, R. E.: Uncertainty Estimation in Regional Models of Long-Term GIA Uplift and Sea Level Change: An Overview, Journal of Geophysical Research: Solid Earth, 125, https://doi.org/10.1029/2019JB018983, 2020.

Smith, L. and Wheatcraft, S.: Handbook of Hydrology, Chapter 6, McGraw-Hill, ISBN 0-07-0039732-5, 1993.

Zhang, X., Liu, L., Zhao, T., Chen, X., Lin, S., Wang, J., Mi, J., and Liu, W.: GWL-FCS30: a global 30m wetland map with a fine classification system using multi-sourced and time-series remote sensing imagery in 2020, Earth System Science Data, 15, 265–293, https://doi.org/10.5194/essd-15-265-2023, 2023.

Zhang, Y. and Li, Y.: Evolution of lake water volume in global closed basins since the Last Glacial Maximum and its implication for future projection, Progress in Physical Geography, 46, 613–629, https://doi.org/10.1177/03091333221079776, 2022.