We thank the referee for the comments and appreciate the effort spent on going so thoroughly through our manuscript. In our opinion, the suggestions from reviewer will improve readability of an improved version a lot.

The reviewer largely agrees with reviewer 1, hence we will not go into too much detail with our general answers, albeit sticking to our points that:

- to address the reviewer's comments would, in our opinion, lead to an an objective decision by the editor to see the paper accepted, with minor revisions.
- we think our manuscript could benefit from a stronger clarification that we are **not** trying to achieve the development of the best possible groundwater model. In the light of this special issue from the EartH2Observe project, this manuscript addresses the collaboration gap between global-scale modellers and catchment-scale modellers by explaining a simple method, originating from a global-scale method, that is computationally effective and results in (New Zealand's first) nationwide results.

We have addressed all comments below (*referees comments in Italic* and our reply in red):

The writing, and therewith the presentation and discussion of the research, should be significantly be improved. In addition to the points razed by R1, I suggest to rewrite the abstract and introduction and specifically focus on logic of the reasoning (meaning is a statement followed by the right argument and is the argument clear) and being as clear as possible. For example, abstract L2-3 reads: Large-scale models are simplified and not used at smaller-scales, because hydrology and water policy are constrained at the catchment scale. This does not make sense. What the author meant to say is that large scale models, are not useful for smaller scale groundwater assessments yet, because of the simplifications (and the coarse resolutions), therewith are not useful for e.g water policy. The next line reads: However, . . … . However, the statement in this line cannot be linked to the previous statement. Something like "for water policy smaller-scale models are more useful. However, ..." should be included. This are just two examples within the first three lines. Also, be careful using "this" "that" "it" without a summary word.

Overall from the abstract and introduction it was not clear for me what the main motivation and goal of this research were and how it will help us to improve current modelling efforts; to improve the EWT model but also be more useful for water managers? The lack of a logical structure and the bad writing are not beneficial for a clear understanding.

That is an excellent suggestion, also to clear up the potential confusion arising from the wording about scales. We will incorporate this into our improvements. Thanks.

- I found the manuscript very limited in discussion of previous work, methodologies, results, and relevance of the work done. For example, on discussion of previous work: P4 L4 "many studies . . .." And then only one reference is a bit limited, as it is not a review paper you refer to. P4 L7 "De Graaf apply a model. . .. Global-scale input data" This is too generalized, it should be a bit more specific what is meant with "a model" and "input data". Especially as you give some details for the Fan et al 2013 model. 1 to 2 Lines extra focusing on the differences between the two models referred to is needed. I know the models are quit difference. I little review here will also connect to the discussion, and will help you getting your point across why your model is better than the large-scale models available currently (see also my points later on)

Thanks for this suggestion. We will incorporate this into our improved version.

P4 L17: How do you know groundwater models are less reliable in data-sparse regions as there is no data to validate the results. In the case of a model calibration, like done in this study, you can say your model performs best for the regions where you do have data to calibrate on (the whole meaning of a calibration).

Instead of the original sentence "Groundwater models are less reliable in data-sparse regions, ……", we will improve this sentence by something like "Because groundwater models cannot be calibrated in data-sparse regions they leads to less reliable results, …..".

Methodology and results: In section 2 it is not explained what happens when water tables hit the surface, nor is it explained that this is not simulated as a head dependent flux and river infiltration (water entering your aquifer) is not included. How realistic is this in the real world? (this should come back in conclusion/discussion as well) Also, your model result look very biased toward shallow water tables, (however not discussed in the manuscript). I think this positive bias can be explained by the way drainage is estimated (see also comment R1). Another aspect I do not understand is the storage and the convergence criterium that is left out. I agree with R1 that 'steady-state' in combination with a timestep is a bit confusing. How I understand it, is that you run the model over 100 years forced with the same climate data until an equilibrium is reached (i.e. a steady-state). I think for this kind of procedures the term 'dynamic steady state' is used often. (I certainly would not call it transient). What I do not understand, for such a dynamic steady state you still need a storage coefficient, so how does that work? Also, it is not yet clear what you used as a criterion to stop your run. It is written that the convergence criterium is not used, as running the model beyond 100 years did not improve model performance. But how did you decide that 100years were enough; did you check your model outcomes, estimate R for water tables and when that looked good you stopped it. Or was it wallclock time driven, or CPU time driven? I think whatever criterium you used is fine, but now it raises questions.

I fully agree with R1 on point C and more extensive sensitivity analysis should be done. From the results it cannot be concluded which model change has the largest impact on the results.

Since the EWT method has already been explained in great detail in other publications, we feel like it is not our task to explain it again. However, we see the point in having to explain some detail, since we change the method to increase speed. We'll also make sure to include the issue of speed better throughout the manusrcipt: how long would it take to model the whole country with an advanced approach; how fast are we doing it, and what price do we pay in terms of uncertainty. We suggest to rely on an improved uncertainty/sensitivity to explain most of our changes.

In my opinion, a relevant aspect of the discussion that is not/not enough elaborated is where we stand now and how it will help is further. How useful is your model in reality, as it is a steady state model approach, not simulating groundwater gradients, calibrated for New-Zealand, under natural conditions only, only unconfined aquifer systems? Are there now model that can do this maybe better, and under real world circumstances (i.e. current climate conditions and human impacts). In other words, if you need to advice the New-Zealand water managers, how should they use the model and what do they need to know about the model structure and uncertainties to interpret the results correctly and use the model to its full potential? It for which purposes can the model not be used, and what should be improved to make the model useful for the more real world simulation (varying climate and human interactions).

We appreciate this comment and will elaborate more on the potential applications where the EWT could be used to solve issues relating to: data-sparsity; national guidelines that cut across regions.

Reading the authors comments on R1 point C I think the authors should be careful in saying that regions where not modelled before (is New Zealand not included in the large-scale models, I think so); stressing the computational efficiency (how efficient is the model, and how does this compare to other large-scale model efforts?).
Good point. We will rephrase this.

Minor comments: In the introduction a bit more details on the modelling should be given: 1 to 2 lines saying it is a flux-based approach, simulating steady-state water table heads, using averaged climate conditions, run for 100 years etc.

OK. It might even be that we then throw even more of the reimaging theory in the Appendix, depending on how clear our message should be, that we want to use an existing method, with improved data, and see how useful that is to solve water management issues that cut across regions or are in data-sparse areas.

P7-L16: "drained by humans"; artificial drainage? P8-L6-7 "who . . .. ." Leave this out, it is not relevant as you do not use the parameters of Gleeson. P6 L5 "the improved NWT"; is this the same at L4 "the NWT" or is there also an improved version (leave out improved). F8: it would be more logic to switch those scatters, so that wte, discussed first, becomes (a) and wtd (b) (same for the other scatters).

Thanks, we appreciate the time taken to even suggest these minor details. We'll improve them.