We thank the referee for the comments and appreciate the effort spent on going so thoroughly through our manuscript.

The reviewer suggests rejecting the manuscript.

We disagree with this recommendation because, in our assessment, minor modifications to the paper will address referee's comments. In our opinion, an objective decision by the editor would see the paper accepted, with minor revisions.

We will stress, in our review of the paper, that the results of our study can be used for national overviews; or to look at inter-catchment (i.e., trans-boundary, across regions) water issues; or as initial estimates for more advanced models. In the light of the special issue from the EartH2Observe project (https://www.hydrol-earth-syst-sci.net/special_issue935.html), this manuscript addresses the collaboration gap between global-scale modellers and catchment-scale modellers by describing 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):

A The presentation of the research is done poorly.

All of the reviewer's comments are addressed in the following. We largely agree with the comments and will revise the paper accordingly.

A1) There are linguistic weakness

(e.g. in the abstract: inconsistent terms and meanings with respect to smaller scale/

small-scale/local models and larger-scale/global models, l9: "because the quality

of their, coarse and global-scale, input data is large", l1: "larger, i.e. global", while larger

here should also refer to national).

We agree with the reviewer's comments and will revise the Abstract so that it provides and easier read.

Specifically:

L9: Instead of 'quality', we should have said 'uncertainty', which solves this sentence.

L1: this sentence refers to global-scale models and is therefore correct.

A2) The scientific terminology is not always used correctly, and some statements seem to be wrong (e.g. p6, l4 "ground-based, satellite-observed and modelled parameters"; p8 l16: g is not the gravitational constant but gravitational acceleration, and it is not 9.90 m2/s but 9.81 m/s2. Why "rainfall recharge" instead of "diffuse groundwater recharge"?

In regards of P6, l4: "ground-based, satellite-observed and modelled parameters". We can see how this sentence is a little confusing to reader. Minor corrections to grammar will make this sentence easier to understand.

p8 l16: “gravitational acceleration”. Yes, that is definitely a typo. Our mistake and we’ll correct it. We’re also not sure how the 9.9 came in. The value in NZ is approximately 9.80, but we’ll change it to the standardised value of 9.81.

A3) The reason for setting parameter values are not always clearly given, e.g. p8 l27: “As cell resolution of the NWT model is 200 m, the values of a, b and fmin of Eq. A2 were changed accordingly, to 75, 150 and 4”. To what extent does the cell size leads to is e.g. setting the value to 75 instead of 120 in the global-scale model?

We see how p8, l27 – l29 could confuse the reader. Minor corrections to grammar will make these sentences easier to understand.

In regards of: *‘To what extent does the cell size leads to is e.g. setting the value to 75 instead of 120 in the global-scale model?’* We assume that the referee requires explanation as to why these values were used, which is given in the next sentence, that says: “These values were also used by a 200 m resolution EWT model in the Amazon basin (Fan and Miguez-Macho, 2010a) .”.

A4) The manuscript does not explain, except in the Appendix, one major simplification of both the global (EWT) and the New Zealand (NWT) model: There is no hydraulic gradient dependent interaction between groundwater and rivers; where the water table reaches the land surface, the groundwater is assumed to flow out. This, however, may be the main reason for the dominant overestimation of groundwater table elevation as compared to observations as river levels may be below the land surface elevation. Also, losing rivers cannot be simulated.

Indeed, there is no hydraulic gradient dependent interaction between depending on river water level. More advanced local models, for example modflow (when using one of the integrated sw processes in modflow) is implicitly coupled at the time step level. That is, modflow will continue iteratively solving the coupled gw and sw flow equations until the exchange budget error is less than the users specified tolerance. But the rigor of incorporating a surface water flow model comes at a high cost of model run time and processing cost. We suggest to adjust some texts (minor revision) so that it clarifies better that the novelty of this paper does not lie in looking for the best model. We use an existing global-scale model (in our case EWT) and, knowing that this model is simplified, we are looking at the interplay between model purpose and the simplifications. As mentioned, when looking at simplified, global-scale, and local, advanced, models, we are looking for a model that ‘meets in the middle’.

We will add a comment mentioned this limitation in the revised paper. To take away some possible confusion, we could only refer to earlier EWT research references without any further explanation. However, we think that highlighting some model descriptions and simplifications, with a general description in the main text, and slightly more explanation in the Appendix, facilitates an easier read.

This paragraph is only intended for further discussion in case the referee would appreciate that. The underlying assumption in the EWT model is that rivers are resolved in the grid used for integration, which is why calculations have to be done at high resolution. Of course, this is a simplistic approximation, and more so in the original global 30” grid; however, the intent of the global model was to capture continental or global scale (first order) patterns in water table depth, and for this purpose the approach was adequate and results fared surprisingly well when comparing with observations (more than one million data points were used in the global result validation). In the present study, the scope is a much smaller area (New Zealand) and hence grid spacing is much reduced. We are fully aware that at about 200m resolution rivers are still not resolved. Notwithstanding, even when the water table depth might be underestimated in the model simulated river cells, results in general would not be substantially changed, unless the actual river in the cell runs in a deep and very narrow canyon. The model result in this latter case would represent conditions at the bottom of the canyon, where the river is, which might not be representative of water table depths in the rest of the cell. At 200m resolution, this case would be rather infrequent.

This paragraph is also only intended for further discussion in case the referee would appreciate that. In regard of: '*Also, losing rivers cannot be simulated':* Losing rivers are only simulated by the model assuming that this is recharge. Losing rivers due to river runoff (of water coming from upstream) is indeed not simulated, similar to many numerical groundwater models who also have trouble with this and to this day there is no comprehensive nationwide data of losing and gaining rivers in New Zealand. As mentioned, the EWT and NWT are simplified models, so we do not expect them to pick up model expertise that is not even contained in advanced models. Please note that if losing rivers would be incorporated in the NWT, then the groundwater level would be even higher

B Methodological weaknesses, combined with confusing presentation: In the manuscript, it is stated at various locations that the model was run in daily time steps for 100 years. However, a steady-state model cannot be run at daily time steps, because per definition there is no time variable in a steady state model, and the change in hydraulic head over time is zero. In addition, to do transient runs, one would need to set a storage coefficient, which is not mentioned in the manuscript.

The text, which we will correct, is confusing in regards of a steady-state that appears to be transient. In earlier versions of this manuscript we have discussed this with other experts in the field. From that perspective we have chosen to use their advice and call it "steady-state, or dynamic steady-state" since that is the common terminology for such models. So from a physics perspective it is definitely a steady state model. If needed, we can adjust that to 'dynamic steady state' if that satisfies the referee.

We were referring with this comment to the strategy used in achieving convergence in the model result, yielding the sought for equilibrium water table. In the original global calculations with the model, iterations where performed with annual recharge values (i.e. yearly time steps). These speeded up convergence in most parts, however causing fluctuations in high slope terrain with deep water table and substantial recharge, as was the case of the mountains of New Zealand. For this reason, we tried iterations representing smaller time-steps, aiming at limiting the aforementioned fluctuations, and run for a number of steps that we quantified in number of years. Perhaps mentioning years was not a very appropriate choice, since as the reviewer points out, it can cause confusion. We will better clarify in the revised manuscript that the model result is indeed steady-state and not transient, and refer to the number of iterations needed to achieve convergence with the raw number, instead of using a time-measurement equivalent, such as years.

C Lack of new information/innovation that is of general scientific interest The analysis is lacking components that would lead to improved scientific understanding. I suggest to analyze the specific reasons for the better fit of the NWT model results to observations. In a type of sensitivity analysis, variants of the NWT model could be run, in which only one "improved" data set is included while the other data sets remain those of the EWT model. Or alternatively where all but one data set is improved. For example, to understand the impact of the new hydraulic conductivity approach, do one NWT variant in which the approach used in the EWT model is applied. This would be a useful analysis to support the suggestion in the last paragraph of the discussion to use the hydraulic conductivity approach used for NWT (Gleeson data) also for the global-scale EWT, to improve it. But it was not shown in the study whether with the EWT hydraulic conductivity approach the fit to observations in New Zealand would have been better. Similarly, the groundwater recharge estimate used for EWT could be used as input of another NWT variant, and the resulting water table elevations could be *compared to the standard NWT results to understand the importance of improved/national groundwater recharge estimates. Then, the presumably large role of the DEM in improving results would be clearer, and your concluding statements would be more firmly based.*

Re: Lack of new information/innovation that is of general scientific interest.

We contend that the paper is of general scientific interest. The issue of the collaboration gap between global modellers and local, catchment-scale, modellers is a real problem to our science, which focusses at large scale. Interest has been shown in the paper we submitted to the Special Issue "Integration of Earth observations and models for global water resource assessment", forthcoming from the EartH2Observe programme. We also addressed the issue in the EGU 2017 EartH2Observe session called HS1.12 (full session description at http://meetingorganizer.copernicus.org/EGU2017/session/23938). This session was well-attended and is clearly an important future direction for international hydrology.

Re: The analysis is lacking components that would lead to improved scientific understanding.

Improved scientific understanding that comes from the paper, and related work, are:
- that there is a possible role for global-scale models for smaller-scale studies (e.g., national or catchment-scale), i.e., to cover data-sparse areas, to provide initial estimates, to be more computationally efficient.
- That this method covers areas of New Zealand that were never modelled before, and it can thus be used in data-sparse areas as initial estimates without extensive model run time and cost;
- That this model, despite its simplicity, correlates surprisingly well with known observations of hydraulic head. We suggest putting in some graphs on EWT correlation (already quite high) and NWT correlations (even higher correlation), if that satisfies the reviewer.
- The model was used in a way that results in a much more computationally efficient way than any other advanced model. We suggest that we define the model run times of the model for our case study areas and the nation.

To satisfy the reviewer's comments, possible suggestions for improvement to the paper are:

A) To analyse the specific reasons for the better fit of the NWT model results to observations
B) To do one NWT variant in which the approach used in the EWT model is applied.
C) show in the study whether with the EWT hydraulic conductivity approach the fit to observations in New Zealand is better.
D) the groundwater recharge estimate used for EWT could be used as input of another NWT variant, and the resulting water table elevations could be compared to the standard NWT results.

In our opinion, A) and B) would add merit to the paper, without adjusting our intended scope, i.e. point out. Suggestions C) and D) would dive into a deeper uncertainty analysis that we consider outside of the scope of this paper and would typically be undertaken with the incorporation of local advanced models, e.g., to see what improves from EWT to NWT to a local advanced calibrated model (where the 'worth' of a simplified model would be quantified in terms of model purpose, model cost and more). We would typically see this as a (very substantial) follow-up study.

I would also suggest adding to Fig. 8 the simulation results of EWT to directly visualize the improvement of NWT over EWT, and adding to Fig. 11 also the results of the standard NWT with a spatial resolution of 200 m, not only the LiDAR-based 100 m variant that is shown (but not indicated in caption).

Thank you. This is a good idea that we will consider to include in the paper, as an addition to the already described comparisons between EWT and NWT in the current manuscript. results (Section 4.2, Table 2, and Section 4.3).

However, we also want the opinion of the editor if this will not provide too much detail. For example, we could also reference to the PhD thesis in which all these additional figures are depicted to keep the manuscript concise [Westerhoff, R. S. (2017). Satellite remote sensing for improvement of groundwater characterisation (Thesis, Doctor of Philosophy (PhD)). University of Waikato, Hamilton, New Zealand. Retrieved from https://hdl.handle.net/10289/10922)].