

In their paper, "Application of global models and satellite data for smaller scale groundwater recharge studies," Westerhoff et al. present the development of a national recharge model for New Zealand. The modeling approach is based on the WaterGAP global model and the authors present a new application for New Zealand that runs at higher 1km spatial resolution. I think the authors have done a thorough job detailing the limitations of both model inputs and the conceptual approach. However, I am not sure what the scientific novelty or the scientific questions of this work are. In its current form, the manuscript covers model development, limitations and ideas for future work, but there are no scientific questions asked or answered. Furthermore, since the modeling approach is based on the WaterGAP model it's not clear that its presenting or validating a novel modelling approach. I also have significant concerns about the application of this global approach, which is based on a large number of empirical relationships and factors, to regional scale analysis. The questions asked by global models are very different from regional models and the authors do not make a strong enough case for the reasons to apply this approach at a smaller scale rather than using a more physically based tool. Finally, in my opinion, the writing needs improvement: there are many sentences that are difficult to follow or are overly vague. I've provided detailed comments below that I think can improve the clarity of the manuscript in its current form. However, even if these changes are addressed it's still unclear to me what the novel scientific contribution of this work would be beyond documenting a model and therefore I recommend rejection.

General Comments:

- Title: The current title is very general and implies that you are considering multiple combinations of global models and satellite data for different domains. A more appropriate title in my opinion would be "Development of a national recharge model for New Zealand".
- It's not clear what the intended uses of this model are and what spatial scale the authors intend for it to be used at. The term 'smaller-scale' is used in the title but from what I can tell from the paper the authors just mean smaller than global scale by this. The authors note the advantages of a higher resolution national model, but they also caution that the model will need to be improved before it can be used for regional analysis (Line 568-570). I thought the purpose of developing this national model (as opposed to the existing global models) would be to provide insights into behaviors within the country so this sentence seems to negate the entire exercise.
- Similarly, the current model is uncalibrated but the authors note that local applications 'might require the model to be calibrated' (Line 598). If the goal is very large scale estimates of recharge, then I'm not sure if a separate model aside from existing global tools is needed. If the goal is a tool that can be used for regional analysis, then the lack of calibration would seem to be a significant shortcoming of this analysis.
- Uncertainty in the goals of this work is reflected in the introduction which is lacking focus. Everything from differences between existing models (line 74) to transboundary disputes (line 55) is covered. I think the introduction should be refocused on the specific applications and motivations for the New Zealand model.

- Line 85: In my opinion it's a big oversimplification to say that the differences between model types are less important than the differences in input datasets. Also this seems to directly contradict the example on lines 74-77 that found significant differences between models. if you are considering only models that use a simple soil balance approach then I can see how differences between models would be small but what about other types of models that incorporate lateral groundwater flow for example?
- Line 110 says that one of the strengths of WaterGAP is that it includes a 'deeper 'geology' layer' but from the NGRM description it says it just has a 'single soil layer' so this doesn't seem like a strength that you can claim. Also from my understanding the 'deeper geology layer' in WaterGap is really a bucket not a physical system so please explain what strength this is adding?
- Line 113: Given the number of empirical correction factors in this approach it would seem necessary to calibrate this model for any regional analysis. It's not clear to me how the authors can expect to get reasonable results without calibrating this tool? This description needs to be expanded significantly.
- Line 169: How did you decide on the '2 times weighting'? Is there any sensitivity analysis to back this choice up or is there someone you can cite?
- I think it would be helpful to expand section 2.1 to include all of the model equations and required inputs before all of the input datasets so it's clear how they are fitting together and what parameters you need to estimate before you cover the datasets.
- Line 252-256: Uncertainty in initial conditions can be significant and I don't think it is sufficient to assume that this 'unknown error is resolved after six months when most soils are at their wettest'. Some verification and sensitivity analysis of this assumption is needed.
- Additional details are needed to support many of the model assumptions. For example:
 - Line 259: Where does the 75% assumption come from? Also I'm confused what you mean when you say "all other soils accept all recharge". Does this mean there is no runoff?
 - Line 261: from my understanding of your approach there is no lateral flow in the subsurface only vertical. Please clarify.
 - Line 298: Why is it assumed that uncertainty in AET decreases with AET/PET ratio?
 - Line 300: Why is uncertainty in storage assumed to be a function of PAW?
 - Line 303: The 10% standard deviation for fsoil appears to be arbitrary. How was this determined?
- Line 461: You recommend coupling NGRM to a groundwater model but I don't understand what the advantages would be of coupling to NGRM as opposed to developing a separate groundwater model.
- Along these lines it's also not clear to me why the authors have chosen to start from the WaterGAP approach as opposed to the existing physically based tools that could be applied at the 1km resolution. WaterGAP was designed for global analysis and as such has to deal with all of the associated data limitations at this scale. For regional and national models where data is available it would seem that this data could inform a

more sophisticated modelling approach. I think the authors need to make a much stronger case for the feasibility of this approach at higher resolution when there is no calibration occurring.

Specific Comments:

- Line 7: Varying water ‘policies’ not ‘policy’
- Line 22: Rework ‘is therefore assumed to be capable to’
- Line 41: Smaller scale what? This sentence seems incomplete.
- Line 46: This sentence is awkward. The phrase ‘diversity of some climates’ should be revised. I think what you are referring to here is locations where sub-grid heterogeneity is an important factor?
- Line 54-55: Not sure what you mean here. It seems that you are confusing model extent with resolution. This sentence seems to imply that for large scale models ‘entire nations and continents’ are lumped together and I don’t think that’s what you mean to say.
- Line 83: ‘showed that the differences between’ not ‘showed that the difference in’
- Figure 1: Clarity of this figure would be improved if you change your scale for elevation to use more of the color range in the plot.
- Precipitation is only at 5km resolution but MOD16 is at 1km
- Line 141: Here its noted that the MOD16 AET: ‘could be used in New Zealand studies, since: they seem to fit expected values and pattern in large parts of New Zealand; and the data already take into account vegetation characteristics’. This sentence is overly vague and qualitative and does not show that these data are sufficiently accurate for the high resolution application proposed here. More detailed discussion and analysis is needed to support the use of this dataset.
- Line 161: ‘but with 3 additional’ not ‘by with 3 additional’
- Line 161: Please explain where these three classes came from.
- Line 163-167: This description is not clear. Please explain where these lookup values come from.
- Line 178: What ‘age’ are you referring to here? I’m not sure why you need age as a parameter at this point.
- Line 215: I would like to see more details about how representative this function is for the land cover types and climates of New Zealand.
- Line 244-248: The term ‘prefers’ is confusing her. Does this mean that both options are possible and it can swap between alternatives or is it just that the slope relation was chosen instead of soil type relationships?
- 323: This would be more compelling if you chose three ‘representative’ locations that you can use to illustrate different behaviors. rather than three ‘random’ locations.
- Line 330: “rainfall in the different analysis periods appears relatively similar” is too vague please describe the quantitative differences in the sentence.
- Line 341: What do you mean by “Mean annual NGRM rainfall recharge estimates for the same period are equal to rainfall recharge observations at the three lysimeter stations”?

Do you mean that they are reporting the same variable or that they are quantitatively equal?

- Line 380: The sentence starting, “If Rushton and NGRM” is confusing. Please rephrase.
- Line 408: Not sure what is meant by ‘relating differences in existing local models’ please clarify.
- Lines 413-426: I think some discussion of the feasibility of less simplified approaches is warranted here. For example, groundwater models that simulate groundwater surface water interactions using lateral groundwater flow.
- Lines 440-441: While I understand the computational costs associated with higher resolution it seems like some additional tests could easily be conducted to determine sensitivity to resolution.
- Line 450: “was therefore in the model equation already clipped to the actual value of K”, this is confusing please reword. Also, what is the ‘actual value of K’ you are referring to here?
- Line 459: what do you mean by “if the groundwater table is deep enough”?
- Line 472: “much finer, and probably better” this language is in imprecise please reword.
- Line 477: I would think it is ‘bias in addition to uncertainty’ or ‘bias caused by uncertainty’ as opposed to bias ‘rather than uncertainty’
- Line 519: Also I would think that the 1km topography has the largest inaccuracies in the steep high elevation mountainous regions. Is this the case?
- Line 543: “Other parts of the model cannot always cope well with irrigation” I’m not sure exactly what this sentence means but it would imply some model inconsistencies that need to be addressed. The authors provide some examples but I think all of the inconsistencies need to be listed here.