Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2018-561-RC2, 2019


Hydrology and Earth System Sciences
Discussions

# *Interactive comment on* "Regionalization with Hierarchical Hydrologic Similarity and Ex-situ Data for the Estimation of Mean Annual Groundwater Recharge at Ungauged Watersheds" *by* Ching-Fu Chang and Yoram Rubin

**Anonymous Referee #2**

Received and published: 9 February 2019

Summary "Regionalization with Hierarchical Hydrologic Similarity and Ex-situ Data for the Estimation of Mean Annual Groundwater Recharge at Ungauged Watersheds" by Chang and Rubin, outlines a new approach to predict groundwater recharge in ungauged watersheds and assess process controls. The approach used is based upon hierarchical hydrologic similarity and uses a nested tree framework. Bayesian Additive Regression Tree models were developed and evaluated. Following this, each of the BART model selection was assessed using Classification And Regression Trees. I believe this manuscript is within the scope of the journal.



I would accept this paper with major revisions to be made. Firstly, I think some time should be spent making this manuscript better organised and easier to follow, especially figures and abbreviations. The General Notes section below highlights a few improvements to make. Secondly, one of the goals of the study was to develop a method to predict recharge in ungauged watersheds. However, it was only tested in a data abundant scenario using modelled baseflow data and therefore may not be practical when faced with more limited observational data. Point 3 of the main suggestions discusses this further and point 4D suggests a potential way forward. However, as a method to help understand which controls for groundwater recharge are being reflected in the recharge estimation outputs, I think the approach you have used is very good. It helps the reader understand the interaction between different controls and could be used in other hydrological studies or research areas. If the authors cannot address the issue concerning whether the approach is applicable in predicting recharge in ungauged watersheds using solely observational data, perhaps focussing on process understanding could be a way forward.

General Notes

1. Organise objectives more clearly.

Currently I am finding objectives in different parts of the paper which sometimes sound only partially similar. Two are at the end of the introduction (where they should probably go) "Estimate mean annual groundwater recharge at ungauged catchments" and "Propose new approach to understand hydrologic similarity with regards to mean annual groundwater recharge". Others are found at Page 7, lines 12-14 ("To understand what behaviour we can expect in ungauged areas") and page 8 line 6 ("To reveal controlling factors of dynamic hydrologic similarity system for mean annual groundwater recharge"). Put all objectives at the end of the introduction section.

2. Organise sections/paragraphs more clearly.

Sections 3.4 to 3.5 are very methodological, describe the method used in testing un-

gauged areas and how to evaluate the models and should be moved to the method section. These sections don't seem to give further details about the case study and should be moved to the method section.

3. Organise sections/paragraphs more clearly.

Page 8, Lines 20-30 (Section 2.4). Explains the data partitioning, but data partitioning is then explained again in section 3.4. Integrate this explanation with section 3.4.

4. Organise sections/paragraphs more clearly.

Page 9 Line 5:" Section 2.3 is not demonstrated in this case study"; Remove section 2.3 as it is not used and therefore is not providing information which is useful for this study.

5. Organise sections/paragraphs more clearly.

Page 10, Lines 20-28: you start discussing transformations made to recharge/baseflow data in a section which is supposed to be about climate variables. Move this to section 3.1.

6. Method section has too many equations.

Instead of relying upon equations, explain your methodology in laymans terms (simply) and then reference the papers which give the more explicit details about the method. The existing text describing the methodology should also be simplified.

7. Describe CART in method section.

CART is used to evaluate the different BART models, but no explanation of CART is made in the method section and some readers may not be familiar with this method. Describing CART may also be a useful step in explaining BART which seems to be a more complex version of CART. A schematic of how the BART models fit within the CART framework should also be provided.



8. Explain the terms ex-situ and in-situ data.

These terms come up quite a lot and I believe describe how the data is partitioned for training and testing. However, the terms are used prior to the section which describes data partitioning. Please provide definitions for these terms in the abstract or introduction where they are first used. This is particularly important as readers could get confused with thinking in-situ means from in-situ observations such as water table fluctuations or tracers, which I don't think is what you mean by the term.

9. Provide maps comparing the benchmark model to the other BART model estimates. This would make for an interesting comparison, showing the spatial variation from each of the models, which the reader cannot learn form figures 4-6.

10. Improve the naming conventions for the predictors and the models.

Some of the predictor names are not intuitive at all i.e. NLCD01_41, therefore when it is shown in a figure, I can't learn anything from this. Additionally, instead of calling the BART models 1-6, call them aridity, P&PET, all climate, Soils, Geology, topo. Then I can understand what each model in the figure represents without having to flip to the tables at the back.

11. Legends explaining colour codes on figures are needed. Colours being used in figures 8 and 9 need explaining in a legend.

Main Suggestions

13. Are you assessing recharge ratio or annual recharge?

Despite often referring to annual recharge in the text as the signature you are evaluating, Page 10 line 22 states that annual recharge was normalized by precipitation. This therefore means you are analysing the recharge ratio, a different signature to annual recharge, which instead quantifies the proportion of precipitation being converted to recharge. I do think the recharge ratio is a good signature to use to evaluate recharge controls beyond precipitation. Therefore, I suggest changing the text to say

you are evaluating the recharge ratio to understand recharge processes and predict the recharge ratio in ungauged watersheds.

14. The method applied here to predict recharge in ungauged watersheds needs to be trialled in a less data abundant scenario, otherwise I don't think you can claim that this is a method which can predict recharge in ungauged watersheds.

You have trialled this method in a data abundant scenario using modelled data (gridded baseflow data) as your test case. Regression tree models typically perform better with increasing observations to learn from and therefore any adequate performance you achieved in this data rich scenario may not be achieved when applied to in-situ observations of recharge.

I suggest that instead of using all the estimations within the two basins for training and testing, sample only a fraction of them and then see how well the models perform. Idea - If you sample only data points (for training and testing) where the streamflow data underlying the baseflow data has been evaluated against observations, this would then give an indication of how well the model works in real-world scenarios. This resultant model could then be used for prediction across the entirety of the two basins. You could then perform a second test on the basin wide modelling results, comparing these to the original gridded baseflow dataset you used. From this you could then learn how well the method trains and tests in a more likely scenario of limited data, as well as see how well the model trained and tested on limited data performs in data abundant scenario.

This would then give a better indication of whether the method can be practically used for prediction in ungauged watersheds.

15. It would be useful to know what number of observations is approximately needed to adequately train and test the BART models for prediction purposes.

16. If you decide not to test your prediction method on a more realistic scenario, where data is not as abundant as modelled estimates, I would suggest focusing on the objec-



tive of learning about process controls at large scales.

17. I like the approach of having separate BART models for different types of environmental characteristics. This provides a very nice way of understanding what are the important controls for recharge in different areas.

18. Topography and Landcover characteristics should have their own models, similarly to geology and soils.

No explanation was given as to why you grouped these two domains together and not the others. I think it would be better to be consistent and have separate BART models for each. You could learn something extra from doing this.

19. Aggregate some of geological predictors into their higher-level lithological groupings.

The number of Geology predictors in comparison to those available for climate, soils etc is too high. This could likely give the Geology BART model an unfair advantage in comparison to the other models. It also means that any BART tree for Geology is unlikely to be very informative as it will be difficult to understand what all the sub lithological categories mean.

20. Don't partition data by which basin it is in.

In the discussion you highlight aridity index was a good model as the data from both basins overlapped with respect to aridity, something which wasn't the case for Mean annual precipitation and evapotranspiration. You argue that this by saying that information can come in ad-hoc and therefore your testing it in what could be a realistic situation. However, I would argue that the modeller always has the opportunity to decide how to partition their training and testing data. Therefore, in order to improve the robustness of the estimations, should be looking to make the distributions of each attribute in the training and testing samples, overlap.

I also think you should be sampling less data for training and testing purposes in line

with point 14.

21. Explain the benchmark a bit further so it is clear what you did.

I think the benchmark was a good way to see how the different models could improve recharge estimates and then learn something about process controls. But I think a bit more explanation of how you determined the benchmark is needed, especially as how we evaluate groundwater systems at large scales is an interesting topic. Would also be good to see how this benchmark then compares to the 6 BART models (not shown in Figure 4 and some maps would be good).

Minor Suggestions

22. Page 4, lines 10-13: do you need to say this?

23. Page 6, Figure 1: Would be good to have a schematic of how the BART models are used within the CART framework.

24. Page 6, Lines 16-17. Don't understand what this means.

25. Page 8, lines 22-23. Does each plausible predictor set i.e. k=1 have multiple BART models? Or does each predictor set have just 1 BART model?

26. Page 8, Lines 9-17: I found the explanation of the two levels of similarity confusing and think it would benefit from rephrasing.

27. page 9 Line 5:" Section 2.3 is not demonstrated in this case study"; Do you need Section 2.3?

28. Page 9, Figure 2: Why not show a map of the baseflow estimates in MRB1 and MRB2. This would show the reader where the two basins are as well as allow them to see the variability of baseflow in the region.

29. Page 10, lines 3-5: I think you may need a reference to show that baseflow analysis is suitable for annual recharge estimation in the eastern US to justify yourself. You



justify it in the discussion (section 5.2.1) but should do it earlier on.

30. Page 10, lines 15-19: If you are already looking at climate variables for the year 2002 which correspond to the baseflow data you are using, why are you also interested in the long term climate variables?

31. Page 10, lines 15-17: For the long-term climate variables it would be better to use data which is consistent with regards to temporal extent (i.e. 1970-1990 for both P and Ep).

32. Page 10, Lines 21-22: If normalizing recharge by precipitation, your analysis is then looking at the recharge ratio rather than annual recharge which is a different signature.

33. Page 10, lines 21-28: Why are you discussing transformations made to your recharge data in a section which is supposed to be about the climate variables you are using? This would be better in the recharge estimate section or as part of your methodology.

34. Page 11, line 3: Why have topography and landcover been aggregated together in one model when all other domains, soil, geology, climate have their own models?

35. Pages 12-14: Sections 3.4 and 3.5 seem to be part of your methodology rather than an explanation of your case study.

36. Page 12, Lines 28-30: Is there one BART tree for each predictor set? Or does each predictor set have multiple BART trees which it can learn from?

37. Page 12, Lines 29:30: I like this approach of having separate models for different types of environmental characteristics.

38. Page 13, Lines 5-9: This description of the benchmark doesn't seem to be about predictor partitioning or data partitioning at all. The theory behind the benchmark estimation should probably be in a method section. Could then show the benchmark model (as a map in the results section and compare it to the 6 BART models and the

original data).

39. Page 13, Lines 5-9: I think one line on how kernel density estimation works would be useful. 40. Page 14, Lines 1-2: sentence doesn't sound right linguistically "how certain a BART model can infer that relationship".

41. Page 14, Lines15-16: If this is the case, is it reasonable to use so many geology predictors. It could unfairly improve the performance of the geology BART model making geology appear to be more important than what it actually is.

43. Pages 14 & 15: Confused as to what the different variances are showing/what you mean.

44. Page 15, Figure 4: Instead of calling models by 1-6. Would be better to call them Aridity, P and Ep, All climate variables, topography, geology, soil. Then reader doesn't need to keep flipping to the appendix to find out what the different model numbers represent. This would be good convention for all other figures as well.

45. Page 15 Lines 4: "regardless of the predictor set, the total predictive variance is always lower than the variance of the benchmark model". Would be good to show the variance of the benchmark model in figure 4.

46. Page 15 Line 12: Why is it surprising that the aridity model improved RMSE the most. Climate is one of the dominant controls for groundwater recharge.

47. Page 16: Would be nice to see maps of RMSE reduction across each of the 6 models. Perhaps this would be needed in supplemental information. This would help understand where the different environmental characteristic types are more or less important.

48. Page 16, Lines 13-18 & Page 18, Figure 7: Figure 7 seems to be a possible example of different conditions. Would be good to have a second figure next to this one which showed what each of the 6 models looked like in comparison to this. Otherwise I'm struggling to understand what the message is.



49. Page 18: I don't think the section title "RMSE labels" is informative as to what this section is about. What does this section actually show?

50. Page 19: Figure 8: âĂć What do the different colours correspond to? âĂć Again, instead of k =1,2,3. . . maybe soil model, Geology model etc would be better. âĂć Code names are not always intuitive "NLCD01_41" "BGEOL_147"- what does that mean? âĂć Tree branch width may be more intuitive if its proportional to the number of watersheds going down each branch. End node impurity is already shown at the bottom of each node.

51. Page 20: I don't think the section title "LPD labels" is informative as to what this section is about. What does this section actually show?

52. Page 21: Figure 9: Similar comments to Figure 8.

53. Page 23 Lines 1-6: I think this suggests that your method of partitioning data, i.e. by their basin, may not have been the best option.

---

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-561, 2019.