Response to Anonymous Referee #1's comments on manuscript hess-2017-679 (Predicting groundwater recharge for varying landcover and climate conditions: – a global meta-study)

We sincerely thank Anonymous Referee #1 for his/her constructive comments which have helped to improve the article. We address each comment in turn below.

Comment: Line 78: ***The fact that the FAO estimates are limited/unreliable is mentioned twice in the paper. How so? It would useful to delve deeper into the limitations of the FAO methodology to help the readers.***

We have added the following to Lines 78-84 to clarify this.

FAO statistics were based on estimates compiled from national institutions. The data estimation and reporting capacities of national agencies vary significantly and raise concerns about the accuracy of the data (Kohli and Frenken, 2015). In addition, according to FAO AQUASTAT reports, most national institutions in developing countries prioritise subnational level statistics over national level statistics, and in most cases data is not available for all sub national entities. This decreases the accuracy of country wide averages and raises concerns about the reliability of using them as standard comparison measures.

Comment: Lines 123-130***: highlights the rationale for selecting the explanatory factors in this study. Were any relevant factors excluded due to data/other constraints?***

It is true that insufficient and poor quality data often limit studies such as ours, and we have amended the relevant paragraph to acknowledge this more clearly. The relevant section of the paragraph now reads:

Line 134 -142: The choice of predictors was made based on the availability of global gridded datasets and their relative importance in a physical sense, as informed by the literature. We employed 12 predictors comprising meteorological factors, soil/vadose zone factors, vegetation factors and topographic factors. However, other factors which could have a sizable influence on recharge were not included in this study because there was insufficient data. Given this, we did not consider the effects of irrigation on recharge, limiting the scope of the study to rainfall induced recharge. Subsurface lithology can be another major factor determining recharge. Once again, due to the lack of lithological and geological datasets at a larger scale, these factors were also eliminated from the study. Better quality information about various predictors would have been desirable to enhance the accuracy of prediction.

Comment: Line 341-343: ***What was the Vopt for the top 10 models? Are the predictors shown in Table 3 equivalent to Vopt? Vopt could also be labelled on Figure 5 to make it clear.***

We have made some changes in terminology improve the clarity of this aspect of the paper. Figure 5 is changed as a result and the discussion is modified as follows.

Line 352-359: The choice of better models was made by considering the PoE of individual predictors (refer section 3.2.1) and the number of predictors in the model (*V*). Figure 5 shows the performance criteria for the top three models for different *V* values. The model performance increased with *V* up to 4 to 6 depending on the different criteria. After that, AICc, CAIC, RMSE and R^2_{adj} values remained almost constant, indicating that further addition of predictors did not improve the model performance. In particular CAIC shows a clear minimum at V=4 and it penalises model complexity more rigorously. Table 3 illustrates the predictors in the top 10 models selected based on CAIC. Nine of the top 10 models had *V* =4, and the remaining model (3rd best) had *V* =3.

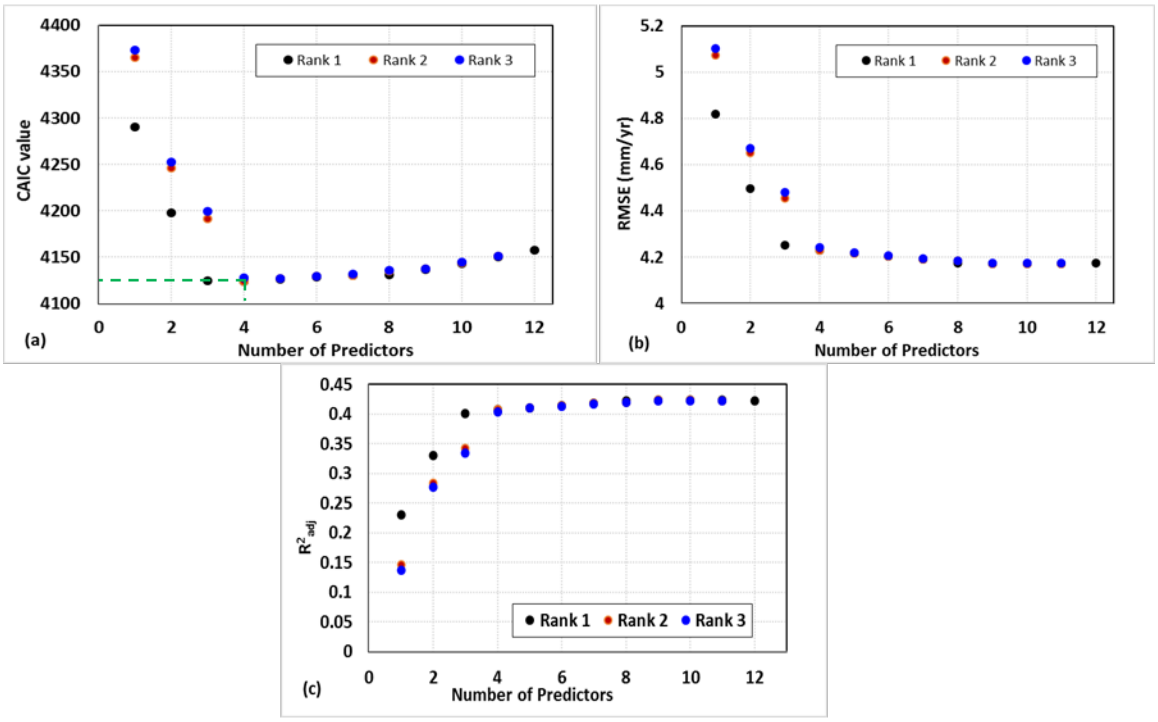


Figure 5. (a)CAIC (b) RMSE, and (c) R^2adj for the top 3 models with different number of predictors upto 12 and the green dotted lines representing the number of predictors for the best performance criteria value.

Comment: Line 366: ***How did the models with R^2 = 0.56 differ from the top 3 models shown in figure 5 which have a R^2 of ~0.42?***

The top three models shown in Figure 5 were built using the entire dataset, whereas the models discussed in line 366 (with R^2=0.56) were built as part of sub sample testing. In the sub sample testing, the entire dataset was randomly divided into 80% and 20% subsets and 80% of the data were used for building the model. The predictive performance developed model was tested against the omitted 20% of data (Reff line 233 -236). The statistics mentioned in Line 366 correspond to only 20% of the data at a time.

Comment: ***Figures 3, 6 and 7 are not very clear Increasing the size of axis text/legend would help. Figure 7 appears stretched.***

We have amended the figures as suggested, as shown below:

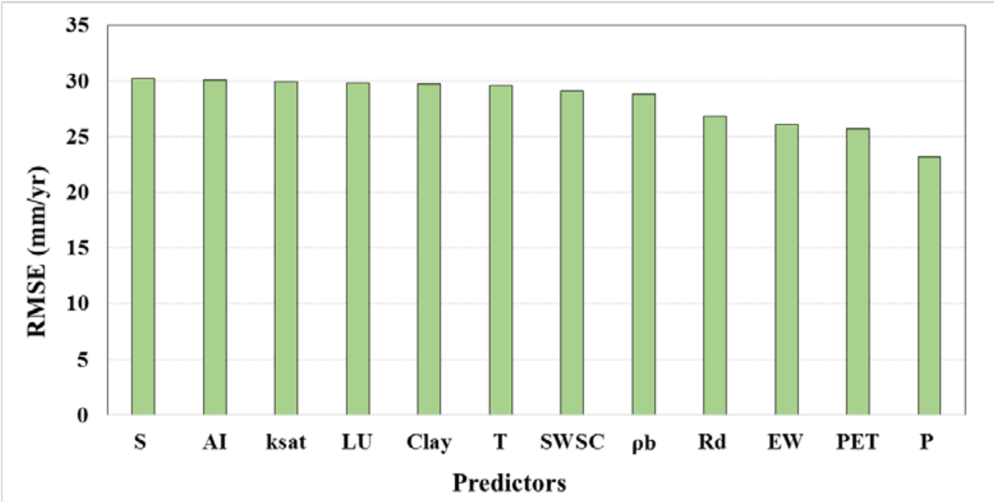


Figure 3. Model fit performance criteria for single predictor regressions.

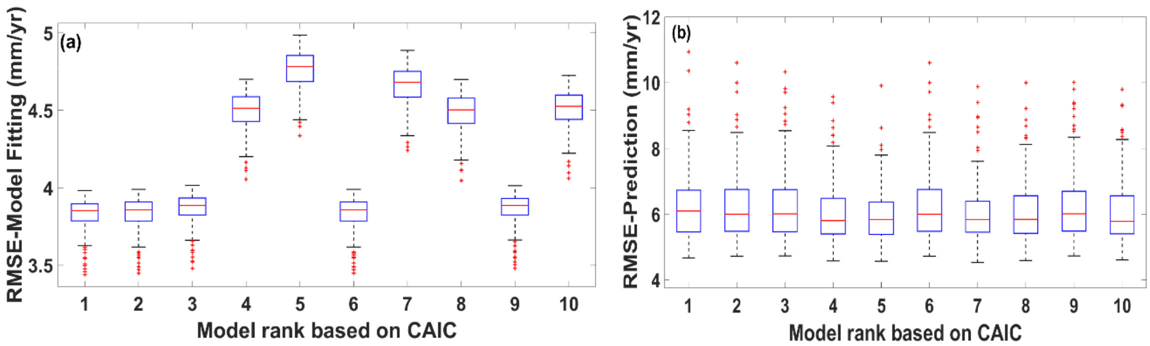
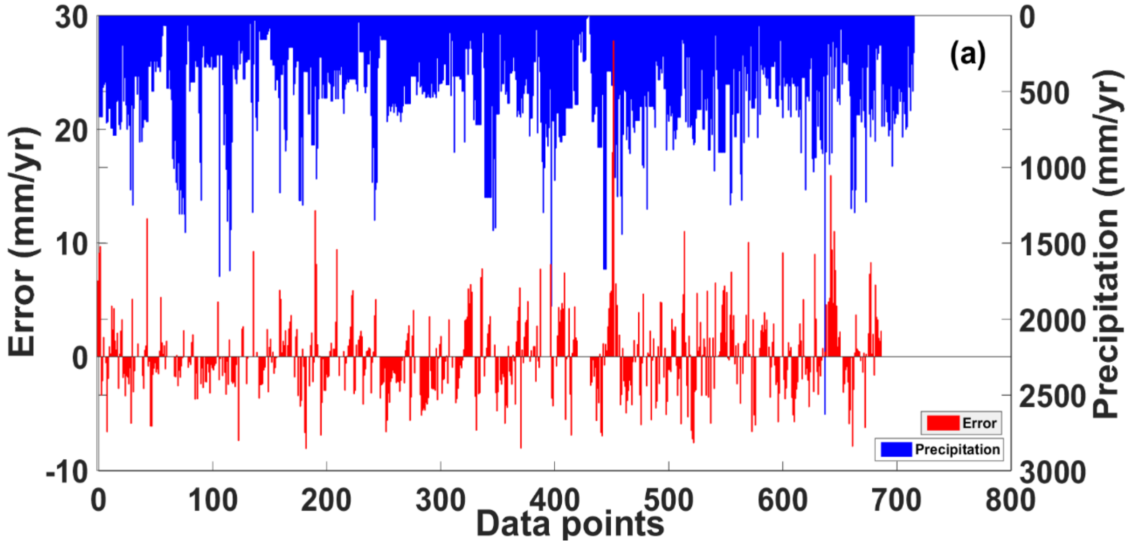


Figure 6. RMSE of sub-sample (a) model fitting and (b) model prediction of top 10 models according to CAIC.



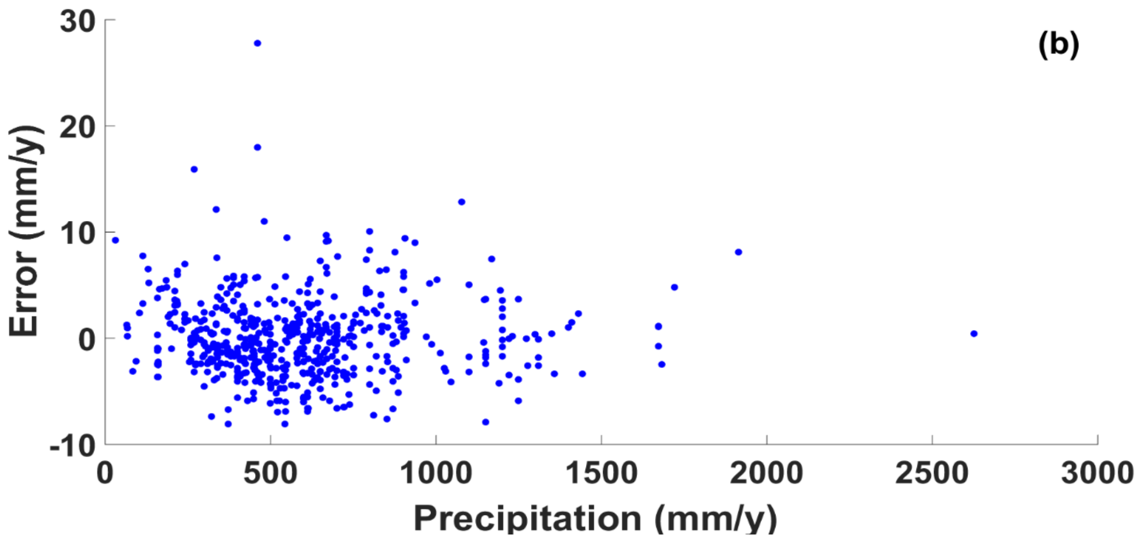


Figure 7 (a) Error at each data point along with the corresponding rainfall obtained using the leave-one-out model testing procedure and (b) Scatter plot between error at each data point and corresponding precipitation.

Comment: Line 385-413: ***The procedure to calculate the recharge values shown in Figures 8-11 is not very clear. Was one of the 'better' models used to calculate the map? Or, were all the 'better' models used and then averaged? Please clarify. It would also be useful to have a table that has the regression coefficients for selected models that includes the R2 values.***

We have tried to clarify the method and the relevant text (lines 398-401) now reads as follows:

In this study, the best model as defined by CAIC (model 1 in Table 3) was used to generate the recharge map. However, due to the similarity in structure of the top 10 models (Table 3), all models were equally good at predicting groundwater recharge and gave similar results (not shown).

We have revised Table 3 by adding model parameter coefficients and Adj R^2 values as shown below:

Table 3. Coefficient of predictors used in the top 10 models, ranked based on CAIC.

Model Rank	P	T	PET	Rd	S	k_{sat}	SWSC	AI	EW	ρ_b	Clay	LU	R^2_{adj}
1	0.008		-0.005			0.008						-0.915	0.407
2	0.009		-0.005								-0.062	-0.887	0.407
3	0.008		0.005									-0.862	0.401
4	0.007		-0.004					0.002				-0.050	0.253
5	0.008		0.020							0.250		-0.003	0.332
6	0.009		-0.005						-0.062			-0.887	0.410
7	0.009	0.003	-0.171									-2.258	0.407
8	0.007		-0.004				0.002					-2.014	0.346
9	0.008		-0.005	-0.032								-0.864	0.335
10	0.008		-0.005		-0.001							-0.052	0.404

Comment: ***Figure 11 compares the model estimated mean annual groundwater recharge for different countries with the FAO estimates. It would be pertinent to see if the countries that are most deviant from the 1:1 line are ones that didn't have study sites (out of the 715) used in the analysis.***

We have added a new figure (Figure 12) and expanded the discussion accordingly.

Figure 12 shows the country wide distribution of errors in model prediction in comparison with FAO statistics. Very high errors were found in countries with fewer model building data points. The model considerably overestimated recharge for Russia, Canada, Brazil, Indonesian Malaysia and Madagascar.

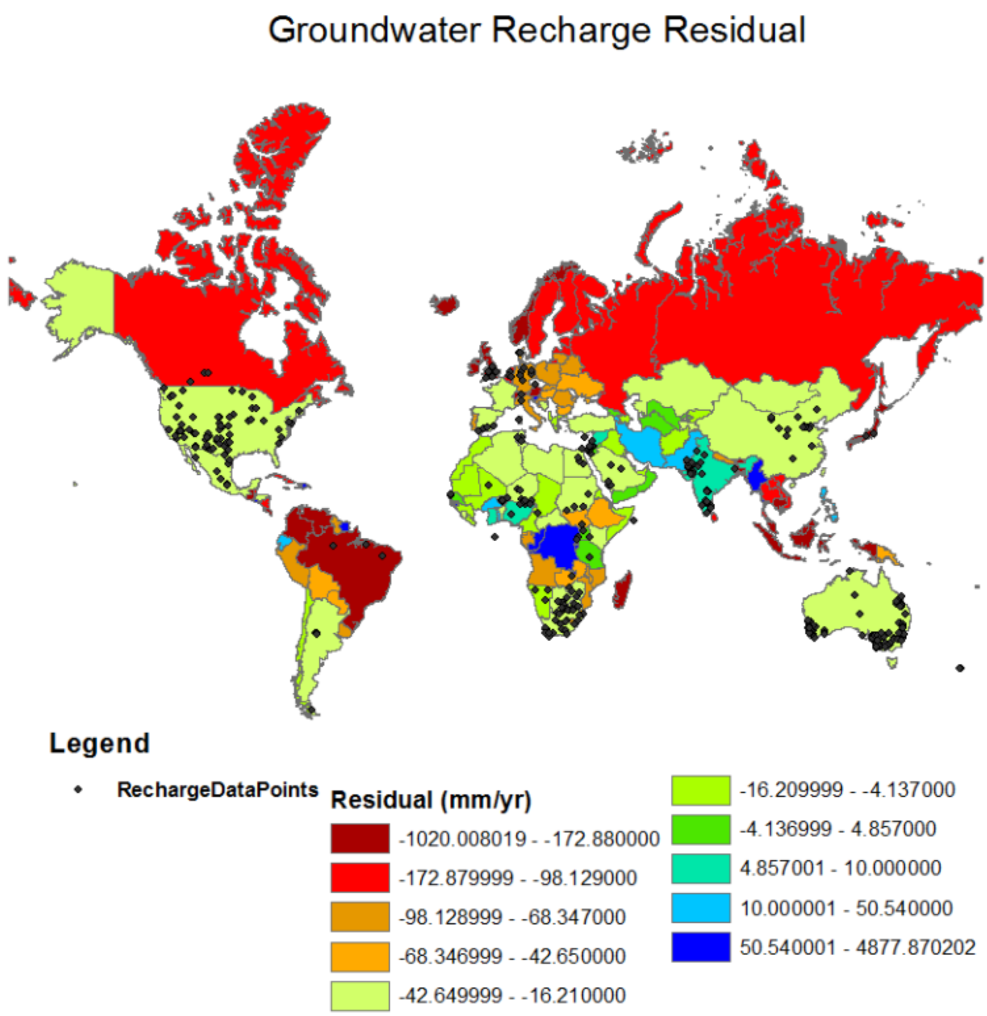


Figure 12. Spatial distribution of groundwater recharge residuals (FAO estimates – Model estimates) along with recharge sites selected for model building.

Comment: Line 412 and Line 480: ***Given that the FAO method is unreliable, how does the country-wide model results compare with estimates from complex hydrological models like PCR-GlobWB and WaterGAP? This is fairly important as it would help solidify the results obtained in the study***

We have added a new figure (Figure 11 (b)) comparing country level recharge estimates from the current model with WaterGAP, and revised the discussion accordingly. We were not able to compare our results with PCR-GlobWB, as its country-wide recharge results are not publically available.

We added the following to Line 428-431: Recharge estimates from the best models in the present study were compared to recharge estimates from the complex hydrological model (WaterGAP). Even though the model in this study overestimates recharge for countries with fewer data points, the scatter shows a smaller spread compared to the FAO estimates.

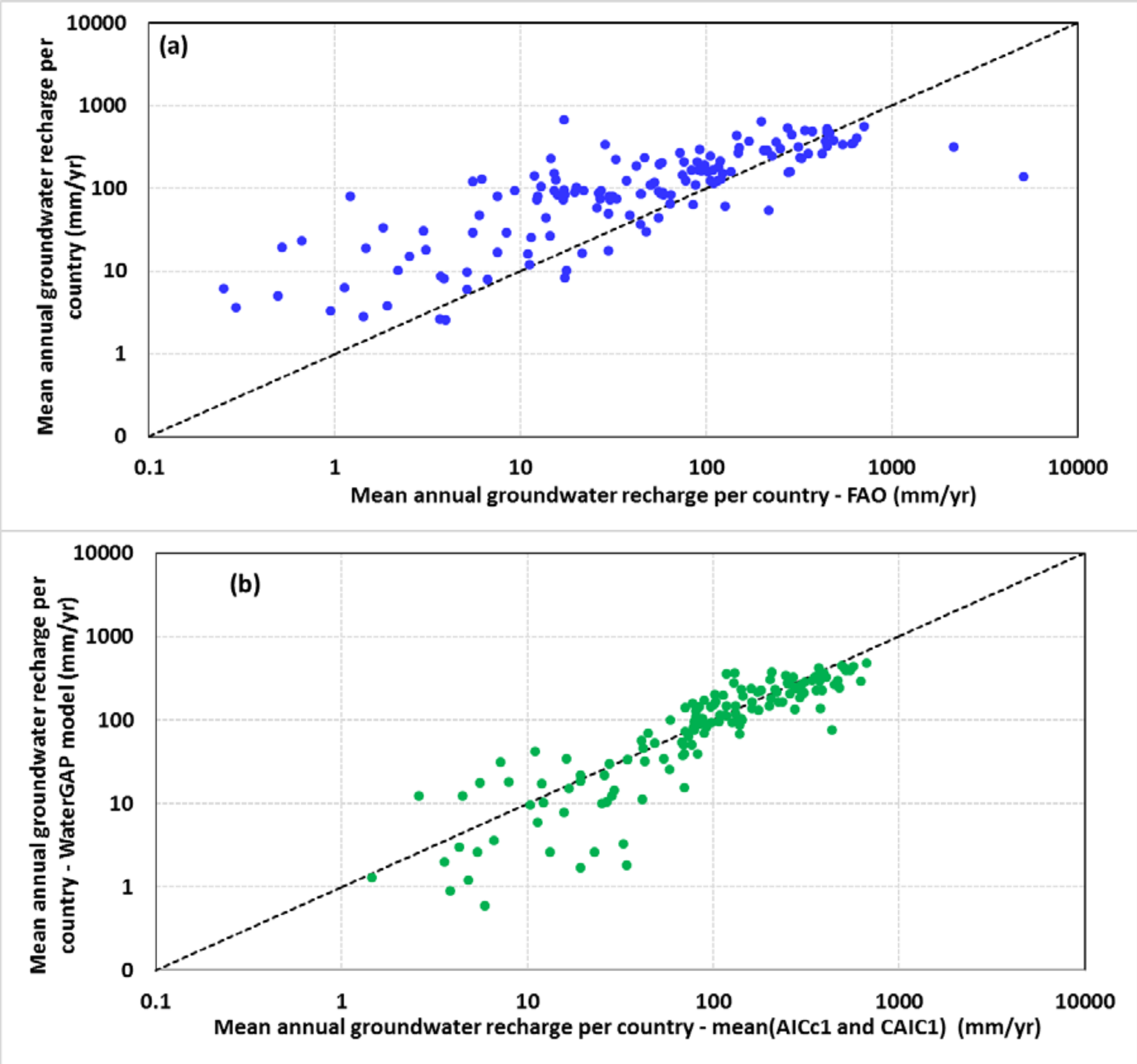


Figure 11. Comparison of predicted recharge against country level estimates from (a) FAO and (b) WaterGAP model

Comment: Line 455-467: ***While this paragraph discusses the influence of vegetation on recharge, the results fail to illustrate this influence. Please clarify how this influence was observed in the results.***

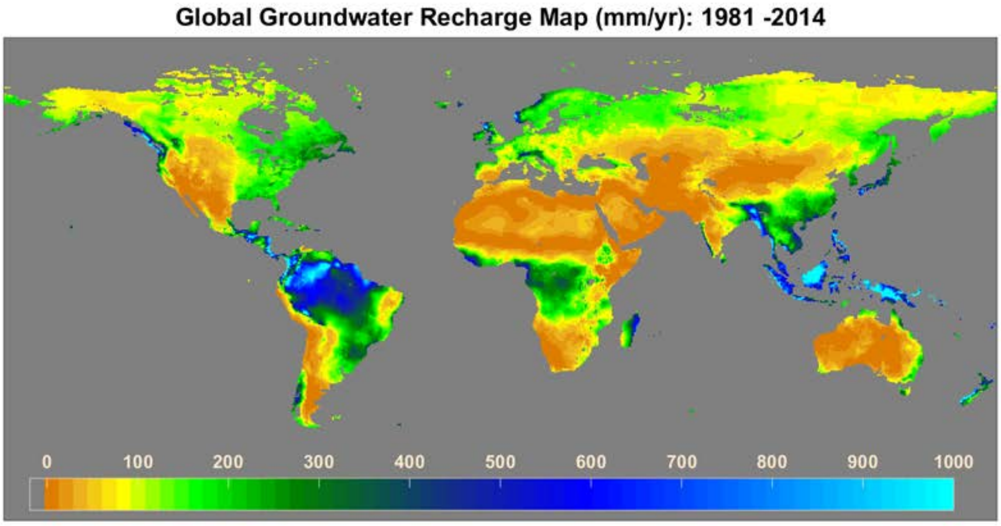
We have modified the discussion by highlighting the importance of vegetation as shown in our results.

Line 493 -497: In this study Land Use (LU) was used as a proxy for vegetation. According to the results, LU was found to be one of the predictors having the highest Proportion of Evidence (PoE) (Figure 4). In addition, all the better performing models included LU as one of the predictors which clearly indicates that vegetation is one of the most influential factors for groundwater recharge.

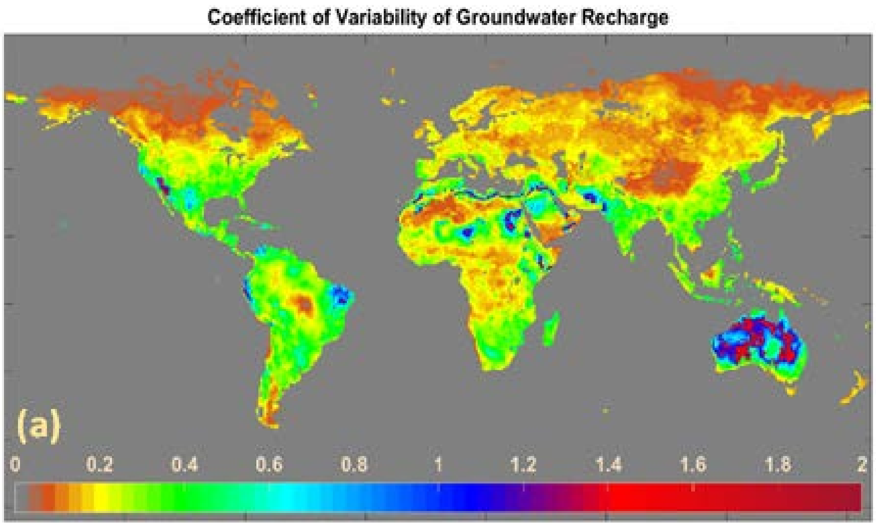
Technical Comments

Figure 5 is a multi-part figure and should be labelled a),b), c) The legends in Figures 8 and 10 are difficult to see Line 643-645: Citation format not consistent

We have improved Figures, 5, 8 and 10 as shown below:



Modified Figure 8. Long-term (1981 -2014) average annual groundwater recharge estimated using the developed model.



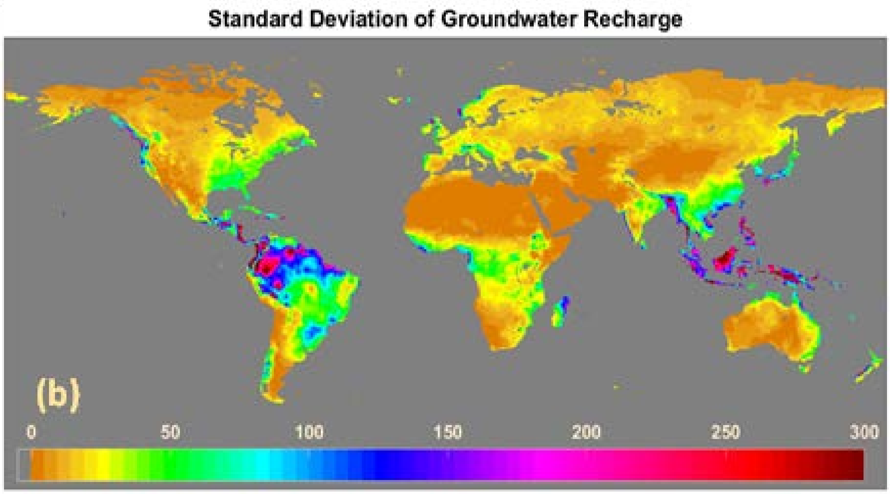


Figure 10. Map showing (a) coefficient of variability and (b) standard deviation of annual groundwater recharge from 1981 to 2014.

We have revised the citation format as indicated below:

Kohli, A. and Frenken, K.: Renewable Water Resources Assessment – 2015 AQUASTAT methodology review, Food and Agricultural Organisation of the United Nations, 1-6 pp., 2015.