We thank the reviewer for his comments and suggestions which will help to improve the manuscript. Below we address the concerns raised point by point in bold font.

Please note: To ensure an unbiased evaluation I have performed this review without reading the other reviewer comments on this manuscript.

Summary:

The manuscript entitled ”The impact of forest regeneration on streamflow in 12 meso-scale humid tropical catchments” summarizes a modeling study performed on a number of meso-scale catchments in Puerto Rico in which the change of the water balance (and the associated Q-measures) by forest regeneration resulting from farm land abandonment was investigated.

The most interesting aspect of the study is that it challenges a widely accepted research paradigm which states that increased forest cover will also increase the evapotranspiration within the water balance of a landscape (Bosch & Hewlett, 1982; Stednick, 1996; Andréassian, 2004; Brown et al., 2005). Following this paradigm, reforestation of larger land areas would lower stream runoff, especially under low flow conditions. However, the results of this study question this paradigm at least for meso-scale tropical catchments by providing results opposing the paradigm. These results are further supported by a detailed comparison with similar studies from other, meso-scale tropical catchments where similar results regarding a low impact of reforestation on the water balance have been stated. Therefore I consider the manuscript as a valuable and unique contribution to the literature of forest-hydrology and very suitable for publication in HESS.

The overall quality of the manuscript is high. The methods are well developed, fully suitable for the research question and well described. The use of the spatially lumped HBV-light model seems justified for the given model purpose. The research hypotheses are well developed from a broad body of literature. The results are described clearly and concise. The discussion brings up the critical parts of the manuscript and discusses them sufficiently. The conclusion summarizes nicely the overall results. Aspects that could potentially be improved concern primarily the model uncertainty, especially the discussion of the sensitivity of different model parameters and the uncertainty of input data. More specifically, the point that I would be the most cautious about it the ‘problem of closing the water balance’ for some of the catchments. It is obvious that the values used for the parameter PCORR are very variable for some of the catchments: F (1.215), G (0.712), I (1.415) and L (1.4). Presumably, this parameter will also have a very high sensitivity, because P dominates the water balance equation. To me this indicates that there must be additional mechanisms that cause these inconsistencies. Whereas some of these mechanisms are named in the results section (see P 3062, L2-4) they are only briefly discussed within the discussion section. I would suggest strengthening this part in the discussion. The points I would suggest for this are:

We thank the reviewer for his positive assessment of the manuscript. We agree that the adjusted PCORR values for catchments F, G, I, and L warrant some discussion and have therefore added the following sub-section to the Discussion:

"HBV-light model performance

The HBV-light performance was good for both the calibration and validation periods (Table 4), suggesting that the simulated Q for the catchments can be used with confidence. Note that strong land-cover effects would have deteriorated performance statistics for the validation period. Several catchments required optimization of the PCORR parameter (cf. Table 4), possibly due to biases in the PRISM P map, uncertainties in the Q and/or catchment boundary data, water extractions, inter-basin groundwater transfers (potentially exacerbated by karst), and/or water recycling (e.g., Ellison et al., 2012), which combined may cancel out or amplify one another. However, the influence of the P scaling on the results is probably limited because the simulated Q was only used to control for climate and storage carry-over effects."

a. Uncertainties in catchment areas (You give USGS uncertainty estimates of +/- 10%). However, these estimates appear very small to me given that at least some catchments are underlain by limestone aquifers with well-developed karst systems. I believe that this geological setting could very well account for the inconsistencies of the water balance.

This is a good point. However, karst is mainly found in the northwestern part of the island (Olcott, 1999) and thus only in catchments L, A, and D. So while the presence of karst may explain the adjusted PCORR value for catchment L, it is probably unable to do so for the other catchments with adjusted PCORR values (F, G, and I). Nevertheless, we now mention karst as a factor that could exacerbate inter-basin groundwater transfers (see above).

b. Another mechanism causing this could be 'water recycling' (e.g. that water that has been evaporated or transpired reoccurs in a catchment as precipitation). I am not sure if such a mechanism is discussed for the tropical regions of PR, but it could well account for some of the imbalances (especially considering that some of the catchments have corrections for P of -30% and others have +40%). An additional paper that may be interesting to add to the discussion regarding this aspect is given by Ellison et al. (2012).

We agree that water recycling may have contributed to the variable PCORR values, as water recycling can cause over-catch in *P* gauges when the gain from upwind forests exceeds the downwind loss or under-catch in *P* gauges when the downwind loss exceeds the gain from upwind forests (see Van Dijk et al., 2012, and references therein). The effect of water recycling is thus related to the broader issue of gauge placement and representativeness (see e.g., Briggs and Cogley, 1996). It is, however, difficult to quantify this phenomenon with certainty. Nevertheless, we have included water recycling as a potential reason for the adjusted PCORR values and made reference to Ellison et al. (2012) as per the reviewer's suggestion (see above).

c. Finally, the spatial uncertainty in P inputs. Even if the IDW regionalization may be powerful, it may simply miss some large convective events that have a small spatial extend, but high P intensities. Some type of uncertainty estimation on the regionalized P maps would also be great (maybe something like a 'leave out' approach for some stations). However, I do note that the manuscript is rather long already, so the authors should consider this as an optional thing.

We emphasize that the IDW-computed long-term mean *P* has been matched to PRISM elevation-corrected long-term mean *P* on a per-pixel basis. Therefore, large errors in the water balance are likely attributable to uncertainties in the PRISM *P* map due to the highly heterogeneous topography. Uncertainties in the PRISM *P* map is included as a potential reason for the variable PCORR values in the new sub-section that was added to the Discussion (see above). We agree with the reviewer that it is possible that some *P* events have been completely missed, and this is stated in the original submission at page 3056 lines 9 to 11. Unfortunately, space does not permit a leave-one-out approach to quantify the *P* uncertainty. We stress that the time series of each rain gauge have been visually checked for artifacts, spurious trends, and outliers.

References:

Andréassian, V. (2004). Waters and forests: from historical controversy to scientific debate. Journal of Hydrology 291(1-2), 1-27.

Bosch, J.M. & Hewlett, J.D. (1982). A review of catchment experiments to determine the effect of vegetation changes on water yield and evaporation. Journal of Hydrology 55(1-4), 3-23.

Brown, A.E., Zhang, L., McMahon, T.A., Western, A.W. & Vertessy, R.A. (2005). A review of paired catchment studies for determining changes in water yield resulting from alterations in vegetation. Journal of Hydrology 310(1–4), 28-61.

Ellison, D., N. Futter, M. & Bishop, K. (2012). On the forest cover–water yield debate: from demand- to supply-side thinking. Global Change Biology 18(3), 806-820.

Stednick, J.D. (1996). Monitoring the effects of timber harvest on annual water yield. Journal of Hydrology 176(1–4), 79-95.

Small in-text edits/suggestions (please note: I am not a native speaker):

Abstract: P 3046, L2: I would remove ‘comparatively’

Agreed, thank you.

Introduction:

P3048 L1-5: is rooting depth also a variable that should be mentioned here?

In this paragraph we explain that the net effect of forestation on *Q* depends on the quantitative trade-off between increases in water infiltration due to enhanced soil infiltration and decreases in flow due to enhanced water use. The former is usually more important than the latter, and hence forestation commonly reduces the water yield (Bosch & Hewlett, 1982). We feel that listing the specific factors contributing to the higher water use of forests (i.e., deeper root systems, and higher leaf area index, aerodynamic roughness, and interception) is unnecessary here.

P3048 L22: remove relatively (or state relative to what: : :)

We agree and have removed “relatively”.

P 3048 L29: maybe use ‘regions’ as a simpler term for physiographic units

We have replaced “physiographic units” by “regions”. Thank you for the suggestion.

Data and Methods:

P3051 L7: any chance to get to know the resolution of the photo interpreters? Appears like an open question for a reader: : :

The sentence at page 3051 line 7 of the original submission has been changed as follows: “Although the maps for 1951 and 1978 were rasterized at a resolution of ~30 and ~11 m, respectively, the actual mapping resolution used by the photo interpreters is estimated at ~300 and ~50 m, respectively.” Thank you for pointing this out.

P 3052 L21: ‘The’ should not be capitalized.

Corrected.

P 3054, L7: I would maybe be good to state why the authors developed this methods. Otherwise it leaves the reader a bit with a method were it is unclear why it should be used.

We agree and have added the following sentence to the start of the paragraph: "Having reliable catchment-mean time series of the climatic variables (*P*, T_{min}, and T_{max}) is important to prevent spurious trends in the simulated *Q* from influencing the results."

P 3055 L7-12: This could be shortened, if only IDW is used. Just state on which basis the authors chose their method.

We agree that listing other spatial-interpolation techniques and writing out the IDW equations (Eqs. 2 and 3 in the original submission) is unnecessary, as this information can easily be found in standard hydrological textbooks. We have therefore shortened the text and removed the IDW equations.

P 3055 L10: remove 'here'

Done.

P 3056 L30: It may be good to state that inter-basin GW tranfers are not considered in the model.

We have added the following sentence to the HBV-light model sub-section of the Data and methods: "Note that HBV-light does not consider groundwater flow within or between catchments."

P 3058 L1: the authors may need to explain what 3D Q is in a lumped model.

This is a typesetting error and should have read "3-day" instead of "3-D". Thank you for pointing this out.

P 3058 L10: how where they combined? Please explain a bit more.

Please refer to Booij and Krol (2010) for an explanation of how the single objective functions were combined to form a single aggregate score. We feel that the explanation is best left out, since the paper is already quite long. However, if the editor believes that an explanation should be added we are happy to do so.

P 3059 L13: the authors could consider removing this equation, given that it is just a linear trend that is assumed.

We agree and have removed the equation, thereby shortening the paper.

P 3060: Interesting: : : using the Jackknife approach.

Results:

P 3061 L 25: I am not sure if ‘degraded’ is the right term here. Maybe use ‘negatively affected’

We agree that “degraded” is not the correct term and have therefore replaced it by “deteriorated”.

Discussion:

P 3063 L20: ‘trends’, plural

P 3063 L21: remove ‘in turn’

P 3064 L1: Maybe replace ‘accept’ with ‘support’ – for me a hypothesis can only be supported: : :

All corrected; thank you for the suggestions.

P 3064 L11-16: this is great!

P 3068, L1: replace ‘outstanding’ with ‘challenging’

P 3069, L2 and L 8: replace ‘errors’ with ‘uncertainties’ P 3069, L15-16: replace ‘is about’ with ‘is estimated to be’

All corrected.

From L 25: great reading.

Conclusion: Good, nothing to add.

A final remark after reading the other reviewers comments: I do agree that the level of detail, especially in the methods section is at the edge of being too detailed. However, this may at least in part be personal preference and I, personally, would rather prefer a detailed and complete description of what was done, than having an ‘incomplete’ manuscript. So I guess it up to the authors and the editor to decide what level of detail is needed.

We also prefer a detailed and complete manuscript over an incomplete manuscript that leaves many questions unanswered. As stated in our response to Reviewer 3, we feel that it would be difficult to substantially shorten the manuscript as all the different elements of the analysis are necessary to arrive at robust conclusions. Nevertheless, the manuscript has been shortened somewhat,

since the explanation about the IDW technique (see above) and the abstract (see our response to Reviewer 3) have been shortened, and because the trend line equation has been removed (see above).

References

Booij, M. J. and Krol, M. S.: Balance between calibration objectives in a conceptual hydrological model, Hydrolog. Sci. J., 55, 1017–1032, 2010.

Briggs, P. R. and Cogley, J. G. Topographic bias in mesoscale precipitation networks. Journal of Climate, 9(1):205–218, 1996.

Olcott, P. G.: Puerto Rico and the U. S. Virgin Islands, in: Ground water atlas of the United States, chap. HA 730-N, http://pubs.usgs.gov/ha/ha730/ch n/index.html, US Geological Survey, Washington DC, 1999.

Van Dijk, A. I. J. M.; Pena-Arancibia, J. L., and Bruijnzeel, L. A. Top-down analysis of collated streamflow data from heterogeneous catchments leads to underestimation of land cover influence. Hydrology and Earth System Sciences, 16(9):3461–3473, 2012.