We thank this reviewer for her/his thorough review and appreciate the comments. Our detailed responses to the individual comments are provided in this bold font (below each comment).

I wish I could say I liked this paper – it oozes statistical methodology, there are some (many?) heroic aspects to the work, and must have taken a (very) long time to put together. I wonder how many million computer files were generated whilst the work was being done? At the same time, the infinite detail overwhelms the reader. I have had a few goes at going through it and always tend to lose my way to some extent in the seemingly infinite detail of what was done. I think the conclusion is valid and relevant (that there is no really clear detectable effect on streamflow) but I am not sure just where that leaves the reader.

We are pleased that the reviewer considers the paper's topic and conclusions to be relevant. We wish to emphasize that it was not our intention to come across as "heroic" or to impress anyone with the methodology, we merely attempt to examine, in the most robust way we can, whether changes in streamflow can be attributed to changes in forest cover. We argue that the number of (HBV-light) model runs used here (30,000) is not excessive. There are numerous studies that have used more model runs. For example, Pinol et al. (1997) used 55,000 model runs, whereas both Zhang et al. (2008) and Marshall et al. (2004) used 100,000 model runs. The reason for using such a large number of runs is to ensure that the multi-dimensional model parameter space is fully explored. We can see that some aspects of the analysis may come across as complicated, yet we believe that all the steps in the analysis are critical to reach the conclusions with sufficient confidence. For example, the spatio-temporal interpolation of the precipitation and potential evaporation time series (Section 3.5 of the original paper) was necessary to prevent spurious trends in these variables leading to erroneous outcomes, while the computation of standard errors (Section 3.7 of the original paper) was necessary to quantify the uncertainty of our estimates. These are both important aspects to comprehensively deal with, and we note both Reviewer 1 and Reviewer 2 had no issues with the comprehensive details (thus ensuring repeatability) of our analysis.

I think, basically, that the paper has a dual aim of impressing the reader and conveying the work done. I think that the former stuff could be substantially expunged. Some sort of diagram illustrating the methodology conceptually? Some sort of simplification?

Our aim was, and is, not to impress readers with statistical methodology; our sole aim is to scientifically support robust conclusions. To ensure that the methodology adopted in this study is more easily understood we added a flow diagram that outlines the specific steps performed to arrive at the scatterplots; thanks for the positive suggestion. See Figure 1 below.

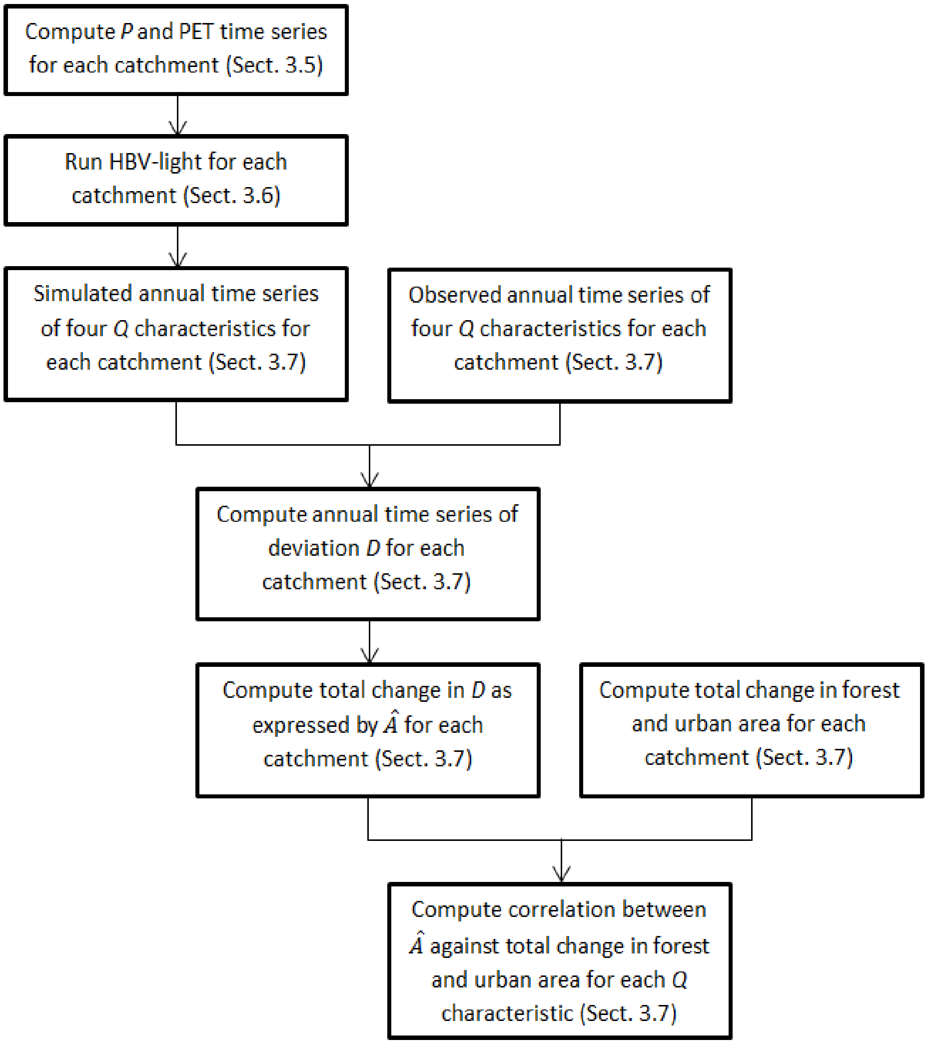

Figure 1: Flow diagram illustrating the steps performed to examine whether changes in forest and/or urban cover influenced the four observed *Q* characteristics.

I gather that streamflow records were used, some sort of "normalising procedure" to overcome problems with individual records, and the results correlated with land use change. The result was either no correlation or a weak negative correlation with the percentage of forest.

My recommendation to make the paper more accessible:
(a) Shorten and simplify the abstract, and avoid stuff like Q metrics.
(b) Simplified account of methodology with detailed work either obtainable on request or in an Appendix.
(c) Gross simplification of Tables.

We agree that the abstract is rather long and have therefore shortened it from 353 to 326 words. However, we do not agree that the methodology can be shortened substantially or simplified, as we believe the steps carried out in the analysis are crucial to arrive at robust conclusions. We also note that the reviewer later on states that the paper has an "*excellent methodology*". We do not agree that the number of investigated streamflow characteristics should be reduced, as forests are likely to impact some aspects of the hydrograph more than others (e.g., dry-season flows). Finally, we note that the other two reviewers did not take issue with the methodology, nor the level of detail it was documented to in our original submission.

I think the question of errors in the primary streamflow data could be addressed a little. The authors seem to disbelieve their own findings ("Possible explanation for the apparent lack of a clear signal..." but equally, perhaps there isn't a clear signal).

All streamflow data were visually screened for artifacts. We agree with the reviewer that it is possible that unaccounted errors in the streamflow data might have confounded the analysis, and did mention this in the Discussion of the original paper (page 3069, lines 1 to 24). We agree that the use of the word "apparent" is not appropriate here and have therefore deleted it.

Specifically
Does the manuscript represent a substantial contribution?
Good, but communication compromised by so much detail

As stated above, we are convinced the amount of detail is needed to fully explore the information in the data we used. In addition, the other two reviewers did not take issue with the methodology, nor the level of detail it was documented to in our original submission.

Scientific Quality
Approach valid? Yes, although some considerations of error in the stream gauging records would be appropriate?

Agreed, noting this was previously mentioned in the Discussion (page 3069 lines 1 to 24) of our original submission.

Results discussed in an appropriate way? Reader tends to be cascaded with detail.

The current level of detail provided in our results section is needed to allow us to draw firm conclusions. Additionally, the other two reviewers did not take issue with the level of detail provided in the results section.

Presentation Quality
Graphics and tables have too much detail for most readers.
Writing is good but "dense"

We consider the level of detail in the paper necessary to provide the reader with sufficient background and argumentation. If we made writing style less dense then this would increase the overall word count of our revised manuscript, and as Reviewer 3 already seems concerned by the length of our original manuscript we decided to maintain the current writing style, which is, to some degree, a subjective choice and some specialist readers might prefer less details. However, as others might require more rather than less explanation we prefer to err on the side of caution. As it is easy for an experienced, specialist reader (such as Reviewer 3) to skip material that is provided, as opposed to fully understand material that is not provided, we tried to strike the right balance, and note that the other two reviewers did not suggest the text was too detailed or dense. Additionally, similar studies have equal levels of detail (e.g., Wilk et al., 2001; Adnan and Atkinson, 2011; Silveira and Alonso, 2009).

Overall

The paper is an appropriate piece of work for the journal with excellent methodology. To some extent the methodology overwhelms both the data and the reader. I cannot imagine most readers doing anything than going through the abstract and even that is hard-going). So my suggestion is a substantial simplification of the paper with more concentration of the overview and less on the detail

Thanks for your assessment that "*The paper is an appropriate piece of work for the journal with excellent methodology*". To ensure repeatability of the method developed, and as both Reviewers 1 and 2 had no issue with the level of detail provided, and noting that experienced, specialist reader may wish to skip some of this material, we have simplified our revised manuscript where possible.

Comments Noted While Reading the Paper

1: The abstract is quite long.

We have shortened the abstract in the revised version (from 353 to 326 words) but are hesitant to shorten the abstract more to avoid loss of essential information.

2: A little jargonish – "Q metrics", "mesoscale",

We agree that the term "streamflow metric" is a bit unclear by itself and have therefore replaced it throughout the paper with "streamflow characteristic". Furthermore, to avoid confusion we have explicitly defined "meso scale" as catchments ranging in size from 1 to 10,000 km^2 in our revised paper. Thank you for pointing this out.

3: Not sure that logging for timber is deforestation.

The word “logging” was never used in our original submission (nor in our revised paper) so we are not sure exactly what the reviewer is referring to. However we agree that ‘logging’ usually refers to selective timber harvesting and ‘clearing’ to full-blown forest conversion to other land uses.

4: It is a very high rainfall zone so that results from less humid areas may not be very applicable.

This is correct and this is why Table 6 of the original paper lists only studies in humid tropical, subtropical, and warm-temperate climates.

Page 3048: Line 10. Don’t the reductions in stormflow also go the same way as increased forest water use (i.e. less streamflow?).

This is correct; the increased water use of forests generally leads to reductions in total water yield as well as small stormflows. However, forest cover changes are considered unlikely to impact the magnitude of large stormflows (see FAO, 2005, and references therein). Furthermore, the impact of changes in forest cover on dry-season flows depends on the quantitative trade-off between increases in infiltration due to enhanced soil infiltration and decreases in flow due to the enhanced water use. This was discussed on page 3047 lines 19 to 27 and page 3048 lines 1 to 19 of our original submission.

5: The paper is quite hard to read and seems to have a number of contradictory statements embedded in it.
Hypothesis 3 embraces all possibilities so will always be true and easily proven. Good to have an axiomatic hypothesis.

Actually, hypothesis 3 states that “… depending on the *trade-off* between the changes in vegetation water use and infiltration associated with forest regrowth, Q_{bf} shows either a negative, no, or a positive relationship with the area under regenerating forest.” This hypothesis therefore does embrace all possibilities but the actual result will be determined by the balance in changes incurred for the respective processes (i.e., vegetation water use and infiltration). Because the change in soil infiltration has been relatively minor in Puerto Rico (see page 3064 lines 16 to 24), we expect Q_{bf} to exhibit a negative relationship with the area under regenerating forest.

6: The three driest months receive 1200 mm per year? Does this mean 400 mm in that period? Ditto for the 3100 mm per year – do we divide by 4? Needs clarification.

This is correct and we agree that our notation was slightly cumbersome; thanks for the comment. The numbers should be divided by 4 to yield the seasonal totals. We have changed the numbers to express the seasonal totals in the revised paper.

7: Acronymns – thus we are investing Q and relation to P but socioeconomic changes in PR are at the bottom of it. I presume PR means Puerto Rico (indeed looking back, it does) but probably better spelt out in full.

We agree that the abbreviation PR may be unnecessary and have therefore refrained from using it in the revised paper.

8: So the rationale is to use routine stream gauging records and correlate with land use change. Over the years I've refereed quite a few papers with a similar approach (and usually inconclusive results). The roughness of the streamflow record is usually pretty bad, so that is something that might be looked at. Can you categorise the accuracy of the data somehow?

It is unfortunately difficult to quantify the exact error in the streamflow data, and we recognize that unaccounted errors in the streamflow data may have affected the results (as mentioned in our original submission on page 3069 lines 1 to 24). However, the streamflow error is probably limited, since the gauges are maintained by the USGS Caribbean Water Science Center (http://pr.water.usgs.gov/about/whatwedo.html), which maintains high standards of quality.

7: Daily Q commonly exceeds Daily P when there is significant catchment storage so not sure that this needs filtering.

When the streamflow regularly exceeds the precipitation for multiple events this is likely due to a bias in the streamflow or precipitation data, rendering them unusable for the analysis.

8: Figure 4: I wonder whether a binary classification of forest/agriculture would be better?

We disagree and argue that this would make the figure considerably less informative. For example, it is important to understand the changes in urban area as these also have the potential to alter streamflow. It makes a world of difference in terms of hydrological response whether a forest is replaced by, say, well managed grassland, poorly managed cropland, or indeed urbanization. Excluding the latter would have created problems in the Discussion section with respect to explaining the observed changes in stormflows for instance.

9: Heroic statistics.- 30,000 model runs alone to ensure convergence!

We do not agree that the number of model runs is disproportionately high. Please refer to our initial response to this comment.

10: After reading a few pages of this, one's head spins with hypercubes, parameter spaces, composite objective functions, etc. Heroic (but perhaps heroic in the sense that the Charge of the Light Brigade was heroic but the strategy wasn't great....). For lesser mortals, the stuff on many of the pages can only really be skipped over. For a comprehensible paper I think that this stuff needs condensing.

Latin hypercube sampling (LHS) is a better alternative to the commonly used Monte Carlo technique used to generate candidate parameter sets by exploring the *a priori* parameter space. The objective functions are subsequently used to evaluate the performance of each model run. We would argue that these are well-established concepts in hydrological modeling (e.g., Sun et al., 2012; Kuczera, 1997; Shafii and De Smedt, 2009). We have no opinion on the Charge of the Light Brigade but understand it was the consequence of a misunderstanding in the British military command. To avoid misunderstandings, we have added relevant references for each of the concepts used so that the reader can more easily find background material.

11: A new concept – "per pixel trends"

We agree that the use of "per-pixel" in this context (page 3061 line 8) might be confusing, and we have therefore removed "per-pixel" from the sentence in the revised paper.

12: I think that some sort of block chart illustrating what was done conceptually really would help. The reader gets lost in the world of jargon or "modern" improvements of old techniques.

We have created a flow diagram outlining the different steps of the analysis (see Figure 1 in this response); thanks for the suggestion.

13: The discussion, including alternative explanations, could be dramatically shortened.
14: Tables 1-6 probably contain too much information for most journals.

We re-read the Discussion but do not see obvious opportunities to reduce unnecessary detail, nor did Reviewers 1 or 2 suggest so. However, we welcome specific suggestions on where the Discussion might be shortened. We do not agree that the tables contain too much information. Each table serves an important purpose. Table 1 provides information about the records and climatic situation of the catchments. Tables 2 and 3 are necessary to reproduce the results. Tables 4 and 5 present the results (including estimates of their uncertainty). Finally, Table 6 provides a review of previous catchment-scale studies conducted in humid tropical, subtropical, and warm-temperate climates. Comparable papers dealing with similar issues have used comparable level of detail in their tables (e.g., Wilk et al., 2001; Adnan and Atkinson, 2011; Silveira and Alonso, 2009).

New references cited

Adnan, N. A. and P. M. Atkinson, Exploring the impact of climate and land use changes on streamflow trends in a monsoon catchment, International Journal of Climatology, 31, 815–831, 2011.

Food and Agriculture Organization (FAO), Forests and floods: drowning in fiction or thriving on facts? Forest Perspectives, 2, RAP Publication 2005/03, FAO, Rome, 2005.

Kuczera, G., Efficient subspace probabilistic parameter optimization for catchment models, Water Resources Research, 33(1), 177–185, 1997.

Marshall, L., D. Nott, and A. Sharma, A comparative study of Markov chain Monte Carlo methods for conceptual rainfall-runoff modeling, Water Resources Research, 40(2), doi: 10.1029/2003WR002378, 2004.

Pinol, J., K. Beven, and J. Freer, Modelling the hydrological response of mediterranean catchments, Prades, Catalonia. The use of distributed models as aids to hypothesis formulation, Hydrological Processes, 11(9), 1287–1306, 1997.

Silveira, L. and J. Alonso, Runoff modifications due to the conversion of natural grasslands to forests in a large basin in Uruguay, Hydrological Processes, 23, 320–329, 2009.

Shafii, M. and F. De Smedt, Multi-objective calibration of a distributed hydrological model (WetSpa) using a genetic algorithm, Hydrology and Earth System Sciences, 13, 2137–2149, doi:10.5194/hess-13-2137-2009, 2009.

Sun, A. Y., R. Green, S. Swenson, and M. Rodell, Toward calibration of regional groundwater models using GRACE data, Journal of Hydrology, 422–423, 1–9, doi: 10.1016/j.jhydrol.2011.10.025.

Wilk, J., L. Andersson, and V. Plermkamon, Hydrological impacts of forest conversion to agriculture in a large river basin in northeast Thailand, Hydrological Processes, 15, 2729–2748, 2001.

Zhang, L., N. Potter, K. Hickel, Y. Zhang, and Q. Shao, Water balance modeling over variable time scales based on the Budyko framework - model development and testing, Journal of Hydrology, 360(1–4), 117–131, 2008.