RESPONSE TO REVIEWER #1

We would like to thank the reviewer for the positive general comments on our manuscript and for the specific comments that helped to improve this manuscript. Our point-by-point response to the reviewer comments is given below.

General comments:

Overall this is a very well written manuscript that presents a significant body of work that seeks to identify the key predictor variables explaining interannual variability of water-balance evapotranspiration. Following revision addressing my comments below this manuscript could become a valuable contribution to this area of research. However, for this manuscript to be acceptable for publication the authors need to address the following key issues, which are outlined in detail in the specific comments section below.

1. Assess how important the human disturbance correction is to their total annual runoff data and present results from this assessment.

2. Check whether their definition of a water year is appropriate across all of their catchments and whether a more appropriate definition would significantly impact their results and conclusions.

3. Briefly explain the variance partitioning technique and improve the presentation of the results from analyses using that technique.

4. Check the physical plausibility of the water balance of their key data sets used in the analysis. It is highly likely that physically implausible data are being analysed here and they will add noise to the results.

Once these issues have been addressed it will become possible to assess the results and conclusions presented in this manuscript with more confidence.

We have discussed each of these general comments in detail in the specific comments below.

Specific Comments / Corrections:

1) Page 5741, Lines 17-20: It would be worth mentioning here that different observed runoff data sets have produced different runoff trends. For example, Labat et al (2004), used by Gedney et al (2006), showed predominately increasing trends in runoff. While Dai et al (2009) and Milliman et al (2008) found mainly decreasing trends in runoff. The quality of the Labat et al (2004) runoff data set has been criticised due to the method used to infill and extend each runoff time series (see Legates et al., 2005; Peel & McMahon, 2006; Dai et al., 2009).

   We have noted these points in the revised manuscript (section 1.).

2) Page 5744, Line 3: What method is being used to calculate potential

evapotranspiration here? A reference is provided for how it was calculated, but some basic information about the methodology used is required here. For example, is potential evapotranspiration being estimated based on Penman, Penman-Monteith, Priestly-Taylor, Reference crop, etc or is it a point or areal estimate of potential evapotranspiration along the lines of the complimentary relationship?

> The method used is the Priestley-Taylor potential ET. We have added additional information about the method in the revised manuscript (section 2.1).

3) Page 5744, Lines 11 — 16: The authors correctly note that this observed runoff data set will be impacted by human disturbances and they seek to address this impact by adding water consumption estimates from the WaterGap2 model to the observed runoff to obtain a 'naturalised' (my term, not used by the authors) runoff time series. There is nothing wrong with this approach per se. However, since actual evapotranspiration is estimated in this manuscript using the water balance, ET = Precipitation — Runoff, it is critical to know how important this water consumption correction is to the final runoff series being used in the water balance equation. The authors should calculate the % of total flow that is due to the water consumption correction at each catchment and present a histogram (or some other appropriate figure) of the results. This will provide insight into how important the correction is to the runoff data and hence how important the correction is to the conclusions drawn from this manuscript. If the % of total flow due to the water consumption correction is non-trivial then the correction may play a significant role in the results presented in this manuscript.

> We have included a histogram in the revised manuscript to show the % of runoff due to consumption, separately for wet and dry basins. This correction accounts for a 2% increase in runoff in wet basins and 37% in dry basins, on average. Dry basins are strongly affected by the correction, but as runoff only forms a small part of the water balance in these basins, the correction leads to a smaller relative change in ET (12% reduction). In wet basins, the correction accounts for 2% of ET estimates. As such, we do not believe the correction compromises the conclusions of our study. We have included these figures in the revised manuscript (section 2.2).

4) Page 5744, Line 18: The water year is defined as October to September. Is this definition applied across all catchments? If so, then it will be inappropriate for many catchments. Since this manuscript is investigating the interannual variability of ET using annual time series data, an appropriate definition of the water year should be used at each catchment. One definition frequently adopted is to start the water year in the month with the lowest mean monthly runoff at the catchment. Using an inappropriate water year definition can split the main flow months between two years and introduce a lag between annual precipitation and runoff.

> The same water year definition was applied across all catchments and we clearly state (P5744, L18-19) that this is done to account for snowpack water storage in (mainly) the Northern Hemisphere. This definition has also been applied in other recent papers, e.g. Dai et al. (2009) and Zhang et al. (2012).
>
> We investigated lags between annual runoff and precipitation across wet and dry basins. When no lag was introduced, the mean correlation

coefficient between runoff and precipitation was 0.68 in wet and 0.66 in dry basins. By introducing a one-year lag in precipitation, the correlation coefficients dropped to 0.00 for both wet and dry basins and thus we found no evidence for significant lags in runoff compared to precipitation. Whilst we acknowledge our definition of a water year does not strictly apply to all basins, it does not seem to compromise the relationship between runoff and precipitation and removes the significant effect of precipitation storage in snowpacks over the winter.

5) Page 5745, Lines 21-23: The point-based method is appropriate for large catchments, but can become problematic for smaller catchments. It would be helpful to know the distribution of catchment area (a histogram?) so the reader can assess the likely error introduced by this method.

The same method was used for small basins (500-10,000km^2) in Ukkola and Murray (2013), and despite the downsides of this method, observed basin runoff and precipitation were found to compare favourably against gridded data at 0.5° (the same resolution as used in the current study). As such, we do not believe the extraction method has compromised our results to a significant degree.

6) Page 5746, Line 12: Define "PFT".

This has been corrected in the revised manuscript (section 2.3).

7) Page 5747, Lines 1-5: The authors should note that nutrients are not being modelled here. The response of vegetation to enhanced CO_2 is not solely limited by water; it can also be limited by nutrients (Korner, 2006).

We have noted this in the revised manuscript (section 2.3).

8) Page 5747, Line 22: Results from the variance partitioning methodology form a major part of this manuscript. The authors need to explain this methodology briefly and in particular explain how cross-correlated predictor variables influence the results. I am certain the meteorological variables will show significant cross-correlation, which means they are not independent and hence may violate the assumptions of the variance partitioning technique.

Variance partitioning explicitly accounts for cross-correlation among predictor variables and as such allows the estimation of the *unique* effect of each set of predictors *after* the effects of cross-correlation and random variance have been removed. The method thus partitions variance into four components: the unique contributions of the two sets of predictor variables, the common contribution of the two sets, and a residual component (random variance). The common contribution arises because there is cross-correlation, and the unique contributions are what remain after the cross-correlation has been accounted for. The unique effect of one set of predictors is calculated as the difference between adjusted R^2 values derived from multiple regression analysis using i) all predictor variables and ii) the other set of predictors.

We have added a more detailed discussion of the method in the revised manuscript (section 2.5).

9) End of Section 2: Significant energy has been spent collating and presenting various data sets and outlining models and analyses to be used. However, a presentation has not been made of the physical plausibility of the critical data (precipitation, runoff and potential evapotranspiration). I strongly recommend the authors present a diagram of the data along the lines of Figure 8 in Kauffeldt et al. (2013), Figure 1 of Le Moine et al (2007) or Figure 4 of Peel et al (2010) to assess the physical plausibility of the data set being used in the subsequent analyses. Catchments found to have implausible water balances could be removed from the subsequent analyses, which may remove some of the noise from the results and provide more confidence in the results.

We investigated the physical plausibility of our data using the criteria set out in Kauffeldt et al. (2013) and Peel et al. (2010). Three main criteria undermining physical plausibility were identified: i) negative mean annual water balance ET, ii) mean actual ET > mean potential ET and iii) mean annual runoff ratios > 1. Our findings are shown in the table below:

Criterion:	**CRU-based ET:**	**GPCC-based ET:**
Negative mean annual water balance ET	Nass (-108 mm/year)	Lagarfljot (-206 mm/year)
Actual ET > Potential ET	7 basins (all wet) (AET exceeded PET by 9-416 mm/yr)	6 basins (all wet) (AET exceeed PET by 3-610 mm/yr)
Runoff ratio >1	Nass (runoff ratio=1.1)	Lagarfljot (runoff ratio=1.23)

We re-ran the multiple regression analyses used to attribute trends in wet and dry basins (Tables 1 and 2, respectively) without the offending basins. The slopes, p-values or R^2 values did not change significantly for either CRU- or GPCC-based ET. The same predictors remain significant and the magnitude of significant coefficients is virtually unchanged. We have included a sentence in the revised manuscript to point out some of the data are not physically plausible but that the results are not sensitive to this (section 3.2).

10) Results section: In several places in this section results are discussed that are not shown in either the Tables or the Figures. For example, page 5749, line 21 "36-40 % and 43-49 % of ET" and page 5750, line 21 "42-49 %". These results are interesting, but it would be better to present them in the Tables and Figures so that the reader can see the results and assess how they were derived. Overall I found the discussion of the variance partitioning results difficult to assess as they are not

presented (Tables 1 & 2 and Figure 2) in the form they are discussed.

The percentages (R^2-values) have now been collated into two tables.

11) Figure 2: I recommend a consistent scale is used for Figures 2a, 2b and 2c. The version presented uses a consistent set of colours, but the discretisation of the legend changes for each map, so the same colour represents a different R^2 of adjusted R^2 in each map, which is confusing.

The figure has been redrawn and colour scales standardised across all maps.

References:

Dai, A., Qian, T., Trenberth, K. E., and Milliman, J. D.: Changes in continental freshwater discharge from 1948 to 2004, J. Clim., 22, 2773-2792, doi:10.1175/2008JCLI2592.1, 2009.

Kauffeldt, A., Halldin, S., Rodhe, A., Xu, C.-Y., and Westerberg, I. K.: Disinformative data in large-scale hydrological modelling, Hydrol. Earth Syst. Sci. Discussions., 10, 487-517, doi:10.5194/hessd-10-487-2013, 2013.

Peel, M. C., McMahon, T. A., and Finlayson, B. L.: Vegetation impact on mean annual evapotranspiration at a global catchment scale, Water Resour. Res., 46, W09508, doi:10.1029/2009WR008233, 2010.

Ukkola, A. M., and Murray S. J.: Hydrological evaluation of the LPX-DGVM for small river catchments in the UK, Hydrological Processes (in press).

Zhang, Y., Leuning, R., Chiew, F. H. S., Wang, E., Zhang, L., Liu, C., Sun, F., Peel, M. C., Shen, Y., and Jung, M.: Decadal trends in evaporation from global energy and water balances, J. Hydrometeorol., 13, 379-391, doi:10.1175/JHM-D-11-012.1, 2012.