Earth Syst. Sci. Data Discuss.,
https://doi.org/10.5194/essd-2019-57-AC1, 2019

Interactive comment on "Paleo-hydrologic reconstruction of 400 years of past flows at a weekly time step for major rivers of Western Canada" *by* Andrew R. Slaughter and Saman Razavi

Andrew R. Slaughter and Saman Razavi

a.slaughter@ru.ac.za

Received and published: 12 November 2019

We are extremely grateful for the constructive interactive reviews of our paper entitled: "Paleo-hydrologic reconstruction of 400 years of past flows at a weekly time step for major rivers of Western Canada". The comments are constructive and will help immensely in improving the paper. The format of this author response will be to address each reviewer comment sequentially by first quoting the comment and then providing a response immediately below.

Question 1 "Why should we make sure that the reconstructed flows "properly preserve the statistical properties of the reference flows, particularly, short- to long-term persistence and the structure of variability across time scales", mentioned later in the abstract and used in the methodology"

This is an important comment, and the paper has been changed to make this point clearer. This point has been mainly ignored in the previous work, as discussed in Razavi et al. (2016) and Razavi and Vogel (2018). The reconstructed flows over 'the reference period' were constructed in a way to preserve the statistical properties of 'the reference flows'. This is to ensure that the important statistical properties in the entire reconstructed flow dataset are sufficiently representative. The outcome of any modelling study that uses these flows will depend to a large extent on the presence of these statistical properties. Many performance measures for example in a water management modelling study will depend on the variability of weekly flow or the long-term autocorrelation in the flow.

Question 2 "Similar to Razavi et al. (2016), only the four . . ." needs to be explained more in-depth in this manuscript to make it self-contained."

The reviewer makes a good point by this comment relating to the MLR models for the North Saskatchewan River and the Oldman River sub-basins using the four and eight chronology sites falling within the sub-basins, respectively. It was thought that the MLR models would have less uncertainty if they were constructed using the chronology sites falling within the respective basins. Unfortunately, the Red Deer River and Bow River sub-basins contain few or no chronology sites, so as a way around this, all the chronology sites were used in the construction of their MLR models. As evident in the results shown in Figure 3, the use of chronology sites falling within a particular sub-basin to construct the MLR model for that sub-basin does in fact increase the accuracy of the reconstructed biennial flows. We will make sure that this point is clear in the revised manuscript.

Question 3 “Adding to comment 2 above, the following methods/approaches need to be explained in Section 2 to make the manuscript self-contained: a. Page 5: why MLR is expected to reconstruct flow from tree ring data, adequately. b. Page 5: The “leave-one-out cross-validation strategy” c. Page 6: The “random matching”.

The reviewer brings up very good points by these comments, and addressing them in our paper should improve the description of the methodology considerably. Multiple linear regression (MLR) is a simple but effective method of modelling the direct monotonic (approximately linear) relationship between tree growth rate and flow. Since both flow and tree ring growth depend on soil moisture, we can expect a fairly linear relationship. The “leave-one-out-cross-validation-strategy” maximises the validation of the MLR models given that the overlap period between the tree-ring chronologies and reference flows is relatively short. Here, if the overlap period is n years, n-1 validations are possible by sequentially calibrating using n-1 years of data and validating against the year left out, with the average R2 of the n-1 validations used in the final analysis of the model performance. Random matching relates to step four in Figure 2, where as the reconstructed biennial flows are stepped through, a reference average biennial flow with similar hydrological properties is randomly selected. The randomness of the selection allows an ensemble approach to be adopted as multiple datasets can be generated that are different but retain the same underlying statistical properties. We will make sure that these points are clear in the revised manuscript.

Comment 4 “Authors need to elaborate on the negative regression coefficients in table 1. Are they physically meaningful? Could the regression be constrained to take only non-negative coefficients? In the same table, variables (e.g. WWP and JOLA) have to be introduced.”

This is a very good point. The tree growth (as represented by three ring width) and water availability (as represented by streamflows here) are always positivity correlated in moisture-limited settings (not necessarily true in energy-limited settings). This means that a regression coefficient in a single linear regression should always be positive.

However, in multiple linear regression, the signs of some of the coefficients might become negative because of collinearity, which relates to the dependence of the tree-ring chronologies. The only way to avoid seeing negative signs is to apply principal component analysis (PCA) to the repressors before doing regression. However, applying PCA would not improve the predictive power of the regression in this case, and therefore, has not been conducted. We will make sure to explain this point in the revised manuscript. We will also define the variables in the regression equations more carefully.

Comment 5 “Page 7: What do author mean by naturally in “The biennial reconstructed time series naturally demonstrate smaller variability compared with the biennial flows in the reference period, when MLR models are used for reconstruction.”

The reviewer has highlighted the use of a term which may cause confusion, and the paper has been adapted to remove the use of the word “naturally”. The MLR models fitted by the Least Square Method always produce smaller variance compared with the variance of observations. As such, reconstructed flows will have less variability compared to the reference flows as the tree-ring chronologies will not explain all the variation in the reference flow. The word “Naturally” has been replaced with “As expected”.

Comment 6 “The two year instrumental periods briefly introduced in the abstract need to be explained more in-depth in Section 2. 7.”

The reference to a ”two-year instrumental period” relates to matching the broad properties of the biennial reconstructed flow with those of the biennial average reference flow. In this study, these properties were the hydrological category of wet, average or dry and whether the wetter year in the biennial average occurred in the first or second year. The paper has been adapted to make this point clearer.

Comment 7 “The persistence calculation should be explained in Section 2.”

We appreciate the reviewer drawing attention to the need for further explanation. Further description of the autocorrelation calculation has been added to Section 2.

Interactive comment on Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2019-57, 2019.