Hydrology and Earth System Sciences Discussions

Interactive comment on "Classification and flow prediction in a data-scarce watershed of the Equatorial Nile region" *by* J.-M. Kileshye Onema et al.

Anonymous Referee #3

Received and published: 9 June 2011

This paper shows the results of the application of two different regression analyses to predict mean monthly flow in the data scarce Semliki river catchment. Although the importance of this topic, I have some reservations on the general validity of this research and I would like authors to clarify some main points: How can the used methodology be applied to other catchments? What is innovative in this methodology compared to previous research done in this field? What is the required data for applying this methodology? To me, the structure of the paper is confused and does not present data, methodology and results in a clear way. Paragraph 3 "Methods and materials" to me should be changed into "Data and methodology" and should give a more clear ex-

planation of the available data (maybe the use of a Table to show collected data would help) and the two approaches utilized (in the current manuscript, this paragraph is only listing the names of the two methodologies and not giving any information on the steps to apply them). Part of the description of the methods is spread in the manuscript in different paragraphs. Authors should cluster this information in the "Data and methodology" paragraph. Conclusions are very general.

SPECIFIC COMMENTS Furthermore, I have some specific comments listed here below.

ABSTRACT: Line 11: what about the outcomes of the tree clustering analysis? I have not seen any explanation of this in the manuscript. Line 18: there is no further comment in the manuscript about the ranges of R, R2 values mentioned in the abstract.

INTRODUCTION: A number of other studies have addressed the issue of flow prediction in ungauged basins using a similar approach to the one proposed by this paper (e.g. Castellarin et al. 2004, Advances in Water Resources). Iit would be interesting adding in the introduction a comparison of the proposed approach to the studies already available in the literature.

STUDY AREA: Line 8: where is this station located? Is this the station where the 28 years flow measurements were taken?

METHODS AND MATERIAL: I don't find this paragraph clear. The information that should be given here is missing or partly given. I suggest authors to change the title of the paragraph into "Data and methods" and provide here a detailed description of the original data and an explanation of the methodology used in the study. Page 3602 Line 19: replace "Table 1 shows" with 'Table 1 presents", since Tale 1 is only a list of the characteristics taken into account. Lines 21-22: please, before using an acronyms, always write first the full name (see NDVI or NOAA-AVHHR) Page 3603 Lines 2-4: where were these measurements taken? What sort of flow data was used? Lines 5-10: I got confused reading these lines and had to go through them a couple of times.

Can authors be more clear and explain here in details the two mentioned methods? This paragraph should be devoted to clearly show the reader what approach was used and provide the reader with the necessary tools to understand this study and apply this methodology to other case studies. To me, these lines are not clear enough. It is not even clear what are the steps undertaken and Figure 3 is not helping! Lines 10-15: to be consistent, please provide a short description of R2 and adjusted R2.

RESULTS AND DISCUSSIONS: I am confused by the structure of this paper: shouldn't 4.1 and 4.2 be part of the paragraph on methodology? Page 3603 Line 23: "....high correlations (higher than 0.5)..." where and how do authors set the threshold to identify a high value of correlation? Page 3604 Lines 1-15: A number of values, variables and tests are mentioned here....but there are no comments on what they represent, how to compute them and what is the purpose to use them. Please, re-word this part and provide a more clear explanation. Page 3605 Lines 14-15: this is not clear to me. What do authors mean by "projection of cases?" Lines 16-19: the original sub-catchments are here grouped into two categories, is there any other criteria for doing so, other than simplify the prediction equations? What would be then the outcomes if one single group was to be used? Lines 19-21: this sentence is not clear. What is the "major categorization" authors are referring to? Page 3606 As mentioned above, this paragraph to me has to go into the methodology description. Furthermore, a new paragraph should be added with a clear description and discussion of the results of this study. Did authors perform a cross validation during their analysis? Was the entire set of data used for the regression analysis? Where did authors show and comment on the results of the mentioned (in abstract and conclusions) ranges of values for the multiple R, multiple R2 and adjusted R2? I only see the minimum and maximum performance of the three coefficients in Table 11.....which is not mentioned in the manuscript. Please add in the manuscript a comment to the Table. Figure 6: is this in terms of volume? What is the observed value? Can you describe the observed data in the paragraph "Data and methodology"?

CONCLUSIONS: I have the following recommendations: - Clearly state what is new about this study - Can this methodology be generalized to other case studies? - What is the result of this study that makes authors confident that "the linearity assumption between catchment descriptors and the discharge is adequate for Semliki and hydrologically similar regions"? I have not seen any comment or analysis to support the conclusion that for a catchment with hydrologically characteristics similar to Semliki the linearity assumption is adequate.