Referee Three

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:

The predictions from the two models are in close range with estimates provided by Senay et al. (2009 Hydrological Processes 23, 3675-3681) in their attempt to document the overall basin dynamics of the Nile River. In addition to the fact that this study provides a finer spatial resolution and can be easily used.

 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?

Table 1 that will be included in the revised manuscript indicates the required data for these models, the sources, software and processing involved worth mentioning is the fact that this approach is suitable for humid regions. For the following reason,

The historical monthly flow measurements at the outlet of the Semliki watershed for 1950-1978 were used for the calibration of both the linear and the non- linear models. For each subcatchment, an historical 28-years monthly mean volume was computed proportionally to the subcathment area and was labeled as "control". This approach is supported by the fact that in humid basins like the Semliki as opposed to arid and semi-arid regions, "stream flows increase in the downstream direction, and the spatial distribution of average monthly or seasonal rainfall is more or less the same from one part of the river basin to another, hence the runoff per unit land area is assumed constant over space. In these situations, estimated flows are usually based on the watershed areas, as contributing flow to those sites, and the corresponding streamflows and watershed areas above the nearest or most representative gauge sites" (Loucks et al., 1981 and Loucks and Van Beek, 2005).

These amendments will be included in the revised version of the manuscript.

Table 1

Description	Source	Observation
Landform	SOTERCAF	GIS processing (WINDISP, ARC View3.3, Excel 07-10, Statistica 8.0)
Lithology	SOTERCAF	GIS processing (WINDISP, ARC View3.3, Excel 07-10, Statistica 8.0)
Soils	SOTERCAF	GIS processing (WINDISP, ARC View 3.3, Excel 07-10, Statistica 8.0)
Drainage Density (Dd)	SRTM 90m-DEM, SWAT pre-processor	Subcatchments areas generated from SWAT preprocessor (WINDISP, ARC View3.3, Excel 07-10, Statistica

		8.0, SWAT)
Stream Length	SRTRM 90m-DEM, topographical map (1/50,000), SWAT pre-processor	Generated from the SWAT preprocessor and cross validation with traditional map (WINDISP, ARC View3.3, Excel 07-10, Statistica 8.0, SWAT)
Stream slope	SRTM 90m DEM	Remotely sensed acquired (ARC View 3.3, Excel 07-10, Statistica 8.0)
Rainfall	FEWS NOAA / RFE (2001-2007) rain gauge at Beni (1973 to 2008)	Remotely sensed acquired and Locally corrected and calibrated (WINDISP, ARC view3.3, Excel 07-10, Statistica 8.0)
Elevations: Minimum Maximum Area-weighted average	SRTM 90m-DEM, topographical map (1/50,000)	Remotely sensed acquired and cross validation with traditional map. (Arc View 3.3, Excel 07-10, Statistica 8.0)
NDVI	NOAA-AVHRR (1982- 2008)	Remotely sensed acquired and correlated with rainfall. (WINDISP, ARC View3.3, excel 07-10, Statistica 8.0)

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 explanation 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). **Noted and changes will be made as appropriate**

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.

Noted Changes will be made accordingly

Line 18: there is no further comment in the manuscript about the ranges of R, R2 values mentioned in the abstract.

Noted Changes will be made as appropriate

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.

Noted and this will be included in the revised version of the manuscript, valuable publication have documented the use of landscape attributes for flow prediction, however no such study has been reported in equatorial and humid region like the Semliki where basically no data is available.

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

At the outlet of the Semliki watershed, yes

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

Noted and changes will be made as appropriate

Line 19: replace "Table 1 shows" with 'Table 1 presents", since Tale 1 is only a list of the characteristics taken into account.

Noted

Lines 21-22: please, before using an acronyms, always write first the full name (see NDVI or NOAA-AVHHR)

Noted

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.

Noted and changes will be made as appropriate

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!

Noted and changes will be made as appropriate

Lines 10-15: tobe consistent, please provide a short description of R2 and adjusted R2.

Noted

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.

Noted and changes will be made as appropriate

Page 3605 Lines 14-15: this is not clear to me. What do authors mean by "projection of cases?" Lines 16-19: the original subcatchments 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?

The PCA was performed for all the variables (landscape attributes) (figure 4) (with subcatchments as cases), to identify the redundant variables for the development of the optimal regression-based models, subsequently the PCA results of variables was projected in terms of cases (subcatchments). This allowed the identification of similar subcatchments from the PCA (Figure 5). The PCA identified four groups in terms of similarities of physiographic attributes but these groups could be reduced to two categories (Figure 5). Lumping all the subcatchments into one group would reduce the predictive performance of the model to be developed as it will not be versatile enough to cater for dissimilarities identified in subcatchments at this spatial scale.

Lines 19-21: this sentence is not clear. What is the "major categorization" authors are referring to? Page 3606

The two major groups of similar subcatchments as identified by the PCA (figure 5)

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?

The historical means were used for the development of the regression-based models and subsequently the model was run for a recent period 2001-2007 were observed data are not available unfortunately for cross validation. Performance reported in the manuscript are linked to the calibration process.

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.

Noted and changes will be made as appropriate

Figure 6: is this in terms of volume? What is the observed value? Can you describe the observed data in the paragraph

Yes these are volumes and the observed values are the monthly historical means.

"Data and methodology"?

CONCLUSIONS: I have the following recommendations: - Clearly state what is newabout this study - Can this methodology be generalized to other case studies? **This approach is suitable for humid regions and this observation will be included in the revised version of the manuscript.**

– 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.

Noted and the referee observation will be addressed as appropriate