Anonymous Referee #4
Received and published: 1 August 2012

1) Overview

Utilizing a top-down approach, this study aims at exploring the dominant processes that control the variations of runoff regime behavior over the continental United States. The MOPEX dataset containing over 50 years of daily climatic and flow data from 197 catchments located in a range of climatic and physiographic settings in the continental United States was used. The authors start with a simple two-bucket model (representing fast flow and slow flow processes) and systematically increase the complexity through addition of new processes on the basis of model performance assessment in relation to observed runoff regime curves. Using the complete model, a sensitivity analyses was performed to explore the dominant processes and the required minimum model complexity based on a performance metric (Aikake information criterion). The authors conclude that systematic regional trends exist in dominant processes across Continental U.S. The identified dominant processes were: fast and slow runoff components, snowmelt, subsurface-influenced fast flow, interception loss and seasonal phenology. Only at the end of the manuscript, the authors did somehow a weak attempt to link process understanding explored through the regime curve to the flow duration curve.

2) General Comments:

I think the manuscript is interesting for the HESS readership and makes a good contribution in the subject area of top-down approach to process understanding. The authors performed a detailed modeling study with a diverse dataset to investigate the dominant controls and required model complexity to reproduce the regime curve in over 150 catchments in the U.S. I read the manuscript with a great interest. Due to a lot of work that was performed, the manuscript is very lengthy. I think it should be, and can be, shortened. Below, I provided a few suggestions in this regard. My main concern about the manuscript is that no quantitative results were provided about the model calibration and performance assessment steps. These should be included in the manuscript as summary plots together with appropriate discussions. Only then the reader will have an understanding about satisfactory model and the required processes. I also think that the title of the manuscript should be changed. Current title puts more emphasis on FDC rather than the regime curve; almost 95% of the manuscript is about reproducing regime curve. The manuscript should be revised based on the comments provided in this section and sections listed below before it could be accepted for publication.

We appreciate the reviewer's constructive suggestions. We have made efforts to address these comments: adding the discussion regarding model validation and uncertainty estimation, substantially shortening the manuscript as the reviewer's advises, and modifying the title and abstract, as well as the beginning of the introduction. We hope the reviewer finds this adequate and satisfactory.

Main Comments:

1) The main goal of this paper is to understand the process controls underpinning the runoff regime behavior of some 197 watersheds in the continental U.S through a model-based top-down approach. Therefore the title of the manuscript does not reflect the content of the manuscript and should be modified. The current title misleads the reader as if the manuscript is devoted to investigation of the physical controls of the flow duration curve. Also, flow duration curve provides magnitude frequency of flow within a specified period and does not contain any time information. Therefore the use of "role of seasonality" and "flow duration curve" in the title is also misleading.

We agree with the reviewer that this paper is focused on the regime curves rather than the flow duration curves, and the "role of seasonality" is misleading in the title. We have now modified it to emphasize our focus on the regime curve. This manuscript is the second of a four-part series, motivated by the study of the flow duration curve by Yokoo and Sivapalan (2011). Although our exploration began from the perspective of the regime curve, the goal of the entire series is to unravel the vast amount of information contained in the flow duration curve. Since regime curve represents the middle part of the flow duration curve (Yokoo and Sivapalan,2011)), we chose the regime curve as the signature to understand the underpinning processes controlling the flow duration curve. The ultimate goal is still to understand the physical controls of regional patterns of Flow Duration Curves. A brief discussion about the linkage between the regime curve and the flow duration curve was included in this manuscript, and a detailed synthesis was developed in the fourth paper of this four-part work (Yaeger et al, 2012).

Therefore we think it would be better to keep the series title for consistency with the other three papers (i.e. "Exploring the physical controls of regional patterns of Flow Duration Curves"), and change the specific title of this manuscript as follows. We hope the reviewer finds these changes adequate:

Exploring the physical controls of regional patterns of Flow Duration Curves - Part 2: Role of seasonality and associated process controls, from the perspective of Regime Curves

2) A clear definition of the runoff regime curve should be provided in Section 1 or Section 2.1. The definitions provided state that "mean seasonal variation of within-year runoff variability". Does this mean the authors take the 50-year average of runoff for each day of the year? Clearly state in the manuscript. If daily values are used than the authors should make a distinction between "runoff" and "streamflow" because the latter includes the channel routing. I believe MOPEX dataset contains the streamflow observations.

We appreciate the reviewer's clarification of "runoff" and "streamflow". We have now re-defined the regime curve as the "50-year average of streamflow for each day of the year," and this has been included in Sect. 2.1.

3) What is the advantage of using a Bayesian method for parameter estimation with assumptions of normally distributed regime curves and fast and slow flows? Why not use a more straightforward calibration algorithm such as "Dynamically dimensioned search" or "Shuffled Complex Evolution"? If uncertainty bounds were provided Bayesian method would be more understandable.

We agree with the reviewer that one of the advantages of a Bayesian method is to facilitate uncertainty analysis, which is necessary and helpful and is also the reason we chose it for parameter estimation. We apologize for the missing of the uncertainty estimation in the manuscript, we have now conducted it as follows: given the best fit parameter set for each catchment, the minimum, mean, maximum and standard deviation values for each parameter represent the distribution across catchments (these best fit sets). The upper and lower bounds are defined from the plot of likelihood and parameter values. For each catchment, along the MCMC sampling, there is a chain of likelihood values which are summed cumulatively from the value of smallest parameter value; the upper and lower bounds are then defined when the sum of the likelihood values just exceeds 95% and 5% of the total. The relative error is calculated as half of the range between the upper and lower bounds as a percentage of the parameter with the maximum likelihood value. Median relative error is the median level of the uncertainty among the catchments. The results will be presented in Table 1 in the revised manuscript.

4) Page 7056, Line 11: Explain how the MSE threshold of 0.53 was selected to classify satisfactory/unsatisfactory models. Is this value selected to choose 75% of the watersheds based on fast and slow runoff components? Explain in detail and justify, as this is an important value.

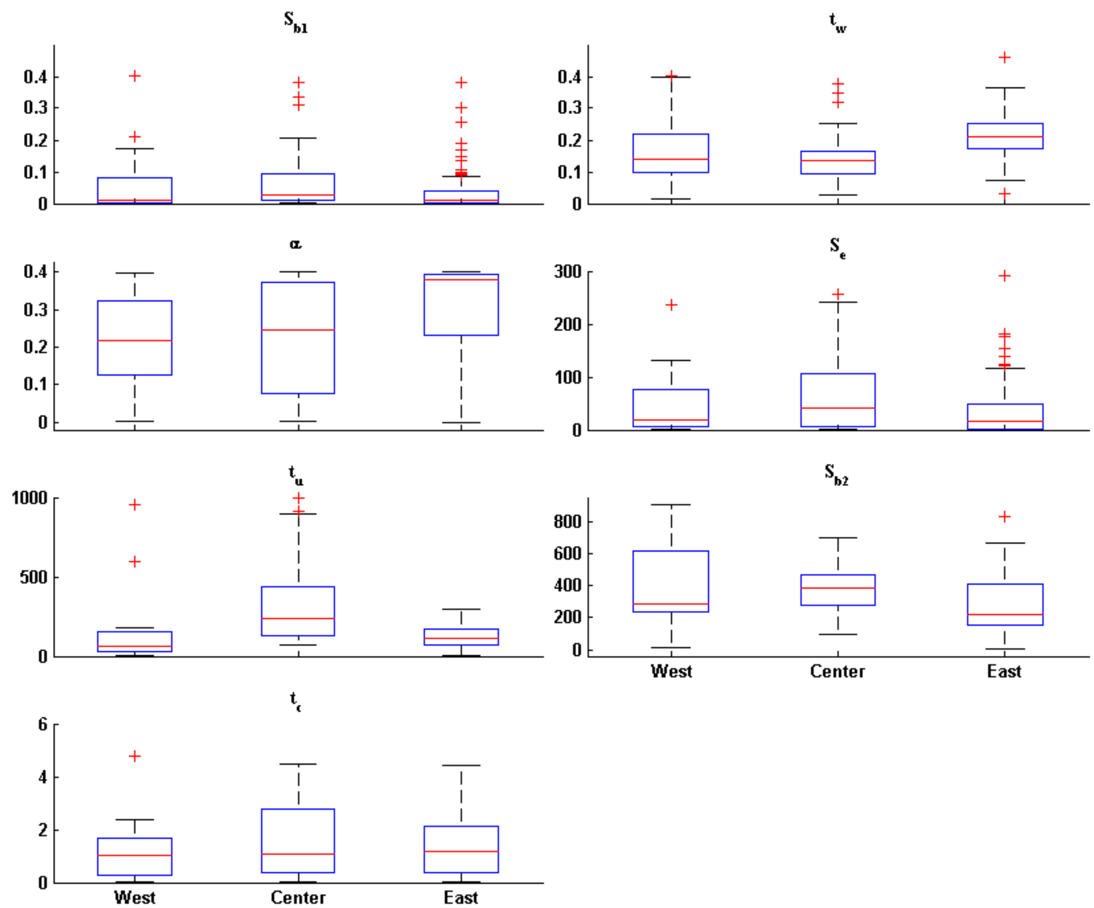
We agree with the reviewer that it is sudden to give 0.53 without an explanation; we have now described how we arrived at it as follows. We hope the reviewer finds it satisfactory.

The initial screening of the model's simulations suggested that even the complete model was insufficient in certain catchments, such as those in the Midwest, where human impacts cannot be ignored. In some catchments, the flow regime curves were bimodal while the model was only able to capture one of the flow peaks. Because it is a simple model, we would not expect it to accommodate anthropogenic activities; therefore, we need to eliminate these catchments where the model performs poorly. To ensure that the model captures the dynamics as well as the volume of the flow, we use MSE as our criterion. The decomposition of the MSE (or Nash-Sutcliffe efficiency) shows that the MSE consists of three components: mean, variance and correlation coefficient (Gupta et al., 2009). However, as the error is scaled by the standard deviation, it could be problematic for comparisons among catchments. To avoid this, we standardized the flow before the MSE calculation. We selected the 90% of the catchments with lowest MSE in fast flow, slow flow and total flow separately and then obtained the intersection of these three sets to determine those catchments that had the lowest MSE in fast flow, slow flow and total flow simulation. These catchments were then considered as satisfactory catchments. We apologize for the usage of the 0.53 without a clear explanation, we have now switch the emphasis from this 0.53 value to the selection processes described here.

5) Table 1. Instead of reporting the mean values of the parameters, their distribution should be provided, which gives more information about the catchment groups. Box whisker plot is a good way to show distributions together with the means.

We agree with the reviewer that a box plot could provide a better presentation of the parameters' distribution across catchments. We have now updated Table 1 with standard *deviation, for the sake of brevity, we will not include the boxplot in the manuscript, but we present it in this response as follows. We hope the reviewer is satisfied with these updates.*

		S_{b1} (mm)	t_w (days)	α	S_e (mm)	t_u(days)	S_{b2} (mm)	t_c (days)
East	**Mean**	0.065	0.218	0.306	36.846	120.260	281.858	1.469
	SD	0.158	0.078	0.128	49.540	64.644	163.704	1.268
	Median Rel. Error (%)	31.47	30.35	13.87	42.65	21.10	9.49	24.26
Center	**Mean**	0.068	0.140	0.221	78.007	323.567	350.640	1.763
	SD	0.098	0.084	0.147	101.615	282.408	160.895	2.049
	Median Rel. Error (%)	11.32	17.50	20.93	32.46	16.78	9.34	14.39
West	**Mean**	0.062	0.159	0.225	56.099	189.287	394.281	1.447
	SD	0.094	0.100	0.132	81.326	351.256	262.644	1.826
	Median Rel. Error (%)	29.19	23.86	20.34	51.94	29.30	7.71	27.28



Where the eastern catchments are located near the east coast and within the Appalachian mountain region, while the western catchments are those located on the west coast and in the Rocky Mountains area; the remainder of the catchments forms the central US group (after removal of catchments deemed "not satisfactory").

6) Page 7042, Lines 7-9 vs. Page 7049, Lines 5-8: In Page 7042, authors state that base model is calibrated first and then new processes are added. Not clear whether model is calibrated again after each process is added. In Page 7049, the authors state that full model (including all process improvements) was calibrated first than components were removed while fixing the parameter values to the calibrated parameter values of the full model.

We apologize for the confusion. The model development (Sect. 2.2) and the parameter calibration (Sect. 2.3) are two procedures. In the model development phase (Sect. 2.2), since our goal is to capture the regime curve as well as possible in order to explore the possible dominant processes in the study catchments, each model was calibrated again after the addition of each process. After the complete model was constructed, we applied this complete model to all 197 catchments, and used the Bayesian method along with MCMC to calibrate the parameters (Sect. 2.3). The parameter set calibrated in Sect. 2.3 was then used in the subsequent comparative model performance assessment, where the parameter values were fixed when components were removed. We hope the reviewer finds the explanation clear and adequate.

7) Model calibration and model performance assessment are critical steps in identifying dominant processes and minimum model complexity. The link between the model and the hydrologic processes can be established only when the calibrated parameters are physically tied to the watershed processes. I think both calibration results and performance assessment results should be provided in a quantitative manner in this manuscript. How the MSE values varied across regions? What were the improvements in AIC criterion between model improvements? A summary of the quantitative values of these measures should be provided in a figure or two and discussed in the manuscript.

We agree with the reviewer that the model calibration and performance assessment are critical in this work. We have conducted the uncertainty analysis for the model calibration as described in comment #3:

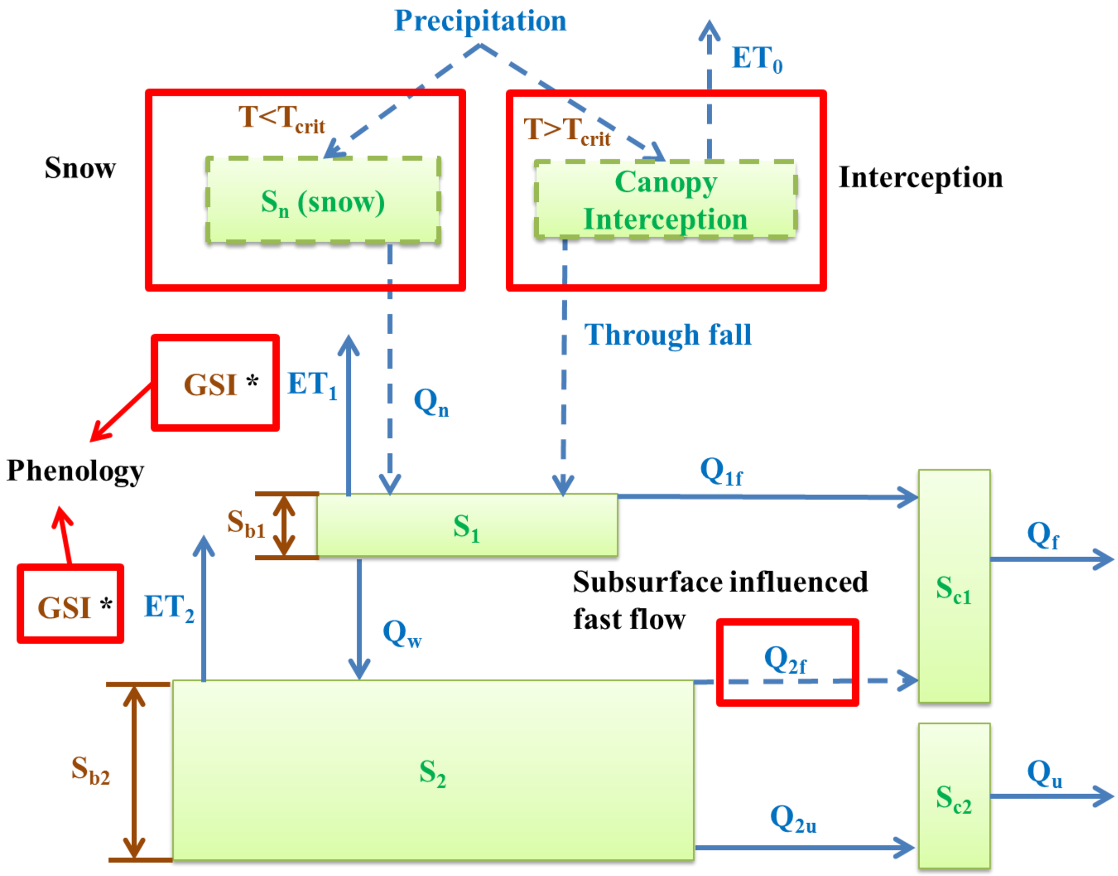
	S_{b1} (mm)	t_w (days)	α	S_e (mm)	t_u(days)	S_{b2} (mm)	t_c (days)
Minimum	0.001	0.013	0.000	0.037	1.548	4.184	0.073
Mean	0.069	0.189	0.274	49.756	187.987	326.358	1.538
Maximum	1.013	0.533	0.300	339.181	1301.191	879.561	9.659
SD	0.14	0.09	0.14	69.44	221.68	183.98	1.51
Median Rel. Error (%)	33.57	33.31	23.74	46.73	24.05	11.54	29.19

We think the discussion section has presented model performance assessment from different perspectives. The MSE value has been presented in the term of satisfactory/not satisfactory *catchments in Fig. 10 as we described in comment #4. While presenting a map for AIC improvement in each model would be too much, as we have 16 models in total, for the sake of brevity we have shown the results of AIC improvement among every model in Fig. 11 in terms of the most important process (the one with most AIC improvement with single addition of each process) and in Fig. 12 in terms of the necessary processes (the one that has most AIC improvement with minimum processes).*

8) Figure 3: Figure 4 contains all the information provided by Figure 3. Figure 3 should be removed and the base model should be highlighted in Figure 4.

Thank you for the suggestion about these two figures. We cut Fig. 3 and re-sketched Fig. 4, hoping the reviewer finds this one appropriate:

Reservoirs are represented in solid green boxes; green is used for state variables, blue for fluxes and brown for model parameters. Red boxes show the four added processes and dashed lines denote the fluxes from these added processes.



9) Section 2.2 and Section 3 should be merged and condensed to reduce the manuscript length.

We agree with the reviewer that for the top-down approach, the model development and the result are closely related, and that by merging them, we can substantially shorten the manuscript. We have now merged Section 2.2 into Sect. 3 in the revised manuscript (for the *brevity of the manuscript, we will not quote the new section here). We hope the reviewer find the revised manuscript solid and concise.*

10) Figure 13: Figure 13a is not necessary and can be removed to shorten the manuscript. All the information contained in Fig 13a is already provided by Fig 13b.

We appreciate the reviewer's suggestion in helping shorten the manuscript; we have now removed Fig. 13a. We hope the reviewer finds this satisfactory.

Minor Comments:

We thank you for these comments, and we have addressed them in the revised manuscript.

Page 7039, Line 21-22: State which "vegetation type" data is used and the number of classes.

We used the UMD (University of Maryland) vegetation type from the Mosaic vegetation dataset. There are 14 classes in total, including: open water, evergreen needleleaf forest, evergreen broadleaf forest, deciduous needleleaf forest, deciduous broadleaf forest, mixed cover, woodland, wooded grassland, closed shrubland, open shrubland, grassland, cropland, bare ground and urban and built-up. We have included the statement in the revised manuscript.

Figure 1: Explain AI in the figure caption. Describe whether "calendar year" or "water year" is used. If water year is used describe the period.

Thank you for the suggestion, we have now included the explanation of AI ("ratio between potential evapotranspiration and precipitation") in the figure caption. We also clarified that we used the "calendar year" in the manuscript. We hope the reviewer finds this adequate.

Figure 1 & 2: Describe how mm/day runoff was calculated.

We have described the calculation of mm/day runoff in Fig. 1 as "50-year average of streamflow for each day of the year" and the mm/day runoff in Fig. 2 was calculated by sorting the 50-year daily streamflow and plotting it against the frequency of occurrence.

Page 7041, Line 13: Explain clearly how the FDC was constructed in this study.

The FDC was constructed as follows: the 50-year daily streamflow was first sorted, then plotted against the corresponding frequency of occurrence, calculated as (i/N+1), where i was the rank and N was the total number of days.

Page 7042, Line 10: Explain "critical assessment" of model performance. How does it differ from classical model performance assessment?

The model performance assessment we did in this work is diagnostic in a top-down approach (Klemes, 1983; Sivapalan et al., 2003;Thompson et al., 2011), outlined as follows. The base *model was initially applied to the study catchments; in catchments where parameterization cannot improve the prediction, we incorporated processes that we hypothesized would be able to fill the gap between prediction and data. Through this systematic assessment of model prediction, model updating, and model re-assessment, we used the model as a tool to explore the catchments' characteristics. The goal of this assessment was not to achieve perfect prediction, but rather, to determine the controlling processes. We have added this explanation in the manuscript in the hopes of clarifying this idea.*

Page 7043, Line 10-20: Refer to model schematic figures (Figs 3 and 4) where appropriate.

Thank you for the comment; we have now referred to the new model schematic figure in the revised manuscript.

Page 7044, Line 19-20: Clarify this sentence together with the implications for modeling. Also define "Mean residence time" parameter in relation to parameters already explained.

We have updated this sentence as follows; we hope the reviewer finds this description more clear:

"The parameter t_c is the mean residence time – the catchment-scale-averaged time raindrops need to travel from hillslope to catchment outlet. It relates to the drainage area, river network structure, topographic gradient, etc.; here we will calibrate it. Since the fast flow and slow flow are routed together in the river network, they share same mean residence time, even though we treated them separately."

Sections 2.2.1 – 2.2.5: In all these modification steps do the authors perform model calibration and performance assessment steps? These steps should be clearly explained before going through the modification steps. Even after reading "Section 2.3.1 Model calibration" it is not clear to me whether parameter calibration is performed after each modification listed in these sections.

We apologize for the confusion we have caused. We have included the following sentences in the beginning of Sect. 2.2, and we hope the reviewer finds this statement clear and adequate.

The top-down approach was conducted in two steps: firstly, we updated the base model with hypothesized processes, calibrated each model after the addition of each process, thus improving model prediction (Sect. 2.2); after the full model was constructed, we applied the full model to all 197 MOPEX catchments, and used the Bayesian method along with MCMC to calibrate the parameters (Sect. 2.3). The parameter set calibrated in Sect. 2.3 was then used in the following comparative model performance assessment, where the parameter values were fixed when components were removed.

Page 7050, Line 25:Replace "prediction" with "performance" since the study deals with reproducing historical flows but not forecasting future flows.

Thank you for pointing this out; we have made the replacement.

Page 7050, Line 26: Explicitly state how many of the 197 catchments were removed from the dataset?

We have included the number of catchments (which is 45) that were removed; thank you for helping us clarify this.

Page 7051, Line 6-10: Explain each term in Equation 16 clearly. There is no "SQobs" variable. And what is "Qmean". Is it "Qobsmean".

We apologize for the confusion caused during the print, the correct equation is:

$$SQ = \frac{Q - mean\,(Q_{obs})}{std\,(Q_{obs})}$$

where Q represents the time series of flows (observed flow for SQobs or model-predicted flow for SQsim), Qobs is the time series of observed flow, SQobs is the standardized observed flow, and SQsim is the standardized simulated flow. Since SQobs and SQsim were calculated from Equation 16, we use SQ to represent both of them. We have now clarified this in the revised manuscript.

Page 7051, Line 8: SQsim is estimated from Eqn. 17? Please explain.

Again we apologize for this confusion. As explained in the previous statement: both SQobs and SQsim were calculated from Equation 16, thus we use SQ to represent both of them. We have updated the manuscript to better explain this.

Equation 17: MSE: Mean squared error. There is no "Mean" term in the equation.

Thank you for pointing this out; we have now corrected it:

$$MSE = \frac{\sum (SQ_{obs} - SQ_{sim})^2}{N}$$

where N is the length of the simulation.

Page 7051, Line 17: Define "satisfactory".

As we explained in major comment #4, since the model fails in some of the catchments that have experienced major anthropogenic impacts, we needed to remove these catchments, since nothing meaningful could be said about their dominant processes. Thus, we conducted a catchment selection process (outlined in the response to comment #4) to remove them.

Page 7052, Line 1: Replace "used perform" with "used to perform".

We have made the correction, thank you!

Page 7052, Line 6: The use of "discharge". "Runoff" has been used until now. Please be consistent through the manuscript.

We have replaced "discharge" with "runoff"; thank you for pointing this out.

Page 7052, Line 9: Remove "use".

We have removed it, thank you!

Figure 5. Caption: Replace "CA" with "Northern CA".

"Northern" has been added in the revised manuscript.

Page 7054, Line 4: Replace "subsurface-induced" with "subsurface influenced".

We have made the replacement as suggested.

Page 7054, Line 13: Replace "an further enhanced one" with "a further enhanced".

We have made the suggested replacement, thank you!

Page 7055, Line 1: Explain "GSI", was not explained before.

GSI is the "Growing Season Index". It is used to improve the estimates of actual evapotranspiration and account for the effects of plant water-use patterns, i.e. phenology. The calculation was introduced in Sect. 2.2.5. We have now included the full name.

Page 7056, Line 13: State the number of catchments that were satisfactory /unsatisfactory.

Thank you for the comment; we have included the number of catchments that were satisfactory or unsatisfactory now.

Page 7056, Line 19: Explain "satisfactory".

We apologize for the vague expression. The "satisfactory" here refers to the number of catchments remaining after the elimination in Sect. 4.1, where some of the catchments for which the model failed were removed according to the procedure we explained in major comment #4.

Page 7058, Line 17: Replace "decrease AIC" with "decrease in AIC"

We have made the suggested replacement, thank you!

Page 7058, Line 17: Replace "less 3% " with "less than 3%" or with "3% less".

We have replaced the original with "less than 3%".

Page 7058, Line 22: 197 catchments: I thought a number of catchments have been eliminated earlier in the analysis.

We have updated the number; thank you for catching this oversight.

Page 7059, Line 1: Replace ". snowmelt" with ". Snowmelt".

We have made the suggested change, thank you!

Page 7059, Line 22: Replace ". snowmelt" with ". Snowmelt".

We have capitalized "S".

Page 7062, Line 5: Less than 150 catchments: As mentioned earlier, provide the exact numbers. Earlier stated 197 catchments (see also above).

We now use the exact number, as suggested earlier.

Page 7062, Line 20: Difficult to see the catchments in Southern California in Figs 11 & 12.

Since the Southern California catchments are small, and thus not very visible in the figure, we have now made the boundaries bold to make them more visible. We hope the reviewer finds the new figures more clear.

Page 7062, Line 24-25: "phenology decreases" Is the arrow for phenology in Figure 13 pointing the right direction? In addition, I suggest using elongated triangles instead of arrows. Triangle narrows toward reduced process direction.

We have changed the arrows to elongated triangles in Fig. 13, thank you for the suggestion.

Figure 13: Figure 13b: There seems to be a problem with the "base model" regime curve plots.

We apologize for the confusion. To facilitate comparisons between figures, we use the same scale (0-10mm/day) for all the figures. Since the depth of flow is much higher (0-20mm/day) in in Northern California (where the "base model"was sufficient) than in the other catchments used to illustrate the processes, we have to cut it to keep the scales consistent.

Page 7064, Line 13: It seems from Fig. 14 that in none of the catchments full model captures the extreme flows. Therefore no need to mention NY, GA, FL, TX and ID. Alternatively, mention all regions.

We appreciate the reviewer's suggestion and have removed this parenthesis.

Page 7064, Line 27: Yes, timing information is lost in FDCs however, information on extreme values and frequencies becomes rich. Whereas in regime curve extremes are averaged out. Therefore both FDC and regime curve contain different and complementary information.

We agree with the reviewer that the extreme flow information is richer in the FDC than in the regime curve due to the averaging. We have now rewritten the sentence as follows; we hope the reviewer finds this informative:

In general, because of the connection between the RC and the FDC, seasonality is present in the FDC, though not as obvious as in the RC, due to the loss of temporal information. While the time element is lost in the FDC, information on extreme values and frequencies, which are averaged out in the RC, is gained.

Page 7065, Line 20: Replace ". snowmelt" with ". Snowmelt".

We have made the change, thank you!

Page 7066, Line 10: Close the parenthesis.

The missing parenthesis has been added; we thank you for the correction.

References:

Klemeš, V.: Conceptualization and scale in hydrology, J. Hydrol.,65, 1 –23, 1983.

Sivapalan, M., Bloschl, G., Zhang, L. and Vertessy, R.: Downward approach to hydrological prediction, Hydrol. Processes, 17, 2101 –2111, 2003.

Thompson, S. E., Harman, C. J., Konings, A. G., Sivapalan, M., Neal, A. and Troch, P. A.: Comparative hydrology across AmeriFlux sites: The variable roles of climate, vegetation, and groundwater, Water Resour. Res., 47, W00J07, doi:10.1029/2010WR009797, 2011.