General Response:

Thank Prof. Zappa very much for your careful review and helpful comments. We agree with most of your suggestions for this paper and will do a major revision accordingly. First, we agree with your idea that the hydrograph partition in this study is based on "water sources for runoff generation" rather than the "dominant runoff processes" identified by Schmocker-Fackel et al. (2007) and Uhlenbrook et al. (2004), and we will modify the relevant concepts. Second, we will redesign our calibration experiment using multi measures of agreement instead of the single one of NSE, the new measures contain RMSE, RMSEln and NSEln and different measures of agreements will be used in different calibration steps. Third, we will develop a benchmark model: inter-annual mean value for every calendar day (Schäfli and Gupta 2007) to evaluate and compare the improvement of the proposed calibration method and the automatic calibration method. The detail replies for each comment are shown as below:

Remarks:

(1) I have read this manuscript and find in many aspects sound but in some aspects rather poor. I really think that this manuscript was not yet ready for HESSD in the form it has been submitted and will therefore request major revisions for it.

Thank you very much for your careful review of the discussion paper. We will do a major revision for the paper and submit a revised manuscript.

(2) As the authors mentions, the ideas of separating the hydrograph in order to confine equifinality is focus of current research and different approaches have been proposed. This contribution aligns within these efforts and has some merits, since it is simple and, as also noticed by B. Schäfli in her review, potentially easy to be transferred.

Thanks, the contribution of this paper aligns within the efforts to confine equifinality using hydrograph partition. The new contributions here is that hydrograph partition is done based on water sources for runoff generation which reflects the spatiotemporal variability in snowpack, glaciers, and temperature. Model parameters are grouped and related to different hydrograph partitions and are calibrated separately via a stepwise approach.

(3) I also appreciated the field effort that is declared here in order to validate the estimation of temperature gradients in the region.

Thanks, monthly temperature lapse rates are estimated and validated using temperature data series gauged in automatic weather stations set up in the upstream mountain area.

(4) Concentrating on the poor aspects, when I read in an abstract "dominant runoff processes" (DRP) I expect a paper dealing with DRP (e.g. Schmocker-Fackel et al., 2007; Uhlenbrook et al., 2004). What I find here is interesting, but in my opinion should be declared as dominant mechanism leading to water availability for runoff-generation. This being snow-melt, glacier-melt and (storm) rainfall (and combinations). The separation according to the "Date-Index" DI should then be maintained in order to discriminate the low-flow season, where runoff occurs by water release from the subsurface and this because deep percolation is occurring in the periods where DI is equal "1". Here I am surprised, that among the four parameters selected for calibration there is none linked to the groundwater-flow. I think this is because the processes leading to groundwater recharge occurs outside the season where groundwater-flow governs runoff-generation.

We agree with the idea that the hydrograph separation in this study is based on the water source for runoff-generation rather than using the dominant runoff processes defined in Schmocker-Fackel et al. (2007) and Uhlenbrook et al. (2004), in which the dominant runoff processes refer to "HortonianOverland Flow", "Saturated Overland Flow", "fast Subsurface Flow" and "Deep Percolation". The runoff water source in the study area is composed of storm-rainfall, glacier melt and snowmelt. The subsurface runoff in the winter is also generated by the storm-rainfall in wet period. Here the "Date-Index" DI is defined as 1 or 0 to discriminate the two periods. It equals to 1 indicating the rainfall runoff in the wet period, while it equals to 0 indicating the groundwater baseflow out of the wet period. To simulate the groundwater baseflow better, we will add two more parameters dominating groundwater baseflow to be calibrated in the revised manuscript.

(5) It is also surprising, that while the author make efforts in order to separate the hydrograph in different sub-samples, they trust a single measure of agreement in

order to evaluate the model performance and they make also the frequently made assumption that the Nash-Criterion (NS) is the universal measure for this (see also the comment of B. Schäfli in this respect). NS is dominated by the SM+GM+R period and as Schäfli and Gupta (2007) demonstrated the low reliability of NS as measure of agreement in areas with strong seasonality in the runoff hydrograph. In Viviroli et al. (2009) we were also thinking about how to consider different processes in the model calibration and we propose a step-wise calibration guided by multiple objective functions and by iterative (and sequential) pair-wise calibration of tuneable parameter selected under consideration of the process they are associated with (snow-melt, glacier-melt, infiltration, surface-runoff, interflow).

We reread these reference papers (Viviroli et al. 2009; Schäfli and Gupta 2007) and learned more from them. Considering the low reliability of NSE value in the study basin with strong seasonality in the runoff hydrograph, we will redesign the calibration experiment, in which new measures of agreement (RMSEln and RMSE) will be used to estimate parameters in each step, and different measures will be used in different steps. RMSEln will be used for the estimation of groundwater baseflow and snow melt due to the low magnitude of the two runoff components, and RMSE will be used for the calibration of glacier melt and storm runoff considering their relative higher magnitude. We will also develop a simple benchmark model: the inter-annual mean value for every calendar day to evaluate our simulation using multi measures including NSE, NSEln and BE value (Schäfli and Gupta 2007).

(6) Remove all the links to "runoff generation processes" and replace it with "source of water available for runoff generation".

Thanks. We agree with this and the relevant concepts will be modified. We will replace the concept of "runoff generation processes" with "runoff components" and "water source for runoff generation".

(7) Table 5: Is this the magnitude of improvement you were expecting when designing this study? What if you take instead of NSE a Benchmark efficiency, where you compare the simulation against the seasonal runoff (Schäfli Gupta, 2007). This might be sufficient to lead your parameters to be right for the right reason

(Kirchner, 2006).

We will redesign the comparison experiment here in which we compare the results between the two calibration methods using the benchmark model: inter-annual mean value for every calendar day. We will also evaluate the seasonal performance of the two methods using regime curves, NSEln and seasonal contribution of different runoff components.

(8) 1262– 3,6: The visual inspection confirms your statement. Maybe you have some place on Figure 2 next to the legend to declare a measure of agreement for the red and dotted-red lines with respect to the black line.

To describe the fit between the estimated and observed temperature series, we will add a measure of agreement (RMSE) in Figure 2 which can show that the monthly lapse rate performed better than the annual constant rate, especially in the hot months (April to August).Temperature has significant effect on melt runoff which mainly occurs in the hot months (April to August) in this study basin. For the simulation of melt runoff, the estimated monthly temperature lapse rates are sufficiently good.

(9) 1272-1273: You declare that you reach good simulation results except for some large storm runoff events in summer. I inspected figure 10 and I have to admit, that I was not able to find any event characterized by rapidly rising and falling peak that was simulated with your model. Again, you speak of dominant runoff generation, but your model fails in simulating any situation linked with storm-runoff triggered by storm rainfall. I think that your current perceptual model of this area has some missing components that you should investigate. The hydrographs you simulate merely reacts to weather periods characterized by rising and sinking temperatures.

We re-inspected the simulation results both in calibration and validation period. We indeed find that the simulation didn't capture the rapidly rising and falling peak well. The model used here is similar to that used by Tian et al. (2012), in which the model had simulated the storm runoff in blue river basin successfully. The reasons for the low performance may fall in that the values of parameter WM and B which control the storm-rainfall runoff are unreasonable. Parameters controlling both rainfall runoff and melt runoff were calibrated using the NSE value for measure of agreement. According to Schäfli and

Gupta (2007), the reliability of NSE value in the study basin with strong seasonality is relatively low. In this revised manuscript, the calibration experiment will be redesigned by using different measures of agreement in each calibration step. We will use RMSEln for groundwater baseflow and snow melt runoff, RMSE for glacier melt runoff and storm-rainfall runoff. In the redesigned experiment, the reliability of the simulation will be evaluated using regime curves, seasonal contribution of different runoff components. In this way, the reliability of parameter value should be improved and the simulation of storm runoff triggered by storm-rainfall can get an improvement.

(10) Final considerations:

I think this manuscript has potential, but work is needed to make it more ripe. I think that the design of the experiment can be improved by selecting multiple measures of agreement. I think also that the model should demonstrate to be able to cope with storm runoff before declaring success of this experiment. I would be happy if the authors can do a big effort and submit revised version of this manuscript.

Thanks. We would like to do a major revision and will submit a revised version of this manuscript soon afterwards. We will redesign the calibration experiment using multi-measures of agreement including RMSE, RMSEln, NSEln and a BE value comparing to a benchmark model used in Schaefli and Gupta (2007). The model used in this study has been applied to a dozen of watersheds with varied climatic/geographic characteristics (e.g., Tian et al., 2012), we have confidence to say that the corresponding model structure reflects the-state-of-the-art modeling approach. The low performance of the rapidly rising and falling storm runoff should be attributed to the unreasonable parameter values calibrated using NSE here.

Reference

Schaefli, B. and Gupta, H. V.: Do Nash values have value, Hydro. Process., 21 (15), 2075-2080, 2007.

Schmocker-Fackel, P., Naef, F. and Scherrer, S.: Identifying runoff processes on the plot and catchment scale, Hydrol. Earth Syst. Sci., 11, 891-906, 2007.

Tian, F. Q., Li, H. Y., Sivapalan, M.: Model diagnostic analysis of seasonal switching of runoff generation mechanisms in the Blue River basin, Oklahoma. J. Hydrol, 418, 136-149, 2012.

Uhlenbrook, S., Roser, S. and Tilch, N.: Hydrological process representation at the meso-scale the potential of a distributed, conceptual catchment model, J. Hydrol., 291, 278-296, 2004.

Viviroli, D., Zappa, M., Schwanbeck, J., Gurtz, J. and Weingartner, R.: Continuous simulation for flood estimation in ungauged mesoscale catchments of Switzerland - Part I: Modelling framework and calibration results, J. Hydrol., 377 (1-2), 191-207, 2009.