General Response:

Thank Prof. Schaefli very much for the excellent job in reviewing the discussion paper and giving the helpful comments. We will do a major revision on the paper considering her comments. The revision mainly includes: First, we will rewrite the abstract section by adding more details on the partition method and the model results. Second, we will add more reference relevant to the step-wise calibration and runoff generation mechanism modeling in alpine area in the literature review and remove the viewpoint of "the signatures in common use today insufficiently exploit the hydrograph information in the time dimension or in relation to the dominant runoff generation mechanisms". Third, we will add two calibrated parameters that dominate the groundwater baseflow and use multi-measure of agreement including RMSE, RMSEln, NSE and NSEln to evaluate our modeling. Fourth, we will develop a benchmark model: the inter-annual mean value for every calendar day to evaluate our modeling and discuss the results by comparing with the fellowship studies in the same study area. Fifth, we will add a section of parameter sensitivity modeling in the revised manuscript. The detail responses to each comment are as below:

General comments

This paper presents a step-wise calibration approach for a precipitation-runoff model (THREW) applied to a case study in the Tianshan Mountains (China). The paper is well written and structured. The identification of the different dominant runoff processes for the step-wise calibration is very simple (a separation by date to distinguish snow influenced periods from others, two separations based on air temperature) but sound and certainly transferable to other case studies. The literature review of the paper should however be strengthen to support the viewpoint of the authors that "the signatures in common use today insufficiently exploit the hydrograph information in the time dimension or in relation to the dominant runoff generation mechanisms" (p. 1256).

Another potentially weak point of the study is the fact that only 4 model parameters are calibrated. Accordingly, the overall model performance, after calibration, is rather low. This should be discussed in more detail.

In conclusion, the paper is suitable for publication in HESS after major revisions considering namely the detailed comments hereafter:

31   *Thank you very much for your appreciation on some aspects of the paper. We will do a*

32   *major revision on the paper according to your valuable comments: First, we review more*

33   *references on the precipitation-runoff modeling and calibration in alpine areas and remove*

34   *the viewpoint of "the signatures in common use today insufficiently exploit the hydrograph*

35   *information in the time dimension or in relation to the dominant runoff generation*

36   *mechanisms". To the second potentially weak point of the study, we will do a new job on*

37   *parameter sensitivity and we will expand the number of calibrated parameter to six based on*

38   *the sensitivity results by adding two parameters dominating the groundwater baseflow. And*

39   *we will redesign our calibration experiment as suggested by Prof. Zappa, in which, the*

40   *simulation performance is evaluated using multi-measures of agreement including RMSE,*

41   *RMSEln, NSEln, NSE and a BE value comparing a benchmark model: the inter-annual mean*

42   *value for every calendar day. We will also discuss our model results by comparing with the*

43   *fellowship researches in the same study area.*

44

45   **Detailed comments**

46   **1. Abstract**

47   •   **Does not give any details on how the process separation is achieved and no**

48   **conclusion on how the method performs.**

49   *Thanks. We will add details on the hydrograph separation method and final results*

50   *conclusion in the abstract.*

51   **2. Case study**

52   a) **It would be useful to shortly discuss the hydrological regime and tell the reader**

53   **why it is qualified as "alpine"**

54   *To qualify our study area as "alpine", a detail description of the hydrological regime in*

55   *the study area will be added in Sect. 2.1 considering the geographical location, altitude,*

56   *temperature and precipitation, and glacier coverage of the study area.*

57

58   b) **What do you define as "storm" water (p. 1260)? Runoff that is resulting from**

59   **rainfall that has fallen during this event? How do you know where the water**

60   **actually comes from?**

The "storm" water is not linked to event water in our study. It is a complex problem to identify where the water actually comes from (i.e., event water or pre-event water), which is still the unsolved research question in scientific hydrology. In our study, the "storm" water is just used to define the rainfall in the study mountain area. As the rainfall type in the study mountain area is primarily consisted of convective type caused by topography. To highlight this precipitation property, we used the concept of "storm" water here.

c)   Resolution of the used MODIS data (p. 1260)?

The spatial resolution of the MODIS data (MOD10A2 and MYD10A2) is 500m, and the temporal resolution is eight-day, daily snow cover data is obtained by linear interpolation of the eight-day data, which will be discussed in Sect.2.2.

d)   I do not understand how the glacier area is derived.

The glacier cover area is derived from China Glacier Inventory which is digital shapefile (open by Arcgis software package) and assumes to be stable during the study period. MODIS remotely sensing snow cover products (SCA) are used to describe the dynamic of both snow and glacier coverage in TRB. To the authors' understanding, most snow would melt off during the warm summer and thus the lowest snow/ice coverage in summer season could be considered to represent the glacier coverage. Based on the analysis of filtered MODIS SCA (see Sect. 2.2.3 in the manuscript), the lowest values of snow/ice coverage in summer in the study period (2003-2012) are almost the same, which is also very similar to the value (0.33) calculated by China Glacier Inventory (CGI) (0.35, 0.34, 0.39, 0.36, 0.34, 0.41, 0.35, 0.38, 0.39, respectively, see the figure below). It may indicate that glacier coverage is relatively stable during the study period. Then, we assume that the glacier coverage maintains a constant level (using the CGI level), and glacier mass dynamic is not updated in our paper, partly due to the main concern of the paper is the diagnostic calibration method of hydrological model. We acknowledge that this assume can affect our results in some degree, and we will improve the method to consider the dynamic of glacier mass balance which is left for future work.
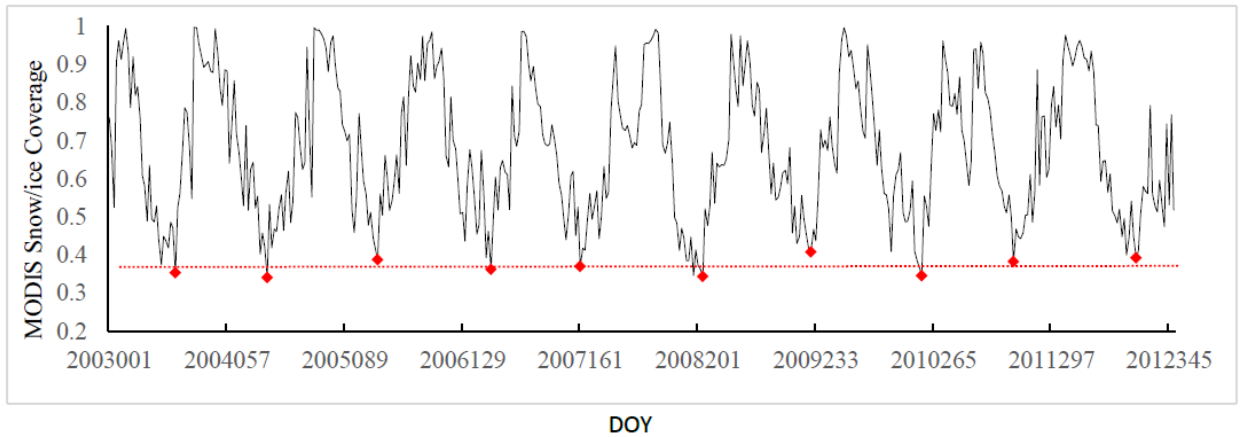
89

90   *Figure1: Dynamic of snow/ice coverage in TRB for 2003 to 2012 from filtered MODIS SCA*

91   *products. The lowest coverage values (red dots) in each summer show the glacier cover area*

92   *in each year which maintains at a relative constant level*

93

94   **3.   Methodology**

95   **a) It should be mentioned somewhere that ice melt has a separated degree-day**

96   **factor (now mentioned only in calibration section, p. 1269)**

97   *Thanks. We will add a sentence to mention the separated degree-day factor for snow and*

98   *glacier melt in Sect.3.2.*

99

100   **b) Should ice melt periods not be identified also on the presence / absence of snow?**

101   *Snowfall in the glacier zone is assumed as ice in our study. To the authors'*

102   *understanding, snowmelt in the glacier zone is much smaller than the glacier melt. So the*

103   *melt water of this kind of snow is calculated using the same degree factor as the glacier melt*

104   *in our study.*

105

106   **c) Model: how is the water balance closed if the snow-covered area is updated**

107   **based on observed MODIS data rather than computed from simulated snowfall**

108   **and melt?**

109   *The basin water balance is closed as the sum of rainfall and melt water equals to the*

110   *sum of runoff and evaporation in this study. This balance is checked for each simulation. In*

111   *the snow-covered zone, snowfall is calculated according to precipitation and two temperature*

4

threshold values, while melt water is calculated using MODIS snow cover area data directly.

The water balance in the snow zone is not closed using the depletion curve (Luce et al., 2004)

which has several parameters to be calibrated. To demonstrate the performance of the

proposed calibration method, we using MODIS and CGI data to update the snow and glacier

area directly to reduce the model parameter dimension. The water balance in the snow and

glacier zone is not concerned in this study.

d) **P. 1270: I do not understand the sentence on low influence of infiltration on stream discharge; where does the baseflow come from?**

The soil infiltration capacity in TRB is very high. Factor that controls the rainfall runoff process is the basin water storage capacity but not infiltration. On the other hand, rainfall in the wet period can infiltrate into soil and form groundwater baseflow. And the baseflow will be calculated using two additional calibrated parameters KKA and KKD in the revised manuscript. The sentence of "infiltration has a minimal influence on stream discharge" in the manuscript should be modified as "infiltration has a minimal influence on storm runoff process and is not concerned here"

e) **Interception p. 1270: it does not only occur on trees (see the work of H. Savenije); how can interception be negligible in a catchment where precipitation is locally as low as 180 mm?**

We agree that it is unreasonable that only taking the interception occur on trees into account. In the revised manuscript, we will inspect the interception occur on all kinds of vegetation which primarily contains tree and grass in the TRB basin. However, according to Sun et al. (2012) and Kang et al. (1980), rainfall on grass and bare soil area penetrates into ground quickly. So the interception on grass area is not significant and can be negligible. Considering the low area of trees, the interception on trees is also not concerned here.

4. **Results**

a) **The calibrated degree-day factor for snow is extremely low (0.9 mm/C/day), is this realistic for other discharge periods (other than the one used for calibration) or should you have a time-variable degree-day factor?**

142   *In the revised manuscript, the calibration experiment is redesigned using multi-measures*

143   *of agreement and a larger number of calibration parameter which will be increased to 6. The*

144   *snow melt degree-day factor will be recalibrated. We have do some new work and obtain a*

145   *new factor as 2.1 mm/(℃.day), a similar value to the previous studies (Ma and Cheng, 2003;*

146   *Liu et al., 2012)in the Xinjiang mountain regions in China. To focus on the partition method，*

147   *we did not using a time-variable degree-day factor.*

148

149   **b) The obtained Nash values for daily discharge are very low for such a regime**

150   **with a strong annual cycle (Schaefli and Gupta, 2007). A value of 0.79 for the**

151   **calibration period is already low, 0.61 for the validation period is very low (in**

152   **my experience, a model that predicts half of the observed flow every day still**

153   **can still give a Nash of 0.6 for such regimes in the Alps); it is also visible from**

154   **Fig.10b that the model seems, overall, to not do a very good job. Any more**

155   **detailed comments on this (considering also the cross-validation)?**

156   *In the revised manuscript, we will add more discussion of the model performance in*

157   *Sect.4. And also, we will redesign the calibration experiment selecting multi-measures of*

158   *agreement (RMSEln, RMSE, NSEln). Parameters will be determined by different objective*

159   *functions in the new experiment. We will also compare our results to a simple benchmark*

160   *model to demonstrate the improvement of our model. The low performance in validation*

161   *period is further discussed in the result section: There are many abrupt runoff events in the*

162   *wet period (May to September), which are mainly driven by storm-rainfall. These extreme*

163   *runoff events are difficult to capture in alpine area where gauged station is scarce on the daily*

164   *scale (Aizen et al., 2000; Jasper et al., 2002). The Nash value is sensitive to the simulation of*

165   *high flow. The low performance can be attributed to underestimation of peak flows. However,*

166   *our result is still better than the simulation by Li and Williams (2008) who do a similar work*

167   *in a basin that near to TRB in Tianshan mountain area. Their Nash values for daily discharge*

168   *varied from 0.51 to 0.78, and also failed to simulate the peak flows in summer. They attributed*

169   *the low efficiency to the heavy precipitation events which is similar to our study.*

170   **c) P. 1274: it is argued that the calibrated parameter dimension is sufficiently low**

171   **to have identifiable parameters; but the low number of calibrated parameters**

172   **leads to such a low degree of freedom that the model cannot do a good job, any**

173   **comments on this?**

174   *The calibrated parameter number in our study is determined by priori information and*

175   *model sensitivity. In the redesigned calibration experiment, the number of calibrated*

176   *parameter is 6 which control the main runoff components (i.e. groundwater baseflow,*

177   *snowmelt, glacier melt and rainfall direct runoff). We will make a comparison between our*

178   *simulation and some fellowship studies in the same area which shows that our model perform*

179   *better than Li and Williams (2008) and Liu et al. (2012), all of their models have a calibrated*

180   *parameter dimension higher than six. Their studies demonstrate that the dimension of*

181   *calibrated parameter is not the significant sensitivity factor that controls the simulation. The*

182   *low efficiency of discharge simulation is attributed to the low performance of peak flows, as*

183   *the melt runoff is simulated well both in calibration and validation period.*

184

185   **d) P. 1274: why should the step-wise calibration method be less sensitive to the**

186   **chosen calibration data than an automatic method?**

187   *The automatic calibration method used here is a benchmark method that uses the single*

188   *whole hydrograph to evaluate the model parameters, while the proposed calibration method*

189   *using multi hydrograph partitions as signatures for parameter identification. The comparison*

190   *between the two methods here is to demonstrate the calibration efficiency of using multi*

191   *hydrological information and the simple stream hydrograph respectively. By selecting*

192   *different calibration data (usually simple time series) for the automatic method, we can*

193   *usually get different groups of parameter producing similar simulations as the information*

194   *hidden in the hydrological process is not sufficiently used . So parameter values determined*

195   *by an automatic method are sensitive to calibration data. However, in the proposed step-wise*

196   *calibration method, time series data are not used for calibration directly. Information of*

197   *hydrological processes is extracted firstly and used to partition hydrograph. The calibration*

198   *data in the proposed method is not simple discharge series, but hydrograph partitions which*

199   *relates to the hydrological process physically (may subject to uncertainty). The relationship*

200   *between parameter and corresponding hydrological process is distinguished, and each*

201   *parameter is determined by the runoff component it controls separately. The role of parameter*

202 *on discharge simulation is distinguished in the proposed method, calibration data for this*

203 *method is more hydrological meaningful than simple data time series used in automatic*

204 *methods. So this method can be less sensitive to the chosen calibration data. The cross*

205 *validation in the manuscript also show that the parameters calibrated by the partitioned*

206 *method are relatively stable except the parameter B which controls the peak flows. The*

207 *simulation of peak flows can be affected by the rainfall measurement data significantly. The*

208 *variation of B value should be attributed to the insufficient of rainfall measurement here.*

209

210 **e) What criteria did determine the number of calibration iterations (p. 1271)?**

211 *The number of calibration iterations is an important issue in model calibration. To study*

212 *the factors that have effects on the iteration procedure is very necessary for the simulation but*

213 *is not concerned in our study. In this paper, we just determine the number of iteration by*

214 *manually inspecting the different of calibration parameter values between two consecutive*

215 *calibrations. If the relative difference of calibration parameter values between two*

216 *consecutive calibrations is lower than 10%, we simply assume that the parameters get stable*

217 *and the iteration procedure stops.*

218

219 **f) It would be nice to have an idea of the model sensitivity rather than a single**

220 **simulation; given the low number of calibrated parameters, this should easily be**

221 **possible and would help understanding the model behavior better.**

222 *Thanks. We will add a section of model sensitivity analysis in the revised manuscript.*

223 *The sensitivity analysis of model parameter will be carried out by a "one-at-a-time"*

224 *approach. Parameters belong to different group are selected for sensitivity analysis, including*

225 *saturated hydraulic conductivity for u-zone K_s^u, saturated hydraulic conductivity for s-zone*

226 *K_s^s, subsurface flow coefficient KKA and KKD, Manning roughness coefficient for hillslope n^t,*

227 *spatial heterogeneous coefficient for infiltration capacity α^{IFL}, ground surface depression*

228 *storage capacity $Fmax^b$, shape coefficient to calculate the saturation excess runoff area from*

229 *the Xin'anjiang model B, spatial averaged tension water storage capacity in the Xin'anjiang*

230 *model WM, degree day factor used to calculate glacier melt GDDF and degree day factor*

231 *used to calculate snowmelt SDDF. Parameter values are set from low 50% and up 50% of the*

value in the calibrated model, and the relative change of simulation measure of agreement

values for different hydrograph partitions are used to evaluate the sensitivity (Eqn.1). The

sensitivity simulation results are shown in Table 1. The results indicated that KKA, KKD, WM,

B, SDDF and GDDF are the most sensitive parameters.

$$R_{MS} = \left| \frac{MS_+ - MS_-}{MS} \right| \times 100\% \qquad (1)$$

Table 1. Parameter sensitivity for multi runoff components. Most sensitive parameters are labeled in red

Measures(period)	Subsurface				Routing	Infiltration	Interception	Storm runoff		Melt	
	K_s^u	K_s^s	KKA	KKD	n^t	α^{IFL}	$Fmax^b$	WM	B	SDDF	GDDF
RMSEln(SF)	9.70	11.14	38.44	44.39	15.70	0.12	0.08	1.07	18.51	7.53	2.88
RMSEln(SM)	4.41	0.54	0.79	1.16	7.17	0.21	0.14	5.24	11.65	25.76	1.07
RMSE(SM+GM)	0.22	0.21	0.62	0.64	10.00	0.17	0.25	7.92	0.29	26.28	40.79
RMSE(SM+GM+R)	0.17	0.85	0.57	0.97	1.84	0.08	0.06	19.35	22.48	10.78	11.57

5. Conclusion

- **Given that the proposed method only separates between rainfall and snow / ice melt driven processes, why would it a priori be limited if applied to catchments with Hortonian overland flow?**

We didn't mean that the proposed method is limited to the catchments with Hortonian overland flow. We will clarify it in the revised manuscript.

6. Literature review

a) **The literature review seems incomplete with regard to step-wise calibration in general and of precipitation-runoff models for high mountainous catchments in particular. The paper by (Schaefli et al., 2005) presents a step-wise calibration method with a similar objective as in the present paper. There are certainly other papers that proposed such a step-wise approach (check e.g. (Huss et al., 2008) or the work of (Pellicciotti et al., 2005).**

We will add some reference in the introduction section focusing on the step-wise

259 *calibration and runoff generation mechanism modeling in alpine areas.*

260

261 **b) There is one reference (van Straten and Keesman) for the ability of regression**

262 **based calibration methods to identify the roles of various model components**

263 **(p.1255). Could you give a more "precipitation-runoff modeling" oriented**

264 **reference?**

265 *We will add several references for the ability of regression based calibration methods to*

266 *identify the roles of various model components (Zhang et al., 2008; Gupta et al., 2008; Yilmaz*

267 *et al., 2008; Hingray et al., 2010), for more "precipitation-runoff modeling" oriented*

268 *references please refer to the reply to the above comment.*

269

270 **c) FDCs have also been used for model calibration, see (Westerberg et al., 2011).**

271 *Thanks. Westerberg et al. (2011) selected several evaluation points on the flow-duration*

272 *curves to calibrate models, and compared two selecting ways to evaluate the effect on*

273 *calibration. We will discuss it in the revised manuscript.*

274

275 **d) There are probably many more references for step-wise calibration and**

276 **calibration on dominant runoff mechanisms (rather than just Boyle et al. 2000).**

277 **It would be interesting to have a more complete discussion of the statement that**

278 **"the signatures in common use today insufficiently exploit the hydrological**

279 **information in the time dimension or in relation to the dominant runoff**

280 **generation mechanisms.".(p. 1256).**

281 *Please refer to the reply to comment a) in literature review, and we will remove this*

282 *viewpoint. Thanks.*

283

284 **7. Other detailed comments:**

285 **a) P. 1256: what is the measurement dimension "M"? Is it equal to the number of**

286 **time series? If yes, why?**

287 *No, it is equal to the number of independent hydrology data which relate to independent*

288 *hydrological processes. Although we usually have a large number of data time series for*

289 *parameter calibration, they are often not hydrological independent, and cannot be related to*

290 *hydrological processes separately, the measurement dimension are generally low.*

291

292   **b) Table 3 does not highlight the calibrated parameters**

293   *Thanks, we will modified the Table and highlight the calibrated parameters.*

294

295 **Reference**

296 Aizen, V., Aizen, E., Glazirin, G. and Loaiciga, H. A.: Simulation of daily runoff in Central Asian
297   alpine watersheds, J. Hydrol., 238, 15-34, 2000.

298 Gupta, H. V., Wagener, T. and Liu, Y.: Reconciling theory with observations: elements of a
299   diagnostic approach to model evaluation, Hydrol. Process., 22, 3802-3813, 2008.

300 Hingray, B., Schaefli, B., Mezghani, A. and Hamdi, Y.: Signature-based model calibration for
301   hydrological prediction in mesoscale Alpine catchments, Hydrolog. Sci. J., 55 (6), 1002-1016,
302   2010.

303 Jasper, K., Gurtz, J. and Herbert, L.: Advanced flood forecasting in Alpine watersheds by
304   coupling meteorological observations and forecasts with a distributed hydrological model, J.
305   Hydrol., 267 (1-2), 40-52, 2002.

306 Kang, E., Zhu, S. and Huang, M.: Some Results of the Research on Glacial Hydrology in the
307   Region of MT. Tuomuer, Journal of Glaciology and Geocryology, 2, 18-21, 1980(in Chinese).

308 Li, X. G. and Williams M. W.: Snowmelt runoff modelling in an arid mountain watershed, Tarim
309   Basin, China, Hydrol. Process., 22 (19SI), 3931-3940, 2008.

310 Liu, T., Willems, P., Feng, X. W., Li, Q., Huang, Y., Bao, A. M., Chen, X., Veroustraete, F. and
311   Dong, Q. H.: On the usefulness of remote sensing input data for spatially distributed
312   hydrological modelling: case of the Tarim River basin in China, Hydrol. Process., 26 (3),
313   335-344, 2012.

314 Luce, C. H. and Tarboton, D. G.: The application of depletion curves for parameterization, Hydrol.
315   Process., 18, 1409-1422, 2004.

316 Ma, H. and Cheng, G.: A test of Snowmelt Runoff Model (SRM) for the Gongnaisi River basin in
317   the western Tianshan Mountains, China,Chinese Science Bulletin, 48 (20), 2253-2259, 2003.

318 Sun, M., Yao, X., Li, Z. and Li, J.: Estimation of Tailan River Discharge in the Tianshan
319   Mountains in the 21st Century, Advances on Climate Change Research, 8, 342-349, 2012(in
320   Chinese).

321 Westerberg, I. K., Guerrero, J. L., Younger, P. M., Beven, K. J., Seibert, J., Halldin, S., Freer, J.
322   E. and Xu, C. Y.: Calibration of hydrological models using flow-duration curves, Hydrol. Earth
323   Syst. Sci., 15 (7), 2205-2227, 2011.

324 Yilmaz, K. K., Gupta, H. V. and Wagener, T.: A process-based diagnostic approach to model
325   evaluation: Application to the NWS distributed hydrologic model, Water Resour. Res.,
326   44,W09417, doi:10.1029/2007WR006716, 2008.

327 Zhang, Z. X., Wagener, T., Reed, P. and Bhushan, R.: Reducing uncertainty in predictions in
328   ungauged basins by combining hydrologic indices regionalization and multiobjective
329   optimization, Water Resour. Res., 44 (W00B04), doi:10.1029/2008WR006833, 2008.