Reviewer 3:

1. General comments: The study of He and his co-authors presents novel insights into tracer-based hydrograph separation using a comparative approach of evaluating traditional against Bayesian EMMA. In this context, the study aims at filling this important research gap in tracer hydrology both from a methodological and process-oriented point of view. The study shows that the Bayesian approach estimates smaller uncertainties and is less sensitive to sampling uncertainties. The study approach also accounts for isotope fractionation, when using EMMA. Beside only minor comments, I think that the study is mature and presents a concise story line to the readership. The references are with up-to-date and a good use of English can be attributed. After revision of few comments, I can recommend this manuscript for further acceptance in this journal.

Reply: Thanks a lot for the positive comments. We have addressed all your concerns in this revised manuscript.

2. Page 6, Line 153: Please use the PALMEX reference (see below).

Reply: Done. Thanks.

3. Page 6, Line 175: Please clarify if the measurement precision is the same for both LGR and Picarro instruments, otherwise add this details.

Reply: Both measurement precisions of $\delta^{18}O$ and $\delta^{2}H$ are ±0.25 ‰ and ±0.4 ‰, respectively. Specified in the revised manuscript. See line 207.

4. Page 6, Line 178: How did you define 'obvious evaporation'? Did you use a deuterium excess threshold? Please insert further details here. Please add also at which EC limit you discarded samples.

Reply: We used threshold values to identify abnormal values of $\delta^{18}O$ and EC located far away from the sample clusters. For $\delta^{18}O$, sample values higher than 5‰ were excluded. For EC, sample values higher than 210 μs/cm were excluded. We specified that in the revised manuscript. See lines 214-217.

5. Page 6, Line 181: Please correct to 'cold season'; Page 15, Line 438: 'In average'.

Reply: Done. Thanks.

6. Page 8, Line 225: Eqs. 1 -5 hold for 3-components and 2-tracer mixing models. Please provide further information on how you inferred 4 components using 3 tracers.

Reply: When quantifying four runoff components using three tracers, four conservative equations for $\delta^{18}O$, $\delta^{2}H$, EC, and water volume are used (similar to Eq.1). The values of $\delta^{18}O$ and $\delta^{2}H$ are typically correlated for each runoff component. However, the coefficients representing the correlation between $\delta^{18}O$ and $\delta^{2}H$ vary among the runoff components in glacierized catchment (see Figure 2), thus providing a basis for the TEMMA_4 to quantify four runoff components using four conservation equations. The contributions of runoff components (f_i) as well as the partial derivatives used to calculate the uncertainty are solved from the four conservative equations using Matlab. However, the solutions are too lengthy to show in the text. We specified these in the revised manuscript. See lines 267-274.

7. Page 10, Line 293 – 295: Why did you not analyze the snowmelt uncertainty in the snowmelt period? Besides, the sentence is not clear to me: snowmelt is indeed more difficult to sample in the glacier melt season but easier to sample in the snowmelt period. Also its spatio-temporal variability is much higher in that period of time when most of the melting occurs.

Reply: We investigated the effects of sampling uncertainty only in the glacier melt season because of the following two reasons: (1) Runoff in the glacier melt season contributes the largest part to annual runoff in our study basin. Accurate quantification of each runoff component in this season is extremely important for the understanding of dynamics of water availability in the study area. (2) In this season more meltwater samples are available (15 snowmelt samples and 23 glacier melt samples) than in the snowmelt season (only 15 snowmelt, Table 1), thus providing a good observation data basis for the investigation experiment. Snowmelt sampling in the snowmelt season in the study basin is also difficult due to the heavy snow accumulation in March to April and the spring flood in May to June. However, we believe the effects of snowmelt sampling uncertainty on the end-member mixing approaches in the snowmelt season should be similar to those of meltwater sampling in the glacier melt season. We explained this issue in the revised manuscript. See lines 368-375.

8. Page 11, Line 308: Please provide more information on the fractionation effect and how you represented it in your analysis.

Reply: The water sources for runoff, such as rainfall and meltwater, are subject to evaporation before reaching the basin outlet, especially in summer. However, the isotopic composition of stream water was measured at the basin outlet, and the contributions of runoff components are *quantified for the total runoff at the basin outlet. After the long routing path from the sampled sites to the basin outlet, the isotopic compositions of rainfall and meltwater mixing at the basin outlet could be different from those measured at the sampled sites, caused by the evaporation fractionation effect. The isotopic composition of water sources at the sample sites are assumed to be normally distributed in Eqs. 6-7, and the changes in the isotopic compositions of water sources caused by the evaporation fractionation effect are represented by the modification variables* $\xi^{18}O$ and $\xi^{2}H$ *in Eq. 10. Parameters describing the prior distributions of isotopic compositions at the sample sites in Eqs. 6-7 are estimated by the likelihood observations of isotope signatures of water samples. The modification variables* $\xi^{18}O$ *and* $\xi^{2}H$ *are estimated by the likelihood observations of isotope signatures of stream water. The fractionation effect on the estimated CRC is quantified by comparing two Bayesian scenarios. In the first scenario (using Bayesian_3_Cor and Bayesain_4_Cor), the isotopic compositions of water sources at the basin outlet are assumed the same as those measured from the sample sites even though the water sources have suffered evaporation before reaching the basin outlet (using Eqs. 6-9). In the second scenario (using Bayesian_3_Cor_F and Bayesian_4_Cor_F), the evaporation fractionation effect on the isotopic compositions of water sources is considered, and the mixing of water tracers of stream water is represented by Eq.10. We added these explains in the revised manuscript. See lines 377-393. Figure 9 illustrates the effects of fractionation on the quantification of runoff components in all three seasons. The estimated changes in* $\delta^{18}O$ *of each water source are presented in Figs. 9a-c, and the contributions of runoff components quantified by the two scenarios are compared in Figs. 9d-f*

9. .Page 11, Line 319: It seems that this sentence contradicts with the one in line 326-328. How can glacier melt have high EC if it has low interaction with mineralized surfaces? Please rephrase both parts accordingly.

Reply: Line 319 has been modified as: "Among the water sources, snowmelt and glacier melt tend to have the lowest EC values, due to low interaction with mineral surface." Lines 326-328 have been rephrased as: "The highest CV value of EC for glacier melt indicates large variability in the glacier melt samples. This is because the glacier melt water samples were collected from a rather clean location (EC value is only 1.5 μs/cm) and a relatively dusty location (EC value is 33.4 μs/cm)." See lines 411-412 and 425-428.

10. Page 14, Line 379 –381: This sentence should be moved to the discussion part.

Reply: Modified as: "The TEMMA_3 estimated the largest uncertainty ranges and Sd values for CRC in all the three seasons, followed by the Bayesian_3."

11. Page 16, Line 469: Please clarify. How can samples taken occasionally lead to sharp changes of the isotopic composition? Moreover, randomly taken samples may be part of a strategy to represent tracer variability.

Reply: Modified as "Due to the limited accessibility of the sample sites caused by snow cover, the water samples of meltwater and groundwater are often collected sporadically. The small sample size and strong variability in sampled tracers likely lead to a large Sd value." See lines 566-568.