Reviewer 2:

1. This is a very interesting and well written manuscript that compares the traditional tracer-based end-member mixing model approach with different versions of a Bayesian mixing model to quantify water sources to runoff in a glacierized catchment in Kyrgyzstan. The findings of this work may have practical implications when applying these approaches to other catchments and are therefore surely interesting to the readers of HESS. The manuscript is logically organized, it is nicely illustrated, the interpretation is well supported by the data, and the discussion is coherent and with relevant and updated references. However, there are some moderate and minor issues that need to be clarified and that I invite the Authors to consider. Please, find these comments, suggestions, and a few corrections in the attached annotated manuscript. I hope they can be useful to the Authors to improve their work.

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

2. Lines 29 and 143: 'water tracer' to 'tracers' or 'hydrological tracers'; line 38: 'were' to 'was'; line 181: 'clod' to 'cold'; line 418: 'show' to 'shows'; line 490: 'rely' to 'relies'; line 726: Change the sentence into "CV stands for coefficient of variation"; line 734: 'snowmlet' to 'snowmelt'.

Reply: All done, thanks.

3. Lines 37 and 57: No need to make up a new acronym 'TEMMA'. EMMA is enough, there is no risk to confound it with the other approach.

Reply: The used Bayesian method is also a type of end-member mixing approach (EMMA). To avoid the confusion, we used TEMMA to represent the traditional end-member mixing approach.

4. Line 70: These are sources of uncertainty that are important in any catchment, not necessarily glacierized catchments. Please, specify why the latter are particularly prone to difficult application of HS (e.g., multiple water sources, high spatio-temporal variability of water sources etc.).

Reply: The glacierized catchments are challenging for application of the end-member mixing approaches because of the following reasons: (1) The catchment elevation generally extends over a large range, leading to strong spatial variability in climate forcing (precipitation and *temperature) and the tracer signatures of water sources; (2) The number of end-member water sources for runoff is high, additionally including snow and glacier meltwater; (3) Water sampling in high-elevation glacierized catchment is difficult due to the logistical limitations, resulting in small sample sizes for the application of (T)EMMA. We specified these in the revised manuscript. See lines 67-73.*

5. Line 77: But only the statistical uncertainty! Please, specify.

Reply: Specified as the "statistical uncertainty" in this manuscript.

6. Lines 83-87: This two issues are important but not very clearly explain. Please, clarify.

Reply: We refined these sentences as follows: These include (1) inappropriate estimation of the variability of tracer signatures of water sources when only few water samples are available. The used Sd values of the measured tracer signatures likely fail to represent the variability of water tracer signature of individual water source across the basin, due to the small water sample sizes; (2) The correlation of tracer signatures and runoff components are inevitably ignored, due to the assumption of independence of the multiple uncertainty sources. The correlation between $\delta^{18}O$ and $\delta^{2}H$ of each water source, as well as the interaction between runoff components could provide additional constraints on the uncertainty in the quantification of runoff components, which however are typically ignored in the Gaussian error propagation technique. See lines 88-97.

7. Line 93: In this paragraph it's important, in my opinion, to add a description on how uncertainty is treated in the Bayesian approach. This is particularly important for the research question #2.

Reply: In the Bayesian approach, both the statistical and model uncertainty are represented by the posterior distributions of parameters. The parameter uncertainty is estimated based on likelihood observations using a Markov Chain Monte Carlo procedure. This explanation has been added in the revised manuscript. See lines 106-109.

8. Line 109: How do Bayesian mixing models estimate the isotopic fractionation? I suggest to add a sentence here.

Reply: Modified as "Benefiting from the prior assumptions for changes in isotope signatures during the mixing process, the Bayesian approach bears the potential to estimate the fractionation effect on isotopic signatures, which however, has not been investigated either." See lines 122-124.

9. Line 113: In the two research questions outlined here it is not adequately stressed/explained why a glacierized catchment has been chosen for addressing these questions. Indeed, they can be applied to any catchment. Please, specify this.

Reply: We added a more detailed explanation here: "In Central Asia, glacierized catchments provide important fresh water supply for downstream cities and irrigated agriculture. Quantifying the contributions of multiple runoff components to total runoff is important for understanding the dynamics of water resource availability at the regional scale. However, uncertainty in the quantification of runoff components in the glacierized catchments are particularly large because of the following reasons: (1) The catchment elevation generally extends over a large range, leading to strong spatial variability in climate forces (precipitation and temperature) and the tracer signatures of water sources; (2) The number of end-member water sources is large, additionally including snow and glacier meltwater; (3)Water sampling in high-elevation glacierized catchments is difficult due to the logistical limitations, resulting in small sample sizes to represent the tracer signatures of water sources." See lines 127-131.

10. Line 143: As we know, EC is not as conservative as tracers. However, due to its easy use it has been often applied in catchment studies. Please, include a short discussion on the possible issue related to the lack of conservative behaviour (e.g., not so relevant at the catchmen scale, or at the runoff event scale etc.)

Reply: We added related discussion on this issue as follows: "EC data has been widely used for hydrograph separation, due to its easy use and quick measurement. While EC is not a conservative tracer, this may have only small effects on the application of hydrograph separation at the catchment scale." See lines 210-213.

11. Line 175: Any procedure to minimize memory effect (carry over effect) was performed?

Reply: Added: "A regular re-calibration procedure has been carried out for the isotope analysis." See line 206.

12. Line 176: First time it's mentioned...define electrical conductivity.

Reply: Defined in line 61.

13. Line 177: Can you quantify the term "abnormal"?

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.*

14. Line 227: It's not clear to me how 4-component HS can be performed using two tracers only. Indeed, due to the collinearity of 18oxygen and deuterium, these two tracers cannot be treated independently. So, how are mixing approaches TEMMA4, Bay4 and Bay4cor defined? Please, this parts need to be extremely clear to the readers.

*Reply: Yes, the values of $\delta^{18}O$ and $\delta^{2}H$ are typically correlated for each water source. However, the coefficients representing the correlation between $\delta^{18}O$ and $\delta^{2}H$ vary among the water sources in glacierized catchment (see Fig. 2), thus providing a basis for the TEMMA_4 to quantify four runoff components. 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 contributions of runoff components (**f**), 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. As expected, results in Table 4 show that the TEMMA_4 failed to distinguish snowmelt and glacier melt runoff, due to the similar tracer signatures of these two runoff components, but succeeded in quantifying the contributions of rainfall and groundwater. The Bayesian_4 and Bayesian_4_Cor estimated the contributions of four runoff components based on the prior distributions of $\delta^{18}O$, $\delta^{2}H$ and EC. The correlation between $\delta^{18}O$ and $\delta^{2}H$ is ignored in Bayesian_4. We used independent prior distributions for $\delta^{18}O$ and $\delta^{2}H$ of each water source. In Bayesian_4_Cor, parameters describing the correlation between $\delta^{18}O$ and $\delta^{2}H$ of each water source were estimated by likelihood observations of the corresponding water source, which also vary among the water sources, thus providing a basis for the quantification of four runoff components using four mixing equations of tracer signatures (similar to Eq.9). The four-components approaches are developed in our study to investigate the following two questions: (1) Is the TEMMA able to quantify four runoff components just using $\delta^{18}O$, $\delta^{2}H$, and EC? (2) Does the correlation between $\delta^{18}O$ and $\delta^{2}H$ help to reduce the uncertainty in the quantification of runoff components? We added these explains in the revised manuscript. See lines 267-274 and 337-346.*

15. Line 288: The three scenarios are not immediately clear. Does the mean refer to the spatial value or the temporal value, or the spatial-temporal value? The same question applies to sd. Then, different compared to what? Please, specify.

Reply: Meltwater sampling in glacierized catchments is typically difficult due to the logistic limitations. Thus, a small number of samples from a few sites are usually used for hydrograph separation. The uncertainty in the representativeness of meltwater samples implies an additional uncertainty source for quantification of runoff components. To investigate the effects of this type of sampling uncertainty, we set up three virtual sampling scenarios. Scenario I is used to evaluate the effects of meltwater sample size, in which four groups of meltwater sample are tested. The four sample groups have the same mean value and Sd of $\delta^{18}O$ or EC, but different sample sizes. Mean and Sd values of $\delta^{18}O$ or EC are calculated for all used meltwater samples in each group, referring to the spatio-temporal variability (same in the following two scenarios). Scenario II is used to evaluate the effects of sampled mean value of $\delta^{18}O$ (or EC) of meltwater. The four sample groups have the same sample size and Sd, but different mean values. Scenario III is used to investigate the effects of Sd values of sampled $\delta^{18}O$ (or EC). The four sample groups have the same sample size and mean tracer signature, but different Sd values. See lines 348-362.

16. Line 330: This is not clearly understandable from the table. Consider replacing it with a boxplot.

Reply: Done. See Figs. 3 in the revised manuscript.

17. Line 346: So, do the bars represent the spatio-temporal standard deviation?

Reply: The bars just represent the minimum and maximum values of each tracer signature.

18. Line 356: This sentence is not clear. Please, specify.

Reply: Modified as "Tracer signatures of rainfall are assumed as the same as the tracer signatures of precipitation samples in all the three seasons". See line 227.

19. Line 466: This holds true for this specific study and perhaps for other catchments (not only glaicerized) but not necessarily for all. This should be noted in the discussion.

Reply: Modified as "Sd values are likely overestimated in this study due to small sample sizes, and thus insufficiently representing the variability of the tracer signatures of the corresponding water sources across the basin." See lines 564-566.

20. Line 469: Sampling occasionally not necessarily lead to sharp changes! Please, explain.

Reply: Modified as "Due to the limited accessibility of the sampled sites caused by snow cover, the 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.

21. Table 1: This table is quite long and dense. Please, consider replacing it with box-plots.

Reply: This table has been split into three sub-tables. Boxplots have been added to present the variability of tracer signatures.

22. Table 4: Perhaps reporting the mean and the SD is clearer than reporting the mean and the range. Please, consider this possible change.

Reply: The ranges of minimum and maximum contributions are used to represent the uncertainty ranges. Sd values have been added in the table. See Table 4.