This is a well-written and interesting manuscript, although it mostly confirms what has been known for quite a while: Most hydrological models do a reasonable job in predicting water fluxes, but are pretty bad at quantifying source contributions. Rainfall-runoff modelling is “easy”, because it converts a known rainfall input to an output of streamflow (with some modifications and buffering). Thus, a model may be able to predict a streamflow response, but that does not mean that the processes are inferred correctly because they cannot be constrained with the available hydrometric data (I believe that this was discussed in a commentary by James Kirchner in 2006, for example, but there are others, including those cited in the discussion section of this manuscript). This manuscript does have value, as it assesses this question rather systematically, and without increasing the dimensionality of the model substantially. However, a more thorough description of the state of the art should be included in the introduction.

I also want to echo the comment made by the first reviewer, regarding the lacking description of the tracer data. Having more information on solutes and isotopes collected is essential to be able to assess the validity of the approach. How many samples were used to quantify the end-member concentration for each period? What is the analytical uncertainty? These are important questions and would allow assigning uncertainties to the calculated end-member contributions (Table 1), which are currently lacking. I noted the reference to Payeur-Poirier, but some of the information is important enough to be repeated in this manuscript. The contributions from different end-members can be highly uncertain, but whether this uncertainty can be captured depends on the type and amount of sampling. It appears to me (but I cannot be certain since the information is not provided) that only one sample was used to quantify the end-member concentration of each period (and hence the lacking uncertainty in end-member contributions). This is rather problematic, because spatial variability in concentrations within the same end member can be large (for example, Kendall et al., 2001). Considering the changes in vegetation in the catchment (from coniferous trees in the lower parts to deciduous trees in the upper parts) it is likely that soils and weathering profiles differ, and that shallow groundwater concentrations are thus heterogeneous in the catchment.

Finally, if I understand the manuscript correctly, only the hydrometric part of the model was validated, because only streamflow data was available for the validation period, but no tracer data and thus no end-member information. If this is the case, this is a major caveat, and should be pointed out clearly in the manuscript. This means that you can tune the model to re-create the observed end-member contributions, but there is no certainty whether it can actually predict the processes occurring. (Being modellers, I am certain the authors understand the implications much better than me)

Specific comments:

Page 1, line 21: “using a simple framework” Can you be more specific here? This leaves the reader wondering what was done. Something like “Using a modified version of the HVB model” might be better.

Page 4, line 11: This is the only place where you mention sulphate. Why was it not used for the further analysis?

Page 4, line 17: the fourth assumption of conservative behavior is rather dubious for some of the tracers (e.g. nitrate), that may behave quite differently along different parts of the flowpaths, depending on the absence or presence of oxygen. This should be pointed out.

Page 5, line 1: (1) If you had sufficient samples to calculate a mean, you could also quantify the uncertainty of each end-member contribution. (2) I question the approach of using mean end-member contributions. Figure 5 shows highly variable end-member contributions during some of the periods, casting doubt on the validity of using mean values. Why did you not simply assign end-members to sampling times and fit to that, rather than artificially defining different periods?

Page 6, line 14: "the first storage" Is this actually correct, i.e. is it really the reservoir marked "1" in Figure 2 and not the soil storage above?

The description of the model uses the term "soil storage" quite frequently. It would be helpful to identify this (I assume this is the upper box with dashed lines) in Figure 2. Adding a short explanation in the figure caption of reservoirs 1 & 2 might be helpful.

Page 7 line 17 – page 8 line 2: Do I understand this correctly, that you cannot validate the tracer-part of the model (because there is no tracer data for the validation period), and thus are only validating the discharge model? If this is the case, please state so explicitly, as this is a major caveat.

Page 14, line 2: "and streamflow components". I assume this refers to end-member contributions. If that assumption is correct, then this directly contradicts page 8, lines 1-2 where it is stated that for 2014, no tracer data was available. Could you elaborate on what you exactly did during the validation period and how you assessed the model performance during this period?

Page 14, lines 18-19: This is not true if you account for the uncertainty in the end-member contributions.

Page 16, lines 8-10: "The uncertainty … show considerable uncertainty…" This sentence is not overly clear. Also, are you referring to uncertainty or variability here?

References:

Kendall, C., McDonnell, J.J. and Gu, W., 2001. A look inside 'black box'hydrograph separation models: a study at the Hydrohill catchment. Hydrological Processes, 15(10), pp.1877-1902.

Kirchner, J.W., 2006. Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology. Water Resources Research, 42(3).