Dear Referee #1,

thanks for your constructive and detailed feedback on the manuscript hess-2021-213. By considering your points I believe that the quality of the manuscript will be further enhanced.

Please find my reply to your comments below.

> Although the title and abstract do not mention this, an important contribution is the matrix formulation Ax=b for the author's method (equation 20 and appendix B), which is much easier to understand, and much more readily extensible, than the equations in the 2019 paper. The application of the condition number, however, is not the conventional one (equation 22), but one that is much less widely known (equation 23). Unfortunately this is introduced with the ambiguous phrase "a similar estimation can be derived" (how? by whom?), leaving readers to wonder where this comes from – a reference is sorely needed. Likewise equation 24 is introduced by "it can be shown that" (how? by whom?) and readers are given no clue what "f" is, making the equation uninterpretable.

The reference for equation 23, which describes the relative error in vector x in case of small disturbances in the coefficients of matrix A, would be Stoer (1994). It is true that equation 23 is not that often mentioned in the literature, maybe I can replace it by a more popular form that would be:

\[
\frac{\text{norm}(\Delta x)}{\text{norm}(x)} \leq \frac{\text{cond}(A)}{1-\text{cond}(A)\text{norm}(\Delta A)/\text{norm}(A))\text{norm}(\Delta A)/\text{norm}(A)}
\]

Equation 24 can be found at Deuflhard and Hohmann (2019) and holds for disturbances in matrix A, in which f'(A) is the derivation of matrix A, also known as the Jacobian matrix. With equation 24, I wanted to emphasize that the norm of the first derivative of the matrix A is also limited by its condition. This is interesting in terms of the Gaussian error propagation, in which the entries of f'(A) are applied to the uncertainties of the known variables. This way, I wanted to point out analytically that a basic Gaussian error analysis of the system A*x = b would not produce different results than an analysis based on the condition number. Certainly, I could better describe this in the corresponding section. However, in section 4.2, I have written that an "ill-conditioned system will lead to larger gradients in a Jacobian matrix and to potentially higher errors in a Gaussian error
propagation”.

> More generally, there would seem to be three important (and possibly disqualifying) limitations in the use of the condition number as a guide to the reliability of this analysis. First, equation 23 holds only for infinitesimal errors in the matrix A, but there is no guarantee that realistic errors in the tracer concentrations (which, along with 1’s and -1’s, make up the non-zero elements of A) are small enough that equation 23 is still realistic. Second, it is hard to see how equation 23 can give useful guidance in the case of equation 20, where some elements of the matrix A are dimensionless constants (1’s and -1’s), and others are dimensional quantities (the tracer concentrations), so their relative magnitudes (and thus which ones have greater influence on the matrix norm) are dependent on the arbitrary choice of units for the tracer concentrations. Since the 1’s and -1’s have no uncertainty (and thus their corresponding elements of deltaA will be zero), it would seem that we could make the ratio of norm(deltaA)/norm(A) as big or small as we want (and thus get any answer that we want from equation 23) just by changing the units of the tracer concentrations. Third, it is not clear how this works when the vector b (which will be multiplied by the inverse of A) is mostly composed of zeroes, as it is here. This part of the paper badly needs a numerical demonstration, with realistic input values (and realistic errors in those input values).

As indicated already in the Introduction section, the condition number is used only as a measure that describes how strongly an input error can affect the calculated output in the worst case. The condition number is not used to calculate absolute or relative errors, maybe I should even better emphasize this in the text. Instead, I exactly calculate absolute and relative errors on the basis of a comparison between the separated event water response with the simulated event water response for different test scenarios in the results section 3. The absolute errors Δ [mm/h] and relative errors Δ [%] are visualized and described for every test scenario. My intention for using the condition number is to evaluate the mathematical constraints that arise directly from the applied linear equation system 20 (A*x=b) by detecting ill-conditioned and well-conditioned situations. It is a very elementary approach, but necessary for every kind of mathematical model, since there is a point at which even exact knowledge of physical input parameters might not lead to reliable results any more. Imagine a catchment in which the bulk mass flow and isotope concentration of water would be exactly known at each point in time. Then, in case of an ill-conditioned linear equation system A*x=b, the proposed iterative separation method might not produce reliable event and pre-event water fractions any more, regardless of any other factor. I have mentioned this in section 2.3 “Condition number and error estimation”. Later in the results section 3, I have shown this by simply violating Criterion 1.

> The benchmark tests presented here assume (as far as I can tell) that there are no errors in the event and pre-event tracer concentrations. This is fundamentally unrealistic, and makes the estimates of the errors in the event water fractions (Table A1, Figures A4, A5, A7, A8, A10, A11, A13, A14, A16, A17, A19, A20, A22) meaningless as guides to the real-world reliability of the method. By contrast, the author’s previous paper (Hoeg, 2019) showed relative errors of up to 100% or more in event water fractions estimated from real-world data from an experimental catchment. A realistic analysis requires realistic simulated errors in the tracer concentrations. These errors go well beyond the analytical errors in the measurement, and should include the likely sampling errors (i.e., the rainfall that is sampled may not be the average of the rain that falls over the whole landscape). In any case, the errors are certainly not zero, and it is not helpful to assume that they are.

> The benchmark tests presented here also assume that there is no isotopic fractionation of either the event water or the pre-event water. This again makes the results unrealistic as guides to what one might expect in the real world. In the real world, evapotranspiration (including interception losses) is often the dominant term in the water balance (rather
than zero, as assumed here), and can significantly alter the isotopic composition of the water reaching the surface, relative to the sampled precipitation, and may also alter the composition of the pre-event water over time. Any change in the isotopic composition of either the event water or the pre-event water would seem to pose serious challenges for the approach presented here.

I agree that the paper (Hoeg, 2019) appears far more descriptive and familiar to experimental researchers, because it is based on data from a field study. Even I personally prefer analyzing experimental lab or field data. However, in the current study I present the results of an elementary benchmark test for the iterative separation model introduced in the year 2019, which from my point of view is an absolutely necessary exercise. In the discussion section "Design and applicability of the rainfall-runoff model", I have described why the applied synthetically generated data do not reflect the complexity of a natural hydrological system.

> The benchmark simulations are unrealistic in other ways as well. The precipitation events are large and regularly spaced, with very long rainless intervals in between, and the event water fractions are large compared to those that are typically observed in many real-world studies (including the author's 2019 study). And the behavior of the benchmark model itself is unrealistically simple; since it consists of two linear reservoirs with a constant partitioning coefficient, the forward transit time distributions of all precipitation events are identical. Readers would be more confident in the results if they were based on a benchmark model that is nonlinear and nonstationary, as real-world hydrologic systems are.

The study validates the iterative separation method on the basis of numerical simulations. The applied rainfall-runoff model has one important advantage, it enables an exact determination of the event water response across an arbitrary number of subsequent rainfall runoff events. Therefore, the focus of the current investigation can be on the capabilities of the iterative separation method, and not on the exactness of the benchmark model.

I am aware that hydrologists prefer realistic rainfall-runoff models, but I am sure that they also prefer reliable separation methods. Here, I had to find a suitable compromise and have chosen a rather elementary setup that will be replicable for everyone. Actually, the events are not regularly spaced in section 3.2 "Random rainfall and 18O input", but they are indeed regularly spaced for technical reasons in section 3.3 "Random rainfall and 18O input and delayed response of event water". If desired, I can put more randomness in section 3.2 regarding the event intervals and the partition coefficient, it will not touch the outcome in the results section. However, in section 3.3, random event intervals would definitely lead to different results, since for technical reasons the rapid mobilization of pre-event water would not be in sync with the rainfall impulses any more, which from my point view would appear physically unrealistic.

> It seems that in some ways the benchmark tests have been designed to conform to the assumptions of the method. But to the extent that this is the case, the benchmark tests only show that the method would work in a world that conformed to the method's assumptions. Readers will be far more interested in whether the method is reliable in the real world, which requires benchmark tests under more realistic conditions.

I have done thousands of simulations with very different parameter sets. For the manuscript, I have picked some examples to demonstrate where the iterative separation model shows good results, but also where it should not be applied from a mathematical and numerical point of view. Maybe I should make this more clear in the conclusions. One the one hand, it is a violation of the model criteria, on the other hand, it may be large fractions of rapidly mobilized pre-event water. The latter was also a surprising result for
me. Nevertheless, on the basis of Niemis theorem it is then possible to find the exact event water response by applying the corresponding age functions, respectively recalculated tracer concentrations. However, in the field practice, where the event water response, respectively travel time distribution, is normally unknown, this might require the use of Monte Carlo simulations.

> The method is presented as being "based on an iterative balance of catchment input and output mass flows along the time axis" (line 84). This is not the case. As with conventional hydrograph separation, there is no mass balance of inputs and outputs in the sense of equations 1 and 2, but just conservative mixing of the event water and pre-event water.

I think as long as all the criteria (1-4) are fulfilled, the above sentence should hold. For instance, I could reword the sentence to "based on an iterative balance of catchment event and pre-event mass flows along the time axis"?

> It is odd to see all the figures put into an appendix, but on the other hand they are too numerous and repetitive to all go in the main text. Presumably this could be straightened out in an eventual revision.

If some of the figures might appear too redundant, then I could certainly thin out a bit.

> This review has not considered the additional material that has been supplied as author comments, because the HESS review process is – as far as I know – based on the assumption that submitted manuscripts are complete and final, not drafts that are still undergoing revision.

This is fine for me. Thank you for your feedback.

With kind regards,
Simon Hoeg