Comment on hess-2021-213
Anonymous Referee #1

Referee comment on "Benchmark tests for separating n time components of runoff with one stable isotope tracer" by Simon Hoeg, Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2021-213-RC1, 2021

This manuscript presents, as the title makes clear, benchmark tests of the author's approach (originally proposed in WRR in 2019) for identifying the fractions of runoff from several previous rainfall inputs, using a single isotope tracer.

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.

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).

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.

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.

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.

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.

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.

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.