

My thanks to Reviewer #4 for a detailed range of comments. The various reviewer comments are set out below in italic, responses follow in plain text.

The manuscript does not indicate much familiarity with the extensive literature on transit time estimation, or with the fundamentals of transit time models.

There is certainly an extensive literature on transit time estimation and transit time models. The omission is deliberate. Such analyses are invariably parametric and model-based, and review of that large literature would be peripheral to the nonparametric model-free theme of a short technical note. My personal view is that for brief communications of this type references should strictly focus only on the relevant, even though inevitably a number of reviewer publications may not be cited.

The proposed method may be a useful contribution if a) its underlying assumptions can be shown to be correct (at least to a good enough approximation), or if, nevertheless, b) its results can be shown to be robust and reliable under realistic benchmark tests, which will include both realistic data errors and deviations of the real world conditions from the idealized assumptions of the method.

Unfortunately, condition (a) is not met, because the behavior of real-world catchments is widely recognized to be inconsistent with the linearity assumptions that underlie the method presented, and the illustrative example presented in section 5 falls far short of the requirements of condition (b). Specific comments on each section follow.

With respect to (a), real-world transit time distributions do always exist, regardless of their variability with time and the nature linear or nonlinear hydrological processes which give rise to them. In its most general form as a time sequence of arbitrary discrete transit time histograms, the nonparametric model can conceptually match any possible time sequence of transit time distributions within the range of the discrete histograms, while not explicitly modelling any nonlinear hydrological process. The lower bound should therefore still be applicable because any system of constraints incorporated to reflect nonlinear processes would have the effect of making the model less flexible and the lower bound would then be greater than when the minimisation process is applied to the unconstrained model. In practice there will probably need to be some realistic constraints applied to time-varying nonparametric histograms such as restricting the degree to which the nonparametric distribution means are permitted to differ from one distribution to the next in sequence.

With respect to (b) there is no argument that a more realistic illustrative example is required – also noted also in responses to other reviewers. As was also noted in other responses, however, the aim of the brief technical note is not to carry out exhaustive evaluations to “prove” the applicability of the method. Rather, the intention is simply to use (inevitably somewhat idealised) examples to encourage individuals with local catchment knowledge to evaluate the lower bound approach when applied to their recorded data sets. Local knowledge is a critical factor because the greater the degree of constraints that can be placed on the histogram sequence, the higher the lower bound will be for a given degree of fit.

While it is true that equal probability of exit is SUFFICIENT to generate an exponential transit time distribution, it is not the case that this is the ONLY way that such a distribution can arise (one can imagine, hypothetically, a series of flowpaths whose lengths are exponentially distributed, or an aquifer whose permeability decreases systematically with depth. In both cases one could obtain an exponential distribution without equal probabilities of exit for every particle in the system.

Equal probability of exit is in fact a necessary condition for a well-mixed water store to give an exponential transit time distribution. This is because the definition of a well-mixed water store is equal probability of exit.

The reference is the paper was clearly only to the lack of general applicability of the exponential distribution with respect to its justification in terms of well-mixed stores – which will only arise in rather restricted conditions. It is regretful however, that the explicit statement was not made in the paper that outside of the well-mixed model there

is also no theoretical justification in the real world for exponential transit time distributions – nor any other parametric transit time distribution for that matter.

Considering now the two supposed alternative exponential models proposed by the reviewer:

With respect to the flowpath model, the proposed series of flowpaths whose lengths are exponentially distributed is a special case of a stream tube model. It is indeed a *very* special case because it is required that nature should for some reason create a situation such that a large number of stream tubes should (i) have an exponential distribution of lengths, (ii) have exactly the same tracer particle speeds, and (iii) have zero dispersion of tracer particles regardless of length of flow. It is hard to conceive of a more contrived and unrealistic flow scheme. The transit time distributions in stream tube models are actually quite complex because of the combined effect of varying stream tube length distributions and varying dispersion characteristics (Bardsley, 2003).

With respect to the aquifer whose permeability “increases systematically with depth”, this represents an entirely undefined entity. What is the nature of the “systematic” increase in permeability that will give rise to a mathematical derivation of an exponential transit time distribution for an aquifer? No reference is offered in support of the assertion and it is worth noting that a strong hydrological argument has been made that exponential transit time distributions are not to be expected from aquifers (Kirchner, 2017).

With reference to “The statement about gamma distributions is likewise misplaced”, what is the particular statement that is supposedly misplaced? There seems no argument with the Nash Cascade gamma model – which gives of course integer values of the gamma shape parameter because any sum is composed of an integer number of components ≥ 1 .

Further aspects of the gamma distribution are also cited by the reviewer:

In contrast, the empirically determined shape factors among the 20 catchments analyzed by Godsey et al. (2010) ranged from about 0.3 to 0.8. None were even close to 1. Approximate gamma distributions with shape factors near 0.5, consistent with the Godsey et al. findings, can potentially arise from advection and dispersion with spatially distributed inputs (Kirchner et al. 2001).

The Godsey et al (2010) authors elected to fit gamma distributions to transit time data and obtained gamma distribution shape parameters from the fitting process in the range of about 0.3 to 0.8. The reviewer seems to attach much significance to these gamma shape parameter values as somehow supporting his assertion that my offending statement (which is presumably “gamma distributions do not warrant special consideration as transit-time distributions outside of idealised situations”) is misplaced.

In fact, whether gamma distribution shape parameters as obtained from fitting to data are much less than 1, near to 1, or much greater than 1 is irrelevant to any argument seeking to justify the gamma distribution. The cited reference (Kirchner et al. 2001) deals with a special case where a perfectly uniform slope segment with uniform tracer increments along its length gives rise to a particular form of mixed inverse Gaussian transit time distribution which has similarities to the gamma distribution with shape parameter < 1 . Even then the gamma model only holds for the special case of this special case where advection and dispersion are of roughly equal effectiveness in transporting the tracer to the stream. To upscale from such an ultra-idealised local model to argue that there is therefore theoretical justification for real-world catchments to have gamma-distributed transit times (of whatever shape parameter) is simply unrealistic. On the other hand, what can be usefully taken from the data of the Godsey et al (2010) paper is that catchment transit time distributions will often be L-shaped, though certainly not necessarily gamma-distributed. Given that a gamma shape parameter range of 0.3-0.8 corresponds to a skewness range of 2.2-3.7, there is no reason why any L-shaped probability distribution within this skewness range could not serve as well to describe the data.

As an aside unrelated to the paper, in the case of groundwater systems the usual approach to making first estimates of transit time distributions for groundwater systems would be by utilising particle tracking options in numerical groundwater models. It is surprising then that there should have been such longevity of analytical approaches to transit time distributions in the arguably more complex case of catchment hydrology. Perhaps there is scope for agent-based numerical catchment modelling to provide a means of simulating catchment transit time distributions.

The previous negative comments concerning parametric distributions do not of course imply that the nonparametric lower bound approach has value. That remains to be determined by application to real data. There is also clearly a need for any revised version of the paper to include an L-shaped transit time distribution, as noted by the reviewers.

Sections 2 and 3 (definitions and steady-state case):

From a purely theoretical standpoint I do not see any problems here, but I am not an expert in linear programming techniques. From a practical standpoint the steady-state case is relevant only as a (potentially rather poor) approximation to the real world.

Section 4 (non-steady-state case):

The time-varying case presented here, in which the shape of the transit time distribution can change but its mean must stay the same, is inconsistent with the entire literature on catchment nonstationarity. Even theoretically, it is very difficult to imagine any catchment that could possibly work this way. This section may be interesting from a mathematical standpoint but is completely irrelevant to the real world.

The gamma PDF's (with shape factor of 5) that are used to generate the test time series bear no resemblance to real-world catchment PDF's (with shape factor well below 1). There is no reason to assume that a method that works with such an unrealistic test time series will necessarily work with a more realistic one.

Other reviewers have made the same points. The steady state case was only included as it has been around for some time and, being a simple case, seemed a logical means to illustrate the linear programming approach. Similarly, the constant mean with varying distributions was just the next level of complexity up from that. Given reviewer comments, any revised paper would be best to focus just on nonstationary L-shaped transit time distributions with different forms and different means and see how they go. Obviously the lower bound approach needs to be robust enough to handle L-shaped transit time distributions if it is to have practical value.

The nonparametric TTD is unrealistically truncated at 23 months. Of course, given that the generating distributions – see (a) above – have trivial tails beyond 23 months, the truncation effects are small. But in the real world, where one cannot know this in advance, what could be the justification for not extending the nonparametric TTD to much longer lags? In that case, of course, the computations would become more difficult and, more importantly, the solutions would become much less constrained.

It has to be remembered that a lower bound to the mean is being sought and not the mean itself. The critical issue is not whether the true transit time distribution has a long tail beyond the upper bound to the nonparametric distribution, but whether the choice of upper bound will have any influence on determining the value of the lower bound to the transit time mean (for a given degree of data matching). This can only be determined after the minimisation process. This is noted also on p.9 (line 30). As can be seen from Fig 5 (c) (d), the lower bound to the mean would have been the same value if the nonparametric upper bound had been set to, say, 2,300 months instead of 23 months (but at the expense of much more calculation). These solutions are therefore not less constrained by increasing the upper bound to the nonparametric distribution.

The truncation of the nonparametric TTD automatically imposes bounds on the mean – between 0 and 23 months in theory, but in practice, with any dispersion (either physical or numerical), the range will be narrower and the tendency will be more toward the center. Thus it is guaranteed that the solution will not be that far from 6 or 12 months, which are known in advance to be the correct answers.

The tendency will certainly not be more toward the centre in general. As noted on page 9 (line 7) the single gamma transit time distribution behind the data was well-discovered because of the linear nature of the data simulation process, not because of some tendency toward the centre. In any revised version of the paper some long-tailed distributions will be included to illustrate that the lower bound to the mean may sometimes be much less than the true mean transit time.

The simulated time series are unrealistically long (not many sites have 21 years of tracer data).

A good point. An example with a shorter simulated time series should be included in a revision.

The simulated time series are unrealistically PERFECT. There are no measurement errors in either the input or output time series. Random numbers are added to the square wave in order to create the X values, but then the Y values are convolved with zero error, and then the X and Y values are used – with zero error – in the proposed LP inversion technique. This bears no resemblance to the real-world inversion problem, where both the inputs and outputs will have errors. Furthermore, those errors are potentially catastrophic for an poorly constrained inversion technique like the proposed LP method

In a revised version a measurement error will be added to both the input and output time series. It isn't obvious though that adding errors will have a "catastrophic effect" effect provided the errors are not too large. It will be interesting to see the outcome.

In summary, is such a contrived test case that it gives no useful information about the robustness or the practical utility of the proposed method. There is no guarantee that, in a more realistic test case, the proposed method would give a meaningful lower bound estimator. Indeed, the result could potentially even exceed the true mean.

As noted above, the proposed more realistic example should clarify such issues – at least to the extent that, if successful, others may be encouraged to try out the method.

References

Bardsley, W.E. 2003. Temporal moments of a tracer pulse in a perfectly parallel flow system *Advances in Water Resources* 26, 599–607.

Kirchner, J. 2017. Interactive comment on "Aggregation effects on tritium-based mean transit times and young water fractions in spatially heterogeneous catchments and groundwater systems, and implications for past and future applications of tritium" by M. K. Stewart et al. *Hydrol. Earth Syst. Sci. Discuss.*, doi:10.5194/hess-2016-532-SC4, 2017