Reply to Reviewer # 1

My thanks taking the time for the review. Reviewer comments are in italic and responses follow.

This paper proposes using a simple nonparametric linear programing approach to improve the actual used methods to obtain mean transit time, calculating instead the lower bounds of the mean transit time.

There is no suggestion of offering an improvement over any previous methods that have been used to obtain mean transit time. The paper is concerned only with an approach to estimating a lower bound to mean transit time. The lower bound in some cases might be orders of magnitude less than the true mean transit time.

I would like to clarify to the author that mean transit time and mean residence time are not the same.

I guess we would all agree that the mean residence time for a hotel is the average of between when guests check in and when they check out. A quick overview of the literature reveals that this definition of mean residence time presently extends over many different fields. However, mean transit time serves perfectly well for the purposes of the paper so the very few mentions of mean residence time will be removed.

Following the previous point, it grows the concern if the author has read enough of the available literature and methods proposed by other authors trying to obtain as well better transit time estimations.

As noted earlier, this technical note is not concerned with seeking better transit time estimation, but rather with proposing a model-independent nonparametric lower bound to mean transit time. That is, the aim is not to find where the mean is, but to define a region where it is not. There is certainly a proliferation of papers relating to transit time concepts, hydrological models leading to various transit time distributions, as well as papers relating to *ad hoc* transit time distributions like the gamma distribution. This extensive literature is not cited because it is unrelated directly to the specific issue of nonparametric lower bound estimation of mean transit time. Perhaps the best approach would be in the Introduction to direct readers to, for example, the cited references in Kirchner 2016a&b . A search for methods directly connected to lower bound estimation for mean transit times did not turn up anything in the hydrological literature. I would of course welcome notification of any such references in journals or texts from hydrology or other fields.

Would the author be able to use this approach to consider as well the water that was in the system previous to the input studied?

The analysis in the paper is not concerned with water as such but with some form of tracer particle passing into the system, observable as a given tracer input time series. A requirement of the method is that the tracer input data extends far enough back in time so that all tracer particles in the system can be mapped back in principle to the recorded tracer input time series. The method would not be applicable with significant amounts of tracer particles present that could not be related to the recorded tracer input time series.

As a suggestion for better understanding that the author might take or not. I think that some parameters could have more user friendly names. For example μ* *could have an L or low as sub-index. That would force to change the name of D to a subindex like 'u', 'upp' or 'h' from highest. These small changes could make smoother for the reader to follow up the terms coming up throughout the paper."*

My thanks for these suggestions – they will be incorporated in the event that the paper is advanced to accepted status.

I am positively impressed that with this approach there is no need of catching the tail while using a gamma distribution. That is an upgrade for the gamma distribution methods.

Many thanks for the support here. However, I am a little confused as to the meaning. A specific gamma distribution was used as an entirely arbitrary choice of a travel time distribution, but just for the purposes of simulating example data with a known mean transit time. However, the method itself is non-parametric and not in any way connected to the gamma distribution or any other parametric transit time distribution.

But it would be interesting as well to test different N values to see if it would provide different μ values. Meaning, is μ* dependent on the size of N chosen?*

As noted in the paper, the choice of *N* is not critical as long as it is sufficiently large not to have influence later in the minimization when seeking a lower bound for μ_T. In the LP minimising operation the probability distribution shifts to the left to minimize the distribution mean value, subject to the constraints. When the distribution thus created has probabilities that are zero in the upper range of the distribution, this confirms that increasing *N* will not affect the minimization outcome. This can be seen, for example, in Fig. 5 c & d.

Would it be possible for unknown catchments to know that a μ with* r=0.9 *is* 2/3 *of the mean transit time? Or was this just a casual coincidence?*

This is with respect to Fig. 4. Sadly, it is just a coincidence. Each real-world catchment will have its own characteristics and data quality can also influence *r*.

Page 2 Block 25: what are the X values with negative sub-indexes physically? Am I correct if I assume them to be the previous inputs to the input I am studying? If that would be the case, then it would be other precipitation as well? Or is it stream water, or groundwater? Let's assume the case where my X values with positive sub-index are precipitation values, would that mean that all my X values with negative sub-indexes are precipitation values from days prior to the positive ones?

The negative subscripts for X indicate the recorded time series of tracer input prior to the time of the first Y output variable. This prior period of off-lap is required because the first Y value (t =1) will be influenced by all the prior X values as far back as is determined by the transit time distribution. It seems convenient to have negative subscripts for the X values to emphasize the off-lap. The physical interpretation of the X values depends on the situation – they could, for example, be flux-weighted ^{18}O values obtained from a rainfall time series for the catchment. So, yes, the X values with negative indices are from prior to the positive ones.

P4 Eq 1: if we assume t=0 you would obtain an X-N which is not defined before

There is only a requirement for seeking to match the available set of Y values, that is Y_1 , Y_2, …Y_K and there is no Y_0 . That is, t in Eq 1 can be any of t = 1, 2, … K but there is no t = 0 involved in Eq.1. My thanks for pointing out the lack of clarity here. The potential for confusion will be avoided by having " $1 \le t \le K$ " to the right of Eq. (1).

P6 B15 eq 5: Same as before, with $t = \varepsilon$ there would be an $X_{\varepsilon\text{-}N}$ not defined, that without saying that none of all the X obtained in that equations are defined due to the epsilon. I understand the author takes them as very small, but the nomenclature is confusing since those are sub-indexes. Perhaps it would be better if it's properly stated that for this case we will assume that the indexes with ε are as if there was no ε.

Again, the issue has arisen because Eq. (5) should have had $1 \le t \le K$ inserted to the right. My apologies for this omission.

The nomenclature has been reworked somewhat (see suggested new text following the **** on the next page).

P6 B25: the progression of this sum is a bit confusing, I think it would go from $\omega_{K-\tau+1}$, $\omega_{K-\tau+2}$,: : : to ω_K.

This portion of the paper could certainly have been written more clearly. My suggestion for a re-write of both text and equations for this part of the paper are as below (there would also need to be a corresponding modification to Eq. (1) for consistency). Hopefully this new text and symbolism will read better:

****

Because the different $P(\tau)$ distributions have different time origins, a given discrete distribution is now symbolised as $P_i(\tau)$, where $0 \le \tau \le N$ as before. That is, $P_i(\tau)$ is the transit time distribution for tracer particles which arrive in the catchment at time i, where $1\text{-}N \le i \le K$. At a given time t the model-predicted tracer output $\hat{Y}_t$ at the recording site is given by:

$$\hat{Y}_t = \theta\sum_{\tau=0}^{N} X_{t-\tau} P_{t-\tau}(\tau) \quad = \sum_{\tau=0}^{N} X_{t-\tau}\,\omega_{t-\tau}(\tau) \qquad 1 \le t \le K \tag{5}$$

The data fit expression giving the least mean absolute deviation from the Y data is therefore found from the minimization operation:

$$\text{MINIMIZE} \quad K^{-1}\sum_{t=1}^{K}\left| Y_t - \sum_{\tau=0}^{N} X_{t-\tau}\,\omega_{t-\tau}(\tau)\right| \tag{6}$$

subject to the equality constraints:

$$\sum_{\tau=0}^{N} \omega_{1-N}(\tau) = \theta \quad , \quad \sum_{\tau=0}^{N} \omega_{2-N}(\tau) = \theta \quad , \quad \ldots \sum_{\tau=0}^{N} \omega_{0}(\tau) = \theta$$

$$\sum_{\tau=0}^{N} \omega_{1}(\tau) = \theta \ , \quad \sum_{\tau=0}^{N} \omega_{2}(\tau) = \theta \ , \ \ldots \sum_{\tau=0}^{N} \omega_{K}(\tau) = \theta$$

where the scale parameter θ is obtained from the minimization, which consists of finding numerical values for each $\omega_i(\tau)$ such that Eq. (6) is minimised. The individual $P_i(\tau)$ probabilities are then found by rescaling with θ as before. If an acceptable data fit cannot be achieved at this point then a more linear flexible model could be evaluated, allowing some degree of variation of the respective μ values.

****

P8 B25: I think the author meant β, instead of α in the text "The gamma scale parameter 'β' is specified: : :"

Yes – my thanks for picking the use of the wrong symbol.

P10 B0: Could the author explain better how is it that higher value on P(0) an illusion? If I were to not consider that value there would be a shift on the μ. I could not follow the author's idea.

What was meant here is that having a mode at P(0) is not a sensitive factor for determining the calculated value of the lower bound. For example, if additional constraints are imposed so there cannot be a mode at P(0) from the minimisation process, there is in fact minimal change to the value of the calculated lower bound in this instance. This is a good illustration for showing that the P(τ) values arising from the minimisation should be regarded as uninformative. Some text making the clarification will be added

Fig 3: X axis is missing units (months I assume)

Yes .. there are units missing as noted. The horizontal axis needs to read " τ (months) "

It would be nicer if Fig 1, Fig2 and Fig 4 are done for black white printing and as well for color blind readers.

Yes – the colour figures were for review purposes.