hess-2023-141 - Response to Reviewer Comments

Electrical conductivity fluctuations as a tracer to determine time-dependent transport characteristics in hyporheic sediments

by: Jonas L. Schaper, Olaf A. Cirpka, Joerg Lewandowski, and Christiane Zarfl

Corresponding author: Jonas Schaper; *jonas.schaper@uni-tuebingen.de*, Department of Geosciences, Eberhard Karls University of Tübingen, Tübingen, Germany

We thank all reviewers *and the editor* for their time spent on our manuscript and their constructive comments that considerably improved our manuscript.

General Remarks

In the following, you find our replies to the reviewer comments on the manuscript "Electrical conductivity fluctuations as a tracer to determine time-dependent transport characteristics" submitted as a research article to HESS.

The main concerns of the reviewers were related to

- statements on the transport nature in hyporheic sediments (Fickian vs. non-Fickian),
- the inclusion of an electric-conductivity (EC) offset term in the presented model,
- the way we determined weighting factors by visual inspection of the L-curve, and
- the general clarity and structure of the manuscript.

The main purpose of the manuscript is to discuss how the time series of a natural tracer, namely specific electric conductivity (EC), can be analyzed at shallow depths of hyporheic sediments. Hyporheic-exchange flows are known to be highly dynamic so that the assumption of stationary (that is, time-independent) travel-time distributions is not valid, at least not for short travel distances over which the velocity fluctuations don't average out. We thus present a method to estimate non-stationary (that is, time dependent) travel-time distributions between rivers and fairly shallow points in hyporheic sediments from time series of EC. The main target is the mean travel time which we parameterize via a continuous time-function of advective porewater velocity for a fixed travel distance equaling the depth of the observation point within the sediment. We parameterize the spread of the travel-time distribution by assuming a constant dispersivity. Choosing the advection-dispersion equation with spatially constant and temporally varying coefficients as model is solely a choice of parameterization. This does not imply that we believe 1-D Fickian transport to be fully correct. We are convinced that estimating time-varying (!) coefficients of travel-time distributions from natural-tracer time series beyond metrics of their mean and spread is unrealistic. For the latter, the input signals of river EC are already too smooth. We will thus remove all statements on a potentially Fickian nature of transport from the manuscript, as they are distracting the readers from the main message.

There appeared to be some confusion of the reviewers caused by our comparison to results of non-parametric deconvolution. The latter approach assumes a stationary travel-time distribution, so that flux transients cannot be resolved. Travel-time distributions derived by deconvolution convert comparably narrow time-dependent travel-time distributions to stationary, broad or multi-modal distributions. We decided to keep this comparison to discuss effects of transient fluxes on deconvolution-derived stationary travel-time distributions, as the latter – established – approach has gained some popularity in the hyporheic-research community. The key message is that a stationary approach leads to artifacts resulting from transient flow. We are sorry that we were not able to convey this message clear enough in the original submission.

Similarly, the offset in EC needed in our model seemed to have confused this particular group of reviewers. EC can be continuously measured at low costs, which makes it a popular natural tracer in rivers and hyporheic systems. But it is not perfectly conservative because some major ions that make up a substantial fraction of the EC signal (particularly calcium and bicarbonate) undergo precipitation/dissolution reactions that depend on temperature and small pH variations. This limitation of EC as natural tracer is well known, but it might have got lost in the introduction of the original manuscript. Anyway, when working with EC as natural tracer you have to deal with systematic offsets. Thus, the question is: how? In response to the reviewer comments, we performed a variety of additional model runs to investigate the effect of several types of trend models in the EC offset (constant, linear, one – and two knots per day used in spline interpolation), which we want to share in a revision of the manuscript. The key result is that reaction-related EC offsets between river water and pore water in the shallow hyporheic zone are unlikely to be constant in time.

Besides that, we intend to include a mathematical way of determining the optimal regularization weights, and improve the clarity and structure of the manuscript throughout. We trust that these improvements will help overcome concerns expressed by the reviewers and hope to get an opportunity to submit a revised version of the manuscript to HESS.

Reviewer 1

The study utilizes electrical conductivity (EC) as a natural tracer to evaluate water transport in the hyporheic zones of urban rivers in Germany and South Australia. By employing a time-dependent advection-dispersion equation (ADE) fitted to EC time series through Bayesian parameter inference, the research demonstrates that porewater velocities are highly variable, experiencing up to a six-fold increase within a 24-hour period. The study purports to validate the Fickian nature of transport in three out of four datasets, thereby affirming the applicability of ADE-based models. However, it recommends caution in interpreting Travel Time Distributions (TTDs) derived from EC, particularly when these distributions display tailings and multiple secondary peaks. The work is well-suited for HESS, and both the modeling and dataset are relevant to the community.

Reply: We thank the reviewer for the summary and assessment of the manuscript.

Comment 1

The paper's primary issue lies in the lack of clarity in its presentation and the insufficient contextualization of the work. Regarding clarity, the derivation and presentation of the model are inadequate. Specifically, it is challenging to comprehend the rationale behind various aspects involving the time series data, such as: why are EC measurements offset only once a day? Why choose 1-8 velocity values per day instead of a continuous velocity change? How are the weight values determined? Is it merely by visually inspecting the ratio from the L curves? What constitutes a hypothetical flow line? Additional examples will be presented later. Furthermore, the frequent references to figures and tables in the supplementary material necessitate constant toggling between the supplementary material and the main paper, suggesting that these should be integrated into the main paper.

Reply: We will restructure the paper to improve both its clarity and the contextualization of its content. Maybe it did not become clear enough in the original submission, but the velocity change in our model is indeed continuous. It is based on a smooth interpolation between knots whose values are estimated; the differences between the values at the knots are further regularized by a smoothness constraint (first-order Tikhonov regularization). The appropriate temporal resolution of these knots is one of the issues addressed in our study: If you choose a high resolution in conjunction with little smoothing by regularization, you simply map noise in the two time-series onto each other. Conversely, if you choose a low resolution or imply very strict smoothing, you miss the information on transient flow contained in the data. The appropriate resolution and smoothing are case dependent because input signals at different streams or at different times have different power spectra. We want to show how to approach the question of the right resolution.

In the revised version, we will discuss the effects of different temporal resolutions of knots for the interpolation of the EC offset, including a linear trend model. We will demonstrate that even when the EC offset is assumed constant, large temporal fluctuations of mean travel time (expressed by fluctuations of porewater velocities) can be estimated from the presented EC time series. Furthermore, we will use the maximum curvature of the L-curve as defined in Hansen (1999) as criterion to define the optimal value of the smoothing weights.

Comment 2:

In terms of context, the authors seem to overlook a substantial body of literature on anomalous transport in the hyporheic zone. This omission is surprising, especially considering that one dataset in their study is non-Fickian. Moreover, the complex processes involved in setting daily velocity values, and extracting weighting times from a hypothetical flow line could potentially result in overfitting the data to appear as Fickian flow. Therefore, I recommend exploring and acknowledging other non-Fickian possibilities and referring to the extensive non-Fickian literature in the hyporheic zone. A few select references are provided, but there are many more:

Singha, Kamini, et al. "Electrical characterization of non-Fickian transport in groundwater and hyporheic systems." Water Resources Research 44.4 (2008).

Boano, Fulvio, et al. "A continuous time random walk approach to the stream transport of solutes." Water Resources Research 43.10 (2007).

Roche, Kevin R., et al. "Effects of turbulent hyporheic mixing on reach-scale transport." Water Resources Research 55.5 (2019): 3780-3795.

Berkowitz, Brian, and Erwin Zehe. "Surface water and groundwater: unifying conceptualization and quantification of the two “water worlds”." Hydrology and Earth System Sciences 24.4 (2020): 1831-1858.

Drummond, J. D., et al. "Effects of solute breakthrough curve tail truncation on residence time estimates: A synthesis of solute tracer injection studies." Journal of Geophysical Research: Biogeosciences 117.G3 (2012).

Sherman, Thomas, et al. "A dual domain stochastic lagrangian model for predicting transport in open channels with hyporheic exchange." Advances in water resources 125 (2019): 57-67.

Haggerty, R., Wondzell, S. M., and Johnson, M. A.: Power-law residence time distribution in the hyporheic zone of a 2nd-order mountain stream, Geophys. Res. Lett., 29, 18-1–18-4, (2002).

Reply: We are grateful to the reviewer to point out the missing literature and included those articles suggested by the reviewer that deal with travel times between surface-water bodies to individual points in groundwater/hyporheic sediments. Some of the references listed by the reviewer, however, deal with travel-time distributions from the river through the hyporheic zone back to the river as needed in reach-scale transport models. These distributions summarize the effects of transport along a wide distribution of path lengths and velocities, which is not comparable to a travel-time distribution from a stream to a single point within its sediment. (This is like confusing groundwater travel times to a point and those observed in a pumping well, where the latter is an integral and shows substantial tailing caused by geometric effects.)

We agree that assuming Fickian transport is debatable in many applications. However, the aim of our study and the presented model is to estimate the time-variability of (mean) travel times from natural EC fluctuations, where we are restricted to what can be extracted from the data. Most tools to fit nonlocal transport models (CTRW, fADE, dual-domain transport, MRMT, …) assume time-stationarity of the transport coefficients, resulting in stationary travel-time distributions. Extending such tools to account for time-dependent parameters would require data that allow extracting higher-order features of travel-time distributions by some kind of deconvolution (e.g., regular Dirac pulses in the inflow would allow to study tailing). We doubt that our natural EC data (or the natural EC data of any other river that we are aware of) are suitable to inform nonlocal transport models with dynamic coefficients. In particular, the EC time series collected as part of our study show distinct diurnal fluctuations, so that they don't allow estimates of long-time (>24 h) transport behavior in a way as time series collected after the traditional pulse injection of an artificial tracer (with the latter having the disadvantage of representing only the conditions at the time of the artificial-tracer experiment).

In summary: Is the transport between the rivers and observation points analyzed in our study Fickian? We don't know. But we doubt that the natural-tracer time series contain the information needed to answer this question. We simply stick to the simplest model that can explain the existing data – according to the modeling rule "as simple as possible, as complex as necessary".

Detailed Comments

Comment 3: Line 143: The sentences, "Thus, we use equation 1 only as a parameterization to obtain time-dependent transfer functions, and we consider the coefficients determined upon calibration as apparent ones. In particular, the time variable velocity may in reality reflect effects of both changes in the true porewater velocities and shifts in travel paths," are unclear. If the ADE is merely used to parameterize the coefficients, how can the study claim the flow is Fickian? This appears tautological. Clarification is needed on why this procedure and the associated ADE are superior to other methods.

Reply: We are sorry for the confusion. The problem does not so much lie in the cited sentence, but in other – misleading – statements elsewhere. We will rewrite the introduction of the model and emphasize even more that the ADE is used exclusively as a parameterization tool to obtain time dependent travel-time distributions from a stream to a single observation point in the sediment, where the travel-time distributions are defined by a few time-dependent coefficients. As noted above, the ADE is used because of its simplicity and because long-time transport behavior (> 24 h) cannot be extracted from diurnal EC fluctuations so that long tails in travel-time distributions that would require nonlocal transport models cannot be extracted from the data.

Comment 4: Furthermore, the authors repeatedly mention throughout the manuscript that travel paths may shift, but they do not elucidate the mechanism responsible for these changes in flow paths. This is crucial, as understanding the mechanism could constrain the variations in flow paths in alignment with the proposed model.

Reply: In the revised manuscript, we will discuss mechanisms that may cause spatial shifts of flow paths from the surface water to hyporheic sediments in the discussion section. Particularly, the topmost layer of streambed sediments (i.e., the first cm) is highly dynamic due to sedimentation and erosion processes changing the geometry of the boundary and the hydraulic properties in the topmost layer. These changes cause changes in the spatial arrangements of flow paths. However, to decipher the exact magnitude and cause of shifts in a specific case would require a detailed 4-D analysis of flow and sediment transport, which is beyond the capability of standard experiments in the hyporheic zone. Thus, the remaining statement is: The standard conceptual model of hyporheic flow paths being fixed tubes is debatable, you will see effects of shifts (the exact nature of which will remain hidden) on solute transport if you observe over a sufficiently long period of time, whereas you may miss it altogether in a single artificial-tracer experiment with pulse injection. Such shifts would have of course implications for reactive-transport models that assume spatially variable reactive properties of the sediment matrix, but these aspects are beyond the scope of the current study.

Comment 5: Line 152: What is the rationale for having only one EC offset per day? Given the data for the stream stage, there could be two offsets per day or a common trend line. An explanation for this choice is needed.

Reply: We have explored the effects of two knots per day in the interpolation of the EC offset, a constant offset, and a linear trend, and want to include the results in the revised version of the manuscript. In all cases, the fitted porewater velocities show similar temporal variability.

Of course, if you allow high-frequency EC offsets (many knots, no smoothing), you can “blame” all differences in EC time series between the input (river) and output (point in the sediment) to offsets that are independent of transport. Thus, the goal should be to allow as little transient behavior in the EC offset while still fitting the data.

Comment 6: Line 155: Similarly, why decide on 1, 2, 4, and 8 velocity values per day? Is there a marker in the data that suggests this? Are there known changes in the head value that necessitate this range of change?

Reply: In all datasets reported in the present study, stream stages show diurnal fluctuations, a finding that motivated the use of more than one knot per day in the temporal interpolation of velocity. As in most regularization problems the choice of the knot resolution is arbitrary and needs to be chosen by the modeler. With increasing number of knots, computational efforts involved in parameter estimation via DREAM increase dramatically. While the goodness of fit initially increases with the number of knots, it may reach a plateau value where additional knots will cease to have positive effects on the goodness of fit. In the revised version of the manuscript we improve the discussion on the number of knots in the velocity interpolation.

Comment 7: Line 159: It is assumed that bold font indicates a vector for all variables, yet this is not explicitly stated in the text.

Reply: We will clarify this in the revised version.

Comment 8: Line 166: The sentence, "subject to a constant that does not depend on the parameters," is unclear. What does this mean in the context of the study?

Reply: The constant comes from taking the logarithm of the Gaussian likelihood function. The scaling factor in front of the exponential in the Gaussian function depends on the assumed measurement error, but not on the magnitude of the residual, and neither on the fitted parameters. When taking the logarithm, this factor becomes a constant that is not altered by modifying the parameters. We will clarify this in the revised version of the manuscript.

Comment 9: Paragraph 174-181: While the method of finding weights through L-curves is described, the reason for doing so is not clear. What purpose does the weighting serve? Is it only to establish how well the model captures the measurements? If so, why is it just “an additional measure of the goodness of fit”?

Reply: In the intended revision, we will extend the above-mentioned paragraph to explain the purpose of the weights and the L-curve method in more detail. In brief, the purpose of the present study and its novelty primarily lie in determining a continuous function of (apparent) flow velocities over time $v(t)$. To avoid overfitting and to determine the optimal number of knots needed to construct a continuous velocity function, regularization is needed, i.e., large “jumps” in consecutive velocity knots are penalized to gain a relatively smooth continuous function of velocity (and EC offset) values. Too much smoothing, however, will lead to a decrease in the goodness of fit if the true velocity function exhibits strong temporal variations. The purpose of the weights is to navigate between meeting the measured EC values in the sediment as well as possible versus allowing as little variations in the fitted coefficients. The approach is common in classical geophysical inversion and can be interpreted as a multi-Gaussian prior of the fitted parameters with linear covariance function if wanted (Kitanidis, 1992, The minimum structure solution to the inverse problem, Water Resour. Res., 33(10): 2263-2272). In essence, a metric of the smoothness is added to the sum of squared residuals in equation 4 (see for instance: Hansen 1999: The L-curve and its use in the numerical treatment of inverse problems).

Comment 10: Line 184: How is the hypothetical flow line established? Given that velocity seems to be unknown due to the unknown path, how many possibilities are there?

Reply: Given that the only knowns are that the trajectory starts somewhere at the river-riverbed interface and ends at the observation point, the number of potential trajectories is infinite. In the simplified case that vertical velocity is spatially uniform and that the riverbed surface is flat, the horizontal flow component would be irrelevant to establish a relationship between depth and travel time. As we will clarify in the revision, the assumed flow line serves the sole purpose of providing a parameterization for the travel-time distribution. For this purpose we assume the simplest case, that is, a straight, vertical flow line from the surface water to the measurement point in the streambed sediment, which we will explicitly mention. Because the exact trajectory and the hydraulic properties along it are unknown, the estimated velocity is an apparent parameter.

Comment 11: Line 201: The paper is laden with specialized jargon that, in my opinion, detracts from its accessibility. For instance, the term "homoscedastic epistemic model error" could be simplified. "Homoscedastic" could be replaced with "homogeneity of variances," and "epistemic" could be substituted with "model uncertainty," resulting in the phrase "variance homogeneity uncertainty due to measurement errors." My potential misinterpretation of these terms underscores the need for clearer explanations rather than reliance on specialized jargon, especially given the broad readership of HESS. I recommend clarifying the terminology to make the paper more accessible to a wider audience.

Reply: We agree and will replace/omit specialized jargon in the revised version.

Comment 12: Line 266-270: The authors transparently enumerate all potential processes that could influence the EC measurements and introduce errors, which is commendable. However, they do not specify how they address these issues. Is this accounted for in the "homoscedastic epistemic model error"? Is there a methodology to estimate the impact of each process relative to the measurement? In line 279, they state that all these ranges of uncertainty should be considered as model uncertainties. Yet, there are distinct approaches to handling model uncertainties (via ensemble methods) and measurement uncertainties (by calculating the potential range of influence). While the authors do acknowledge this by discussing the correlation between coefficients, they conclude by stating, "It is thus likely that the temporal dynamics of EC offset are predominantly related to measurement error." If so, why substitute one form of uncertainty for another when they stem from different sources? This is particularly perplexing given that changes in flow paths are consistently cited as the reason for broad peaks in travel-time distribution and other discrepancies, yet this form of uncertainty is not addressed in the study.

Reply: We are not sure which approaches to handling model uncertainties the reviewer is referring to. The ensemble methods that we know of would imply a set of different conceptual models leading to different mathematical descriptions forming an ensemble, which then is analyzed for instance by Bayesian model comparison. That of course requires that the different models are explicitly formulated (e.g., defining models in which the pathlengths of trajectories vary over time, or in which the hydrogeochemistry of solutes and relevant minerals are explicitly calculated to understand the offsets). To really decipher where discrepancies stem from, much more data would be needed (for instance time series of individual ion concentrations; for shifting flow paths we even don't know what kind of detailed information would be attainable), which neither we nor the majority of other researchers working with EC time series have. We consider it pointless to setting up more complex models without the corresponding data to inform them. At the end, we want to provide a manageable way to interpret data that are easy to obtain in riverbed sediments. But we cannot provide a fully mechanistic explanation of all errors and uncertainties occurring. This is a pretty common situation in environmental monitoring, and it is also common that residuals are an undecipherable mixture of measurement and model errors.

Comment 13: Figure 2: Why is there a discrepancy between the "measured" velocity peak and the mean advective travel time peak? It appears that the maximal residence peaks are misaligned with the corresponding porewater velocity, which is perplexing since one is a consequence of the other.

Reply: The difference between the estimated continuous velocity function and the residence time function arise from our definition and calculation of the mean advective travel time as the time period, where the integral of the past velocity function equals the travel path distance. Thus, there is a time lag between the travel time function and the porewater velocity function. The travel time is always the integral of the inverse velocity over the travel path.

Comment 14: Figure SI-1: Should there be a variation in the dimensions of the regularization weights for porewater velocity (), and the EC offset dimension, ()?

Reply: Yes, there should be. The dimensions of the weights are mentioned in the figure caption.

Comment 15: The term "Stream stage" is frequently mentioned but not defined. This is a recurring issue in the paper and is often the result of using specialized jargon in papers aimed at a specific audience. Please define all terms and refrain from using specialized terminology where possible.

Reply: We will define the term stream stage in the revised version of the manuscript. It is the water table of the stream (typically measured in meters above sealevel).

Reviewer 2

This MS concerns the inverse problem of inference of point-to-point transfer functions for short travel distances beneath streambeds. Although some calibrated hyporheic flow time series are presented and a few remarks made concerning the nature of the transport uncovered, this is not the focus of the paper. This is presented as a paper introducing a new calibration method, and I am considering it primarily on that basis.

Comment 1

I found the presentation confusing and it difficult to determine just what was being proposed, based the information provided in the manuscript. This is obviously a major problem in a document aiming to outline a new method. In particular, it is not at all clear what the relationship is between Equations (4) and (8). Many times, reference is made to use of the non-parametric deconvolution algorithm of Cirpka (2007), and (8) is naturally applicable without specifying a functional form of g(). But elsewhere there is reference to whether transport is or is not Fickian, and to the underlying dispersivity and velocity, as shown in (1). This of course implies a parametric calibration. The two formulations differ in their interpretation of the primary source of mismatch (measurement vs. model error), and in what time series' quadratic variations they penalize (latent variables vs. outcome). Surprisingly to me, the non-parametric (8) appears to be used in a context where the realism and physical interpretation of the underlying parameters are of interest: where the Fickian or non-Fickian nature of the transfer functions is concerned. It seems like it would be ideal to identify the best-fit Fickian transfer function via (4) and compare it with the empirical result.

Reply: The purpose of the present study is to determine time-dependent transport characteristics from EC time series. We do this via the parameterization provided in equations 1 & 2, leading to the objective function of equation 4. We will make this clearer in the revision.

The non-parametric deconvolution (equation 8) is only included for comparison purposes. It is an established technique with the advantage that it does not prescribe the shape of the travel-time distribution, but also with the strong limitation that it relies on stationarity, that is, transport characteristics are assumed to remain identical over time. We want to keep this comparison in order to show that uncommon features in stationary travel-time distributions (such as multiple peaks) can be the result of neglecting the transient flow-and-transport characteristics.

We will remove statements on the transport nature in hyporheic sediments throughout the revised version of the manuscript as this was distracting the reviewers from the main message.

Comment 2

I am also concerned about the introduction of the physically unmotivated "offset" $o(t)$ that fudges the difference between the EC predicted by the transient ADE and the observed EC, and which is allowed to change every day. It is not clear why this function is needed at all. It is possible to simply find the best-fitting calibrated model against a time series by a least squares plus penalty functional procedure similar to the ones shown in the paper. It appears o(t) might have been introduced so that part of the mismatch can be categorized as measurement error in (4). I generally expect model error to dwarf measurement error in these sorts of applications, and in any event, a coarse temporal resolution of $o(t)$ is considered, so the first term of (4) inherently contains some model error. And furthermore, the two regularization terms in (4) do not have a probabilistic foundation: they are determined from the L-curve approach, which is rooted in the idea of minimum MSE. It seems like the complexity of $o(t)$ can be dispensed with from the point of view of parameter identification.

I believe the authors should demonstrate the superiority of the calibration approach in (4) relative to a straightforward approach that does not include the offset and/or time-varying velocity by computing AICc. Furthermore, it is not clearly shown how well the model (1) fits the data, and how much work $o(t)$ is doing to fudge the difference between model prediction and observed data, and how much it is being allowed to vary, ad hoc, from day to day. This should be shown.

Reply: There are good chemical reasons for the EC offset, which has been observed at practically all riverbank-filtration sites. EC results from the concentrations of dissolved ions. If the only ions were Na^+ and Cl^-, EC would be a conservative tracer. However, a substantial fraction of EC is caused by Ca^{2+} and bicarbonate (HCO_3^-) and, to a minor extent, other ions that undergo precipitation/dissolution reactions. It is normal that river-borne water parcels increase in mineralization while being transported through sediments. The factors influencing the increase in EC include the partial pressure of CO_2, temperature, and microbial activity, which vary over time. On top of these chemical reasons the data loggers recording EC time series are known to drift over time. That is, EC is not an ideal tracer. But it is easy to measure and therefore readily available at many sites. When analyzing EC time series, one cannot neglect offsets. The only question is how to deal with it.

We agree that the inclusion of the EC-offset term warrants a more thorough investigation on its effect on the estimated velocity values (representing mean travel time) and the goodness of the fit. In the revised version of the manuscript we will i) use AIC to compare model runs based on both their likelihood and the number of involved parameters and ii) thoroughly discuss the effects of the EC offset by including model runs that a) have no EC offset, b) have a constant EC offset, c) include a linear EC offset trend model and d) have two and one knots per day in the interpolation of the EC offset. The temporal variations of the inferred apparent velocities are very similar in all model runs. The smoothing regularization term is a standard method used in geophysical inversion. As it has the functional form of a sum of squares it can easily interpreted as the logarithm of a Gaussian prior. Specifically, the 1-D smoothness constraint is mathematically identical to a linear generalized covariance function for a multi-Gaussian prior distribution of the parameters (Kitanidis, 1992). While there are Bayesian techniques to obtain the weights (with poor convergence behavior), we suggest following methods that are well established in geophysical inversion based on the curvature of the L-curve (Hansen, 1999) and will apply these techniques more rigorously.

Comment 3

Figure 3b appears to show a comparison of measured and simulated time series, but there is a very obvious delay visible between the two time series. Why did this not result in a differently identified velocity?

Reply: There is almost no difference between the modelled and measured EC time series in the hyporheic zone and thus the line (simulated) and measured (points) values closely overlay. As shown in the legend, the grey dots represent measurements of EC in the surface water of the respective streams, and the delay is actually the signal that we are after.

Comment 4

Statements about the seemingly Fickian / non-Fickian nature of the travel time distributions seem to be based on eyeballing the non-parametric distributions shown in Figure 4. In my view, there is not enough evidence given to support these statements.

Reply: We agree and will remove statements on the nature of porewater transport in the revised version of the manuscript. We plan to keep Figure 4 to discuss effects caused by the violation of the assumption of steady-state flow inherent in the applied non-parametric deconvolution method.

Comment 5

Finally, it would be helpful for the authors to highlight the novelty in the presented results. The model-free deconvolution approach is previously published, and other major aspects---Bayesian framing, quadratic penalty functional, use of L-curve to trade off bias and variance---are all well established in the literature. Is the particular way they are combined original? (Again, this is hard to evaluate because of the confusing presentation.) Or is it the use of these classic techniques in the context of hyporheic flow that is new? Whatever the claim to originality, it should be made clear and contextualized relative to existing literature.

Reply: The novelty of the present manuscript lies in the combination of the above-mentioned methods (L-curve regularization, Bayesian framework) to determine the transient behavior of apparent velocities (and thus mean travel time) in hyporheic sediments. The previously established method of non-parametric deconvolution, which assumes stationarity, primarily serves as a reference and is included to highlight the effects of a violation of the stationary flux assumption in non-parametric deconvolution. We will highlight the novelty and the main goal of the present study more clearly in the revised version of the manuscript.

Reviewer 3

Comment 1: As the earlier reviewers state, the article concerns an interesting and relevant topic but is full with what seem to be arbitrary choices and ad hoc solutions. Something like the time-varying EC off-set, $o(t)$, is such an artefact. No serious physical explanation is provided. To keep things from pure noise-fitting, a regularization is applied but the choice of the weights is based on visual inspection, which is difficult to replicate.

Reply: As listed above, the chemical nature of the EC offset is pretty clear, and we may have missed to explain it in the original submission because we thought that everybody in the hyporheic-zone community knows about it. We will add that information. We have already performed the calculations for a mathematically tractable approach of obtaining the optimal set of weighting factors (see Reviewer I, comment 1) and investigated the effects of the time-varying EC off-set, $o(t)$, in more detail (i.e., use a constant offset value, a linear trend model and two knots per day in the interpolation of the EC offset). We can show that these are neither arbitrary choices nor ad-hoc solutions.

Comment 2: The reason to accept a Fickian model seems to be necessary but not sufficient. What would be the results if non-Fickian models were applied throughout?

Reply: Our primary emphasis is on the temporal variation of apparent velocity (which primarily determines the mean travel time). We need a metric of spread in the travel-time distribution, for which we choose a constant dispersivity. These are parametric choices to keep the inverse problem manageable. A non-Fickian approach with time-dependent coefficients would imply estimating more parameters, which are poorly constrained by the data. The latter is caused by the type of input data: comparably smooth, mainly diurnal variations of EC in the river water, that lack both high frequencies (which you would have in artificial-tracer tests with pulse injection) and distinct information on time scales > 24 h.

As mentioned above, the comparison of our results with the results obtained by non-parametric deconvolution primarily serve the purpose of investigating the effect of transient flow on deconvolution approaches that assume stationarity. We plan on clarifying these issues in the revised version of the manuscript.

Comment 3: Why is maximum likelihood used for sigma_ep and expectation maximization for Theta? And so forth.

Reply: Maximum likelihood is used in the model developed and discussed as part of the present manuscript, because the approach is readily reconciled with the Bayesian approach used to determine posterior parameter probability distributions. Expectation maximization (EM) is part of the previously published approach of non-parametric deconvolution (Cirpka et al., 2007 Groundwater). Specifically, EM is used to obtain the “measurement” error σ_{ep}. Interested readers are referred o the original paper on that method, which is used only for comparison.

Comment 4: It would probably be difficult to go through everything in such detail that the reader becomes convinced of the reasonableness of it all, also because a lot has been covered in an earlier article by Cirpka. A possible way forward is to accompany the article by something like a Python Notebook with annotated code and prepped data sets. That would allow readers to get a better idea about the visual inspection of the steepness of the L-curve, etc. Presently, the code is available on request, which is a good step but it could be better and the impact of the article would be much stronger.

Reply: We are grateful for the suggestion and will publish the python scripts alongside with the data of the present manuscript with the revised version.