Exploring tracer information in a small stream to reduce the uncertainty and enhance the process interpretation of transient storage models

Enrico Bonanno\textsuperscript{1,2}, Günter Blöschl\textsuperscript{2}, Julian Klaus\textsuperscript{3}

\textsuperscript{1} Catchment and Eco-Hydrology Group, Luxembourg Institute of Science and Technology, Belvaux, Luxembourg.
\textsuperscript{2} Institute of Hydraulic and Water Resources Engineering, Vienna University of Technology, Vienna, Austria.
\textsuperscript{3} Institute of Geography, University of Bonn, Bonn, Germany.

Correspondence to: Enrico Bonanno (bonanno@hydro.tuwien.ac.at)

Abstract. The transport of solutes in river networks is controlled by the interplay of processes such as in-stream solute transport and the exchange of water between the stream channel and dead zones, in-stream sediments, and the hyporheic zone. Transient storage models (TSMs) are a powerful tool for testing hypotheses related to solute transport in streams. However, TSM parameters are often non-identifiable leading to an unclear understanding of the processes controlling solute transport in streams. In this study, we increased parameter identifiability in a set of tracer breakthrough experiments by combining global identifiability analysis and dynamic identifiability analysis. We compared our results to inverse modelling approaches (OTIS-P) and the commonly used random sampling approach for TSMs (OTIS-MCAT). Compared to OTIS-P, our results informed about sensitivity and identifiability of TSM parameters on the entire feasible parameter space. Our results clearly improved parameter identifiability compared to OTIS-MCAT that often indicated non-identifiability of TSM parameters. Non-identifiable results led to wrong solute retention times in the storage zone and the exchange flow with the storage zone, with a difference respectively up to four and two orders of magnitude compared to results with identifiable TSM parameters. The severe differences in the transport metrics between results obtained from our proposed approach and results from OTIS-MCAT model also resulted in contrasting interpretation of the hydrologic processes controlling solute transport at the study site. Thus, our outcomes point to the risks of interpreting TSM results when even one of the TSM parameters is non-identifiable. Our results showed that there is clear potential for increasing parameter identifiability in random sampling approaches for TSMs and for advancing our understanding of hydrological processes controlling in-stream solute transport.
1 Introduction

Modelling of stream water movement is pivotal for understanding how nutrients, solutes and pollutants are transported downstream and ultimately affect stream water quality along river networks (Krause et al., 2011; Rathfelder, 2016; Smith, 2005). Experimental studies of water flow and solute transport in the stream channel commonly rely on the observation of tracer breakthrough curves (BTC, i.e., the measurement of the concentration over time of a tracer released in an upstream section). Such a BTC reflects stream discharge (Beven et al., 1979; Butterworth et al., 2000) and longitudinal tracer advection and dispersion (Gooseff et al., 2008). A milestone in the study of solute transport was that in-stream solutes and water are exchanged with slowly-moving channel waters, the dead zones (Hays, 1966), and with the saturated area between the stream channel and the adjacent groundwater (i.e., the hyporheic zone, Triska et al., 1989; White, 1993). This hydrologic exchange results in a skewed non-Fickian BTC with a pronounced tail, which makes the advection-dispersion equation (ADE) unable to correctly describe the observed tracer transport in stream channels (Bencala & Walters, 1983; Castro & Hornberger, 1991). Despite the large amount of studies, the numerous contradictory outcomes (Ward & Packman, 2019), coupled with uncertainties of the model parameters (Ward et al., 2017), raise the question about how informative such modelling results are (Knapp & Kelleher, 2020).

Considerable potential in reducing uncertainty of the processes controlling solute transport in streams lies in modelling the tail of the BTC, since it contains information on the transient storage inside the stream channels (Bencala et al., 2011). For simulating the retentive effect of dead zones on solute transport, Hays (1966) modelled the tail of the BTC by introducing a second differential equation in addition to the ADE. Following a similar approach, (Bencala & Walters (1983) described the solute transport in streams as a pure advection-dispersion transport, coupled with a hydrologic exchange term between the stream channel and a single, homogeneously mixed volume that delays the solute movement downstream (Transient Storage Model - TSM). The estimation of TSM parameters often rely on the use of inverse modelling approaches via nonlinear regression algorithm that can return precise estimation of TSM parameters with a narrow 95% confidence interval (OTIS-P; Runkel, 1998). While this approach found extensive application in past decades, it does not allow a comprehensive assessment about the identifiability of the TSM parameters (Ward et al., 2017; Knapp & Kelleher, 2020). This is because inverse modelling approach do not provide information on performances and parameter identifiability on the entire feasible parameter space (Ward et al., 2017). Also, the parameter sets obtained after inverse modelling approach do not necessarily indicate meaningful results, as non-identifiable parameters can provide good BTC fitting despite being uncertain and non-identifiable (Kelleher et al., 2019). These considerations, coupled with the lack of knowledge of the modeller over the distribution of parameters and their performance after OTIS-P results, led to a progressive increase of studies addressing identifiability in TSMs via random sampling of parameters and global identifiability analysis (OTIS-MCAT model; Ward et al., 2017; Knapp & Kelleher, 2020; Kelleher et al., 2019).

Random sampling approaches provide information on parameter identifiability and accuracy on the feasible parameter space, however they rarely show identifiability for all the TSM parameters (Knapp & Kelleher, 2020). Kelleher et al. (2013) found that the parameters associated with the transient storage process are not identifiable for a large variety of the stream reaches and experiments they investigated. Other studies have shown that TSM parameters are often not-independent and poorly identifiable (Camacho & González, 2008; Kelleher et al., 2019;
Knapp & Kelleher, 2020; Wagener et al., 2002; Ward et al., 2017; Wlostowski et al., 2013). Despite these findings and the crucial need for parameter identifiability, only few studies have explored the reliability of TSM results obtained via inverse modelling approach, and model interpretation is often based on a single set of parameters without testing their robustness (Knapp & Kelleher, 2020).

Addressing the identifiability of TSM parameters is a pressing issue, since we are still unable to link specific physical processes with the parameters derived from BTC studies (Ward & Packman, 2019). This problem is commonly related to the over-simplistic approach of TSM, which is unable to distinguish between the effects of eddies, pools, and the hyporheic zone (Gooseff et al., 2008; Zaramella et al., 2006). To overcome this limitation, the TSM has been modified to include multiple storage zones (Choi et al., 2000), sorption kinetics for reactive tracers (Gooseff et al., 2005; Kelleher et al., 2019), and changes of residence time distributions in the storage zone (Haggerty et al., 2002). While these changes increased the quality of the model fit, they also came at the cost of increased dimensionality with a further reduction of parameter identifiability and certainty, leaving the open question what physical processes exactly are associated with the transient storage modelling (Kelleher et al., 2019; Knapp & Kelleher, 2020).

The observed strong non-identifiability for TSM parameters in random-sampling studies may have three causes. First, there is no common strategy for selecting parameter ranges and the number of parameter sets in TSM simulations. To obtain reliable results, Ward et al. (2017) indicated that modelling studies need to apply TSM on a large number of parameter sets (between 10,000 and 100,000) over a parameter range spanning at least two orders of magnitude. While for some studies, the non-identifiability of parameters might be explained by the low number of parameter sets (less than 10,000) and the relatively narrow selected parameter range (Camacho & González, 2008; Wagener et al., 2002; Wlostowski et al., 2013), non-identifiability was also found when a rather large number of parameter sets and wide range were used (Kelleher et al., 2013; Kelleher et al., 2019; Ward et al., 2017). This is bringing up the question if and when TSM parameters are actually meaningful (Knapp & Kelleher, 2020).

A second cause related to uncertain results in the random-sampling approach for TSM parameters relates to the selected parameters chosen for TSM simulations. The parameters describing the advection-dispersion process (stream velocity, cross-sectional area of the stream channel, and the longitudinal dispersion) are known to be the best identifiable in the TSM (Ward et al., 2017) and once they are kept constant in the random sampling of the parameters, they can drive strong changes in the parameters describing the transient storage process (Knapp & Kelleher, 2020). This dependency indicates that the use of a constant value for an identifiable parameter may result in a mis-estimation of the other TSM parameters. While it is not recommended to keep constant a rather identifiable parameter in random-sampling approaches for TSMs, no study investigated the role that a variable velocity has on identifiability of TSM parameters. This leads to the question on how meaningful, sensitive, and uncertain the transient storage parameters are when stream velocity varies or is kept constant.

A third cause for non-identifiable TSM parameters relates to the selected approach for addressing parameter identifiability. The identifiability analysis used in most studies is based on the Generalized Likelihood Uncertainty Estimation that assesses parameter certainty across the entire observed BTC (GLUE, Beven & Binley, 1992; Camacho & González, 2008; Kelleher et al., 2013; Kelleher et al., 2019; Ward et al., 2017). However, such global identifiability analysis is unable to find informative sections of the simulated BTC with respect to a certain parameter and to unequivocally link a given parameter to a specific process (Wagener et al., 2003; Wagener et
al., 2002; Wagener & Kollat, 2007). This information is particularly important for BTC modelling, since advection-dispersion parameters are physically responsible for the bulk solute transport in the stream and they are therefore expected to act on the rising limb and peak of the BTC (Goosseff et al., 2008). Contrary, the parameters describing the exchange between the stream channel and the transient storage zone are responsible for delaying solute transport compared to the advective-dispersive transport, acting on the falling limb and tail of the BTC (Runkel, 2002). By investigating parameter sensitivity and identifiability across the entire BTC, global identifiability analysis is unable to capture an increase in parameter identifiability towards the tail of the BTC. However, studies addressing the identifiability of TSM parameters over time in different sections of the BTC reported an increase of identifiability for transient storage parameters on the tail of the BTC (Wagener et al., 2002; Scott et al., 2003; Wlostowski et al., 2013; Kelleher et al., 2013). We hypothesise that this information is key in designing a successive parameter sampling in a constrained parameter space – ultimately reducing the uncertainty affecting parameters describing solute retention in streams.

A robust assessment of transient storage parameters would not only improve the model fit of tracer transport and decrease parameter uncertainty, but it might also lead to stronger interpretation on the physical processes controlling solute transport in streams. TSM parameters are often used to calculate metrics on the solute exchange between the stream channel and the transient storage zone and the residence time of solutes in the coupled system (Thackston and Schnelle, 1970; Hart et al., 1999; Morrice et al., 1997; Runkel, 2002). These metrics are pivotal to address the potential for nutrient cycling, microbial activity, and the development of hot-spots in river ecosystems (Mulholland et al., 1997; Smith, 2005; Triska et al., 1989; Krause et al., 2017). However, no study so far indicated and evaluated if and how much the interpretation of hydrologic processes changes when TSM parameters are identifiable and when they are not, due to the enunciated challenges in TSMs.

To address these challenges, we have organised this contribution around three questions related to the key challenges of parameter identifiability in transient storage modelling:

1) How does the identifiability and the information content of model parameters associated with transient storage processes change by using fixed and varying velocity in the random-sampling of TSM parameters?

2) Does the identifiability analysis on specific sections of the BTC reduce the parameter uncertainty in random-sampling of TSM?

With the outcomes of these questions we will address:

3) How does the residence time of solute in the transient storage zone and hydrologic process interpretation vary when TSM parameters are identifiable and when they are not?

2 Study site and methods

2.1 Study site and data

The studied stream reach (49°49'38"N, 5°47'44"E) is located in western Luxembourg, downstream of the Weierbach experimental catchment (Hissler et al., 2021). The stream channel is unvegetated with a slope of ≈6% and consists of deposited colluvium material and fragmented schists (up to 50 cm depth) with local outcrops of fractured slate bedrock in the streambed. The flow regime is governed by the interplay of seasonality between precipitation and evapotranspiration (Rodriguez et al., 2021; Rodriguez & Klaus, 2019) with a persistent discharge.
between autumn and spring, and little to no discharge during summer months (discharge arithmetic mean equal to 6.5 l/s, median of 1.7 l/s, St.Dev. of 11.52 l/s between Aug 2018 and Feb 2020; Bonanno et al., 2021). To test our objectives, we have carried out three tracer experiments with an instantaneous tracer injection at three different flow (Q) conditions: 6th December 2018, Q = 2.52 l/s (E1); 23rd January 2019, Q = 9.05 l/s (E2); 28th January 2019, Q = 22.79 l/s (E3). For each experiment, we prepared an NaCl solution using 2 l of stream water and 100 g of reagent-grade NaCl. We injected the solution in a turbulent pool at the beginning of the stream reach to assure complete mixing in the stream water. Electric conductivity (EC) was measured via a portable conductivity meter (WTW) 55 m downstream of the injection point and converted into Cl− concentration via an EC−Cl− regression line (R² = 0.9999).

2.2 Advection-Dispersion equation and Transient Storage Model formulation

The one-dimensional Fickian-type advection and dispersion equation describes the joint effect of flow velocity and turbulent diffusion on solute transport (Beltaos & Day, 1978; Taylor, 1921, 1954). The differential form of ADE reads:

\[ \frac{\partial C}{\partial t} = -v \frac{\partial C}{\partial x} + \frac{1}{A} \frac{\partial}{\partial x} \left( A D \frac{\partial C}{\partial x} \right) \]  

Eq. 1

Where \( t \) is time [T], \( x \) is the distance from the injection point along the stream reach [L], \( A \) [L²] is the cross-sectional area of flow, \( v \) [L/T] is the average flow velocity, \( D \) [L²/T] is the longitudinal dispersion coefficient, and \( C \) is the concentration of the observed tracer above background levels [M/L³]. The solution of the differential form of ADE for an instantaneous solute injection at \( x = 0 \) [L] reads:

\[ C(t) = \frac{M A}{4(Dt)^{1/2}} \exp \left[ -\frac{(L-vt)^2}{4Dt} \right] \]  

Eq. 2

Where \( M \) is the injected solute mass [M], \( t \) is time [T], and \( L \) is length of the investigated reach [L].

The TSM describes the solute transport in streams by combining the advection-dispersion process in the stream channel through a hydrologic exchange with an external storage zone. The model equations read (Bencala & Walters, 1983):

\[ \begin{aligned} \frac{\partial C}{\partial t} &= -v \frac{\partial C}{\partial x} + \frac{1}{A} \frac{\partial}{\partial x} \left( A D \frac{\partial C}{\partial x} \right) + \frac{q_A}{A} (C_L - C) + \alpha (C_{TS} - C) \\
\frac{\partial C_{TS}}{\partial t} &= -\alpha \frac{\partial}{\partial x} (C_S - C) \end{aligned} \]  

Eq. 3

where the hydrologic exchange with the transient-storage zone is driven by the exchange coefficient \( \alpha \) [1/T] and the area of the transient storage zone, \( A_{TS} \) [L²]. Here, we will refer to \( A \), \( v \), and \( D \) as “advection-dispersion parameters” and to \( A_{TS} \) and \( \alpha \) as “transient storage parameters”; the five parameters are referred to as “TSM parameters”. The solute concentration in the main channel and the transient storage zone are \( C \) and \( C_S \) [M/L³], respectively. The performances of both ADE and TSM results are evaluated using the Root Mean Squared Error objective function (RMSE), which is the most commonly used objective function in solute-transport studies (Ward et al., 2018; Wlostowski et al., 2017; Zaramella et al., 2016).

2.3 Iterative modelling approach to obtain TSM parameters

Several sampling approaches were previously used to estimate parameter uncertainty in TSMs, such as Monte Carlo sampling (Wagener et al., 2002; Wagner & Harvey, 1997; Ward et al., 2013), Latin hypercube sampling.
(LHS, Kelleher et al., 2019), and Monte Carlo coupled with a behavioural threshold (Kelleher et al., 2013; Ward et al., 2017). Here, we use LHS to sample from the selected parameter space, due to LHS higher efficiency compared to the classic Monte Carlo approach (Yin et al., 2011).

We simulated our tracer experiments with the ADE by sampling advection-dispersion parameters via LHS to avoid initial assumptions that could impact the parameter estimates (Figure 1). The RMSE value of the best-performing ADE parameter set was indicated as RMSE_{ADE}. Similar to the Monte Carlo approach coupled with behavioural threshold (Kelleher et al., 2013; Ward et al., 2017), we simulated the three tracer experiments with the TSM through a step-wise approach with \( n \) TSM iterations (\( n \) is number of iterations, Figure 1). To obtain reliable TSM results, Ward et al. (2017) suggested a minimum amount of parameter sets between 10,000 and 100,000. Thus, in each TSM iteration we simulated 115,000 parameter sets. Results of each TSM iteration include RMSE values for the 115,000 parameter sets, and results of global identifiability analysis of TSM parameters. Global identifiability analysis was conducted through parameter vs RMSE plots (Wagener et al., 2003), parameter distribution plots (Ward et al., 2017), regional sensitivity analysis (Kelleher et al., 2019; Wagener & Kollat, 2007), and parameter distribution plots (Wagener et al., 2002; Ward et al., 2017). Globally identifiable parameters satisfy the following criteria: univocal peak of performance in parameter vs RMSE plots and in parameter distribution plots (Ward et al., 2017) and CDF corresponding to the best 0.1% of the results deviating from the 1:1 line and from parameter CDF corresponding to the best 10% of the results (Kelleher et al., 2019). To evaluate the degree of identifiability of a certain parameter, we also evaluated the two-sample Kolmogorov-Smirnov (K-S) test which calculates the maximum distance \( K \) and the corresponding \( p \)-value between two cumulative distribution functions, \( F(P_{0.1}) \) and \( F(P_{10}) \), by:

\[
[K, p] = \max\{F(P_{0.1}) - F(P_{10})\}
\]

Eq. 4

Where \( F(P_{0.1}) \) and \( F(P_{10}) \) are the cumulative distribution function of a parameter \( P \) respectively for the best 0.1% and the best 10% of the results. Following the approach of Ouyang et al. (2014), we grouped parameter identifiability in four categories: highly identifiable (\( K > 0.25, p \leq 0.05 \)), moderately identifiable (\( 0.1 \leq K \leq 0.25, p \leq 0.05 \)), poorly identifiable (\( K < 0.1, p \leq 0.05 \)), and non-identifiable (\( p > 0.05 \)).

100 best performing parameter sets for each iteration were analysed with the DYNamic Identifiability Analysis (DYNIA, Wagener et al., 2002) to address the role of TSM parameters on the BTC and the change of information content. The dynamic identifiability analysis identifies regions of the observed data that are sensitive (or not) to the investigated model parameter, and it can be used to test model structure, to design specific experiments, and to relate the model parameters to a specific simulated model response (Wagener et al., 2004). The dynamic identifiability analysis yields the distribution of the likelihood as function of the parameter values and the information content of the parameters over time. The information content is expressed as one minus the width of the 90% confidence interval over the entire parameter range (Wagener et al., 2002). A wide 90% confidence interval indicates that various parameter values are associated to equally good performances resulting in a low information content. Conversely, narrow 90% confidence intervals and corresponding high information content values suggest that the best performing parameters are contained in a relatively narrow range compared to the feasible range. A detailed description on how to read the plots used to address the global identifiability analysis and the description of the dynamic identifiability analysis algorithm are reported in Appendix A.
results of the previous TSM iteration. Namely, if the global identifiability analysis from the previous TSM iteration indicated that the investigated parameter is identifiable, the best 1 % of the results were used to define its parameter space in the successive TSM iteration. When the identifiability criteria were not met, the parameter space investigated in the successive TSM iteration was increased or, for the case of $A_{TS}$ and $\alpha$, it was constrained based on the dynamic identifiability analysis result (information content above 0.66 on the BTC tail). This condition was chosen by the evidence that transient storage parameters $A_{TS}$ and $\alpha$ are often non-identifiable via global identifiability analysis (Camacho & González, 2008; Ward et al., 2013; Ward et al., 2017; Kelleher et al., 2019), but they are identifiable on the BTC tail (Kelleher et al., 2013; Wagener et al., 2002; Wlostowski et al., 2013).

While the first TSM iteration was conducted to investigate the identifiability of all the possible combinations in the feasible parameter space reported in literature and from results of ADE (Table 1), the successive iterations excluded pairs of $v$ and $A$ whose product was outside the value of the discharge evaluated via dilution gauging ±10 %. This condition was chosen to respect results from Schmadel et al. (2010), who reported that the discharge error from dilution gauging method is ≈8 %. The approach reported in Figure 1 for TSM was used also in the case where $v$ was assumed fixed and equal to $v_{peak}$. The modelling was finalized once every TSM parameter indicated global identifiability via the enunciated criteria and the Kolmogorov-Smirnov test resulted in $K > 0.1$ and $\rho \leq 0.05$ for each TSM parameter.

We compared our results with both inverse modelling results (OTIS-P), and the most-common random sampling approach in TSMs (OTIS-MCAT). OTIS-P is an inverse-modelling scheme that minimise residual sum of squares between the modelled BTC and the observed BTC. OTIS-P model estimates the best-fitting TSM parameter values and their 95 % confidence interval. As indicated in Runkel (1998), we carried out multiple OTIS-P iterations and interrupted them when parameter values calibrated via OTIS-P changed less than 0.1 % between subsequent runs. OTIS-MCAT solves the TSM for the selected number of parameter sets and addresses their identifiability with a global identifiability analysis (Ward et al., 2017). Compared to our approach, OTIS-MCAT considers Monte Carlo parameter sampling instead of LHS, velocity equal to $v_{peak}$ and it does not foresee iterative parameter sampling from results of dynamic identifiability analysis. Thus, we here indicate as “OTIS-MCAT results” the results we obtained after the first TSM iteration when $v$ was assumed fixed and equal to $v_{peak}$.

### 2.4 Hydrologic interpretation of TSM results

The TSM parameter sets obtained after OTIS-P, OTIS-MCAT, and the proposed iterative TSM approach were used to compute some hydrologic metrics to interpret solute transport in streams. We here computed the average distance a molecule travels in the stream channel before entering the transient storage zone ($L_a$ [L], Mulholland et al., 1997):

$$L_a = \frac{v}{\alpha} \quad \text{Eq.5}$$

The average time spent by a molecule in the transient storage zone ($T_{sto}$ [T]) is evaluated as (Thackston and Schnelle, 1970):

$$T_{sto} = \frac{A_{sto}}{\alpha A} \quad \text{Eq.6}$$
We computed the average water flux through the storage zone per unit length of stream channel to interpret the magnitude of flux between the stream channel and the transient storage zone \( q_s \) (L/T), Harvey et al., 1996):

\[ q_s = \alpha A \]  

Eq.7

However, \( L_s \), \( T_{ss} \), and \( q_s \) metrics are not able to encompass both the role of advective transport and of the transient storage. Thus, we also evaluated \( F_{MED} \) that accounts for the median travel time due to advection-dispersion and transient storage and for the travel time only due to advection-dispersion (Runkel, 2002):

\[ F_{MED} = \left( 1 - e^{\left(-\frac{x}{vD}\right)} \right) \frac{DS}{AV + D} \]  

Eq.8

Increasing values of \( F_{MED} \) have to be interpreted as increasing relative importance of the storage zone in the solute transport downstream (Gooseff et al., 2013; Runkel, 2002).

3. Results

3.1 Identifiability of TSM parameters and comparison with OTIS-P and OTIS-MCAT results

The identifiability of TSM parameters was studied for the three tracer experiments injections (indicated as E1, E2, E3, cfr. paragraph 2.1) for two distinct cases: stream velocity considered as a variable parameter, and velocity considered equal to \( v_{peak} \) (\( v = v_{peak} \)). Global Identifiability analysis results are here reported via only parameter values plotted against the corresponding RMSE values.

3.1.1 Transient storage modelling with stream velocity as varying model parameter

After the first TSM iteration the global identifiability analysis indicated that \( v \), \( D \), and \( \alpha \) parameters are sensitive with a unique and identifiable performance peak \( K \) of K-S test always > 0.22 and \( p < 0.05 \) for each tracer experiment). However, \( A \) and \( A_{TS} \) appeared non- or poorly-identifiable for the three investigated BTCs (Figure 2, green dots, \( p \)-value of the K-S test for \( A_{TS} > 0.05 \) for each tracer experiment). The dynamic identifiability analysis provided clearer insights into the effect of the TSM parameters on the BTC and their identifiability ranges compared to the global identifiability analysis. \( v \) and \( \alpha \) were confirmed to be the most identifiable and informative parameters in the rising limb, the peak and the tail of the BTC (information content > 0.66; Figure 3b, h). \( A \) and \( D \) were mostly identifiable and informative during the rising limb and the tail of the BTC (information content > 0.50; Figure 3c-f). \( A_{TS} \) was uncertain and non-informative in most sections of the BTC (information content < 0.33; Figure 3i, j). However, the identifiability of \( A_{TS} \) increased in the tail of the BTC, where the information content was above 0.66 for \( A_{TS} < 5.356 \text{ m}^2 \) for E1 (Figure 3i, j), and for \( A_{TS} < 5.4315 \text{ m}^2 \) and 4.6404 \text{ m}^2 respectively for the BTCs of E2 and E3 (results not shown).

The global identifiability of TSM parameters increased through the iterative model approach and when \( A_{TS} \) or \( \alpha \) were poorly or non-identifiable \( p \)-value of the K-S test for \( A_{TS} > 0.05 \) TSM performances approached at best \( \text{RMSE}_{add} \) (Figure 2, green, yellow and blue dots). After four (for E1 and E2) or five (for E3) TSM iterations, the parameter values plotted against the corresponding RMSE values showed univocal increase of performance toward unique values for \( v \), \( A \), \( D \), \( \alpha \), and \( A_{TS} \) (Figure 2, orange dots), and the RMSE of the best performing parameter sets decreased below \( \text{RMSE}_{add} \) (Figure 2, black horizontal line). Also, the CDF corresponding to the best 0.1 % of the results deviated both from the 1:1 line and from parameter CDF corresponding to the best 10 % of the results (results not shown). These conditions, coupled with the \( K \) of K-S test always larger than 0.1 (average
K for all the TSM parameters equal to 0.36, and p-value < 0.05) indicated parameter identifiability and the finalization of the iterative TSM approach.

The dynamic identifiability analysis for the last TSM iteration showed that the advection-dispersion parameters were important in controlling the rising limb and the tail of the BTC (results reported only for E1, Figure 4a-f), while α was particularly important for controlling the tail (Figure 4g, h) and Aτ for controlling the rising limb and the tail of the BTC (Figure 4i, j).

3.1.2 Transient storage modelling with stream velocity set equal to $v_{\text{peak}}$

The OTIS-MCAT results produced low p-values for each TSM parameter after the K-S test (p < 0.05, K > 0.12) indicating parameter identifiability. However, global identifiability analysis showed that the distribution of TSM parameters did not converge towards univocal and optimal parameter values suggesting that identifiability of TSM parameters was rather uncertain with TSM outcomes performing worse than the ADE (Figure 5, green dots). The global identifiability of TSM parameters increased considerably through the iterative model approach. After the third TSM iteration, the best performing parameter sets approached unique parameter values (Figure 5, blue dots) and the CDF corresponding to the best 0.1% of the results deviated from 1:1 line and from the CDF of the best 10% of the results (results not shown). These conditions, together with K of K-S test always > 0.25 and p-value < 0.05 for each TSM parameter and tracer experiment, showed a clear increase of identifiability compared to the initial OTIS-MCAT results. The increase of parameter identifiability was followed by a sharp increase of model performance, with the best performing parameter sets at the end of the iterative approach having RMSE values below $\text{RMSE}_{\text{ADE}}$ for all the investigated BTCs (Figure 5, blue dots and black line). Dynamic identifiability analysis for the last TSM iteration indicated that A and D control respectively the falling limb and the rising limb of the BTC (Figure 6a-d, results of E1). α controlled both on the rising limb, falling limb and tail of the BTC (Figure 6e-f) and Aτ controlled both the falling limb and the tail of the BTC (Figure 6g, h).

Results from OTIS-P showed parameter identifiability with narrow 95% confidence range for the $A_{\tau}$ and A, while D and α parameters were estimated with higher uncertainty due to larger 95% confidence range (Figure 2, 4). The parameter sets obtained via OTIS-P (Figure 2, 4, red vertical dashed line) were close to approach the best fitting results at the end of the used iterative approach, regardless the fact the velocity was considered as a variable parameter (Figure 2), or was it considered equal to $v_{\text{peak}}$ (Figure 5, Table 2).

3.2 Variation of transport metrics with increasing identifiability of TSM parameters

The investigated transport metrics showed high uncertainty as long TSM parameters were poorly or non-identifiable (Figure 2, 5, green and yellow dots). This was particularly evident after the first and second TSM iterations, when the 100 best performing parameter sets showed $T_{\text{int}}$ values spanning over nine orders of magnitude (Figure 7d-f), while both $L_s$ and $q_s$ spanned over three orders of magnitude (Figure 7a-c, g-i). When TSM parameters were poorly identifiable, the values of the metrics obtained when stream velocity was considered as a variable parameter (Figure 7, blue boxplots, first TSM iteration) and when stream velocity was considered equal to $v_{\text{peak}}$ (OTIS-MCAT, Figure 7, orange boxplots, first TSM iteration) showed relevant differences. When $v$ was considered variable together with the others TSM parameters, the best performing parameter set after the first TSM iteration showed a non-negligible role of transient storage in solute transport for the investigated discharge conditions. This was indicated by the high values of $L_s$ (from ~2 km for E1 to ~69 m for E3), by the rather low
exchange flux $q_t$ (from 0.06 l/s for E1 to 8.8 l/s for E3), and by the long solute residence time in the storage zone $T_{ro}$ (ranging from ~140 days for E1 to ~15 hrs for E3). Conversely, very different values were obtained for the transport metrics when $v$ was fixed equal to $v_{med}$. The results from OTIS-MCAT showed a rather fast exchange flux of the active stream with the transient storage zone ($q_t$ ranged from ~23 l/s for E1 to ~121 l/s for E3), a rather similar $L_s$ for the three tracer experiments (~10 m), and $T_{ro}$ increased with increasing discharge (from ~12 sec for E1 to ~3 sec for E3).

However, once the TSM parameters were identifiable all the transport metrics converged toward constrained values and consistent with OTIS-P results (Figure 7). This was achieved whether stream velocity was kept fixed or was variable in the modelling procedure. Results of the last TSM iteration showed that the investigated transport metrics have low dispersion around the median, and that the median almost coincides with the result of the best performing parameter set (Figure 7, red dots). When all TSM parameters were identifiable for the three tracer experiments, the transport metrics showed increasing $q_t$ (from ~2.7 l/s for E1 to ~23 l/s for E3), increasing $L_s$ (from ~50 m for E1 to ~100 m for E3), and decreasing $T_{ro}$ (from ~150 s for E1 to ~33 s for E3) with increasing discharge conditions (from E1 to E3). $F_{med}$ did not change widely between the TSM iterations since the median of the best performing 100 parameter sets varied always between 0.04 and 0.2 (Figure 7j-l). However, together with the other investigated transport metrics, the dispersion of $F_{med}$ values around the median decreased with increasing identifiability of TSM parameters.

4. Discussion

4.1 The importance of the identifiability of TSM parameter for correct interpretation of hydrological processes

Our results demonstrated that poor or non-identifiability of TSM parameters can result in a wrong hydrological interpretation of the processes controlling solute transport in streams. Our results showed that with increasing discharge conditions $L$, and $q_t$ increased, $T_{ro}$ decreased, and $F_{med}$ was rather stable for simulations where the TSM parameters were identifiable (cfr paragraph 3.2). The low uncertainty and the values of the investigated transport metrics suggested that the transient storage at the experimental site was most probably controlled by in-stream dead-zones (Boano et al., 2014). The observed link of $L_s$, $q_t$, and $T_{ro}$ values with discharge (Figure 7) also suggested that the transient storage at our site became less important in controlling solute transport with increasing discharge conditions. This could be explained as with increasing discharge conditions the wet stream areas and the water depth increased more than the wetted perimeter that embeds dead-zoned and streambed heterogeneities (Gooseff et a., 2008). This would have caused a progressive increase of piston-flow transport and a reduced role of low-flow areas for in-stream solute retention with increasing discharge condition.

However, if we would have based the process interpretation on simulations before we reached identifiability of TSM parameters, the conclusions would have been different. The values for the transport metrics obtained when $v$ was considered variable together with the others TSM parameters, together with published results about solute residence time in the hyporheic zone and in in-stream channel (Boano et al., 2014; Gooseff et al., 2005) could have been interpreted in a way that the transient storage was controlled by in-stream dead-zones during high-discharge events and by a low rate hyporheic exchange that lasted several weeks at low flow conditions (Figure 7, blue boxplots and first TSM iteration). Conversely, results from OTIS-MCAT might have been interpreted in a way that transient storage of the studied stream channel was controlled by dead-zones at the lowest flow...
conditions and by in-stream turbulences that caused solute retention in the transient storage zone to last ~3 seconds during high-flow events (Figure 7, orange boxplots and first TSM iteration).

Compared to the inverse modelling approach (OTIS-P), we were able to interpret modelling results without the uncertainty typically associated with inverse modelling outcomes that do not assess the identifiability of the parameters and of the associated hydrological metrics across their feasible spectrum of values (Ward et al., 2017; Kelleher et al., 2019). Furthermore, by reaching identifiability of TSM parameters, we also obtained a low uncertainty of the transport metrics and a more robust process interpretation compared to the standard random sampling approach, which showed poor identifiability of TSM parameters (Figure 2, 5, green dots). The results presented in this work offer new insights into the role that identifiable TSM parameters have on the interpretation of solute transport processes in streams. The limitations of relying on poorly identifiable parameters in TSMs are generally well known (Wagner & Harvey, 1997; Ward et al., 2017; Kelleher et al., 2019), however the majority of TSM studies did not address parameter uncertainty and/or draw conclusions based on non-identifiable transient storage parameters (see the commentary of Knapp & Kelleher, 2020). Identifiability of TSM parameters is nowadays crucial, since the amount of contradicting interpretation of TSM results hamper our ability to link specific physical processes and hydrologic conditions to an increase or decrease of transient storage parameters (Ward & Packman, 2019). Contradicting hydrologic interpretation and modelling outcomes can be driven by different hydrologic and morphological setting of the investigated stream reach (Gooseff et al., 2005, Kelleher et al., 2013). Our study highlights how the poor or non-identifiability of TSM parameters can result in a wrong estimation of solute transport metrics in streams and can play a major role on the hydrologic interpretation of modelling results.

4.2 Challenges associated to parameter identifiability in TSMs

We showed that non-identifiability of α and $A_{TS}$ in TSM can result from the assumption $v = v_{\text{peak}}$, the selected number of parameter sets, and the parameter space used for the random sampling. Our results indicated that $v$ interacts with α and $A_{TS}$ in transient storage models. This was particularly evident when $v$ was variable together with the other TSM parameters, and the non-identifiability of $A_{TS}$ was coupled with identifiability of $v$ and α (Figure 2, green and yellow dots). On the contrary, $A_{TS}$ was found to be identifiable and α to be non-identifiable when $v$ was fixed equal to $v_{\text{peak}}$ (Figure 5, yellow dots). It is known that a separate evaluation of the advection-dispersion parameters from the transient storage parameters can result in mismeasurement of transient storage parameters due to the high parameter interaction (Knapp & Kelleher, 2020). However, no study so far investigated the role of the assumption $v$ equals $v_{\text{peak}}$ on the non-identifiability of α or $A_{TS}$ in random sampling approach for TSMs. Despite the observed interaction between $v$, α and $A_{TS}$, our study also showed that when all TSM parameters are identifiable the best performing parameter sets showed similar values no matter if the stream velocity was fixed equal to $v_{\text{peak}}$ or was considered a variable parameter (Table 2).

Our results showed that non-identifiability of transient storage parameters might indicate inaccurate TSM results. This was evident from TSM iterations showing non-identifiability of α and $A_{TS}$, with the best model performances approaching the $\text{RMSE}_{\text{ADE}}$ (Figure 2, 5, black line). This outcome indicated that non-identifiability of α or $A_{TS}$ is linked to an underestimation of transient storage process with the optimal modelled BTCs mimicking the ADE. Similar to our results, many authors showed non-identifiability of TSM parameters in random-sampling approach. Previous research found identifiable $A_{TS}$ coupled with non-identifiable α (Camacho & González, 2008; Kelleher
et al., 2013; Wagener et al., 2002; Wlostowski et al., 2013), while other TSM applications found $\alpha$ to be identifiable coupled with non-identifiability for $A_{LS}$ (Kelleher et al., 2019), or $\alpha$ or $A_{LS}$ to be both non-sensitive and non-identifiable (Camacho & González, 2008; Ward et al., 2013; Ward et al., 2017). Random-sampling approach are generally considered more informative than inverse-modelling approach (Ward et al., 2017; Knapp and Kelleher, 2020), however our results indicate that model outcomes showing non-identifiability of transient storage parameters should be used with particular caution for model interpretation due to the rather different parameter estimation when TSM parameters were identifiable and non-identifiable (Figure 2, 5).

Identifiability of TSM is commonly studied via random sampling approaches using between 800 and 100,000 parameter sets sampled from a parameter space spanning several orders of magnitude (Table 1). Our study demonstrated that it is unlikely to reach parameter identifiability via random-sampling approach using less than 100,000 parameter sets when investigating a rather large parameter space of TSM parameters (Table 1). Our results showed identifiability only after the third TSM iteration, between 230,000 and 345,000 parameter sets and by narrowing the investigated parameter range twice (Figure 2, 5, blue dots). While the range and the order of magnitude of advection-dispersion parameter can be estimating by using the ADE, the ranges where $\alpha$ and $A_{LS}$ are identifiable are never known a-priori and random sampling approaches need to target a parameter space large enough to capture the distribution of transient storage parameters on their entire feasible space. Thus, the peak of performance for the transient storage parameters can be so narrow that it can be missed by the random sampling approach or by only a low number of selections. Similar conclusions have been obtained by Ward et al. (2017), who found by using the OTIS-MCAT model via 100’000 parameter sets that the TSM parameters were identifiable only for one of the three investigated BTCs.

Other studies coupled random sampling approaches with behavioural thresholds to derive the most sensitive range of TSM parameters and reduce parameter uncertainty, yet this was done to constrain only the range of $A$ (Kelleher et al., 2013; Ward et al., 2017). Here, we demonstrated the importance of coupling the behavioural threshold approach with the global and dynamic identifiability analyses in successive TSM iterations (Kelleher et al., 2013; Ward et al., 2017). The high information content ($> 0.66$, eg. Figure 3j) of $\alpha$ and $A_{LS}$ on the tail of the BTC provided valuable information to constrain the parameter space in successive TSM iterations. This approach eventually allowed us to identify TSM parameters.

4.3 How TSM parameters control the rising limb, the peak, and the falling limb of the BTC

The relative high information content of transient storage parameters on the rising limb and the peak of the BTC, coupled with the high information content of advection-dispersion parameters on the falling limb and tail of the BTC (Figure 4, 6 a, c, e) showed that all TSM parameters control solute arrival-time and solute retention in stream channels. This outcome is in contradiction with the common interpretation of TSM parameters, where it is assumed that the advection-dispersion parameters control the solute arrival time, while transient-storage parameters are assumed to control the tail of the BTC (Bencala, 1983; Bencala et al., 2011; Bencala & Walters, 1983; Smith, 2005; Runkel, 2002). Following this common interpretation of the role of TSM parameters on the BTC, some authors decomposed the BTC in an advective part and a transient storage part (Ward et al., 2019; Wlostowski et al., 2017). This decomposition allows to quantify the role of advection-dispersion and transient storage embedded in the BTC. However, it also implicitly assumes a negligible role of advection-dispersion
parameters on the BTC tail, and of transient-storage parameters on the rising limb and peak of the BTC which is in contradiction with our findings (Figures 4, 6, 8).

Our work demonstrated that the assumption \( v = v_{\text{peak}} \) might not be representative of the advection role on solute transport in streams. The assumption of stream velocity equalling \( v_{\text{peak}} \) implies that \( v_{\text{peak}} \) can encompass the effect of advection on the entire BTC or at least in the rising-limb and peak of the BTC. However, when \( v \) was used as calibration parameter, our results showed that \( v \) is one of the least meaningful parameters for simulating the peak of the BTC at low discharge (Figure 4a, b, 8a), while higher information content for \( v \) is obtained at higher discharge rates for values larger than \( v_{\text{peak}} \) at the peak of the BTC (Figure 8c, e, dynamic identifiability plots not shown). Our results also highlighted how assuming \( v = v_{\text{peak}} \) caused a stronger influence of \( \alpha \) and weaker influence of \( A_{TS} \) on the BTC compared to the case when \( v \) is variable. Indeed, the dynamic identifiability analysis for the case \( v = v_{\text{peak}} \) underestimated the role of \( A \) and \( A_{TS} \) on the rising limb and peak of the BTC and overestimated the role of \( D \) and \( \alpha \) on the rising limb of the BTC compared to the case when \( v \) was considered variable together with the other TSM parameters (Figures 4, 6).

Several studies have addressed the identifiability of the TSM parameters for different sections of the BTC and showed both similar and contrasting outcomes to our findings (Figure 8g-l, Wagner & Harvey, 1997; Wagener et al., 2002; Scott et al., 2003; Wlostowski et al., 2013; Kelleher et al., 2013). The sensitivity of TSM parameters in different sections of the BTC might be driven by different physical settings or discharge condition of the study sites. For example, the sensitivity of the TSM to \( A \) and \( A_{TS} \) is expected to increase for dispersive streams and alluvial stream channels, compared to mountain reaches with low or null hydrologic exchange with the hyporheic zone (Gooseff et al., 2005; Kelleher et al., 2013). However, our analysis also suggests that the different results on the importance of TSM parameters for certain sections of the BTC reported in Figure 8 could be driven by the selected random sampling approach and the non-identifiability of TSM parameters.

Results from Wagner & Harvey (1997) and Scott et al. (2003) are in partial agreement with our results for the case \( v = v_{\text{peak}} \) (Figure 8b, d, f, g, h) suggesting a non-negligable influence of \( D \) on the rising and the falling limb of the BTC and a significant role of \( \alpha \) and \( A_{TS} \) on controlling the peak, the falling limb, and the tail of the BTC. However, our results also highlight a non-relevant role of \( A \) on the rising limb and the peak of the BTC in our experiments (Figure 8b, f). Our results also show the role of \( A \) on controlling the rising limb of the BTC (Figure 8b, d, f), a result that was not found by other studies (Scott et al., 2003; Wagner & Harvey, 1997). This difference might be driven by different hydrologic conditions and physical settings of the study sites, by the methodologies used for accounting parameters sensitivity, by the parameter sampling procedure, or the strategy used to obtain the best-fitting parameter set.

Consistent with our results, the dynamic identifiability analysis of TSM parameters by Wagener et al. (2002) and Wlostowski et al. (2013) indicated high identifiability of \( A \) on the rising limb of the BTC, while \( A \) and \( A_{TS} \) controlled the falling limb and the tail of the BTC (Figure 8h, j). However, they also report lower information content for \( D, A_{TS} \) and \( \alpha \) on the BTC compared to our study. This difference may arise from the parameter range and number of simulations chosen by the authors that could affect the TSM results. Plots of the parameter values against the corresponding objective function in Wagener et al. (2002) and the regional sensitivity analysis in Wlostowski et al. (2013) do not indicate parameter identifiability for \( A_{TS}, D \) and \( \alpha \). These results together with our identifiability plots when TSM parameters were poorly identifiable (Figures 2, 5, green and yellow plots) suggest that the range and the number of the parameter sets chosen by the authors could have been insufficient to obtain
global sensitivity and identifiability of $D, A_{TS}$ and $\alpha$ parameters. Similar to Wagener et al. (2002) and Wlostowski et al. (2013) results, our dynamic identifiability analysis showed no influence of $A_{TS}$ on the majority of the BTC when it was non-identifiable (Figure 3).

Eventually, results from Kelleher et al., (2013) emphasize the roles that $A$ and $D$ have on large sections of the BTC (Figure 8k, l). While this is consistent with our findings (Figure 8a, c, e), Kelleher et al., (2013) also indicate that transient storage parameters have a rather weak influence only on the tail of the BTC (Figure 8k, l). Different sensitivity of TSM to transient storage parameters could be driven by the different approach used for evaluating the sensitivity (i.e. Sobol’ sensitivity analysis). However, our results suggest that the number of parameter sets (42,000) selected by Kelleher et al. (2013) might be too small to obtain identifiability of TSM parameters compared to the rather large parameter range chosen for their Monte Carlo sampling (Table 1). Results by Kelleher et al., (2013) are very similar to our TSM iterations for cases where $\alpha$ was non-identifiable ($v$ equals to $v_{peak}$, Figure 5 yellow dots, dynamic identifiability plots not shown). We also demonstrated that our results after the first and second TSM iterations are not useful for interpreting transient storage process, because of the non-identifiability of the TSM parameters and the low model performances ($RMSE \geq RMSE_{ADE}$ (Figure 5a-p, green and yellow dots).

5 Conclusion

There is a clear need in stream hydrology to better identify TSM parameters for simulating solute transport in streams. Here we address parameter identifiability by combining global identifiability analysis with dynamic identifiability analysis in an iterative approach to reduce parameter uncertainty in TSMs. Our results show that $v$ interacts with the transient storage parameters. Namely, when $v$ was variable together with the other TSM parameters, we found non-identifiable $A_{TS}$ coupled with identifiable $\alpha$. On the contrary, when $v$ was assumed equal to $v_{peak}$, $A_{TS}$ was found identifiable and $\alpha$ non-identifiable. We proved that non-identifiability of transient storage parameters can result in modelled BTC mimicking the ADE. Non-identifiable TSM simulations also severely misevaluated the solute retention time in the storage zone and the exchange flow with the storage zone, with a difference respectively up to four and two orders of magnitude compared to the results when the TSM parameters were identifiable.

We here validated our initial hypothesis that the BTC tail contains critical information on transient storage parameters, since we clearly reduced parameter uncertainty compared to the standard random-sampling approach coupled with the global identifiability analysis. Results obtained via inverse modelling approach are generally considered less informative about identifiability of TSM parameters compared to random-sampling approaches. However, random sampling approach rarely achieved identifiability in TSMs. As a complement to the existing body of literature, our work shows that the non-identifiability of $\alpha$ or $A_{TS}$ occurring in prior TSM studies might be related to a lack of modelled transient storage exchange due to the narrow peak of performances that can be easily missed by the rather small number of simulations compared to the investigated parameter range. When all the TSM parameters were identifiable, the best performing parameter sets and the evaluated transport metrics converged toward unique values, regardless the fact the velocity was considered as a variable parameter or equal to $v_{peak}$. This allowed us a robust assessment of the hydrological processes governing the solute transport in the investigated site. Our work highlights how both parameter evaluation and process interpretation in TSMs should be used with particular caution even if one parameter between $\alpha$ or $A_{TS}$ is found uncertain and non-identifiable. Stream
hydrologists are currently unable to obtain univocal physical process interpretation from modelling results due to contradictory interpretation of TSM parameters and lack of parameter identifiability in the published studies. Our work casts new lights on the opportunity to increase parameter identifiability and achieve stronger hydrologic interpretation of the processes governing solutes transport in streams.

6 Acknowledgements

This work was supported by the funding from the Luxembourg National Research Fund (FNR) for doctoral training (PRIDE15/10623093/HYDRO-CSI). We would like to acknowledge the financial support of the Austrian Science Fund (FWF) as part of the Vienna Doctoral Programme on Water Resource Systems (DK W1219-N28). We thank Ginevra Fabiani, Adnan Moussa, Laurent Pfister, Rémy Schoppach for their fruitful input and discussions.

7 References


Comparisons between transient storage and water balance analyses. Water Resources Research, 49(6), 3359–3374. https://doi.org/10.1002/wrcr.20148


We reported the used number of parameter sets and the parameter ranges, while in parenthesis it is reported the method used for the parameter sampling. “Double step” indicates that the sampling procedure was divided in two steps. In the first step, A varied across a broad range and in the second step, it was varied across a narrowed range to cover the most sensitive-range of the parameter domain. Each of the two steps has a number of simulations equal to half of the total number indicated in the table.

<table>
<thead>
<tr>
<th>Authors</th>
<th>Number of parameter sets</th>
<th>Range of TSM parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wagner &amp; Harvey, 1997</td>
<td>800 (Monte Carlo)</td>
<td>$A$ (m$^2$) 0.02-0.6  $D$ (m$^2$/s) 0.025-0.8  $A_{TS}$ (m$^2$) 0.01-2  $\alpha$ (1/s) 0.000005-0.001</td>
</tr>
<tr>
<td>Wagener et al., 2002</td>
<td>1,000 (Monte Carlo)</td>
<td>$A$ (m$^2$) 0.3-1.05  $D$ (m$^2$/s) 0.1-0.225  $A_{TS}$ (m$^2$) 0.1-0.5  $\alpha$ (1/s) 0.00035-0.0025</td>
</tr>
<tr>
<td>Wlostowski et al., 2013</td>
<td>2,000 (Monte Carlo)</td>
<td>$A$ (m$^2$) 0.5-1.0  $D$ (m$^2$/s) 0.5-1.5  $A_{TS}$ (m$^2$) 0.05-0.5  $\alpha$ (1/s) 10-4-10-3</td>
</tr>
<tr>
<td>Kelleher et al., 2013</td>
<td>42,000 (Double step Monte Carlo)</td>
<td>$A$ (m$^2$) 0.001-1.0 (in the second step, min and max $A$ values chosen from the best 1,000 results of the first step)  $D$ (m$^2$/s) 0.001-1.0  $A_{TS}$ (m$^2$) 0.001-0.01  $\alpha$ (1/s) 10-5-10-3</td>
</tr>
<tr>
<td>Ward et al., 2013</td>
<td>100,000 (Monte Carlo)</td>
<td>$A$ (m$^2$) +50% $A_{peak}$  $D$ (m$^2$/s) 0.0001-5  $A_{TS}$ (m$^2$) 0.01-10  $\alpha$ (1/s) 10-8-10-1</td>
</tr>
<tr>
<td>Ward et al., 2017</td>
<td>100,000 (Double step Monte Carlo)</td>
<td>$A$ (m$^2$) 0.1-1 (0.3-0.5 in the second step)  $D$ (m$^2$/s) 0.01-10  $A_{TS}$ (m$^2$) 0.01-1  $\alpha$ (1/s) 10-5-10-1</td>
</tr>
<tr>
<td>Kelleher et al., 2019</td>
<td>27,000 (LHS)</td>
<td>$A$ (m$^2$) 1.0 - 3.0  $D$ (m$^2$/s) 0.001 - 10  $A_{TS}$ (m$^2$) 0.01 - 1  $\alpha$ (1/s) 10-6 - 10-2</td>
</tr>
<tr>
<td>This manuscript</td>
<td>Second ADE – 35,000 (LHS)</td>
<td>$v$ (m/s) $v_{peak}$ - 0.8 - velocity of the first recorded increase of concentration in the BTC  $A$ (m$^2$) +20% $A_{peak}$  $D$ (m$^2$/s) 0.0001 - $D_{best}$ - 1.2</td>
</tr>
<tr>
<td>This manuscript</td>
<td>First TSM iteration – 115,000 (LHS)</td>
<td>$v$ (m/s) +50% $v_{ADE}$  $A$ (m$^2$) +50% $A_{ADE}$  $D$ (m$^2$/s) 0.0001 - $D_{ADE}$ - 2  $A_{TS}$ (m$^2$) 0.00001 - 20  $\alpha$ (1/s) 0.00001 - 0.1</td>
</tr>
</tbody>
</table>
Table 2: Summary of the TSM results. OTIS-MCAT results refer to the case $v = v_{\text{peak}}$ without any successive modification of the parameter space via dynamic identifiability analysis results. “Iterative TSM” indicate the best parameter sets obtained after the iterative TSM approach presented in Figure 1 and applied for the cases $v = \text{variable}$ and $v = v_{\text{peak}}$. The best TSM results are indicated with bold font.

<table>
<thead>
<tr>
<th></th>
<th>$v$</th>
<th>$A$</th>
<th>$D$</th>
<th>$\alpha$</th>
<th>$\Delta v_s$</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>E1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OTIS-P</td>
<td>0.0681</td>
<td>0.0395</td>
<td>0.0965</td>
<td>/</td>
<td>/</td>
<td>1.9423</td>
</tr>
<tr>
<td>OTIS-MCAT</td>
<td>0.0739</td>
<td>0.0364</td>
<td>0.0637</td>
<td>0.0006</td>
<td>0.0074</td>
<td><strong>0.6159</strong></td>
</tr>
<tr>
<td>Iterative TSM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$v = \text{variable}$</td>
<td>0.0728</td>
<td>0.0369</td>
<td>0.0522</td>
<td>0.0013</td>
<td>0.0073</td>
<td>0.7229</td>
</tr>
<tr>
<td>$v = v_{\text{peak}}$</td>
<td>0.0728</td>
<td>0.0339</td>
<td>0.0534</td>
<td>0.0013</td>
<td>0.0077</td>
<td>0.7681</td>
</tr>
<tr>
<td>E2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OTIS-P</td>
<td>0.1746</td>
<td>0.054</td>
<td>0.1599</td>
<td>/</td>
<td>/</td>
<td>0.9982</td>
</tr>
<tr>
<td>OTIS-MCAT</td>
<td>0.1774</td>
<td>0.0509</td>
<td>0.1151</td>
<td>0.0016</td>
<td>0.0077</td>
<td>0.4152</td>
</tr>
<tr>
<td>Iterative TSM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$v = \text{variable}$</td>
<td>0.1790</td>
<td>0.0523</td>
<td>0.1131</td>
<td>0.0018</td>
<td>0.0067</td>
<td><strong>0.3377</strong></td>
</tr>
<tr>
<td>$v = v_{\text{peak}}$</td>
<td>0.1774</td>
<td>0.0528</td>
<td>0.1154</td>
<td>0.0015</td>
<td>0.0065</td>
<td>0.3696</td>
</tr>
<tr>
<td>E3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OTIS-P</td>
<td>0.262</td>
<td>0.0874</td>
<td>0.2525</td>
<td>/</td>
<td>/</td>
<td>0.9894</td>
</tr>
<tr>
<td>OTIS-MCAT</td>
<td>0.275</td>
<td>0.081</td>
<td>0.1404</td>
<td>0.005</td>
<td>0.0144</td>
<td><strong>0.2544</strong></td>
</tr>
<tr>
<td>Iterative TSM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$v = \text{variable}$</td>
<td>0.2861</td>
<td>0.0818</td>
<td>0.1286</td>
<td>0.0064</td>
<td>0.0145</td>
<td>0.2697</td>
</tr>
<tr>
<td>$v = v_{\text{peak}}$</td>
<td>0.275</td>
<td>0.083</td>
<td>0.1603</td>
<td>0.0037</td>
<td>0.0123</td>
<td>0.3109</td>
</tr>
</tbody>
</table>
Figure 1: Conceptual modelling workflow. The parameters have the following unit of measurements: velocity $v$ [m/s], cross-sectional area $A$ [m$^2$], longitudinal dispersion coefficient $D$ [m$^2$/s], exchange coefficient $\alpha$ [1/s], area of the transient storage zone $A_{TS}$ [m$^2$].
Figure 2. Parameter values plotted against the corresponding RMSE values for the TSM results conducted for the tracer injections (a-e) E1, (f-j) E2, and (k-o) E3. (a-j) Green, yellow, blue and orange dots indicate results respectively for the first, second, third, and fourth TSM iterations. (k-l) Green dots indicate results for the first and second TSM iterations, while yellow, blue and orange dots indicate results respectively for the third, fourth, and fifth TSM iterations. Each TSM iteration was conducted via 115,000 parameter sets. The red dots indicate OTIS-MCAT results (best parameter set after the first TSM iteration for v equals vpeak) while the black dots indicate the best performing parameter value after the used iterative TSM approach. The horizontal black line indicates the RMSE\textsubscript{ADE} (Table 2). Vertical dashed red line indicates OTIS-P results, while the 95% confidence range for OTIS-P results are indicated via vertical grey areas.
Figure 3. Dynamic identifiability analysis of TSM parameters for the first TSM iteration (E1, v considered as a varying model parameter). (a), (c), (e), (g), (i) likelihood distribution as function of parameter values at each time step. Black line indicates the observed BTC, dashed black lines indicate the 90% confidence limits. (b), (d), (f), (h), (j) indicate parameter information content (red bars) at each time step. Black line indicates the observed BTC.
Figure 4. Same as Figure 3, but reporting dynamic identifiability results for the fourth and last TSM iteration (E1; \( v \) considered as a varying model parameter).
Figure 5. Same as Figure 2, but reporting TSM results when velocity was considered equal to $v_{\text{peak}}$. 
Figure 6. Same as Figure 3, but reporting dynamic identifiability results for the third TSM iteration (E1, $v = v_{\text{peak}}$).
Figure 7. Boxplot of the investigated transport metrics for the best 100 parameter sets for the three simulated experiments. (a-c) $L_y$ (d-f) $T_{m}$, (g,i) $q_s$, (j-l) $F_{med}$. Results are reported for (a, d, g, j, m) E1, (b, e, h, k, n) E2, and (c, f, i, l, o) E3. On the x axis we indicated the n-th TSM iteration. Blue and red boxplots indicate results when velocity was respectively a varying model parameter and when it was kept fixed and equal to $v_{peak}$. Red dots indicate the transport metric values obtained via the parameter sets with lower RMSE. The red and the black horizontal dashed lines indicate respectively the transport metric obtained using the OTIS-P results and OTIS-MCAT results.
Figure 8. Qualitative plots of the TSM parameter influence on different sections of the BTC. (a) and (b) qualitative parameter information content on the BTC for E1, (c, d) E2, and (e, f) E3. (g) Wagner and Harvey, 1997; (h) Wagener et al., 2002; (i) Scott et al., 2003; (j) Wlostowski et al., 2013; (k) Kelleher et al., 2013 for the case of a dispersive mountain stream (Case 1) and (l) Kelleher et al., 2013 for the case of a small low-flow mountain stream (Case 2). In plots (a-f) solid lines indicate an information content above 0.66 while dashed lines indicate an information content between 0.33 and 0.66. Plot (g) has been modified from Figure 7 in Wagener et al., 2002 in order to fit our 0.66 and 0.33 threshold classification in term of information content. Plots (k) and (l) indicate by solid and dashed lines if the parameters influence the model output by itself or through interactions (cfr. Section 6.1 Kelleher et al., 2013). Plots (g) and (i) describe the parameter influence evaluated via sensitivity evaluation (cfr. p. 1733, Wagner and Harvey, 1997) and dimensionless scaled sensitivities (cfr. Table 1 Scott et al., 2002), therefore the parameter influence is here described only using solid line. Plot (j) describes the parameter influence after DYNIA analysis, however information content plots were not reported by the authors, therefore the solid lines indicate the areas for the best performing parameters as indicated in Figure 2 of Wlostowski et al. (2013).
Appendix A - Parameter sensitivity and identifiability

The interpretation of the parameter space is based on the sensitivity and identifiability of the $i$-th parameter on the chosen model (the TSM) via a selected objective function used to compare model results with the observation (the BTC) (Kelleher et al., 2019; Wagener et al., 2003; Wagener & Kollat, 2007; Ward et al., 2017; Wlostowski et al., 2013). A parameter is called sensitive whenever a variation in the parameter value causes variations in the TSM performances (Kelleher et al., 2019). A parameter is identifiable whenever the best-fit value of that parameter is constrained on a relative narrow range across the entire distribution of the possible parameter values (Ward et al., 2017). To assess identifiability of TSM parameters, we used parameter vs likelihood plots, identifiability plots, regional sensitivity analysis plots and parameter distribution plots.

Parameter vs likelihood plots visualize the distribution of the investigated values of a certain parameter plotted against the corresponding values of the objective function (Wagener et al., 2003; Wagener & Kollat, 2007). Identifiable parameters are described in parameter vs likelihood plots by a univocal increase of model performances approaching a certain optimum-value of the parameter (Figure A1a). Non-identifiable parameters are described in parameter vs likelihood plots by a not-univocal increase of performances of the model in certain parameter range (Figure A1b). Parameter distribution plots show probability density function (PDF) divided by behavioural sets (from top 20% to top 0.1% of the results for the selected objective function) (Ward et al., 2017). Identifiable parameters are indicated by narrow range of the PDF relative to the smaller behavioural sets (top 0.1%, 0.5% and 1% of the results) compared to a wider range of the PDF relative to the larger behavioural sets (top 5%, 10% and 20% of the results) (Figure A1c). Non-identifiable parameters are defined by equally wide PDF for the different investigated behavioural sets (Figure A1d). Regional sensitivity analysis plots are obtained after dividing the population of the parameter by behavioural sets (from top 10% of the results to top 1% of the results with 1% step for the selected objective function, Ward et al., 2017; Kelleher et al., 2019). Each objective function population so obtained was transformed into cumulative distribution functions (CDFs) for equal size bins of the parameter range (Kelleher et al., 2019; Wagener & Kollat, 2007). Sensitive parameters are identified by CDF for the top 1% of the results deviating from the CDF for the top 10% of the results (Figure A1e). If the CDFs lay on the 1:1 line, then the objective function is uniformly distributed across the parameter range which indicates parameter unsensitivity (Figure A1f). Identifiability plots display the CDF of the objective function across the selected parameter range (Wagener et al., 2002; Ward et al., 2017). The slope of the CDF will be higher in the parameter interval where the model is more sensitive to that parameter. The measure of the local gradient of the cumulative distribution will be represented by the height of the bar plot in each equally-sized bin across the parameter space. Higher bars and steeper gradients of the CDF line indicate greater model performances in that parameter range and, therefore, parameter sensitivity and identifiability (Figure A1g). On the contrary, equal eight of the bars and similar gradients of the CDF line indicate that the parameter is unsensitive and non-identifiable (Figure A1h).

The plots used to address the global sensitivity analysis indicate parameter identifiability and sensitivity on the entire observed BTC, however they are unable to address if the $i$-th parameter describes the process it is meant to represent or if the role of the $i$-th parameter on the model is constant in time (Wagener & Kollat, 2007). To address identifiability and sensitivity of the $i$-th parameter on the different sections of the BTC we applied DYNIA algorithm which steps are reported in Figure A2 (Wagener et al., 2002).
Figure A1: Examples of the four types of visualizations intended for parameter identifiability and sensitivity with the plots in the first column (a, c, e, and g) reporting an example of plots for sensitive and identifiable parameter and plots in the second column (b, d, f, and h) reporting an example of plots for insensitive and non-identifiable parameter. (a) and (b) parameter vs likelihood plots; (c) and (d) parameter distribution plots for the top 20, 10, 5, 1, and 0.1% of the results; (e) and (f) regional sensitivity analysis plots from the top 1% to the top 10% of the results; (g) and (h) identifiability plots for the top 1% of the model results.
Figure A2. Dynamic identifiability analysis algorithm flowchart. (a) The BTC is subdivided in moving windows (size equal to three times the BTC timestep, Wagener et al., 2002); (b) In each moving window the likelihood (efficiency) of every TSM simulation is evaluated via mean absolute error (Wagener & Kollat, 2007); (c) an efficiency-threshold is chosen (e.g. top 10%); (d) for the chosen model results, the cumulative distribution function is built for each investigated parameter; (e) steps from (b) to (d) are repeated for each moving window and model likelihood for the investigated parameter is plotted over time (white: minimum likelihood; black: maximum likelihood). (f) cumulative
distribution function of the parameter distribution is plot vs the observed BTC together with 90% confidence limits. Narrow limits indicate identifiable parameter while wide limits indicate unidentifiable parameter. (g) a second plot reports the metric of one minus the normalized distance between the 90% confidence limits. Small values of this metric indicate that the selected time window contain a narrow identifiability range for the investigated parameter and, therefore, that it is informative on that part of the BTC (Wagener et al., 2002).