Article:

Sonnenwald, F., Stovin, V.R. and Guymer, I. (2015) Deconvolving Smooth Residence Time Distributions from Raw Solute Transport Data. Journal of Hydrologic Engineering, 20 (11). ISSN 1084-0699

https://doi.org/10.1061/(ASCE)HE.1943-5584.0001190

Deconvolving Smooth Residence Time Distributions from Raw Solute Transport Data

F. Sonnenwald[1], V. Stovin[2], I. Guymer[3]

ABSTRACT

A Residence Time Distribution (RTD) provides a complete model of longitudinal mixing effects that can be robustly derived from experimental solute transport data. Maximum entropy deconvolution has been shown to recover RTDs from pre-processed laboratory data. However, data pre-processing is time consuming and may introduce errors. Assuming data were recorded using sensors with a linear response, it should be possible to deconvolve raw data without pre-processing. This paper uses synthetically generated 'raw' data to demonstrate that the quality of the deconvolved RTD remains satisfactory when pre-processing steps involving data cropping or calibration are skipped. Provided noise levels are relatively low, filtering steps may also be omitted. However, a rough subtraction of background concentration is recommended as a minimal pre-processing step.

Deconvolved RTDs often include small scale fluctuations that are inconsistent with a well-mixed fully turbulent system. These are believed to be associated with over-sampling and/or unsuitable interpolation functions used in the maximum entropy deconvolution process. This paper describes a new interpolation function—Linear interpolation with an Automatic Moving Average (LAMA)—and demonstrates that, in combination with fewer sample points (e.g. 20), it enables smoother RTDs to be generated.

The two improvements, to deconvolve raw data and to generate smoother RTDs, have

[1]Research Associate, Department of Civil & Structural Engineering, The University of Sheffield, Mappin St., Sheffield S1 3JD, UK, e-mail: f.sonnenwald@sheffield.ac.uk

[2]Reader, Department of Civil & Structural Engineering, The University of Sheffield, Mappin St., Sheffield S1 3JD, UK, e-mail: v.stovin@sheffield.ac.uk

[3]Professor, School of Engineering, University of Warwick, Coventry CV4 7AL, UK, e-mail: i.guymer@warwick.ac.uk

been validated with experimental data. Raw solute transport traces collected from a river were deconvolved after background subtraction. The deconvolved RTDs compare favourably with those generated from the more traditional ADE and ADZ models, but provide more detail of mixing processes. A laboratory manhole solute transport data set was deconvolved with and without pre-processing using 40 sample points and linear interpolation. The raw data was also deconvolved using 20 sample points and LAMA interpolation. The two sets of RTDs deconvolved from the raw data show the same mixing trends as those deconvolved from pre-processed data. However, those deconvolved with LAMA interpolation and 20 sample points are significantly smoother.

Keywords: Solutes, Dispersion, Mixing, Hydraulic models, Transfer functions, Residence time

INTRODUCTION

Solute transport traces, or temporal concentration profiles, recorded from complex flow systems (e.g. rivers or manholes) provide a description of the mixing processes occurring and are often analysed using parametrised models, e.g. fitting the Advection-Dispersion Equation (ADE) model or the Aggregated Dead Zone (ADZ) model (Rutherford 1994). Recent work has highlighted the use of Residence Time Distributions (RTDs) as a significantly more flexible approach to modelling solute transport. In this context, the RTD can exactly describe the mixing processes within a specific reach or structure (Guymer and Stovin 2011), and thereby provide additional insight into the mixing processes, e.g. Gooseff et al. (2011); Stovin et al. (2010a).

The RTD is frequently used in chemical engineering to describe reaction mixers (Denbigh and Turner 1984), and is analogous to the instantaneous unit hydrograph (Sherman 1932). It is the system mixing response to a Dirac tracer pulse (instantaneous input) and is often referred to as a non-parametric model. Levenspiel (1972) describes the RTD as the distribution of lengths of time fluid takes to pass through a system. This definition of the RTD, used in this paper, assumes a linear time-invariant system, i.e. steady-state conditions, and therefore stationarity of the flow field. As such, the RTD can be expressed through the convolution integral in Eq. (1), where $E(\tau)$ is the RTD, $u(t)$ is the upstream concentration profile, and $y(t)$ is the downstream concentration profile.

$$y(t) = \int_{-\infty}^{\infty} E(\tau)u(t-\tau)d\tau \quad (1)$$

The Cumulative Residence Time Distribution (CRTD) is the integral of the RTD over time, notated as $F(\tau)$. In other hydrology contexts, the RTD as defined above is instead referred to as a Travel (or transit) Time Distribution, e.g. McGuire and McDonnell (2006). RTDs may also be used to explore catchment-scale processes that are not directly observable, e.g. groundwater transport (Rinaldo et al. 2011).

Given regularly sampled paired time-series concentration data records for $u(t)$ and $y(t)$, solving for the RTD in the convolution integral is an ill-posed problem (Hansen 1998). The general solution is known as deconvolution, i.e. the reverse process of convolution. This is a common problem in many areas, where the identification of the underlying transfer function between two signals is desired. There are multiple approaches to deconvolution; see Madden et al. (1996) for a detailed review. To date, two main deconvolution approaches have been applied to solute transport data, geostatistical deconvolution (Fienen et al. 2006) and maximum entropy deconvolution (Stovin et al. 2010b). This paper presents two improvements to the maximum entropy deconvolution method to further enhance its suitability as a generic approach to the deconvolution of solute transport data (Sonnenwald 2014). These improvements are:

1. The ability to deconvolve raw data, i.e. without the requirement of pre-processing.
2. The ability to produce smoother RTDs, by changing the interpolation function and identifying appropriate numbers of sample points.

After a brief introduction to maximum entropy deconvolution, the potential to deconvolve raw data is investigated. Subsequently, improvements to RTD smoothness are investigated.

Finally, two validation cases are presented showing the benefits imparted by the proposed improvements.

Maximum entropy deconvolution

Maximum entropy deconvolution is a process by which non-linear optimisation is used to refine an estimate of the RTD based on upstream and downstream concentration profiles. Following Skilling and Bryan (1984), a Lagrangian function is created as a combination of an entropy function and a constraint function. By maximising the Lagrangian, a solution for the RTD is derived. This method is outlined below, and detailed in Stovin et al. (2010b), Sonnenwald et al. (2014), and Sonnenwald (2014).

$$S(\hat{E}) = -\sum_{i=1}^{N} \left(\frac{\hat{E}_i}{\sum_{j=1}^{N} \hat{E}_j} \right) \ln \left(\frac{\hat{E}_i / \sum_{j=1}^{N} \hat{E}_j}{r_i} \right) \tag{2}$$

$$C = \frac{\sum_{i=1}^{N} (\hat{y}_i - y_i)^2}{\sum_{i=1}^{N} y_i^2} \tag{3}$$

$$L(\hat{E}, \lambda) = C + \lambda S(\hat{E}) \tag{4}$$

To solve for the estimated RTD $\hat{E}$, Eq. (2)–(4) are implemented in MATLAB (The MathWorks Inc. 2011; Schittkowski 1986) as a minimisation problem and then solved using the `fmincon` function with an active set algorithm. S is the objective function, an entropy function that evaluates shape and helps to encourage a smooth RTD. N is the number of points in the RTD. r is a next-neighbour moving average of $\hat{E}$ (Hattersley et al. 2008). C is a constraint function, which evaluates the goodness-of-fit of the predicted downstream profile $\hat{y}$ compared to the recorded profile using a variation of the R_t^2 function (Young et al. 1980). L is the Lagrangian function. λ is the Lagrange multiplier, which is determined at each iteration by a gradient descent method as part of `fmincon` (The MathWorks Inc. 2011).

The deconvolution problem is computationally simplified by solving only for a sub-sampled RTD in the entropy function, with linear interpolation used to estimate the remainder of the RTD between sample points. Sub-sampling is based on an initial guess of the RTD provided by inverse fast Fourier transform deconvolution, with more sample points being placed where the slope of the guessed RTD is greater. Sonnenwald et al. (2014) additionally recommended the following settings: 40 sample points, 350 iterations, and the R_t^2 constraint function.

Evaluation of RTD quality

Deconvolved RTDs may be evaluated based on their predictive capability and on their smoothness. Predictive capability is evaluated by convolving the deconvolved RTD with the upstream profile used in the deconvolution process. The resulting predicted downstream profile is compared to the original downstream profile, i.e. the output is compared to the data used to generate it. For this comparison, Sonnenwald et al. (2014) suggest the use of the Nash-Sutcliffe Efficiency Index, R^2, where a value of 1.0 indicates a perfect match and $R^2 \leq 0$ indicates no correlation (Nash and Sutcliffe 1970). Smoothness of an RTD may be evaluated by measuring its entropy using Eq. (2) (Sonnenwald et al. 2014). Values closer to zero indicate a smoother RTD.

Where synthetic trace data has been generated from a known RTD, a third evaluation of a deconvolved RTD is possible: a direct comparison between the original and deconvolved RTDs. Sonnenwald (2014) suggests that the Average Percent Error (APE) metric (Kashefipour and Falconer 2000) is more suitable for comparing RTDs as it is significantly more sensitive to differences between profiles than R^2. APE = 0 indicates a perfect correlation, while APE $\geq$ 100 indicates no correlation.

THE DECONVOLUTION OF RAW DATA

Introduction

Raw data is the information collected directly from instrumentation and recorded as-is during experimental laboratory and field work, e.g. voltage readings from a fluorometer. In most cases raw data must be pre-processed before it can be analysed. Saiyudthong

(2003) describes the pre-processing of laboratory solute transport data as a complex chain of operations consisting of calibration, subtraction of background concentration levels, filtering, and cropping the data record (reducing the length, or duration, of the record through data cut-off based on definitions of experiment start and end times).

Researchers can spend significant amounts of time developing pre-processing steps that take into account their specific experimental setup. Guymer and O'Brien (2000) provide a long and detailed description of fluorometer calibration, smoothing, and temporal averaging. Kasban et al. (2010) clearly outline and document several pre-processing steps used when obtaining the RTD using radiotracers. Other work only summarises pre-processing, e.g. Guymer (1998), or effectively ignores it, e.g. Wallis and Manson (2005). While pre-processing is generally not the specific focus of the research, it can have an impact on the quality of the research findings. Joo et al. (2000) show how better pre-processing of training data for an artificial neural network used in predicting coagulant dosing rate leads to a better learning rate, reduced error, and improved predictive capability. Poor pre-processing, e.g. excessive smoothing or cropping, may introduce errors or remove useful information about the system.

Sonnenwald et al. (2014) demonstrated that maximum entropy deconvolution robustly identifies the RTD from pre-processed trace data collected from a variety of mixing systems. Assuming a linear instrument response, deconvolution of raw data should prove to be equally robust, allowing for a reduction in the time spent on pre-processing and potentially reducing sources of errors. This section demonstrates the applicability of maximum entropy deconvolution to raw solute transport data through a sensitivity analysis and, as a result, recommends a minimum required level of pre-processing.

Methodology: Raw solute transport data sensitivity analysis

To investigate how input data impacts on the deconvolved RTD, a sensitivity analysis was carried out. A perfect synthetic trace, i.e. a pre-processed solute transport trace, was generated and then typical pre-processing steps were applied in reverse to create synthetic 'raw' time-series. The raw data were then deconvolved.

The recovered RTDs were scaled according to the mass-balance of the data they were derived from and then evaluated for predictive capability and quality using R^2 and APE respectively. Although Sonnenwald et al. (2014) concluded that 40 sample points should generally be selected for deconvolution, subsequent work (described in the second part of this paper and in Sonnenwald (2014)) has shown that smoother RTDs can be described using only 20 sample points, with no loss of predictive capability. Therefore, 20 sample points were used here.

Synthetic data

To form a perfect synthetic base solute transport trace, an upstream concentration profile has been convolved with a known RTD to create a downstream profile. This trace, Figure 1a, is analogous to pre-processed data. The upstream profile was a Gaussian distribution with $\mu = 24.4$ s, $\sigma = 5.5$ s, and $dt = 0.15$ s. An RTD was synthesised as a Guassian distribution with $\mu = 13.7$ s, $\sigma = 3.1$ s, $\int_{-\infty}^{\infty} E(t)dt = 1$. The downstream profile is created by convolution using Eq. (1). Concentration levels below 10^{-4} were treated as below the limit of detection and set to 0. The synthetic trace is representative of data recorded from an experimental pipe configuration with an 88 mm diameter, 5 l/s flow, and a distance between instruments of 2.7 m (Guymer and O'Brien 2000).

Pre-processing of raw solute transport traces generally consists of four steps: apply a calibration function; determine and subtract background concentration levels; filter noise; and determine the start and end of the signal data (i.e. experimental event), then crop data points before and after. The process of reversing these steps to create synthetic raw data is outlined below. Figure 1b shows an example synthetic raw trace after reversed pre-processing.

Data extension

Laboratory data is often recorded for a longer period than necessary to ensure that the experiment is fully captured. Here, the trace is synthetic and therefore complete. To simulate raw data, extra data points have been added to the start and end of the base trace. Data extension has been added as 0%, 10%, and 20% of trace length before and after the trace. Zeros were used in order to retain mass-balance. Figure 1b has a 10% extension.

Addition of noise

Recorded data is subject to random variation, i.e. noise, either from within the system or due to the instrumentation. The synthetic base trace has no noise, so to simulate realistic raw data, noise has been added according to a truncated normal distribution. The maximum noise level k is defined in terms of the peak upstream concentration, equal to 0%, 5%, 10%, or 20%. Noise is assumed to be normally distributed with $\mu = 0$ and $\sigma = k/3$ between the limits of $[-k, k]$. 20% noise is representative of a maximum of 1 V of noise for a typical 5 V sensor and can be considered a conservatively high value. Figure 1b has 10% noise.

Addition of background

Background concentration refers to a constant or near-constant concentration level measured independently of any experimental event. It is often present in laboratory setups, particularly in those utilising recirculating systems. Subtraction of background is usually carried out to leave only the change in concentration caused by the experiment. This can be done using an assumed mean value or linear function derived from the recorded concentration levels.

To simulate raw data, a background concentration has been added to the base trace, either as a constant value or varying linearly with time (sloped background). Constant background takes the form of a mean background concentration level, defined as a fraction of peak upstream concentration. Values of 0%, 10%, and 20% have been used. Background slope has been applied on top of each mean background level as an additional -2.5% increasing to 2.5% of peak upstream concentration for positive slope or 2.5% decreasing to -2.5% for negative slope. Figure 1b has a 10% mean background with an increasing slope.

Uncalibration

Calibrating raw data for linear sensors consists of multiplication by a known factor to relate sensor reading to concentration level. To simulate raw data, multiplication by an 'uncalibration' factor has been applied to take the base trace out of mass-balance. Factors have been chosen independently for the upstream and downstream profiles so that the peak values are the combinations of 2, 3, 4 or 5 V (16 total). In Figure 1b, both profiles have been uncalibrated to 3 V.

Results: Impact of pre-processing on deconvolution

The combinations of data extension, noise, background (sloped and constant), and uncalibration resulted in 1,728 synthetic raw traces being deconvolved.

Predictive capability of RTDs deconvolved from synthetic raw data

Figure 2a shows R^2 values comparing the base perfect downstream profile with predicted downstream profiles generated using the perfect upstream profile and the scaled recovered RTD. Each individual column corresponds to a different background slope (i.e. negative, no slope, or positive) and contains all combinations of uncalibration. Each group of 3 columns represents a mean background level, while every nine columns represent a specific noise level. All R^2 values indicate extremely good predictive capability, with the overall mean $R^2 = 0.9874$. This indicates a wide range of synthetic raw data can successfully be deconvolved to obtain a reasonable predictive model without any requirement for pre-processing.

There is a clear trend of decreasing predictive capability with increasing noise and increasing mean background level. The greater spread in the columns further to the right indicates that the impact of uncalibration increases with greater background levels and noise, but it does not appear to be systematic.

Background slope and extension have relatively little impact on predictive capability, but do vary systematically and can be explained. A positive background slope leads to lower R^2 values than a negative background slope when mean background level is 0%, independent of uncalibration. The negative portion of the downstream profile with a negative background slope cannot be matched in the deconvolution process, while the greater positive portion due to a positive background slope can be. RTDs deconvolved from the latter will more greatly over predict mass-balance than the former will under predict it. The greater over-prediction results in poorer R^2 values.

The increase of R^2 with extension at no background and no noise may be explained by the wider spacing of sample points that results from the same 20 points being distributed over a longer profile. This reduces the relative potential for noise, leading to an improvement in RTD quality with extension. When there is non-zero background, there is a consistent period of time at the start of the profile when the downstream prediction does not match the recorded synthetic raw data. This period is fixed in length regardless of total duration and therefore, as extension increases, represents a proportionately smaller period of time. The period of poor fit therefore has less negative influence on the R^2 value at greater extension, increasing R^2 values overall.

Quality of RTDs deconvolved from synthetic raw data

Mean APE values for the comparison between the known and deconvolved RTDs are shown Figure 2b. The effects of extension and uncalibration have been combined as they have no systematic impact on predictive R^2 value. The APE results show less variation than the predictive capability results, but can still be grouped similarly. This lower variation suggests the deconvolved RTDs have similar shapes despite the variation in input data quality. The lowest observed mean APE value is 8.21, indicating that the deconvolved RTD will always vary from the actual RTD. Background concentration appears to have a greater impact on RTD quality than noise, as the increase in APE observed when the background level increases from 10% to 20% is generally greater than when the noise level increases by the same amount. APE value generally increases less between 0% and 10% for both noise and background.

Visual inspection of RTDs deconvolved from synthetic raw data

Figure 3 shows representative deconvolved cumulative residence time distributions (CRTDs) for three cases. The first case has 5% noise and no background, the second case has 10% noise and 10% mean background (no slope), and the third case has 20% noise and 20% mean background (no slope). The third case CRTD includes values greater than 1, which in this case indicates a failure of the deconvolution method to cope with raw data that has high background concentration levels and high noise. Overall, the figure shows a reduction in CRTD quality (i.e. increasing APE) with increased noise and background. This confirms the results shown in Figure 2, and together they suggest 10% noise and 10% background levels as limits for deconvolved RTDs. The differences between 0% and 10% noise and background are much smaller than those between 10% and 20%. The 10% limit corresponds to approximate cut-offs of $R^2 = 0.995$ and APE = 35 for this data set. Lower noise and background levels should be preferred to keep RTD quality high.

Discussion: Recommendations for deconvolving raw data

When deconvolving the synthetic raw data, predictive capability of the deconvolved RTD is generally good. Of the four pre-processing steps examined (data extension, noise, background, and uncalibration), extension and uncalibration have been shown to have no systematic impact on the deconvolved RTD, suggesting no pre-processing is necessary for these. However, increased noise and background concentration level both degrade predictive capability and RTD quality in a similar fashion. As a result, 10% noise and 10% background have been suggested as input data quality limits for successfully deconvolving an RTD. These values are applicable to most types of input data since, as the RTD is non-parametric, the deconvolution process is independent of system scales and instead dependent on data characteristics.

Background concentration is a common occurrence. It has a high impact on both predictive capability and RTD quality, and is therefore important to address. Background concentration should be subtracted as part of minimal pre-processing. This subtraction should take into account background slope, as increasing background concentration levels with time particularly influence the deconvolved RTD. However, it need not be overly precise, as at very low background levels noise will have a greater impact on the deconvolved RTD.

Pre-processing for noise is unnecessary provided background subtraction has taken place. At 10% noise with no background, the RTD retains excellent predictive capability and satisfactory RTD shape. In the event of significantly greater noise levels, some filtering should be applied. Additional steps of down-sampling or cropping may be advisable for computational reasons when time-series are of significant length. However, in most cases no significant pre-processing should be required.

Assuming that minimal pre-processing (in the form of subtracting background concentration level, taking into account background slope) is applied, this investigation has demonstrated that raw data can be successfully deconvolved.

ENHANCED RTD SMOOTHNESS

Introduction

To date, RTDs derived with maximum entropy deconvolution have typically been presented in their cumulative form as CRTDs. While this aids interpretation of the underlying mixing processes, the CRTD does not necessarily reveal small fluctuations in the RTD, e.g. those highlighted in Figure 4. These fluctuations numerically cancel out during convolution and so do not impact on the predictive capability of the RTD, but may potentially affect interpretation of the bulk mixing processes.

The presence of fluctuations in deconvolved RTDs highlights a potential issue with the use of maximum entropy deconvolution, namely that a deconvolved RTD might not accurately represent some system characteristics. Considering that the cumulative effect of turbulence in most systems acts to smooth out fluctuations, if the deconvolution process were modified to minimise fluctuations, the quality of the resulting hydrodynamic interpretation should improve. A smoother RTD would aid interpretation as a more convincing representation of mixing processes.

Fluctuations in deconvolved RTDs can in some cases be attributed to over-sampling of the sub-sample points used in the deconvolution process. Over-sampling occurs when too many sample points have been specified so that some points end up tightly clustered, which tends to result in significant fluctuation between adjacent sample point values. This section proposes an enhancement to maximum entropy deconvolution in the form of a new interpolation function to smooth the RTD and a re-evaluation of the number of sample points to reduce over-sampling, both of which should reduce fluctuations. Two alternative interpolation functions are proposed and a sensitivity analysis is carried out.

Interpolation

Interpolation is used by the maximum entropy deconvolution process to generate $\hat{E}$, the estimated RTD. This is a critical part of the goodness-of-fit comparisons that are performed multiple times during each iteration. The interpolation function therefore plays an important role in influencing the deconvolved RTD.

Linear interpolation (currently used), is the simplest type of interpolation. A straight line is drawn between the two closest sample points, and the interpolated data points are evaluated to be on that line. This has the benefit of being conceptually simple and easily executed. There are however, several more complex interpolation functions including Inverse Distance Weighting (IDW) and the Kriging Estimation Method (KEM), which are commonly used functions in GIS applications (Zimmerman et al. 1999). In IDW the point being interpolated is defined to be more closely related to nearby points and less so to further points. In the KEM, the point being interpolated is derived as the result of a statistical model that estimates the relative importance of nearby points.

In cubic interpolation (Fritsch and Carlson 1980), the sample points are used to estimate the derivatives of a cubic function that passes between them. The derivatives are then used to estimate the values at points being interpolated. Splines can also be used for interpolation. They are considered a subset of polynomial interpolation that are specified to have continuous $n-1$ derivatives (de Boor 1978). A cubic spline has continuous first and second derivatives with the result that there are fewer possibilities for the interpolated line than using cubic interpolation.

While any of the above interpolation functions could be used in the deconvolution process to smooth the RTD, a more pragmatic approach to smoothing is to apply a moving average after linear interpolation, i.e. linear interpolation with an automatic moving average (LAMA), outlined below. Initial investigation (Sonnenwald 2014) has shown this, and cubic interpolation, to be the most promising means of smoothing in this context and they are investigated further below.

Methodology: RTD smoothness improvement sensitivity analysis

A sensitivity analysis for evaluating improvements to RTD smoothness as a result of changing interpolation function and number of sample points has been carried out. Linear interpolation, cubic interpolation, and LAMA interpolation have been used to deconvolve three different solute transport traces. They have been deconvolved at between 15 and 45 samples, as Sonnenwald et al. (2014) indicated that this range produced the smoothest results.

The three solute transport traces correspond to: a solute transport trace collected from an 800 mm diameter surcharged manhole with flow at 1 l/s and surcharge at 268 mm (Guymer et al. 2005); a 24 mm pipe trace with transitional turbulent flow at 0.221 l/s (Hart et al. 2013); and a completely synthetic Gaussian trace. The latter was created specifically to demonstrate the effects of over-sampling. Assuming $dt = 1$ s, the upstream profile has $\mu = 25$ s, $\sigma = 5$ s. The RTD has $\mu = 50$ s, $\sigma = 16.67$ s. The area under both curves was normalised to 1 and the downstream profile created using Eq. (1).

Implementing LAMA, linear interpolation with an automatic moving average

The MATLAB `interp1` function (The MathWorks Inc. 2011) has been used for cubic interpolation. However, as there is no convenient moving average function. Eq. (5), describing a moving average, has been implemented. $E_{MA}(x)$ is the RTD with a moving average applied, 2α is the length of the moving average window size, and τ is an integration variable.

In other words, the value $E_{MA}(x)$ is the mean of values of E from $E(x-\alpha)$ to $E(x+\alpha)$.

$$E_{MA}(x) = \int_{x-\alpha}^{x+\alpha} \frac{E(\tau)}{2\alpha} d\tau \tag{5}$$

In terms of the deconvolved RTD, a moving average can be considered a low-pass filter and the window size 2α a frequency cut-off. When applied to an RTD, high-frequency fluctuations shorter than the window size are removed, while the lower frequency mixing response is retained. Therefore, choice of window size is important. If 2α is too long, characteristics of the RTD (e.g. the peak associated with short-circuiting) may be overly attenuated. Conversely, a window size that is too short will not reduce fluctuations in the RTD.

A method of directly estimating a suitable window size from an RTD has been developed so that the moving average filters only the higher frequency fluctuations. This is shown in Eq. (6), where t_p is the time of the peak of the RTD, and t_β is the time at which the CRTD is equal to a fraction β of the CRTD at the peak of the RTD, i.e. $t_\beta = \beta F(t_p)$. As a result of the parameters used in Eq. (6), only the rising limb of the RTD affects the window size estimate. This reduces the risk of an asymmetric distribution (e.g. a non-Gaussian tail) skewing the window size estimate.

$$2\alpha = t_p - t_\beta \tag{6}$$

An initial evaluation of different values of β was conducted by deconvolving a collection of solute transport data for values of $\beta = \{0.05, 0.10, 0.15, 0.20\}$. Table 1 reports average R^2 depending on β. While in many cases there was no difference in performance, for some cases there is a drop in predictive capability when $\beta = 0.05$. This indicates that there is a penalty to predictive capability for using a low cut-off value (i.e. a longer window). All values of β had entropy values with the same order of magnitude and as such a value of 0.10 for β is a reasonable balance between smoothness and predictive performance under a variety of conditions.

Within the deconvolution process, a new estimate of window size is made every time LAMA interpolation is applied. However, as finding the RTD is an optimisation process there are cases where an impossibly large window size can be estimated, which would then cause deconvolution to fail. For these scenarios, a maximum window size estimate ($2\alpha_{max}$) has been specified. If $2\alpha > 2\alpha_{max}$, $2\alpha = 2\alpha_{max}$. $2\alpha_{max}$ has been defined as twice the mean gap in sample point spacing around the peak of the guessed RTD used to sub-sample the RTD.

Results: Impact of interpolation function and number of sample points on RTD smoothness

To investigate the impact of interpolation function and number of sample points on RTD smoothness, 279 deconvolutions were carried out—the combination of 3 traces, 3 interpolation functions, and 31 different numbers of sample points. The mean R^2 value overall was 0.9992 with a minimum value of 0.9816 and maximum value of 1.0000, showing that all deconvolved RTDs form excellent predictive models. Figure 5 presents the predictive R^2 and entropy values, the latter on a log scale, for each combination of interpolation function and number of sample points.

The distribution of R^2 values shows an increasing trend in predictive capability with more sample points. The relatively limited spread of R^2 values at a given number of sample points shows that in most cases interpolation function has a lower impact than number of sample points on predictive capability. The systematic variation in R^2 for the Synthetic data is caused by linear and cubic interpolation treating sample point values as observations through which the RTD must pass, while LAMA smooths these out. Overall there is no clear relationship between interpolation function and R^2 value, which suggests that the choice of interpolation function should primarily be guided by entropy.

Entropy values show increasing smoothness (i.e. values closer to zero) with fewer sample points. This is expected given the results of Sonnenwald et al. (2014) and confirms the impact that number of sample points can have on RTD quality. Independent of the number of sample points, the interpolation function also significantly impacts on entropy. LAMA interpolation performs best, with entropy values significantly and consistently closer to zero. Cubic and linear interpolation both show greater entropy, indicating they are less smooth. This suggests LAMA interpolation as the best choice for a smooth RTD.

Visual inspection of smoothed RTDs

Higher R^2 values and entropy values closer to zero are to be preferred as being representative of smoother, higher quality RTDs. Number of sample points should be chosen (in combination with interpolation function) to provide a balance of predictive capability and smoothness. In this instance, with fewer than 20 sample points there is no appreciable improvement in entropy when using LAMA, and as a result there is no reason to reduce R^2 further by using fewer sample points.

Figure 6 shows RTDs deconvolved with 20 sample points to be visibly smoother than the original 40 sample points. The figure also shows RTDs to be smoothest when using LAMA interpolation, with linear interpolation next smoothest and cubic interpolation least smooth. RTD shape is consistent, independent of the interpolation function and the number of sample points.

Almost all of the 40 sample point RTDs show signs of over-sampling, with variation around the 20 sample point deconvolved RTDs. In the case of the synthetic trace, over-sampling is also visible at 20 sample points using linear and cubic interpolation, but not in the RTD deconvolved with LAMA interpolation. The LAMA interpolated RTD has an APE value of 1.08 indicating it is very close to the original RTD used to generate the synthetic pipe trace. In comparison, the cubic and linear interpolated RTDs have APE values of 10.26 and 6.24 respectively, despite similar predictive capability.

There is the potential that reduced numbers of sample points and LAMA interpolation may constrain the RTD, affecting hydraulic interpretation. However, there is no direct indicator of what RTD provides the “correct” hydraulic interpretation without additional observations. Ideally multiple dye injections should be recorded and deconvolved at both higher and lower numbers of sample points to reveal key system characteristics.

Discussion: Recommendations for improving RTD smoothness

Deconvolved RTDs generated using all combinations of interpolation function and number of sample points result in RTDs with good predictive capability. R^2 decreases in an approximately linear trend with decreasing number of sample points, although the relative differences in R^2 are quite small. Entropy values of the LAMA interpolation function are consistently closer to zero, reflecting smoother RTDs than either linear or cubic interpolation. Visual inspection of the deconvolved RTDs shows that RTD shape remains consistent across interpolation function and number of sample points.

The increased smoothness of the deconvolved RTDs is more consistent with expected system dynamics, and the removal of over-sampling effects is desirable for similar reasons. As the effects of turbulent mixing occur more rapidly than the system time-scale in most cases, the system is expected to be well mixed and therefore have a smooth RTD. Additionally the convolution process acts to average out rapidly changing fluctuations, and therefore they cannot be inferred from the deconvolution process. The result of a smoother RTD is one that more accurately reflects system hydrodynamics. Smoother RTDs are also easier to interpret and cross-compare.

RTD smoothness did not increase at fewer than 20 sample points, while R^2 value in some cases dropped. Therefore, 20 sample points is recommended as a reasonable compromise between predictive capability and entropy performance for obtaining a smooth RTD. More sample points may be necessary when the system the RTD is describing is more complex (e.g. multiple peaks). LAMA interpolation clearly results in the smoothest RTDs for each solute transport trace deconvolved. The synthetic data particularly demonstrates how the impact of over-sampling can be reduced using LAMA interpolation. Fewer sample points and LAMA interpolation have both clearly been shown to improve RTD smoothness and can therefore be recommended.

VALIDATION

Two validation cases have been examined. First, river data has been deconvolved with the proposed improvements. Secondly, the proposed improvements have been applied cumulatively to an experimental manhole data set.

Deconvolution of river solute transport data

The UK Environment Agency has compiled a national database of river solute transport data, including solute traces (Guymer 2002). The traces recorded in the database were done so under varying conditions, e.g. different equipment, background concentration, etc. It presents a unique pre-processing challenge as for most types of analysis, data from each source must be treated differently. Trace data from the national database, recorded from the River Swale (NE17) at approximately an 18 m^3/s flow rate, have been deconvolved.

Figure 7 shows the raw data from the River Swale at five monitoring stations. As the data was recorded at one minute intervals, background subtraction has been done using the first data point as being representative of constant background concentration levels. As the trace data was cut-off at each monitoring station, additional data points have been added before and after (as appropriate) using zeros to form a set of paired temporal concentration profiles of the same duration for each reach. The data were subsequently deconvolved using LAMA interpolation with 20 sample points.

Figure 8 shows the RTDs that describe each of the four reaches, i.e. the RTDs deconvolved using the traces from the first and second, second and third, etc., monitoring stations as the upstream and downstream traces. The national database also includes optimised travel time and dispersion values suitable for use with the ADE model (a Gaussian transfer function, Rutherford (1994)) and with the ADZ model (a delayed exponential decay function, Beer and Young (1983)). RTDs generated from these optimised values are plotted for comparison.

For practical purposes all three models offer good downstream predictions for all four reaches ($R^2 > 0.98$). The deconvolved RTDs show a high degree of comparability with those RTDs predicted by more traditional methods. For rivers, this is expected given that the relevant mixing processes within a long reach are averaged and well integrated. There are, however, details shown in the deconvolved RTDs that may offer additional insights into larger scale effects on the mixing. For example, the secondary peak in Reach 2 may indicate a recirculation zone along a bend. This illustrates how the proposed deconvolution methodology can be used as a flexible approach to the analysis of input data with variable quality. Since only simple pre-processing was necessary, deconvolution could easily be applied to the rest of the database.

Improved deconvolution of manhole solute transport data

A small selection of solute transport traces recorded by Saiyudthong (2003) from an unbenched 400 mm manhole with 30° outlet angle and 4 l/s flow rate has been deconvolved to demonstrate the improvements to deconvolution. First, pre-processed traces were deconvolved as previously recommended by Sonnenwald et al. (2014) using 40 sample points and linear interpolation. Second, the raw data for the same traces were deconvolved after minimal pre-processing, which took the form of a sloped background subtraction based on the mean of the first and last 5 seconds of data as background concentration level estimations, but still using 40 sample points and linear interpolation. Third, the raw traces were deconvolved after minimal pre-processing and using LAMA interpolation with 20 sample points.

3 repeat trials for each surcharge depth have been averaged on a cumulative percentage basis and the resulting CRTDs plotted in Figure 9 using normalised time ($t_{nz} = tQV^{-1}$) to non-dimensionalise manhole volume effects, where t is time, Q is flow rate, and V is volume between fluorometers (Stovin et al. 2010a). The different deconvolution configurations are plotted on the same axes with temporal (x-axis) offsets for easier comparison. The pre-processed traces deconvolved using linear interpolation and 40 sample points, group (i), are plotted from $t = 0_{\mathrm{i}}$. The CRTDs derived from the same experiments, but this time deconvolved from the raw experimental traces, group (ii), are plotted from $t = 0_{\mathrm{ii}}$. The raw traces deconvolved using LAMA interpolation and 20 sample points, group (iii), are plotted from $t = 0_{\mathrm{iii}}$.

All three groups of CRTDs indicate the same bulk mixing characteristics, with two subgroups forming, showing the successful deconvolution of raw solute transport data. One group at lower surcharge depths (darker coloured), shows a cumulative exponential shape, which may be associated with complete mixing. The second cluster is at higher surcharge depths (lighter coloured), with a sharp rise followed by a long tail, which may be associated with a short-circuiting flow field. In detail however, there is variation between the groups that corresponds to differences in RTD shape.

Group (i) shows what appears to be an outlying result, a CRTD whose tail is not clustered with the others of its group. This CRTD does not appear in groups (ii) or (iii) when deconvolution is carried out using raw data. The outlier in this case must be a result of the pre-processing used as it is present in each repeat trial. Previous results (Guymer and Stovin 2011) suggest that such an outlier is inconsistent with the underlying hydrodynamic processes. The differences between groups (ii) and (iii) are minor, but close examination shows that much of the small scale fluctuation has been smoothed out in (iii). Using raw data for deconvolution and fewer sample points with LAMA interpolation both lead to improved quality of the deconvolved RTD.

CONCLUSIONS

Two improvements have been outlined, investigated, and validated for maximum entropy deconvolution as applied to solute transport data. The first is the ability to deconvolve raw data. The second is the application of smoothing within the deconvolution process. Provided minimal pre-processing is performed (subtracting background concentration level), and the instrumentation used to collect the raw data has a linear response, maximum entropy deconvolution can be successfully applied to raw solute transport data to extract the RTD. Furthermore, LAMA interpolation and lower numbers of sample points can be recommended for improving deconvolved RTD smoothness, thereby more accurately representing system hydrodynamics.

Both improvements have been demonstrated with experimental data. Recorded river solute transport data can easily be deconvolved with only minimal pre-processing. The deconvolved RTDs compare favorably to those generated using standard ADE and ADZ models. This opens the door to analysing data from diverse sources with the same methodology that would otherwise require specific pre-processing in each case. Solute transport records from a surcharged manhole have been deconvolved as raw and pre-processed data, showing the same trends in both cases. The raw data deconvolved with LAMA interpolation and 20 sample points not only shows the same trends, but is also noticeably smoother. These RTDs therefore better reflect the bulk mixing conditions of the manhole.

The two proposed improvements to maximum entropy deconvolution function and result in higher quality RTDs. The elimination of the need for advanced pre-processing represents a significant improvement in the efficiency of data analysis and removes sources of uncertainty.

ACKNOWLEDGMENTS

Contains Environment Agency information © Environment Agency and database right

REFERENCES

Beer, T. and Young, P. C. (1983). “Longitudinal dispersion in natural streams.” *Journal of environmental engineering*, 109(5), 1049–1067.

de Boor, C. (1978). *A practical guide to splines*. Springer-Verlag New York.

Denbigh, K. G. and Turner, J. C. R. (1984). *Chemical Reactor Theory*. Cambridge University Press.

Fienen, M. N., Luo, J., and Kitanidis, P. K. (2006). “A bayesian geostatistical transfer function approach to tracer test analysis.” *Water Resources Research*, 42(7).

Fritsch, F. N. and Carlson, R. E. (1980). “Monotone piecewise cubic interpolation.” *SIAM Journal on Numerical Analysis*, 17(2), 238–246.

Gooseff, M. N., Benson, D. A., Briggs, M. A., Weaver, M., Wollheim, W., Peterson, B., and Hopkinson, C. S. (2011). “Residence time distributions in surface transient storage zones in streams: Estimation via signal deconvolution.” *Water Resources Research*, 47.

Guymer, I. (1998). "Longitudinal dispersion in sinuous channel with changes in shape." *Journal of Hydraulic Engineering*, 124(1), 33–40.

Guymer, I. (2002). "A national database of travel time, dispersion and methodologies for the protection of river abstractions." *R&D Technical Report P346*, The Environment Agency.

Guymer, I., Dennis, P., O'Brien, R., and Saiyudthong, C. (2005). "Diameter and surcharge effects on solute transport across surcharged manholes." *Journal of Hydraulic Engineering*, 131(4), 312–321.

Guymer, I. and O'Brien, R. (2000). "Longitudinal dispersion due to surcharged manhole." *Journal of Hydraulic Engineering*, 126(2), 137–149.

Guymer, I. and Stovin, V. R. (2011). "One-dimensional mixing model for surcharged manholes." *Journal of Hydraulic Engineering*, 137(10), 1160–1172.

Hansen, P. C. (1998). *Rank-deficient and discrete ill-posed problems: numerical aspects of linear inversion*. Society for Industrial and Applied Mathematics.

Hart, J., Guymer, I., Jones, A., and Stovin, V. R. (2013). "Longitudinal dispersion coefficients within turbulent and transitional pipe flow." *Experimental and Computational Solutions of Hydraulic Problems*, P. Rowinski, ed., Springer.

Hattersley, J. G., Evans, N. D., Hutchison, C., Cockwell, P., Mead, G., Bradwell, A. R., and Chappell, M. J. (2008). "Nonparametric prediction of free-lightchain generation in multiple myelomapatients." *17th International Federation of Automatic Control World Congress (IFAC)*, Seoul, Korea, 8091–8096.

Joo, D., Choi, D., and Park, H. (2000). "The effects of data preprocessing in the determination of coagulant dosing rate." *Water Research*, 34(13), 3295–3302.

Kasban, H., Zahran, O., Arafa, H., El-Kordy, M., Elaraby, S. M. S., and Abd El-Samie, F. E. (2010). "Laboratory experiments and modeling for industrial radiotracer applications." *Applied Radiation and Isotopes*, 68(6), 1049–1056.

Kashefipour, S. and Falconer, R. (2000). "An improved model for predicting sediment fluxes in estuarine waters." *Proceedings of the Fourth International Hydroinformatics Conference*, *Iowa, USA*.

Levenspiel, O. (1972). *Chemical Reaction Engineering*. John Wiley & Son, Inc.

Madden, F. N., Godfrey, K. R., Chappell, M. J., Hovorka, R., and Bates, R. A. (1996). "A comparison of six deconvolution techniques." *Journal of Pharmacokinetics and Biopharmaceutics*, 24(3), 283–299.

McGuire, K. J. and McDonnell, J. J. (2006). "A review and evaluation of catchment transit time modeling." *Journal of Hydrology*, 330(3), 543–563.

Nash, J. E. and Sutcliffe, J. V. (1970). "River flow forecasting through conceptual models part I - A discussion of principles." *Journal of Hydrology*, 10(3), 282–290.

Rinaldo, A., Beven, K. J., Bertuzzo, E., Nicotina, L., Davies, J., Fiori, A., Russo, D., and Botter, G. (2011). "Catchment travel time distributions and water flow in soils." *Water Resources Research*, 47(7).

Rutherford, J. C. (1994). *River mixing*. John Wiley & Son Ltd, Chichester, England.

Saiyudthong, C. (2003). "Effect of changes in pipe direction across surcharged manholes on dispersion and head loss." Ph.D. thesis, The University of Sheffield, The University of Sheffield.

Schittkowski, K. (1986). "NLPQL: A fortran subroutine solving constrained nonlinear programming problems." *Annals of Operations Research*, 5(2), 485–500.

Sherman, L. K. (1932). "Streamflow from rainfall by the unit-graph method." *Engineering News Record*, 108, 501–505.

Skilling, J. and Bryan, R. K. (1984). "Maximum-entropy image-reconstruction - general algorithm." *Monthly Notices of the Royal Astronomical Society*, 211(1), 111–124.

Sonnenwald, F. (2014). "Identifying the fundamental residence time distribution of urban drainage structures from solute transport data using maximum entropy deconvolution." Ph.D. thesis, The University of Sheffield, The University of Sheffield.

Sonnenwald, F., Stovin, V., and Guymer, I. (2014). "Configuring maximum entropy deconvolution for the identification of residence time distributions in solute transport applications."

Journal of Hydrologic Engineering, 19(7), 1413–1421.

Stovin, V., Guymer, I., and Lau, S. D. (2010a). "Dimensionless method to characterize the mixing effects of surcharged manholes." *Journal of Hydraulic Engineering*, 136(5), 318–327.

Stovin, V. R., Guymer, I., Chappell, M. J., and Hattersley, J. G. (2010b). "The use of deconvolution techniques to identify the fundamental mixing characteristics of urban drainage structures." *Water Science and Technology*, 61(8), 2075–2081.

The MathWorks Inc. (2011). *MATLAB R2011a*. Natick, MA.

Wallis, S. and Manson, R. (2005). "Modelling solute transport in a small stream using DISCUS." *Acta Geophysica Polonica*, 53(4), 501.

Young, P., Jakeman, A., and McMurtrie, R. (1980). "An instrumental variable method for model order identification." *Automatica*, 16(3), 281–294.

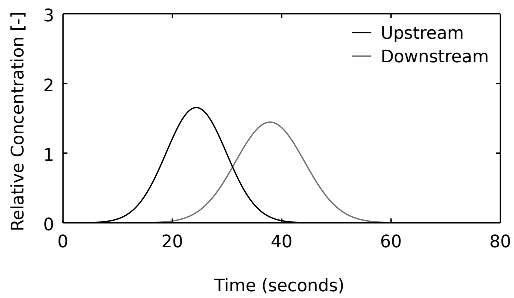
Zimmerman, D., Pavlik, C., Ruggles, A., and Armstrong, M. P. (1999). "An experimental comparison of ordinary and universal kriging and inverse distance weighting." *Mathematical Geology*, 31(4), 375–390.

List of Tables

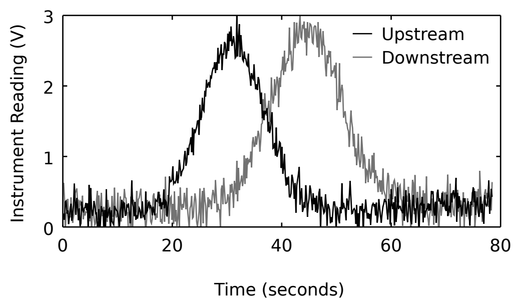
TABLE 1: Variation in predictive capability of RTDs (mean R^2) with respect to different window sizes (values of β)

β	R^2
0.05	0.9269
0.10	0.9321
0.20	0.9333
0.20	0.9309

List of Figures



(a) Perfect synthetic trace before reversed pre-processing



(b) A synthetic 'raw' trace after reversed pre-processing

FIG. 1: Synthetic data before and after reversed pre-processing has been applied

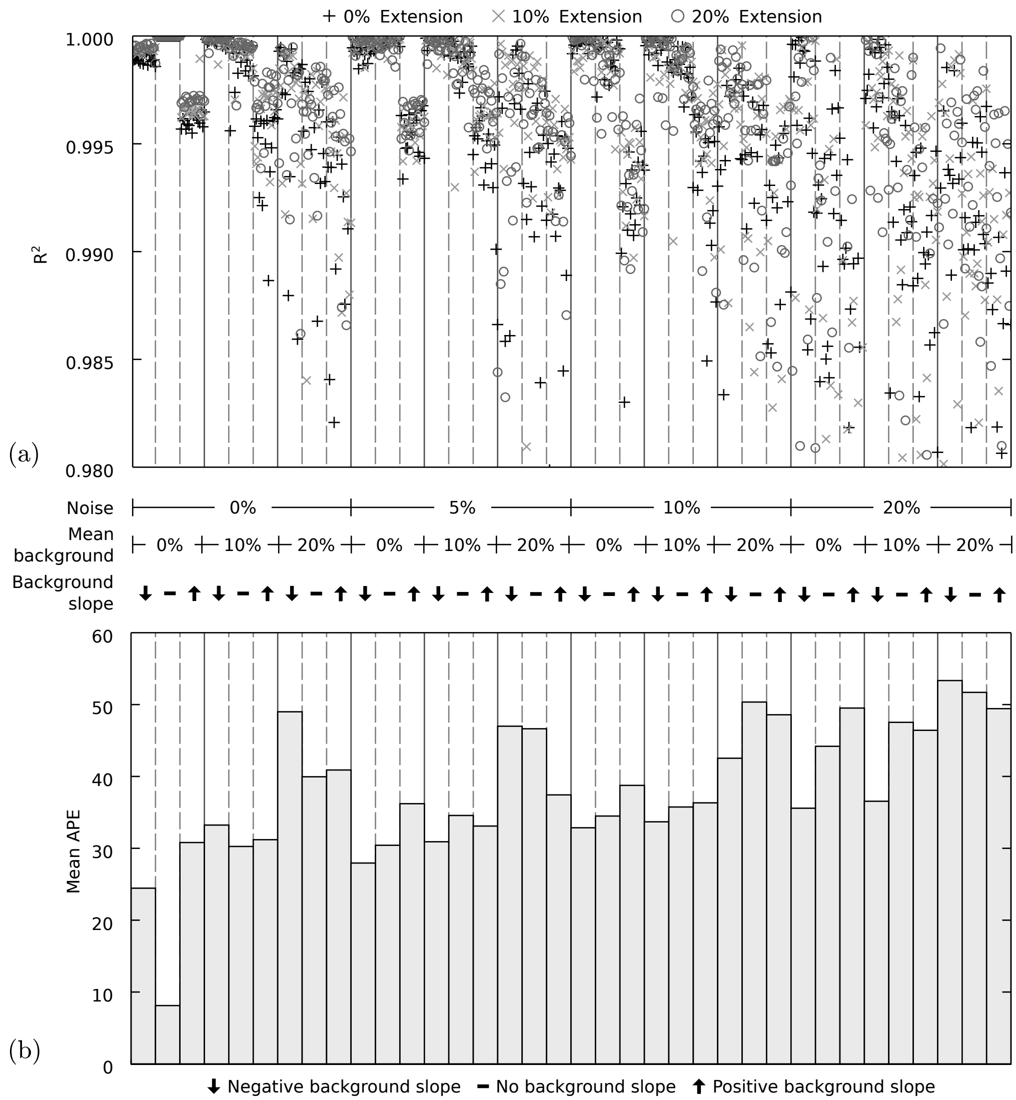


FIG. 2: Impact of raw data characteristics on (a) the predicted downstream profile generated using deconvolved RTD compared to the known downstream profile, evaluated using R^2; and on (b) the deconvolved RTD compared to the known RTD, evaluated using APE

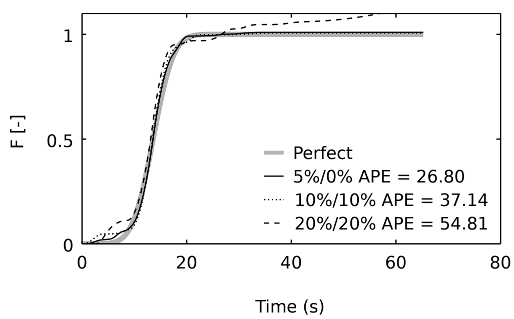


FIG. 3: Representative deconvolved CRTDs for different combinations of noise/background compared to the known perfect CRTD

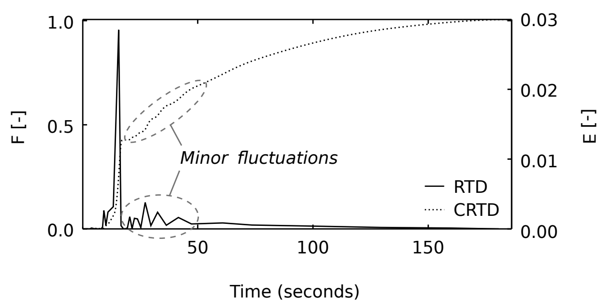


FIG. 4: Example of minor fluctuations observed in a deconvolved RTD and CRTD

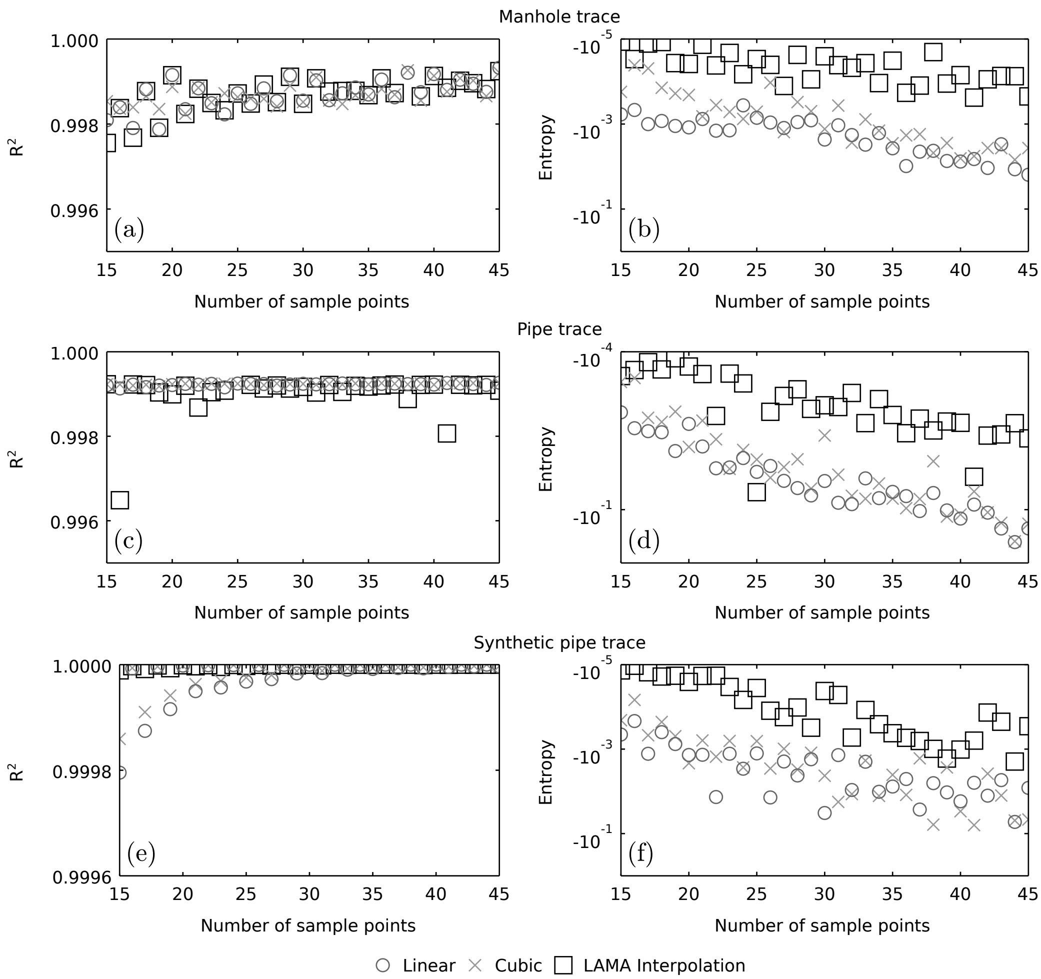


FIG. 5: Predictive capability (a, c, e) and smoothness (b, d, f) of deconvolved RTDs across interpolation function and number of sample points for three different solute transport traces

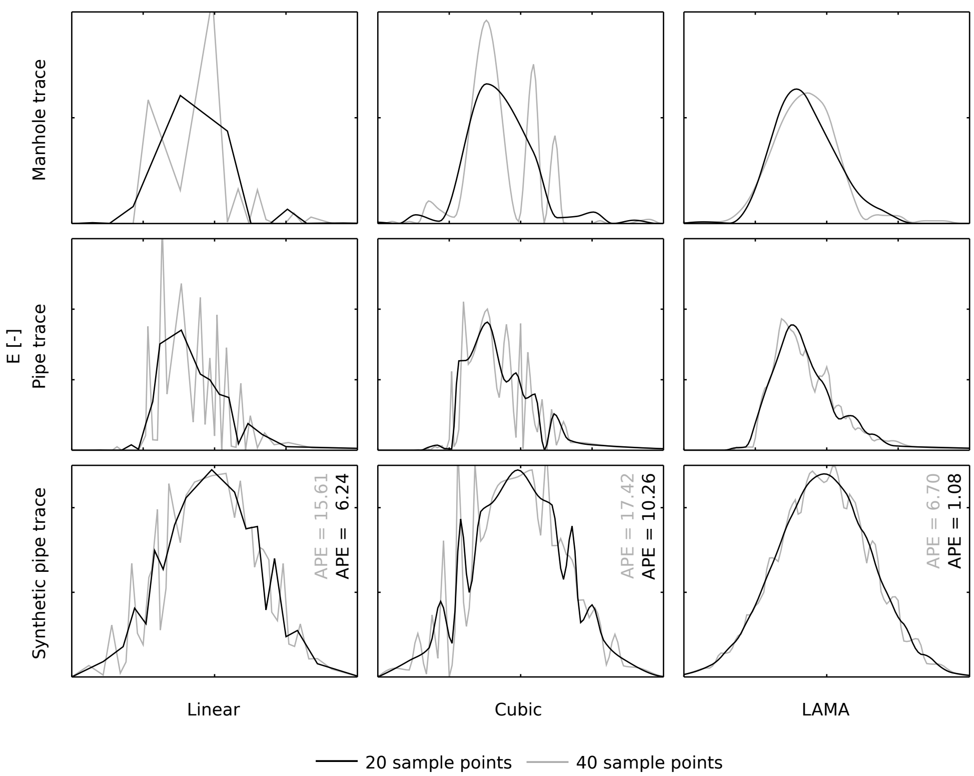


FIG. 6: A visual comparison of RTDs deconvolved with 20 and 40 sample points using linear, cubic, and LAMA interpolation, for the three different solute traces examined

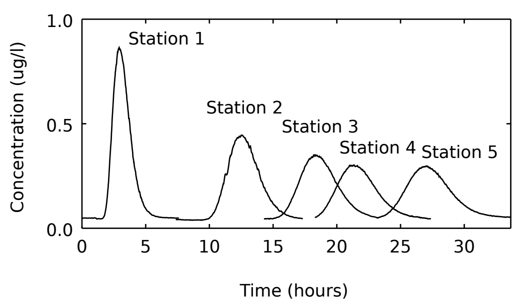


FIG. 7: Raw solute transport data collected at five monitoring stations on the River Swale (NE17) (data from Guymer (2002))

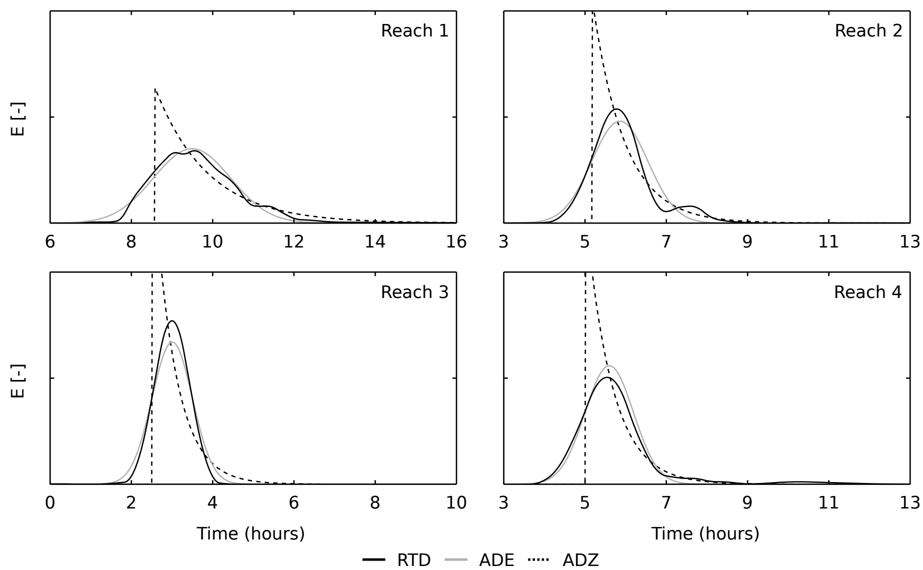


FIG. 8: Deconvolved RTDs (labeled RTD) compared with RTD functions generated by best-fit ADE and ADZ model parameters

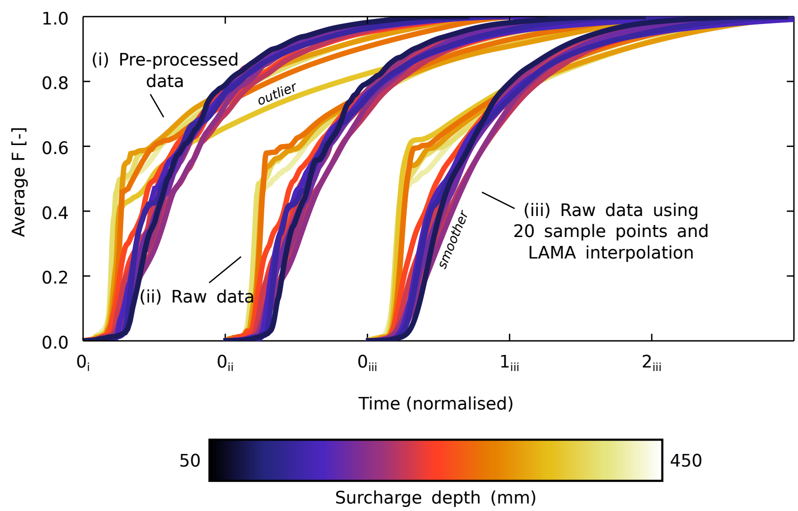


FIG. 9: Comparison of CRTDs deconvolved with and without improvements from unbenched 30° outlet angle surcharged manhole data at 4 l/s