Configuring maximum entropy deconvolution for the identification of residence time distributions in solute transport applications

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

ABSTRACT

The advection-dispersion equation (ADE) or aggregated dead zone (ADZ) models and their derivatives are frequently used to describe mixing processes within rivers, channels, pipes, and urban drainage structures. The residence time distribution (RTD) provides a non-parametric model that may describe mixing effects in complex mixing contexts more completely. Identifying an RTD from laboratory data requires deconvolution. Previous studies have successfully applied maximum entropy deconvolution to solute transport data, with RTD sub-sampling used for computational simplification. However, this requires a number of configuration settings which have to date not been rigorously investigated. Four settings are investigated here: the number and distribution of sample points, the constraint function, and the maximum number of iterations. Configuration options for each setting have been systematically assessed with reference to representative solute transport data by comparing the goodness-of-fit of recorded and predicted downstream profiles using the Nash-Sutcliffe Efficiency Index, evaluating RTD smoothness with a measure of entropy, and through consideration of the mass-balance of the RTD. New methods for defining sample point distribution are proposed. The results indicate that goodness-of-fit is most sensitive to constraint function and that smoothness is most sensitive to the number and distribution of sample points. A set of configuration options that includes a new sample point distribution

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

[2]Senior Lecturer, 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

is shown to perform robustly for a representative range of laboratory solute transport data.

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

INTRODUCTION

Background

Solute transport is affected by mixing processes. As such, improved understanding of solute transport can lead to both new applications in water quality modelling and to improved understanding of the underlying processes that affect mixing. This applies to processes in natural rivers and channels as well as man-made structures such as pipes and manholes.

The advection-dispersion equation (ADE) or aggregated dead zone (ADZ) models have traditionally been used to evaluate or model solute transport (Rutherford 1994). Both are parametric models that apply an understanding of the processes involved to derive a system of equations. They include assumptions and, provided they are met, the models can perform extremely well, e.g. in pipe flow (Taylor 1954). Model performance degrades when the underlying assumptions are not met (Davis et al. 2000; Rieckermann et al. 2005).

In chemical engineering, the residence time distribution (RTD) is frequently used to describe mixing within reactors in response to a Dirac pulse (an instantaneous input) (Levenspiel 1972). Equation 1 shows the relationship between upstream $y(t)$ and downstream $u(t)$ temporal concentration data through convolution with the RTD $h(t)$. The RTD is also known as a transfer function. In hydrology the RTD is analogous to the unit hydrograph (Sherman 1932).

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

Recent research has used the RTD to describe solute transport in urban drainage systems, e.g. Guymer and Stovin (2011). The particular benefit of an RTD is that, as a non-parametric model, no assumptions are made on how the system operates. Therefore, the RTD can exactly describe complex mixing processes in a reach or structure, such as dead-zone

short-circuiting (Stovin et al. 2010a). Unfortunately this benefit incurs a cost, as identifying an RTD is significantly more complex than identifying the parameters of traditional models.

The general method of identifying an RTD from recorded laboratory data is deconvolution. There are many methods and applications for deconvolution. An overview of some common methods is given by Madden et al. (1996). Other applications include noise cancellation (Pandolfi 2010) and gas chromatography (Zhong et al. 2011). Within solute transport research, deconvolution techniques have been used to examine soil transfer functions (Skaggs et al. 1998), bank filtration (Cirpka et al. 2007), and transient storage (Gooseff et al. 2011). We have previously used maximum entropy deconvolution to investigate solute transport in manholes (Stovin et al. 2010b; Sonnenwald et al. 2011; Guymer and Stovin 2011).

Although maximum entropy deconvolution has previously been successfully applied to solute transport data, no rigorous investigation into how the configuration settings affect the quality of the results obtained has been reported. Four maximum entropy deconvolution settings impact on the quality of the deconvolved RTD: the number of sample points; sample point distribution; constraint function; and the maximum number of iterations. Inappropriate configuration options for any of the settings may result in a poor quality RTD. This paper aims to systematically identify a robust set of options that can be used to deconvolve the RTD from typical solute transport data. To this end, a sensitivity analysis has been carried out with a range of data and options.

Maximum entropy deconvolution

Maximum entropy deconvolution is a discrete computational technique that uses regularly sampled paired upstream and downstream temporal concentration profiles to deconvolve the RTD. An estimate of the RTD, $\hat{h} = \{h_1, \ldots, h_N\}$ where N is the number of data points, is to be made as flat as possible with the only exceptions being those implied by the upstream and downstream data (Skilling and Bryan 1984). Flatness of $\hat{h}$ is measured by an entropy function S, Equation 2, which also enforces non-negativity. A constraint function C, Equation 3, ensures that the RTD is valid by comparing the goodness-of-fit of the predicted

downstream concentration profile $\hat{y}$ against the recorded profile y, where $\hat{y}$ is calculated as the convolution of $\hat{h}$ and u. C is typically, as presented here, the chi-squared function, where σ is an error estimate. The RTD is identified by combining both equations in a Lagrangian function L, Equation 4, and maximizing. λ is the Lagrange multiplier determined during the maximization process. Sub-scripts denote specific points in discrete time.

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

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

$$L(\hat{h}, \lambda) = S(\hat{h}) - \lambda C \tag{4}$$

The software and methodology used for maximum entropy deconvolution of solute transport data is an evolution of a pharmacokinetics application (Hattersley et al. 2008). In pharmacokinetics, data points are often collected at uneven time intervals, e.g. by a nurse making rounds. As a result, the entropy function was modified for piecewise data, where the value between points is assumed to vary linearly, and Equation 5 was developed. The r term is added as a base-line prediction of the RTD in the absence of other data. r takes the form of a nearest neighbour moving average where $r_i = ((\hat{h}_{i-1} + \hat{h}_{i+1})/2)$ and at $i = 0$ and $i = N$ the value of the two nearest points, e.g. $r_N = (\hat{h}_{N-1} + \hat{h}_N)/2$. The inclusion of r results in an entropy value that evaluates smoothness; entropy values closer to zero indicate a smoother function.

$$S(\hat{h}) = -\sum_{i=1}^{N} \left(\frac{\hat{h}_i}{\sum_{j=1}^{N} \hat{h}_j} \right) \ln \left(\frac{\hat{h}_i / \sum_{j=1}^{N} \hat{h}_j}{r_i} \right) \tag{5}$$

To obtain $\hat{h}$, Hattersley et al. (2008) converted Equation 4 into an equivalent minimisation problem. This was solved using a Sequential Quadratic Programming (SQP) technique

implemented in Matlab, `fmincon` (The MathWorks Inc. 2011). SQP is an optimisation algorithm that works by minimising a quadratic model of the problem to find the next step towards the solution (The Morgridge Institute for Research 2012).

Maximum entropy deconvolution was further modified for application to solute transport data by Stovin et al. (2010b). The piecewise capability previously introduced was modified to create a simplified deconvolution problem where the RTD is sub-sampled. This reduces computational expense and the impact of noisy data while maintaining the benefits of a non-parametric model. The sub-sampled RTD is defined only at n sample points, spread between the start and end of the concentration data, as the length of the RTD is unknown. Sample points are otherwise placed where more variation is expected in the RTD. A full RTD is reconstructed from the sub-sampled RTD using linear interpolation.

METHODOLOGY

Configuration settings for maximum entropy deconvolution

The first two configuration settings are number and positioning of sample points. As linear interpolation is used to reconstruct the RTD, each sample point defines a change in the slope of the RTD. Therefore, changing the position and number of points is expected to have a high impact on the identified RTD.

Skilling and Bryan (1984) suggest that alternative constraint functions may be preferable to χ^2, hence this configuration setting is also examined here. As C effectively evaluates goodness-of-fit, correlation measures form suitable alternatives. Different correlation measures may place different emphasis on matching the shape, scale, or noise (Sonnenwald et al. 2013).

`fmincon` introduces the fourth configuration setting, maximum number of iterations, which imposes an upper limit on `fmincon` so that it does not enter an infinite loop. Too few iterations, however, will stop the deconvolution process before convergence is achieved, i.e. before the RTD is identified. `fmincon` also introduces convergence criteria to determine

when optimisation stops and an 'initial guess' that is the start point of the optimisation process.

Number of sample points

Stovin et al. (2007) suggested that as few as 7 points are necessary to define an RTD. A minimum of 10 sample points has therefore been used. 20, 40, 80, 120, 160, and 200 sample points have also been evaluated. After 200 points we have observed computational cost to increase significantly. Stovin et al. (2010b) used 40 sample points.

Sample point distributions

Sample points are placed where more variation in the RTD is anticipated by incorporating basic assumptions of the expected RTD. Six sample point distributions have been developed using varying amounts of prior knowledge, described below and shown in Figure 1.

- **Equally spaced (ES):** The sample points are evenly distributed across the input data. This distribution assumes no knowledge of the RTD.
- **Log from zero (LFZ):** The interval between sample points increases logarithmically from the start to the end of the data. This distribution assumes more variation earlier in the RTD and less variation as time goes on, i.e. an exponential decay.
- **Downstream log (DwL):** First arrival time and end of event are defined as 1% of peak concentration. Three sample points are evenly distributed from the start of the data until the difference in first arrival times, after which the interval between sample points increases logarithmically until the end of the downstream event. Three more sample points are evenly distributed until the end of data. From Equation 1 it follows that there must be some delay in the RTD if there is a delay between first arrival times. This is the sample point distribution previously used by Stovin et al. (2010b).
- **Double log (DuL):** Half of the sample points are distributed logarithmically from the start of the data to the difference in time to peak, which is used as an estimate of delay. The other half of the sample points are logarithmically distributed away

from the difference in time to peak to the end of the data. A greater concentration of points around the time the RTD peak is expected allows for more uncertainty in its location.

- **Slope based (SB):** This is a new development. An approximation of the RTD is used to distribute the sample points where slope is expected to be greater. The approximation is computed using Fast Fourier Transformation (FFT) deconvolution (Madden et al. 1996) with Blackman-Tukey Windowing (Blackman and Tukey 1958; Harris 1978) applied to the input data to improve accuracy. The absolute area of the first derivative of the approximation is evenly divided and sample points placed at the division points.
- **Double cubic (DC):** This is a new development. It is the same as the DuL distribution, but using cubic spacing. This results in a more spread out distribution, similar to the log from zero and slope-based sample point distributions, which is expected to allow greater flexibility in capturing complex profile characteristics, e.g. secondary peaks.

Constraint functions

In a previous investigation carried out to identify potentially suitable correlation measures for solute transport model identification (Sonnenwald et al. 2013), twelve correlation measures were examined. Eight measures were found to be sensitive to transformation and transformation intensity while remaining insensitive to noise, and were therefore judged to be suitable as constraint functions. These are: the Burnham-Liard Criterion (BLC) (George et al. 1998); χ^2; Furthest Fitting Cost Based Similarity (FFCBS) (Ye et al. 2004); the Nash-Sutcliffe Efficiency Index (R^2) (Nash and Sutcliffe 1970); Root Mean Square Deviation (RMSD) (Anderson and Woessner 1992); the Coefficient of Determination (R_t^2) (Young et al. 1980); the Integral of Squared Error (ISE) (Ghosh 2007); and Average Percent Error (APE) (Kashefipour and Falconer 2000). They have been converted into equivalent constraint functions for inclusion in the present sensitivity analysis. The error estimate σ of χ^2 is taken

from Stovin et al. (2010b) as 5% of recorded value.

Maximum number of iterations

Maximum number of iterations in practice indicates a maximum amount of effort that should be used in deconvolving the RTD should an optimum RTD not be found earlier through convergence. 50, 100, 150, 200, 250, 300, and 350 iterations have been evaluated. A maximum of 200 iterations was used by Stovin et al. (2010b).

Convergence criteria

Initial testing has indicated no sensitivity to convergence criteria. They have been left at `fmincon` defaults as previous work has used them successfully.

Initial guess

Initial testing has indicated no sensitivity to the initial guess. As the optimisation starting point it does not change the minimization problem, but an initial guess that is closer to the final solution is a 'warm start' and has been shown to reduce the amount of time necessary to reach convergence in SQP algorithms (Fan et al. 1988). Therefore the initial guess is fixed as the result of a FFT deconvolution with Blackman-Tukey windowing (as used in the SB distribution). Stovin et al. (2010b) used a flat line guess based on $\int_{-\infty}^{\infty} h(t)dt = 1$.

Selection of data for sensitivity analysis

We have several datasets from previously published laboratory studies available. Within these, five mixing scenarios are represented; pipe flow (Hart et al. 2013), open channel flow (Guymer 1998), storage tank mixing (Guymer et al. 2002), below-threshold (BT) surcharged manholes, and above-threshold (AT) surcharged manholes (Guymer et al. 2005; Guymer and Stovin 2011). The threshold is the surcharge depth at which hydraulic regime within a manhole switches from a fully-mixed (BT) to a short-circuiting (AT) system.

Two sets of typical solute transport concentration data from each of the five mixing scenarios were selected to ensure that conclusions would not be unduly influenced by a single test within each mixing scenario. The 10 paired upstream and downstream concentration

profiles (henceforth referred to as 'experiments') are outlined in Table 1 and shown in Figure 2. In all cases pre-processing of the raw data (i.e. calibration, smoothing, background removal) applied in the previous studies has been retained.

Analyzing RTD performance

As previously stated, the full RTD is generated from the sample points via linear interpolation. A complete predicted downstream profile can then be generated by convolving the upstream profile with the full deconvolved RTD. A successful deconvolution is defined as one with high goodness-of-fit between the predicted and recorded downstream profiles, as measured by a relevant correlation measure. Sonnenwald et al. (2013) suggested R_t^2, R^2 and APE as suitable for this application. The R^2 correlation measure has been chosen here for its high sensitivity to overall profile shape. With a perfect match, $R^2 = 1$, and for $R^2 \leq 0$ there is no correlation.

We have observed that RTD shape can vary significantly when the difference in R^2 values between RTDs is very small. As a result the entropy function (Equation 5) has been applied to the deconvolved RTD to evaluate smoothness. A smoother RTD is assumed to better represent a natural turbulent system, and therefore entropy values closer to zero are desired.

Mass-balance of the RTDs has also been used for evaluation. Normally $\int_{-\infty}^{\infty} h(t)dt = 1$. When mass recovery is not perfect, e.g. due to calibration error, then instead $\int_{-\infty}^{\infty} \hat{h}(t)dt = \int_{-\infty}^{\infty} y(t)dt / \int_{-\infty}^{\infty} u(t)dt$. RTD quality can also be evaluated as the ratio between the expected and actual sum of the RTD.

RESULTS AND DISCUSSION

The combination of configuration options and experiments resulted in 23,520 deconvolutions. These were carried out using batch processing on the Intel Xeon X5650 nodes of the Iceberg parallel high-performance computing cluster at The University of Sheffield. Processing took approximately 187 days of CPU time. 61.4% of the predicted downstream profiles in comparison to the recorded downstream profiles exceed an R^2 value of 0.95 and 34.6% exceed 0.99 indicating that many combinations of configuration options are acceptable.

Mean and standard deviation of R^2 values

The mean (μ) and standard deviation (σ) of R^2 with respect to each configuration option are shown in Figure 3. Options that result in low mean R^2 values like BLC, χ^2, ISE, and FFCBS, should not generally be used. They have therefore been eliminated from further consideration as robust deconvolution configuration options. The remaining options are evaluated across only the R^2, RMSD, R_t^2, and APE constraints.

Figure 4 illustrates the poor performance of the χ^2 and ISE constraints in contrast to R_t^2, before solution convergence. χ^2 roughly matches the shape but not scale and ISE only roughly matches shape. The performance of these two constraints does not improve with more iterations while the performance of the R_t^2 constraint does, which is typical of the other remaining constraints, R^2, RMSD, and APE.

Figure 3 also suggests that the DwL and ES sample point distributions perform poorly, and therefore these two distributions were eliminated from further consideration. Figure 5 confirms the elimination of DwL and ES by comparison to the SB distribution. Only the SB distribution fits the data for both Experiments 2 and 7. The other two distributions result in approximate fits for Experiment 7 only. For Experiment 2, DwL is mostly flat and ES is almost entirely coincident with the x-axis. This highlights the impact of poor sample point distribution choice.

The difference in DwL performance between Experiment 2 and 7 highlights the potential unreliability of sample point distributions when assumptions made in developing the distribution are not met. For DwL at low numbers of sample points, the 6 fixed points leave too few (only 4) points to characterize the curve. Additionally, due to the lower limits of detection and the effects of noise, the first arrival time identified from the concentration data will be coincident or later than the actual RTD peak. This results in too few points to correctly capture the rising limb of the RTD, leading to the observed poor performance.

After eliminating BLC, χ^2, ISE, FFCBS, DwL, and ES as configuration options, the mean R^2 values indicate improving goodness-of-fit for maximum number of iterations up to

150 iterations and near constant performance thereafter. As such, 50 and 100 iterations were also eliminated, at which point it was observed that mean R^2 also tended to increase with number of sample points. Although this is not evident in Figure 3, R^2 increases until 80 sample points, then remains close to constant. Due to their low mean R^2, 10 and 20 sample points have been eliminated as well.

All 4,000 remaining R^2 values exceed 0.95, and 68.6% exceed 0.99. Differences in mean R^2 value are less than 0.002, and as such there is very little sensitivity of goodness-of-fit to the remaining options. This demonstrates the robustness of maximum entropy deconvolution for most combinations of 40-200 sample points, the LFZ, DuL, SB, and DC distributions, the R^2, RMSD, R_t^2, and APE constraints, and 150-350 iterations.

Entropy values

Entropy values have been examined to further evaluate RTD sensitivity to configuration option. Mean entropy values for each experiment with respect to each option are shown in Figure 6. These are plotted individually as entropy is a dimensional measure. The figure provides insight into the sensitivity of the deconvolved RTD to the different options the configuration settings can take.

40 sample points results in the entropy closest to zero for 9 of 10 experiments, which clearly recommends 40 sample points and therefore other numbers of sample points can be eliminated from consideration. The general trend of entropy values further from zero for increased number of sample points is consistently observed independently of dataset. A greater number of sample points provides increased potential for entropy as each sample point represents a possible change in the slope of the RTD.

The LFZ and SB distributions appear to perform almost identically across all experiments, with entropy values significantly closer to zero than the DuL and DC distributions for almost all experiments. The entropy values further from zero indicate that, although the DuL and DC distributions will generate RTDs with high goodness-of-fit, the shape of the RTDs is less smooth. They are indicated to be less robust and can therefore be eliminated

from consideration.

Number and distribution of sample points have the highest impact on entropy and therefore on the quality of the deconvolved RTD. This is consistent with the problem formulation, i.e. changes in sample point position affect the numerical problem being solved. Although there are multiple RTD solutions for each experiment, improved sample point positioning (and lower numbers of sample points) limits variation and results in smoother RTDs. That R^2 values remain high in these cases demonstrates the robustness of maximum entropy deconvolution as applied to solute transport.

There is no clear trend in constraint function, with high variation between experiments. The smaller changes in entropy with respect to constraint are reasonable considering that constraints are interchangeable measures of error. As all of the constraint functions, R^2, RMSD, R_t^2, and APE, are indicated to be perform similarly they are retained for further examination.

Entropy values are closer to zero as the maximum number of iterations increases for Experiments 2, 5, and 6. The opposite trend is shown by Experiments 7, 9, and 10. Experiments 1, 3, 4, and 8 show no clear trend. Typically, however, more iterations allows for a better solution to be reached, with either entropy closer to zero or increased goodness-of-fit. Therefore, 350 iterations can be recommended and lower maximum numbers of iterations eliminated from consideration. Higher numbers of sample points require comparatively more iterations to reach entropy values closer to zero. Maximum number of iterations has the lowest impact on entropy performance, which indicates that most RTDs reach convergence.

Mass-balance performance

Performance has been further examined by comparing the mass-balance of the remaining deconvolved RTDs. The LFZ and SB distributions have been compared, using 40 sample points, the remaining four constraint functions, and 350 iterations. The SB distribution performs better, with all values close to 1, and therefore LFZ has been eliminated from consideration. The mass-balance performance shows no systematic variation with respect to

constraint function.

Recommended configuration options

There is some evidence in the entropy data presented in Figure 6 that the paired experiments from each of the five datasets responded similarly to the four different constraint functions; this suggests that the optimal constraint function may be linked to dataset characteristics. However, general investigation and consideration of all results suggests that the R_t^2 constraint may perform slightly better. An additional argument in favour of R_t^2 would be that it is already a well used and understood measure in the field of solute transport. Therefore the new SB sample point distribution, 40 sample points, 350 iterations, and the R_t^2 constraint function have been identified as a robust set of configuration options.

VALIDATION

Predicted downstream profiles and CRTDs generated with the robust configuration options (40 sample points, the new SB distribution, the R_t^2 constraint, and 350 iterations) are shown in Figure 7. The lower than expected final value of the CRTD for Experiment 1 is the result of the poor mass-recovery of the laboratory concentration data (Table 1). The predicted profiles give confidence that the identified configuration options are fit for use in deconvolution, with mean $R^2 = 0.994$.

CONCLUSIONS

Maximum entropy deconvolution has previously been successfully applied to laboratory solute transport data to identify the residence time distribution from laboratory data. Here, we have used laboratory data to evaluate the impact of four different configuration settings on the deconvolved RTD. These settings are the number and distribution of sample points, the constraint function, and the maximum number of iterations.

The smoothness of the deconvolved RTD, evaluated by entropy, is particularly sensitive to number and distribution of sample points. A greater number of sample points provides increased potential for noise as each point is a possible change in slope of the RTD. Smaller

numbers of sample points therefore tend to result in a smoother RTD, as well as reduced computational expense. However, too few or poorly positioned sample points will result in a poor quality RTD. A new slope-based sample point distribution, where sample points are positioned based on an Fast-Fourier Transform deconvolution approximation, has been proposed and shown to perform best out of the 6 tested sample point distributions.

The constraint function affects the overall goodness-of-fit between the recorded downstream concentration profile and a predicted profile generated using the deconvolved RTD, here evaluated by R^2. While maximum entropy deconvolution has typically utilized χ^2 as the constraint function, alternative correlation measures place different emphasis on matching profile shape, scale, or noise. The present analysis suggests that χ^2 does not provide a robust constraint for solute transport data, but that the R^2, RMSD, R_t^2, and APE constraint functions do. There is some evidence that the optimal constraint function may be linked to specific data set characteristics, but as it is well understood in the field of solute transport, R_t^2 has been recommended as the most generically applicable constraint function.

Finally, we have shown that a maximum number iterations greater than 200 has a minimal impact on either the R^2 value or RTD smoothness. However, performance in some cases continues to increase maximum number of iterations and so 350 iterations has been recommended here. RTD smoothness results imply that the vast majority of deconvolutions reach convergence before the maximum number of iterations is reached.

Across ten representative laboratory solute transport data, the recommended configuration options – 40 sample points, the new slope-based sample point distribution, the R_t^2 constraint function, and a maximum of 350 iterations – result in a mean R^2 value for the predicted downstream profiles of 0.994. This confirms that maximum entropy deconvolution with the options recommended here provides a robust and effective means of identifying the RTD from laboratory solute transport data.

REFERENCES

Anderson, M. and Woessner, W. (1992). *Applied groundwater modeling: simulation of flow and advective transport.* Academic Press, Inc., London.

Blackman, R. B. and Tukey, J. W. (1958). *The measurement of power spectra, from the point of view of communications engineering.* Dover books on engineering and engineering physics. Dover Publications.

Cirpka, O. A., Fienen, M. N., Hofer, M., Hoehn, E., Tessarini, A., Kipfer, R., and Kitanidis, P. K. (2007). "Analyzing bank filtration by deconvoluting time series of electric conductivity." *Ground Water*, 45(3), 318–328.

Davis, P. M., Atkinson, T. C., and Wigley, T. M. L. (2000). "Longitudinal dispersion in natural channels: 2. The roles of shear flow dispersion and dead zones in the River Severn, UK." *Hydrology and Earth System Sciences Discussions*, 4(3), 355–371.

Fan, Y., Sarkar, S., and Lasdon, L. (1988). "Experiments with successive quadratic programming algorithms." *Journal of Optimization Theory and Applications*, 56(3), 359–383.

George, S., Burnham, K., and Mahtani, J. (1998). "Modelling and simulation of hydraulic components for vehicle applications - a precursor to control system design." *Simulation '98. International Conference on (Conf. Publ. No. 457)*, 126 –132 (sep-2 oct).

Ghosh, A. K. (2007). *Intro. to Linear & Digital Control Systems.* Prentice-Hall Of India Pvt. Ltd.

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

Guymer, I., Shepherd, W. J., Dearing, M., Dutton, R., and Saul, A. J. (2002). "Solute reten-

tion in storage tanks." *Proceedings of 9th International Conference on Urban Drainage, Portland, Oregon, USA*.

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

Harris, F. J. (1978). "On the use of windows for harmonic analysis with the discrete fourier transform." *Proceedings of the IEEE*, 66(1), 51–83.

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.

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.

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.

Pandolfi, L. (2010). "On-line input identification and application to active noise cancellation." *Annual Reviews in Control*, 34(2), 245–261.

Rieckermann, J., Neumann, M., Ort, C., Huisman, J. L., and Gujer, W. (2005). "Dispersion coefficients of sewers from tracer experiments." *Water Science and Technology*, 52(5), 123–133.

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

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

417 Skaggs, T. H., Kabala, Z. J., and Jury, W. A. (1998). "Deconvolution of a nonparametric
418 transfer function for solute transport in soils." *Journal of Hydrology*, 207(3-4), 170–178.

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

421 Sonnenwald, F., Stovin, V., and Guymer, I. (2011). "The influence of outlet angle on solute
422 transport in surcharged manholes." *12th International Conference on Urban Drainage.*,
423 Porte Alegre, Brazil.

424 Sonnenwald, F., Stovin, V. R., and Guymer, I. (2013). "Correlation measures for solute
425 transport model identification & evaluation." *Experimental and Computational Solutions*
426 *of Hydraulic Problems*, P. Rowinski, ed., Springer.

427 Stovin, V., Guymer, I., and Lau, S. D. (2010a). "Dimensionless method to characterize the
428 mixing effects of surcharged manholes." *Journal of Hydraulic Engineering*, 136(5), 318–
429 327.

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

433 Stovin, V. R., Guymer, I., and Lau, D. (2007). "Cumulative concentrations modelling longi-
434 tudinal dispersion - an upstream temporal concentration profile-independent approach."
435 *Proceedings of The 5th International Symposium on Environmental Hydraulics*, Tempe,
436 Arizona.

437 Taylor, G. (1954). "The dispersion of matter in turbulent flow through a pipe." *Proceedings*
438 *of the Royal Society of London. Series A: Mathematical and Physical Sciences*, 223(1155),
439 446–468.

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

The Morgridge Institute for Research (2012). "Sequential quadratic programming, <http://www.neos-guide.org/content/sequential-quadratic-programming> (July. 18, 2012).

Ye, J. C., Tang, Y., Peng, H., and Zheng, Q. L. (2004). "FFCBS: A simple similarity measurement for time series." *Proceedings of the 2004 International Conference on Intelligent Mechatronics and Automation*, 392–396.

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

Zhong, W. J., Wang, D. H., Xu, X. W., Wang, B. Y., Luo, Q., Senthil Kumaran, S., and Wang, Z. J. (2011). "A gas chromatography/mass spectrometry method for the simultaneous analysis of 50 phenols in wastewater using deconvolution technology." *Chinese Science Bulletin*, 56(3), 275–284.

List of Tables

TABLE 1. Summary of laboratory solute transport concentration data used.

Experiment	Description	Flow (l/s)*	Duration (s)	Mass recovery
1	24 mm Pipe[1]	1.084	150.0	84.42%
2	24 mm Pipe[1]	0.221	150.0	98.45%
3	Storage Tank[2]	6.9	240.2	100.00%
4	Storage Tank[2]	6	371.6	100.00%
5	Natural Channel[3]	13.7	157.3	101.96%
6	Trapezoidal Channel[3]	46.1	73.7	105.60%
7	400 mm BT Manhole[4]	1	117.3	100.00%
8	400 mm AT Manhole[4]	1	91.0	100.00%
9	800 mm BT Manhole[5]	1	186.0	100.00%
10	800 mm AT Manhole[5]	1	116.7	100.00%

[1]Hart et al. (2013), [2]Guymer et al. (2002), [3]Guymer (1998)
[4]Guymer et al. (2005), [5]Guymer and Stovin (2011) *As reported

List of Figures

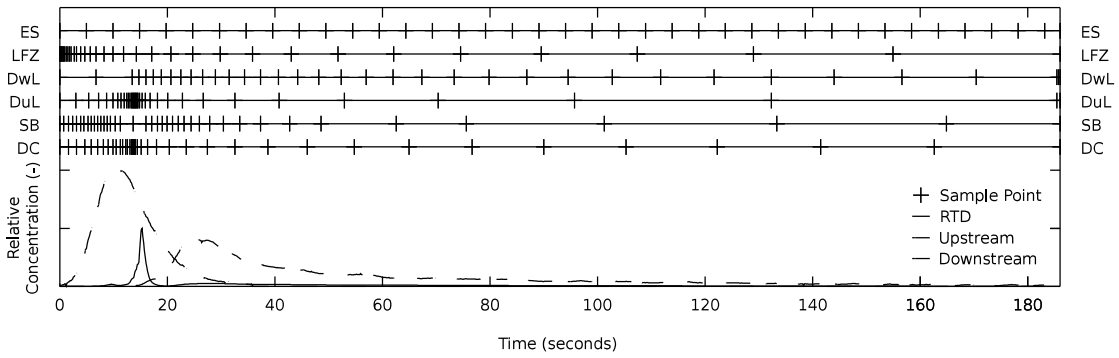

FIG. 1. Example sample point distributions using 40 sample points.

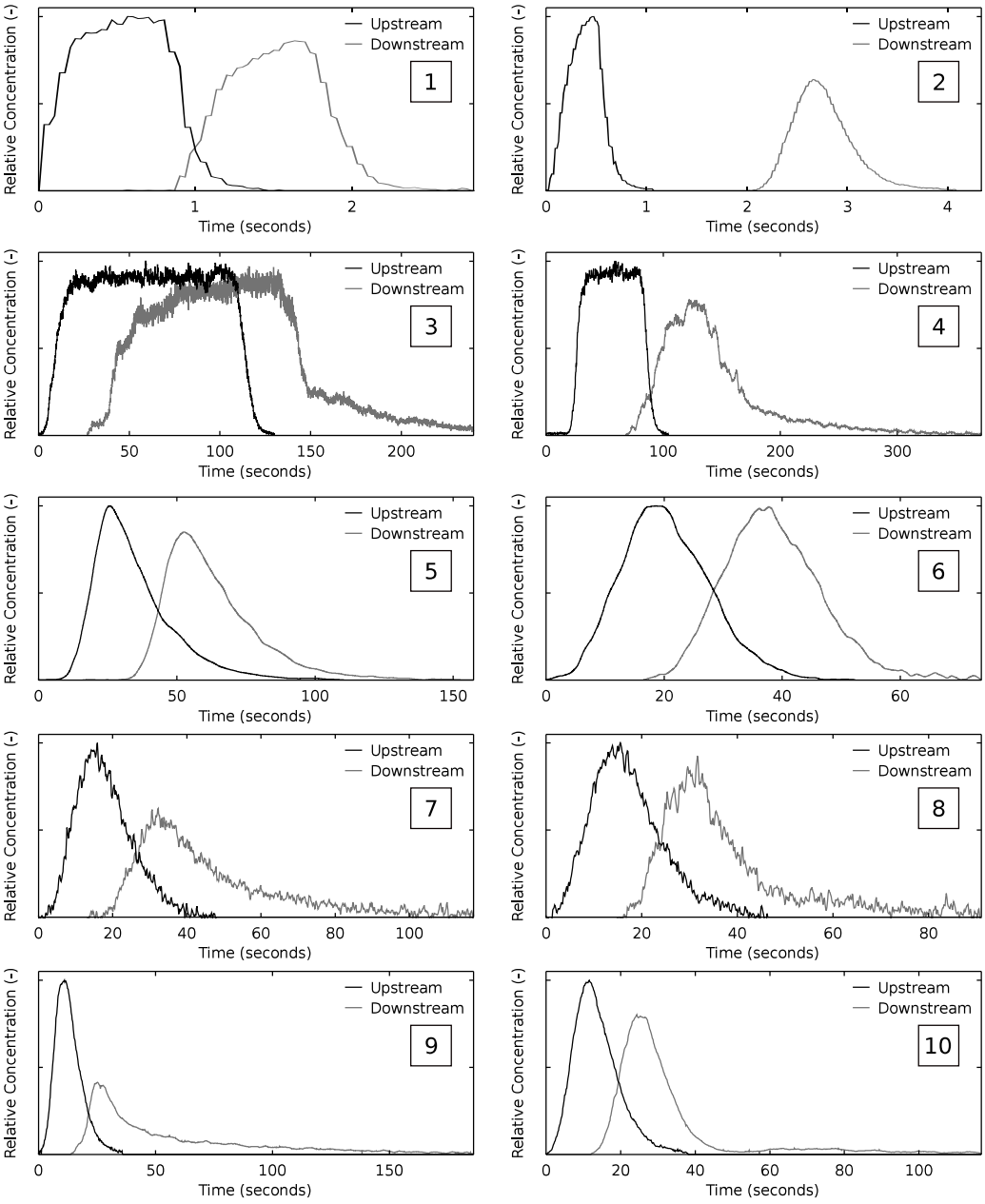

FIG. 2. Upstream and downstream concentration profiles of experiments. Time origin set to 0 and Experiment 1 and 2 zoomed in for display.

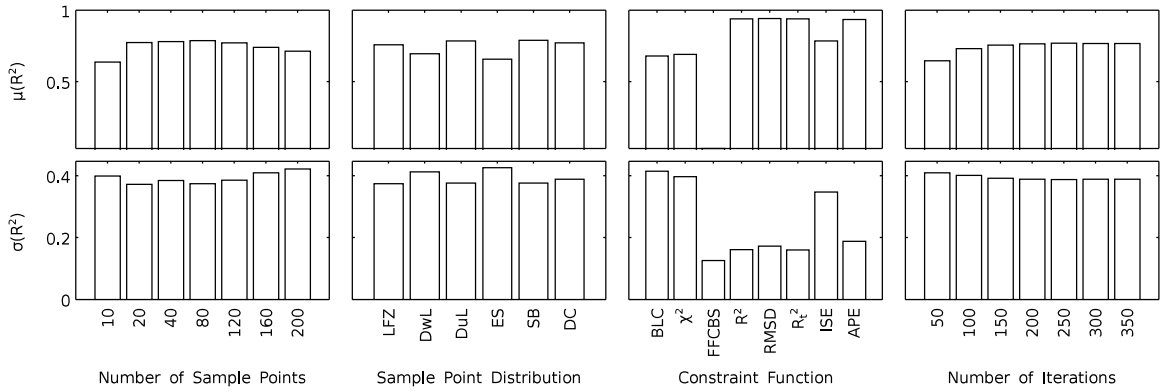

FIG. 3. Mean (μ) and standard deviation (σ) of R^2 values by configuration option.

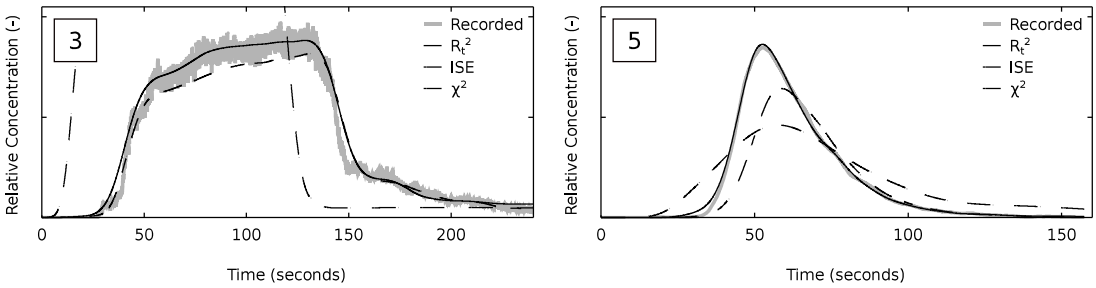

FIG. 4. Predicted downstream profiles with deconvolved RTDs using 40 sample points, the SB sample point distribution, 50 iterations, and the χ^2, R_t^2, or ISE constraint for Experiments 3 and 5.

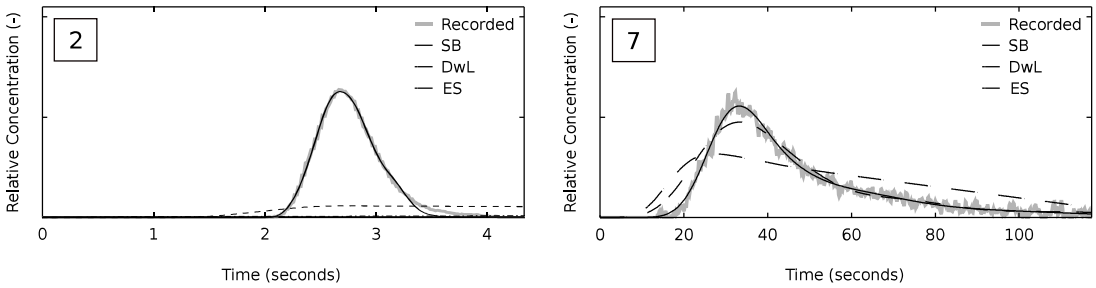

FIG. 5. Predicted downstream profiles with deconvolved RTDs using 10 sample points, the DwL, ES, or SB sample point distribution, 350 iterations, and the R_t^2 constraint for Experiments 2 and 7.

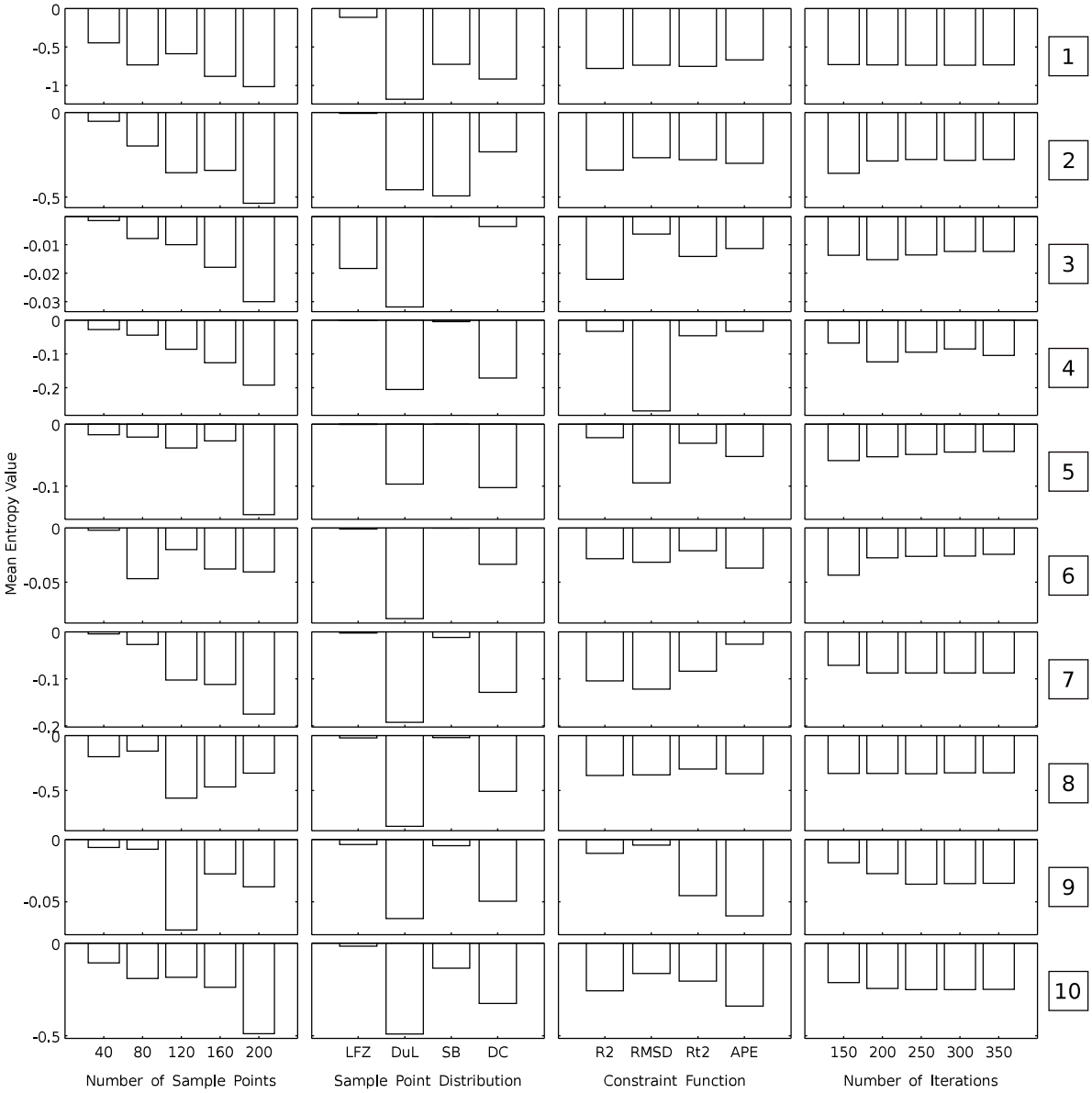

FIG. 6. Mean entropy values by experiment and configuration option. Min $\mathrm{R}^2 = 0.950$, mean $\mathrm{R}^2 = 0.994$.

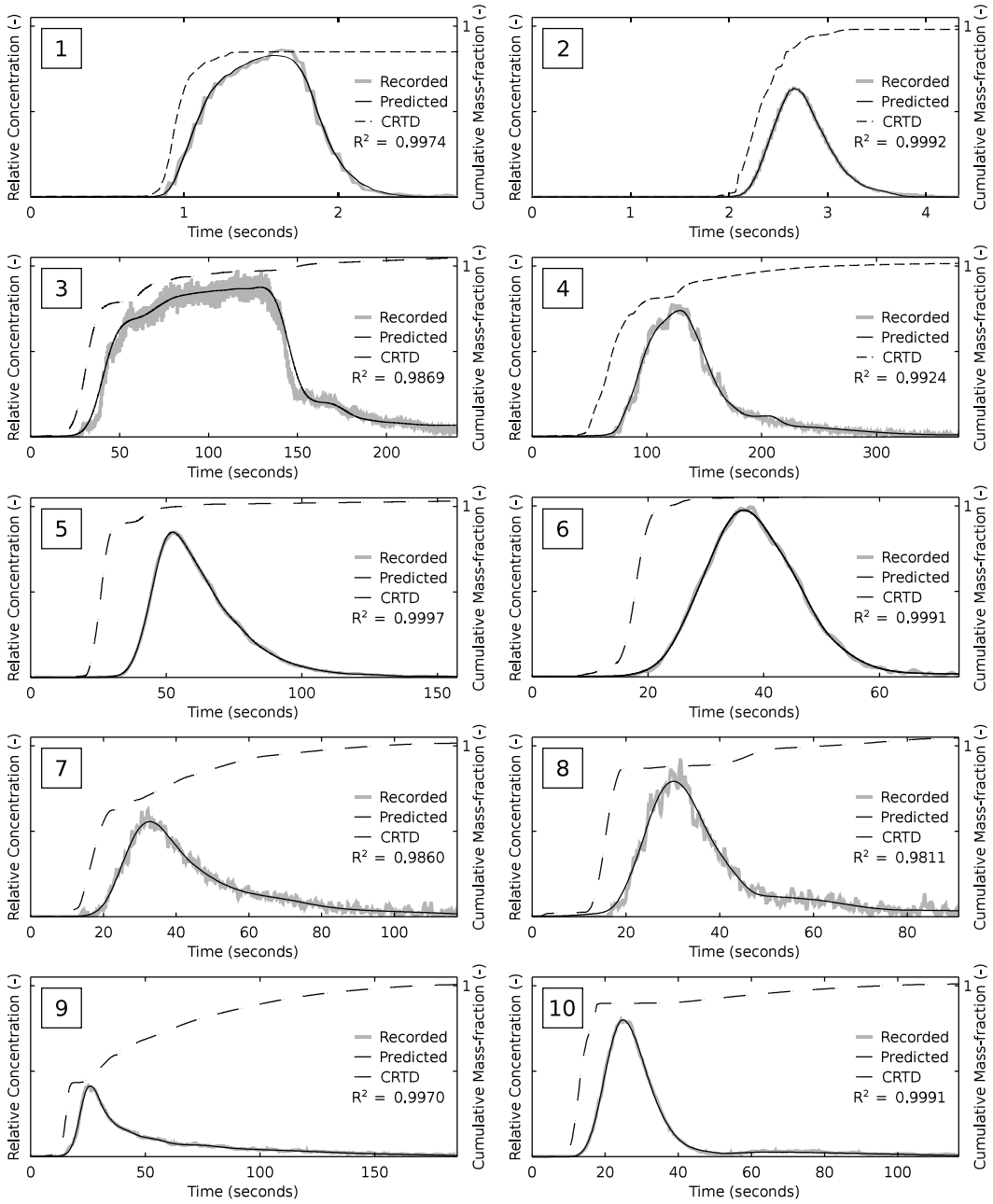

FIG. 7. Predicted downstream profiles and deconvolved CRTDs for each experiment.