
Article:
Camacho Suarez, V.V. orcid.org/0000-0002-8164-7927, Schellart, A.N.A., Brevis, W. et al. (1 more author) (2019) Quantifying the impact of uncertainty within the longitudinal dispersion coefficient on concentration dynamics and regulatory compliance in rivers. Water Resources Research, 55 (5). pp. 4393-4409. ISSN 0043-1397

https://doi.org/10.1029/2018wr023417

Water Resources Research

RESEARCH ARTICLE
10.1029/2018WR023417

Quantifying the Impact of Uncertainty within the Longitudinal Dispersion Coefficient on Concentration Dynamics and Regulatory Compliance in Rivers

V. V. Camacho Suarez[1], A. N. A. Schellart[1], W. Brevis[2], and J. D. Shucksmith[1]

[1]Department of Civil and Structural Engineering, University of Sheffield, Sheffield, UK, [2]Department of Hydraulics and Environmental Engineering and Department of Mining Engineering, Pontifical Catholic University of Chile, Santiago, Chile

Key Points:
- Uncertainty associated with longitudinal dispersion coefficients is quantified and propagated to time-concentration profiles using 1D ADE
- Application to case study found highly variable and considerable levels of uncertainty, influencing assessment of regulatory compliance
- Statistical criteria for evaluation of dispersion coefficient equations are less suitable for water quality regulation compliance evaluation

Correspondence to:
V. V. Camacho Suarez,
v.camacho@sheffield.ac.uk

Citation:
Camacho Suarez, V. V., Schellart, A. N. A., Brevis, W., & Shucksmith, J. D. (2019). Quantifying the impact of uncertainty within the longitudinal dispersion coefficient on concentration dynamics and regulatory compliance in rivers. *Water Resources Research*, *55*, 4393–4409. https://doi.org/10.1029/2018WR023417

Received 1 JUN 2018
Accepted 28 MAR 2019
Accepted article online 1 APR 2019
Published online 29 MAY 2019

Abstract The one-dimensional advection dispersion equation (1D ADE) is commonly used in practice to simulate pollutant transport processes for assessment and improvement of water quality conditions in rivers. Various studies have shown that the longitudinal dispersion coefficient used within the 1D ADE is influenced by a range of hydraulic and geomorphological conditions. This study aims to quantify the impact and importance of the parameter uncertainty associated with the longitudinal dispersion coefficient on modeled pollutant time-concentration profiles and its implications for meeting compliance with water quality regulations. Six regression equations for estimating longitudinal dispersion coefficients are evaluated, and commonly used evaluation criteria were assessed for their suitability. A statistical evaluation of the regression equations based on their original calibration data sets resulted in percent bias (PBIAS) values between −47.01% and 20.78%. For a case study, uncertainty associated with the longitudinal dispersion coefficient was propagated to time-concentration profiles using 1D ADE and Monte Carlo simulations, and 75% confidence interval bands of the pollutant concentration versus time profiles were derived. For two studied equations, the measured peak concentration values were above the simulated 87.5th percentile, and for the other four equations it was close to the 87.5th percentile. Subsequent uncertainty propagation analysis of four diverse rivers show the potential considerable impact on concentration-duration-frequency-based water quality studies, with 1D ADE modeling producing predictions of quality standard compliance which varied over hundreds of kilometers.

1. Introduction

Maintaining good surface river water quality standards for different uses (drinking, recreation, ecological habitat, etc.) is a challenging task due to the extensive list of complex and variable natural and anthropogenic factors affecting water quality conditions. Rivers are subject to variable physical, chemical, and biological processes that may affect their vulnerability to pollution loads (Chapman, 1996). River water quality models provide a tool for simulating such processes to assist in the assessment and improvement of water quality conditions that may not be otherwise obtained from field monitoring. However, lack of knowledge of water quality processes and the river system of interest can limit the reliability of the model predictions. Modeling uncertainties can thus lead to suboptimal water or infrastructure management decisions (Sriwastava et al., 2018). Thus, quantifying and communicating the accuracy of water quality predictions is a key component for improving water quality conditions and managing better water resources (Refsgaard et al., 2006; van Griensven & Meixner, 2006).

While uncertainties have been studied within other areas of catchment modeling, such as rainfall-runoff, groundwater, wastewater treatment, and urban drainage models (Arnbjerg-Nielsen & Harremoës, 1996; Beven & Binley, 1992; Dotto et al., 2012; Freni & Mannina, 2010; Mannina & Viviani, 2010b; Refsgaard et al., 2007; Schellart et al., 2010b; Willems, 2008), relatively few studies have focused on the uncertainties within surface water quality modeling; of these most relate to biochemical processes of specific substances. For instance Van Der Perk (1997) explored the model uncertainty and accuracy using eight phosphate concentration models for the Biebrza River, Poland. Although they found that when increasing model complexity, the accuracy of the model also increased, the parameter identifiability decreased and the parameters became increasingly correlated. Abbaspour et al. (2007) evaluated the capabilities of SWAT (Soil and Water

Assessment Tool) when modeling the Thur River basin in Switzerland. Using the 95% confidence intervals and the ratio of the mean distance in between the 95% confidence intervals and standard deviation, they found excellent predictions for flow and nitrate and good predictions for sediment and total phosphorus. Lindenschmidt et al. (2007) examined the structural uncertainty in modeling dissolved oxygen nutrients, phytoplankton dynamics, sediment, and micropollutants using the WASP5 package (Water quality Analysis Simulation Program) which coupled three models: (1) a hydrodynamic model, (2) a dissolved oxygen, nutrient, and phytoplankton model, and (3) a sediments model. Vandenberghe et al. (2007) utilized a Monte Carlo-based uncertainty propagation approach to examine predictive uncertainties in the ESWAT model due to a selection of water quality parameters and model inputs. Despite its importance in modeling time varying/dynamic river and pollution impacts (Boxall & Guymer, 2003; Mannina & Viviani, 2010a), studies investigating the impacts of uncertainties associated with mixing processes on water quality modeling and decision making are limited (Benke et al., 2008; Tscheikner-Gratl et al., 2018). Existing studies have mostly focused either on comparing the accuracy of calibrated models of varying complexity (Moghaddam et al., 2017) or on the uncertainty in estimating dispersion (or other) parameters in themselves using different methods, without propagating the effect of these parameter uncertainties within water quality modeling predictions (Alizadeh et al., 2017; Noori et al., 2016; Piotrowski et al., 2010; Sattar & Gharabaghi, 2015).

When simulating large (catchment scale) systems, the modeling of the physical transport of pollutants in rivers is commonly implemented within water quality models using the one-dimensional advection dispersion equation 1D ADE (Rutherford, 1994).

$$\frac{\partial C}{\partial t} = -v\frac{\partial C}{\partial x} + k_x\frac{\partial^2 C}{\partial x^2} \tag{1}$$

where C is the concentration (mg/L), t is the time (s), v is the river mean velocity (m/s), x is the distance downstream (m), and k_x is the longitudinal dispersion coefficient (m^2/s). The 1D ADE describes the change in cross-sectional averaged concentration of a solute with respect to time as a result of the advection and dispersion processes in turbulent flows (Fischer, 1979; Rutherford, 1994), dispersion being a product of differential advection (velocity shear) and turbulent diffusion processes. The 1D ADE is applicable in conditions where an equilibrium becomes established between the velocity shear and diffusion processes (Shucksmith et al., 2007). An appropriate solution to equation (1) is dependent on the boundary conditions.

Mixing processes and hence dispersion coefficients are known to be highly variable between rivers and over different hydraulic regimes. The presence of various common riverine features such as irregular bed forms, channel meandering, vegetation, pools, and riffles can largely influence such hydraulic and geometric conditions leading to significant variations in dispersion and mixing processes (Guymer, 1998; Noss & Lorke, 2016; Shucksmith et al., 2010). Despite its widespread use within modeling tools, several sources of uncertainty have been identified within the 1D ADE when applied to river systems. Primarily, the 1D ADE does not represent the asymmetry typically observed in tracer concentration profiles observed in field studies (van Mazijk & Veling, 2005). The persistent skewness in observed concentration profiles has been attributed to a number of processes, including transient storage effects and hyporheic exchange processes (Bottacin-Busolin & Marion, 2010; Briggs et al., 2009; Fernald et al., 2001; Nordin & Troutman, 1980; Zaramella et al., 2003), prolonged lack of equilibrium between diffusion and dispersion effects (Schmalle & Rehmann, 2014), and use of frozen cloud type approximations within field measurements (Rutherford, 1994). A number of modeling tools have been developed to account for profile asymmetries, which generally include additional or replacement terms and parameters to account for storage type effects in river systems (Runkel, 1998). However, this can lead to increased difficulties in parameter estimation due to issues associated with equifinality (Beven & Binley, 1992; González-Pinzõn et al., 2013) and generally requires more complex and well-designed measuring campaigns for calibration (Reichert & Vanrolleghem, 2001). Despite its limitations, the 1D ADE is still the most commonly used type of model for water quality assessments. In addition, most water quality assessments are relatively insensitive to the accurate prediction of distribution tails, instead being based on concentration exceedance frequencies and durations over given thresholds (F. W. R., 2012). A number of studies have shown that the calibrated 1D ADE is able to reproduce field observations of mixing processes with accuracy sufficient for such catchment scale water quality modeling applications, without the inclusion of transient storage/increased skewness effects (Ani et al., 2009; Launay et al., 2015; Marsili-Libelli & Giusti, 2008).

A common aspect of transport and mixing models is the identification of parameters via calibration/fitting of the model to observed data (Fischer, 1979). However, field measurements needed to calibrate mixing models over a range of flow conditions at a study site are often costly and time consuming. Several attempts have therefore been made to empirically and physically quantify the 1D ADE dispersion coefficient in terms of the underlying hydraulic processes and/or general river characteristics. Elder (1959) first derived an equation for this dispersion coefficient based on an analysis of an infinitely wide channel. This method has been generally recognized (Rutherford, 1994; Seo & Cheong, 1998) to underestimate natural dispersion in rivers due to the neglect of transverse shear dispersion processes. Fischer (1979) derived an equation for the dispersion coefficient that included a triple integral to account for the local transverse mixing. However, difficulties in accounting for the transverse mixing coefficient have been encountered mainly due to the absence of information regarding the transverse velocity and depth (Deng et al., 2001). More recently, numerous empirical equations to estimate dispersion coefficient based on geometrical river characteristics have been developed based on regression analyses of published data sets of tracer studies and the resulting fitted 1D ADE parameters (Kashefipour & Falconer, 2002; Liu, 1977; Magazine et al., 1988; Seo & Cheong, 1998; Zeng & Huai, 2014). These equations are commonly based on dimensional analysis of key hydraulic and geometric parameters known to influence dispersion and turbulent diffusion processes including the width, depth, mean velocity, and mean shear velocity. Such empirically based formulations of dispersion coefficient as a function of bulk river properties are often implemented within water quality models to determine longitudinal dispersion coefficient default parameters. For instance, the default longitudinal dispersion coefficient in the Qual2K water quality model is calculated using a regression equation from Fischer (1975). The default value within the D-Water Quality module used within the software packages Delft3D and SOBEK is calculated using a function based on the mean velocity, width, Chezy coefficient, and the total depth (Deltares, 2018). InfoWorks ICM uses a default equation based on bulk river characteristics (shear velocity, channel width, and mean flow velocity) to determine the dispersion coefficient (Innovyze, 2017). This, however, raises important questions regarding the sensitivity of water quality assessments (and associated decision making) to inaccuracies in the estimates of such parameters. Understanding the uncertainties introduced via the use of these methodologies for quantifying the dispersion coefficient based on bulk river characteristics is therefore of importance when considering the accuracy of water quality modeling studies. Little research on error propagation through existing calibrated models for water quality in rivers has been conducted to date (Benke et al., 2008), and work to understand the implications of the longitudinal dispersion coefficient uncertainty within water quality predictions is rare. While some studies have determined the accuracy of some parameter estimation techniques at specific case study sites (El Kadi Abderrezzak et al., 2015; Launay et al., 2015), or investigated the uncertainties resulting from the use of the 1D ADE (as well as an alternate stochastic transfer function-based approach) at a site at different flow rates (Romanowicz et al., 2013), to the authors' knowledge, there is a lack of studies that robustly estimate and propagate parametric uncertainties associated with the determination of the dispersion coefficient. The nature, scale, or significance of this uncertainty, its relationship to model structure uncertainty, or the associated implications for commonly deployed water quality assessments is therefore not currently well understood.

The aim of this paper is to quantify the impact of uncertainty introduced to river water quality modeling as a result of utilizing current state of the art regression equations to determine longitudinal dispersion coefficients. The assessment is based on the 1D ADE due both to its ongoing widespread application and to the availability of a significant number of historical published data sets over a range of field sites with which to robustly characterize parameter uncertainty. This paper first independently evaluates six longitudinal dispersion regression equations by quantifying their statistical accuracy against published data sets of tracer studies. Then, a Monte Carlo analysis is carried out to propagate uncertainty inherent in the empirical formulations of dispersion coefficient to time-concentration profiles for an independent river solute tracing case study. Finally, the paper estimates and discusses the potential impact of this uncertainty on water quality legislation compliance based on a concentration-duration-frequency analysis using rivers of different hydraulic and geometric characteristics.

2. Evaluation of Methodologies to Estimate Dispersion Coefficient in Rivers

This study identified and reviewed a range of methodologies and equations for predicting longitudinal dispersion coefficients in rivers. It was found that most regression analyses have been based on the same

Table 1
Evaluated Longitudinal Dispersion Equations and Number of Data Sets Used in Their Development

Name	Equation	Number of training/ calibration data sets
Deng et al. (2001)	$k_x = \frac{0.15}{8\,E_t}\left(\frac{v}{u^*}\right)^2\left(\frac{B}{H}\right)^{1.67} H u^*$, where	73
	$E_t = 0.145 + \left(\frac{1}{3{,}520}\right)\left(\frac{v}{u^*}\right)\left(\frac{B}{H}\right)^{1.38}$	
Etemad-Shahidi and Taghipour (2012)	if $\frac{B}{H} \leq 30.6$, $k_x = 15.49\left(\frac{B}{H}\right)^{0.78}\left(\frac{v}{u^*}\right)^{0.11} H\,u^*$	149
	if $\frac{B}{H} > 30.6$, $k_x = 14.12\left(\frac{B}{H}\right)^{0.61}\left(\frac{v}{u^*}\right)^{0.85} H\,u^*$	
Zeng and Huai (2014)	$k_x = 5.4\left(\frac{B}{H}\right)^{0.7}\left(\frac{v}{u^*}\right)^{0.13} H\,v$	116
Disley et al. (2015)	$k_x = 3.563\ Fr^{-0.4117}\left(\frac{B}{H}\right)^{0.6776}\left(\frac{v}{u^*}\right)^{1.0132} H\,u^*$	56
Wang and Huai (2016)	$k_x = 17.648\left(\frac{B}{H}\right)^{0.3619}\left(\frac{v}{u^*}\right)^{1.16} H\,u^*$	116
Wang et al. (2017)	$k_x = \left(0.718 + 47.9\frac{H}{B}\right) v\,B$	116

Note. The parameter k_x is the longitudinal dispersion coefficient (m^2/s), B is the river width (m), H is the river depth (m), v is the river mean velocity (m/s), u^* is the river mean shear velocity (m/s), and Fr is the Froude number.

underlying data set which has grown over time as more studies have been added. The data sets consist of published values of bulk river characteristics and "measured" dispersion coefficients. Typically, these values are based on averaged values of cross-sectional river surveys as well as the results of tracer study tests in which some form of parameter identification techniques (e.g., method of moments) have been performed to identify mixing parameters. However, given the size of the database and the unavailability of the raw data, it is not possible to robustly evaluate the accuracy of the underlying data sets. There is considerable overlap in the empirical basis for most published regression-based methods found in the literature. However, a number of statistical and regression analysis methods have been deployed in order to produce numerous formulations to calculate dispersion coefficients. In this study, we focus on six published equations (shown in Table 1) for a more rigorous evaluation and uncertainty analysis. These studies were selected because they contain large and clearly identifiable published data sets which are comparable, thus making a fair comparison of their potential predictive accuracy. In addition, as analysis techniques have progressed, the predictive power and accuracy of the regression equations have tended to grow over time. Therefore, by utilizing relatively recent methodologies, we aim to evaluate the best case in terms of uncertainty levels within water quality predictions. The identified equations are commonly based on regression analysis of key identified parameters such as the ratio between river mean velocity and shear velocity $\left(\frac{v}{u^*}\right)$ and river aspect ratio $\left(\frac{B}{H}\right)$. These parameters have been determined to be influential on calculating the dispersion coefficient by several studies via dimensional analysis and observed correlations (Kashefipour & Falconer, 2002; Zeng & Huai, 2014). The six equations selected for analysis are described below.

The equation presented in Deng et al. (2001) is based on Fischer (1975) triple integral for longitudinal dispersion coefficient. By deriving an expression for the transverse velocity profile in alluvial rivers, they considered the local velocity deviation from the cross-sectional average velocity and solved the triple integral to derive an analytical equation for the longitudinal dispersion coefficient. A regression data set was used to test the proposed equation to determine the suitability of the coefficients in the equation. Etemad-Shahidi and Taghipour (2012) developed a model tree method to produce an alternate equation to derive the longitudinal dispersion coefficient. The method consists of a recursive algorithm that performs the regression analysis on the underlying data sets by reducing a standard deviation factor. Zeng and Huai (2014) used a nondimensional analysis to determine that most equations underestimate the longitudinal dispersion coefficient for rivers with aspect ratios between 20 and 100. They suggested that a more accurate formula for longitudinal dispersion can be found via implementing an additional factor based on the mean velocity. Disley et al. (2015) performed dye-tracing experiments on a small stream in Ontario. Using the collected

Table 2
Summary of Statistical Analysis Based on Data Sets Used in the Construction of Dispersion Equations

Analyses	Wang et al. (2017)	Wang and Huai (2016)	Disley et al. (2015)	Zeng and Huai (2014)	Etemad-Shahidi and Taghipour (2012)	Deng et al. (2001)
Percent accuracy	61.2	63.8	73.2	61.2	62.4	64.4
$\mathrm{RSR}=\frac{\sqrt{\sum_{i=1}^{n}\left(k_{x_i}^{M}-k_{x_i}^{P}\right)}}{\sqrt{\sum_{i=1}^{n}\left(k_{x_i}^{M}-\overline{k_x}^{M}\right)}}$	0.78	0.83	0.39	0.88	0.68	1.41
$\mathrm{PBIAS}=\frac{\sum_{i=1}^{n}\left(k_{x_i}^{M}-k_{x_i}^{P}\right)*100}{\sum_{i=1}^{n}k_{x_i}^{M}}$	20.78	3.49	−3.01	−7.48	31.9	−47.01
$R^2=\left[\frac{\sum_{i=1}^{n}\left(k_{x_i}^{M}-\overline{k_x}^{M}\right)\left(k_{x_i}^{P}-\overline{k_x}^{P}\right)}{\sqrt{\sum_{i=1}^{n}\left(k_{x_i}^{M}-k_x{}^{M}\right)^2}\sqrt{\sum_{i=1}^{n}\left(k_{x_i}^{P}-\overline{k_x}^{P}\right)^2}}\right]^2$	0.41	0.43	0.85	0.44	0.63	0.36
$\mathrm{NSC}=1-\frac{\sum_{i=1}^{n}\left(k_{x_i}^{M}-k_{x_i}^{P}\right)^2}{\sum_{i=1}^{n}\left(k_{x_i}^{M}-\overline{k_x}^{M}\right)^2}$	0.39	0.31	0.84	0.23	0.53	−0.99

Note. Analysis includes percent accuracy, RMSE-observations standard deviation ratio (RSR), Percent bias (PBIAS), Coefficient of determination (R^2), Nash-Sutcliff coefficient (NSC). RMSE = root-mean-square error.

data and a selection of rivers from previous studies, they developed a new regression equation of longitudinal dispersion. Disley et al. (2015) incorporated the Froude number to capture the effect of the slope on dispersion processes. Wang and Huai (2016) based a new equation on an analysis of dispersion in a rectangular flume and applied this understanding to natural rivers. To obtain the longitudinal dispersion coefficient from a rectangular flume, they transformed the nonintegral form of the velocity distribution into a Fourier series to solve the triple integral for longitudinal dispersion. Consequently, they used 80% of their selected data set to train the algorithm developed for predicting the dispersion equation. Finally, Wang et al. (2017) suggested a concise form of the dispersion coefficient equation for various flow conditions. The study developed a general dispersion equation for pipe flows and calibrated it for natural rivers using a genetic algorithm model.

To initially evaluate the predictive accuracy of each of the regression equations presented in Table 1, a statistical analysis was conducted using the original regression data sets employed in the construction of each formulation (Table 2). The corresponding data sets were selected for the analysis to provide a fair evaluation of each model, so that each equation is only compared against its own regression data set. The statistical criteria used to evaluate the equations include (i) percent accuracy, (ii) the root-mean-square error (RMSE)-observations standard deviation ratio (RSR), (iii) percent bias (PBIAS), (iv) coefficient of determination (R^2), and (v) Nash-Sutcliff coefficient (NSC). A definition of these criteria can be found in Table 2. The optimal value of the RMSE-observations RSR is 0.0. RSR standardizes the RMSE using the observations and describes the residual variance (Moriasi et al., 2007). The percentage bias (PBIAS), is used to measure the tendency of the formulations to overestimate or underestimate the observed value. The optimal value of the PBIAS is 0.0. Therefore, the best performing equation is the one with the smallest absolute value of PBIAS (Gupta et al., 1999). Negative PBIAS value indicates that the equation is overpredicting the value of the longitudinal dispersion coefficient, while positive values indicate underprediction. The accuracy is the percentage of predictive ratios (*Pr*) between 0.5 and 2 or its equivalent logarithmic range between −0.3 and 0.3 (Seo & Cheong, 1998; White et al., 1973). The R^2 describes the degree of collinearity between the predicted and observed data. It also indicates the proportion of the observed data that is explained by the variance. Higher values indicate less error variance. However, R^2 does not detect systematic overprediction or underpredictions (Krause et al., 2005). The NSC is a normalized indicator of the performance of the equation. The weakness of the NSC is that, because it is squared, it is more sensitive to high values in the data set than lower values (Krause et al., 2005). These statistical measures were selected because they are commonly used to evaluate, train, or optimize the dispersion equations (Disley et al., 2015; Etemad-Shahidi & Taghipour, 2012; Sattar & Gharabaghi, 2015; Seo & Cheong, 1998).

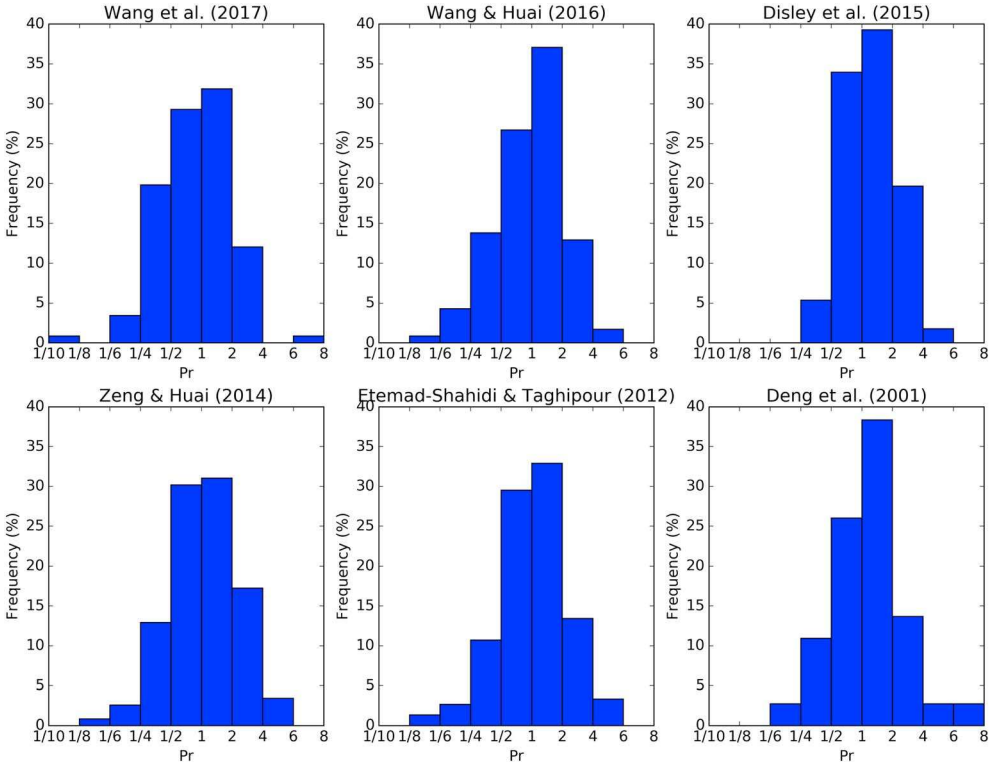

Figure 1. Probability histograms of predictive ratios (*Pr*) obtained from regression data sets used in the construction of dispersion equations.

Table 2 presents the statistical results of the evaluation of the dispersion equations. According to the statistical model evaluation techniques described above, Disley et al. (2015) equation has the highest accuracy, lowest RSR, least absolute PBIAS, highest R^2, and highest NSC. The Etemad-Shahidi and Taghipour (2012) equation has the second lowest RSR and second highest R^2 and NSC. Deng et al. (2001) equation has the second highest accuracy. From the negative PBIAS (Table 2), it is noted that the Disley et al. (2015), Zeng and Huai (2014), and Deng et al. (2001) equations tend to overpredict the dispersion coefficient.

Figure 1 shows the distribution of Predictive ratios (*Pr*) values ($Pr = k_x{}^P / k_x{}^M$; where $k_x{}^P$ and $k_x{}^M$ are the predicted and measured longitudinal dispersion coefficients), grouped into histogram bins, for each of the equations. The equations that have the highest percentages of *Pr* values within 0.5 and 2.0 are the most accurate equations as also defined by the "Percent accuracy" in Table 2. Among the studied equations, it is observed that Disley et al. (2015) equation has the largest amount of *Pr* values between 0.5 and 2.0 (73.2%). The other equations all have similar *Pr* values between 0.5 and 2.0, ranging between 61.2% and 64.4%.

3. Methodology for Uncertainty Propagation

This section presents a methodology to evaluate the uncertainty within water quality modeling due to the selection of the dispersion coefficient using the equations detailed in Table 1, utilizing a data set from an independent dye-tracing experiment conducted in the Chillan River. A general background description and discussion of uncertainty propagation using Monte Carlo methods within environmental modeling can be found in, for example, Benke et al. (2018) and Helton (1993). The Chillan River field study has been selected as the data have not been included in the derivation of any of the studied dispersion coefficient equations and because field survey and solute concentration versus time profiles are available, measured downstream of an instantaneous release of tracer. The Chillan River is located in Chile's 8th Region, approximately 400 km south of Santiago de Chile. It emerges from the Andes Mountains and flows west until it meets the Ñuble River (Brevis et al., 2001). In May and April of 2003, river survey information was

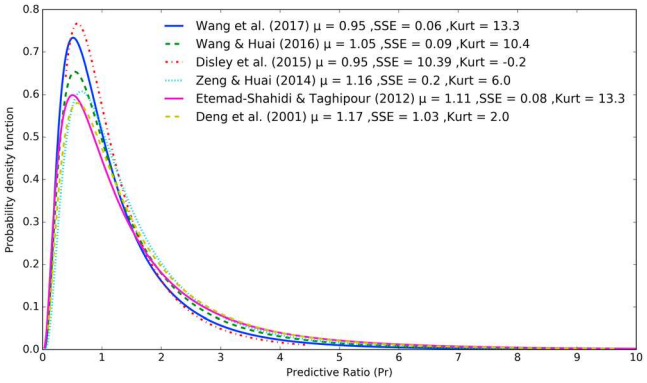

Figure 2. Fitted probability distributions for predictive ratio (*Pr*) using regression data sets with their corresponding mean (μ) sum of square errors (SSE), and kurtosis (Kurt).

collected alongside two tracer experiments at a site close to the city of Chillan. Sixty milliliters of 20% Rhodamine WT tracer was released in the river following the general guidelines described in Hubbard et al. (1982). The tracer experiment carried out in May 2003 was selected because the tracer breakthrough curve was complete and the river hydraulic and geometric data were available. The data were obtained following the same methodology for concentration measurements as the one used in De Smedt et al. (2005), also taken at the Chillan river. A calibrated fluorimeter, Turner Designs Model 10, with a detection limit of 0.01 ppb, was used for the concentration measurements. Samples were analyzed in situ and in the laboratory. The in situ samples were immediately analyzed using the fluorimeter. Due to time overlap between concentration curves at different sampling stations, some of the samples were stored in thermally isolated compartments and later analyzed using the same fluorimeter and the same calibration curves. The samples were periodically checked for changes in the pH of the river. Comparison of the total mass between the sampled concentrations curves confirmed that losses of Rhodamine mass between stations can be assumed negligible. The resulting river concentration versus time profiles in the study reach were obtained from samples taken at measurement points positioned 2.5 and 3.8 km downstream of the release (after full cross-sectional mixing of the solute). A total of 28 cross-section surveys were carried out between the upstream and downstream sampling points containing hydraulic and geometric information. The study reach was divided into several consecutive subreaches between each pair of cross sections. Longitudinal and transverse survey data were collected and then digitized using AUTOCAD 2000 to determine the cross-sectional area (reach mean = 10.6 m^2, standard deviation [std. dev] = 7.5 m^2), wetted perimeter (reach mean = 17.2 m, std. dev = 5.7 m), surface width (reach mean = 16.4 m, std. dev = 6.1 m), depth (reach mean = 0.7 m, std. dev = 0.5 m), sinuosity (reach mean = 1.5, std. dev = 0.2), and average slope (reach mean = 0.005, std. dev = 0.002). During the selected tracer experiment, flow measurement was carried out at the injection site. A current meter (OTT Waterflow) was used to determine velocity over the cross section, and these measurements were integrated over the cross section to calculate mean flow rate according to standard practice described in standard ISO-748:2007(E) (ISO 1100–2, 2010). The flow rate was calculated as 2.6 + 0.05 m^3/s. The mean velocity at each measured cross section was calculated using the flow rate and measured wetted area resulting in a reach mean and standard deviation of 0.45 and 0.34 m/s, respectively. Further details and results of the experiment are presented in Segura (2004).

To quantify and propagate the parameter uncertainty from the various dispersion equations, the following steps were conducted for each equation analyzed:

1. Probability functions were fitted to the distributions of *Pr*, (from Figure 1). The distributions were fitted using the Python package (Fitter) developed by (Cokelaer, 2014). Fitter evaluates 80 function types from the statistical distributions of the Scipy package (Oliphant, 2007). In all cases, the *Pr* distributions were best described by lognormal functions. This concurs with previous studies which evaluate *Pr* (Kashefipour & Falconer, 2002; Seo & Cheong, 1998; Zeng & Huai, 2014). The probability distributions with their corresponding mean, sum of square errors (SSE), and kurtosis are shown in Figure 2.
2. A Monte Carlo analysis was carried out obtaining 2,000 randomly drawn *Pr* values from each logarithmic probability function. These *Pr* values were used to adjust the deterministic dispersion coefficient value calculated using each dispersion coefficient equation for each river subreach using the measured and calculated hydraulic and geometric information (*B,H, v, u* Fr*) derived from the field measurements. This is similar to the method used by Schellart et al. (2010a) for studying uncertainty inherent in coefficients in existing regression equations. For example, Figure 2 shows that using the Disley et al. (2015) equation, the predicted k_x could be anywhere between approximately 0 and 10 times the "possible real k_x," so dividing the predicted k_x by each of the randomly drawn *Pr* values would give 2,000 possible real k_x values. A straightforward Monte Carlo simulation was deemed the most suitable approach, due to its conceptual simplicity as well as its ease of explanation to, for example, regulators (Benke et al., 2018; Helton, 1993; Sriwastava et al., 2018).

3. Using an analytical solution of the 1D ADE, given by equation (2) below (Rutherford, 1994), and the "possible real k_x values" (based on the drawn *Pr* value from step 2), the downstream concentration profile located at 3.8 km was calculated. This was achieved by successively routing the observed upstream concentration profile (at 2.5 km) over each subreach until the concentration profile at the last subreach was obtained (utilizing the geometric and hydraulic data). This resulted in 2,000 possible predicted downstream concentration profiles.
4. The 12.5th, 50th, and 87.5th percentiles of each concentration distribution were then identified and compared with the observed concentration profile and with the routed concentration profile obtained using the deterministic dispersion coefficient. The 75% confidence intervals were selected because these can be estimated more reliably than larger confidence intervals based on the relatively limited available data.
5. To remove errors caused by the field measurements of velocity, the observed concentration profile at the last subreach was used to calibrate the travel time and mean velocity of each subreach; that is, the total travel time adjusted until a match was achieved between the observed and predicted concentration distribution centroids. This was required as the aim of this work is to identify the uncertainty associated with the dispersion coefficient, rather than the initial estimation of velocity caused by field measurement. Consequently, the velocities for each subreach were corrected proportionally based on the recalculated travel time. Steps 2–4 were repeated with the corrected mean velocity and travel time values to produce the final predictions and confidence bands.

$$C(x_2,t)=\sum_{\tau=-\infty}^{\infty}\frac{C(x_1,\tau)\,v\,\Delta\tau}{\sqrt{4\pi k_x T}}\exp\left(-\frac{v^2(T-t+\tau)^2}{4k_x T}\right) \tag{2}$$

Where $C(x_1,\tau)$ is the temporal concentration profile at x_1 (upstream of each subreach) at time τ, $C(x_2,t)$ is the concentration at the location x_2 (downstream of each subreach) and time t, v is the mean velocity over the subreach, k_x is the longitudinal dispersion coefficient, and T is the travel time over the cross section, initially calculated (precalibration) using the cross-section distances and measured velocity. Equation (2) is based on Taylor's analytical solution to equation (1), utilizing the frozen cloud approximation to convert between the temporal and spatial domains. A full discussion of this solution can be found in Rutherford (1994).

3.1. Uncertainty Quantification Results

The mean, SSE, and kurtosis of each fitted lognormal distribution for the corresponding dispersion equations are shown in Figure 2. The mean values of the distributions range between 0.95 and 1.17 (with 1.0 representing the perfect agreement between predicted and measured coefficients). Wang et al. (2017), Wang and Huai (2016) and Disley et al. (2015) have the narrowest distributions and mean values closest to 1.0 as also noted by the narrower histogram in Figure 1, while the remaining dispersion equations have wider distributions and have mean values higher than 1.0 (hence an average overprediction). Disley et al. (2015) has the largest SSE, while Wang et al. (2017) has the smallest SSE indicating a better fit between measured and predicted dispersion coefficients in the database. Wang et al. (2017), Wang and Huai (2016), and Etemad-Shahidi and Taghipour (2012) have the highest levels of kurtosis indicating longer tails. Heavier tails indicate that some predictions heavily overestimate the dispersion coefficients. However, it should be noted that distribution tails are very sensitive to small numbers of outlying data. Disley et al. (2015) equation results in the highest probability density in Figure 2 which is in agreement with having the highest accuracy in Figure 1.

Figure 3 presents the results of the uncertainty propagation methodology when applied to the field data set from the River Chillan. Figure 3 displays the observed concentration profiles at the upstream and downstream measurement stations and the predicted concentration profiles based on the dispersion coefficients calculated using each of the deterministic equations in Table 1 and the analysis presented above. The predicted concentration profiles include the deterministic prediction, the 50th percentile (median), and the 12.5th and 87.5th percentiles (75% confidence interval) resulting from the Monte Carlo analysis. To show the influence of the changes of the river characteristics (e.g., river depth and width) on the deterministic dispersion coefficients, the reach mean and standard deviation values of the river reaches deterministic dispersion coefficients are shown in Figure 3. It is noted that the standard deviation of predicted dispersion

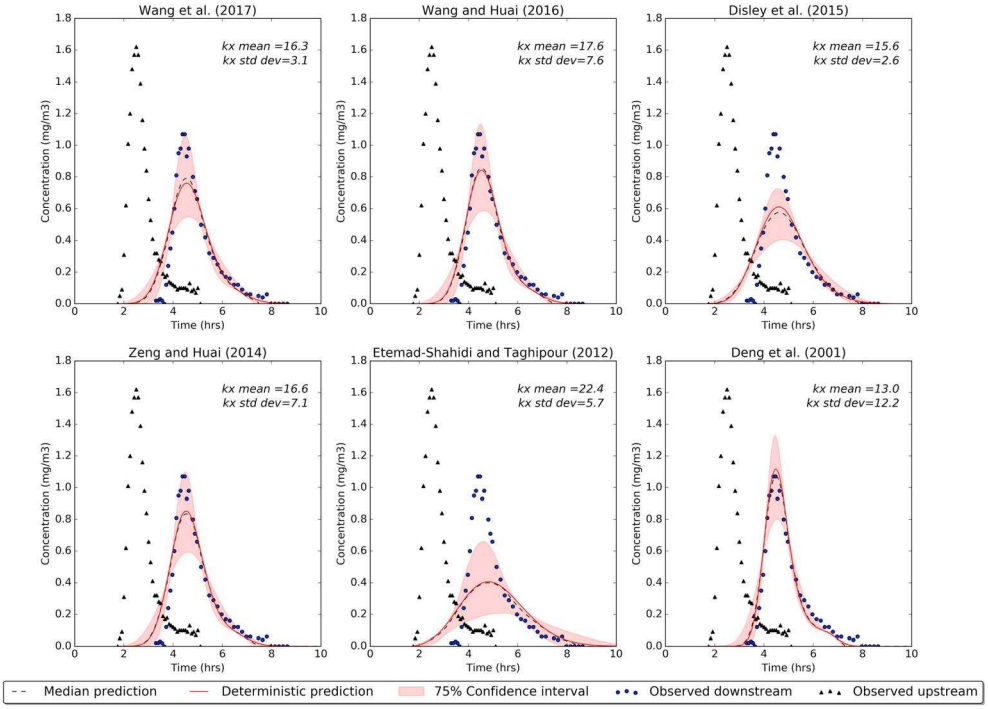

Figure 3. Concentration versus time profiles retrieved when using six different dispersion equations and their corresponding data sets for sampling stations of the river Chillan. Shaded bands represent the 75% confidence interval. k_x mean and k_x std dev are the results of the mean and standard deviation of the dispersion coefficients for the 28 river cross sections.

coefficient over the subreaches varies significantly between the equations, indicating that some equations are more sensitive to longitudinal variations in the river characteristics. The largest mean deterministic longitudinal dispersion coefficient leads to flatter concentration profiles as observed in Figure 3 for Etemad-Shahidi and Taghipour (2012) with lower peak concentration values. The opposite is true for a low dispersion coefficient when using Deng et al. (2001) equation. This results in a taller and narrower concentration versus time profile. It is noted that a significant proportion of observed concentration values fall outside the 75% confidence intervals when using Etemad-Shahidi and Taghipour (2012) and Disley et al. (2015) equations. The concentration versus time profiles obtained using Wang et al. (2017), Wang and Huai (2016) and Zeng and Huai (2014) dispersion equations have similar 75% confidence intervals, median, and deterministic concentrations. The deterministic predictions from these dispersion equations still underestimate the observed concentrations, but the observed concentrations are within the 75% confidence intervals. The simulated concentrations using the deterministic dispersion coefficient predicted by the Deng et al. (2001) equation visually resemble the observed concentrations more accurately than the other equations, with the observed concentrations well within the 75% confidence interval. It is noted that almost all predicted profiles within the 75% confidence interval fail to reproduce the early leading edge of the observed concentration profiles. Overall, the methodology has been shown to provide additional information in regard to concentration predictions over and above the use of the deterministic models, with uncertainty bands encompassing the observed concentration values. Five out of the six studied deterministic equations underestimate observed peak concentration levels (by an average of 29%). Etemad-Shahidi and Taghipour (2012) equation results in the largest underestimation among the studied equations by approximately 64%. Such underpredictions indicate that mixing processes are generally lower in the River Chillan than is predicted by the studied regression equations. Confidence intervals are of considerable size but are approximately equivalent between the equations, indicating the inherent uncertainty associated with the evaluation of dispersion coefficients by using regression equations derived from data from other rivers. The simulated 12.5th percentile concentration profiles resulted in simulated peaks between 26% and 81% of the measured value.

Table 3
Measured and Deterministic Dispersion Coefficient Values Together With Median and 75% Percentile Values Obtained From Monte Carlo Analysis for the Four Selected Rivers

Parameters	John Day River			Monocacy River			Copper Creek			New River		
Width (m)	34.1			92.9			18.6			102		
Depth (m)	2.47			0.71			0.39			4.4		
Mean velocity (m/s)	0.82			0.16			0.14			0.17		
Shear velocity (m/s)	0.18			0.046			0.116			0.008		
Aspect ratio $\frac{B}{H}$ (–)	13.8			130.8			47.7			23.2		
Mean to shear velocity ratio $\frac{v}{u^*}$ (–)	4.6			3.5			1.2			21.3		
Measured k_x (m^2/s)	65.0			41.4			9.9			22.4		
k_x Equation (m^2/s)	Deterministic k_x	k_x 50	k_x 12.5, k_x 87.5	Deterministic k_x	k_x 50	k_x 12.5, k_x 87.5	Deterministic k_x	k_x 50	k_x 12.5, k_x 87.5	Deterministic k_x	k_x 50	k_x 12.5, k_x 87.5
Wang et al. (2017)	117	110	19, 420	16	14	6, 34	4	4	2, 11	48	43	16, 110
Wang and Huai (2016)	118	127	20, 501	14	15	6, 34	4	4	2, 10	67	66	25, 178
Disley et al. (2015)	91	107	42, 298	35	41	24, 76	8	9	5, 18	105	123	65, 249
Zeng and Huai (2014)	84	91	16, 344	22	25	11, 54	5	5	2, 12	54	59	22, 147
Etemad-Shahidi and Taghipour (2012)	63	68	8, 329	26	28	10, 70	8	8	3, 24	9	9	3, 26
Deng et al. (2001)	71	81	15, 344	26	30	13, 68	4	4	2, 11	93	100	40, 268

4. Impact of Dispersion Coefficient Uncertainty on Concentration-Duration Threshold-Based Standards

Section 3 presented a propagation methodology to estimate uncertainty within surface water quality predictions associated with the dispersion coefficient derived using the regression equations based on river characteristics. To understand the potential implications of this uncertainty, this section evaluates the propagated uncertainty from the dispersion coefficient taking into consideration water quality standards and regulatory guidance in a site-specific context. Such guidelines and water quality standards have been developed and improved over the years to protect aquatic life from situations that may cause stress in river environments (Milne et al., 1992). One methodology widely used in the United Kingdom to regulate rainfall-driven time-varying releases (e.g., from urban drainage systems) into receiving waters is the intermittent standards approach. This consists of defined concentration-duration-frequency thresholds for specific substances (F. W. R., 2012). With this approach, dissolved oxygen and un-ionized ammonia concentrations must not exceed given thresholds for longer than specified durations, with values based on the return period of the storm event.

To evaluate the uncertainty due to the empirical dispersion equations with regard to concentration-duration-frequency water quality regulation, an analysis of four rivers of different geometrical and hydraulics properties (Table 3) obtained from the data set in Wang and Huai (2016) is conducted. This data set was selected because it was the most extensive data set with the most overlapping data among the evaluated studies. The measured, deterministic (from each equation in Table 1), and the upper, median and lower quantiles of the dispersion coefficients for the four rivers as calculated using the method described in section 3 are also shown in Table 3. John Day River represents a deep (2.5 m) river with one of the lowest aspect ratios $\left(\frac{B}{H}\right)$ of 13.8 in the data set. The measured dispersion coefficient of 65 m^2/s was the largest among the studied rivers. The Monocacy River is a shallow river with one of the largest aspect ratios (130.8) and largest widths (92.9 m). Its measured dispersion coefficient was 41 m^2/s. The Copper Creek and New River show the contrast between a low versus a high mean to shear velocity ratio $\left(\frac{v}{u^*}\right)$. Copper Creek has a mean shear velocity of 0.116 m/s and thus a low mean to shear velocity ratio (1.2) and a measured dispersion coefficient of 10 m^2/s. The New River has a lower shear velocity (0.008 m/s), high mean to shear velocity ratio (21.3), and a measured dispersion coefficient of 22 m^2/s. In each river, we utilize a pseudoconcentration distribution of ammonia and route it downstream, utilizing the same methodology as presented in section 3 to estimate confidence intervals. A constant cross section and flow were applied to the simulated rivers. At discrete positions (every 200 m) downstream of the release, the duration over a specified

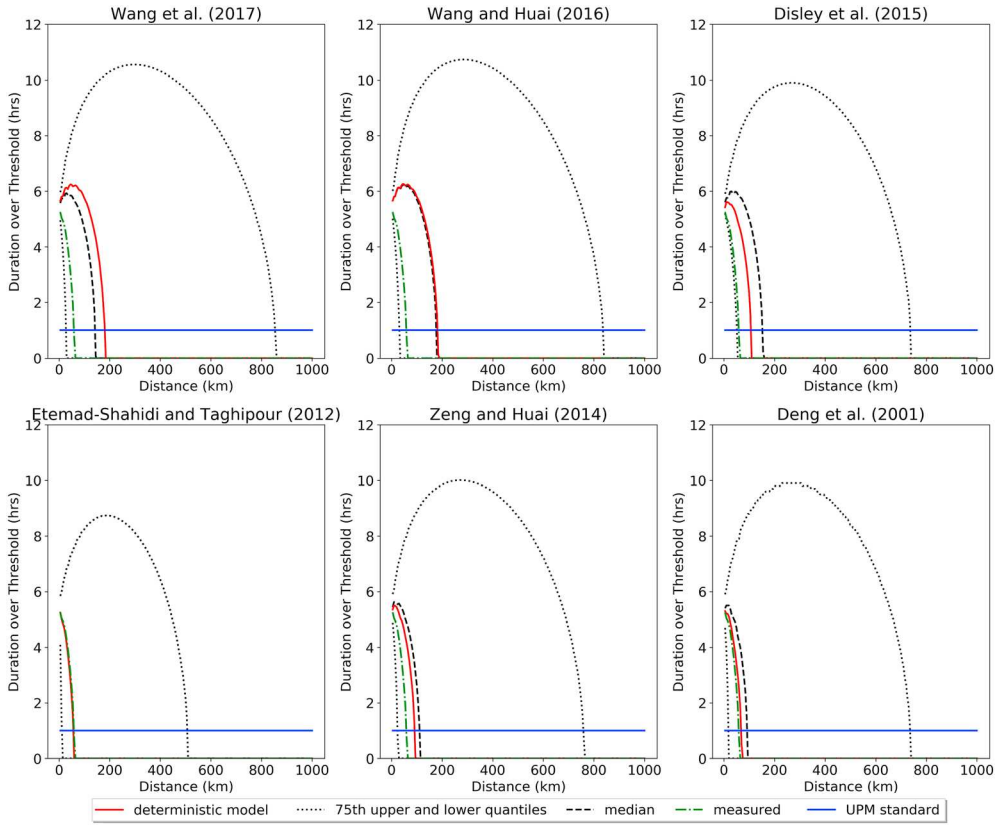

Figure 4. Duration over threshold versus distance for John Day River. Dotted lines present upper and lower 75% confidence intervals, red continuous line displays the deterministic k_x equation used, black dashed line presents the median of Monte Carlo simulation, blue continuous line presents the water quality set threshold, and green semidotted line represents the duration over threshold values obtained when using the measured k_x.

threshold is determined and evaluated in light of U.K. concentration-duration water quality standards for ammonia. The analysis illustrates how uncertainty in the use of dispersion coefficient regression equations has the potential to influence the degree of compliance with water quality regulation within modeling studies.

The initial concentration distribution (upstream boundary condition) is based on an integrated modeling study presented by Norris et al. (2014) where an integrated catchment in the United Kingdom was modeled using Infoworks ICM. The peak of the simulated initial concentration profile was 0.61 mg N/L, and its duration was 15 hr. The distribution is considered only as an indicative description of a potential concentration and is used in the analysis to determine the confidence intervals of predictions using 1D ADE modeling. The water quality threshold was obtained from the Urban Pollution Drainage Manual (F. W. R., 2012) for Salmonid Fishery Standards for 1-hr/1-year events, specified as 0.105 mg N/L unionized ammonia. It is noted that to model ammonia concentrations, it may require to incorporate other processes such as biological oxygen demand decay, nitrification, uptake by plants and bacteria, and heterotrophic respiration as applied in other studies (Lopes et al., 2005; Radwan et al., 2001). However, as more parameters are added to the transport model, the parameter uncertainty is less identifiable. This study aims to analyze the uncertainty from the dispersion coefficient only; thus, the ammonia concentration is assumed conservative within the river reach.

4.1. Duration Over Threshold Analysis Results

Figure 4 presents the resulting simulated durations that the ammonia concentrations exceed the water quality threshold (0.105 mg N/L) as a function of distance downstream of the simulated initial distribution within the John Day river when utilizing (i) observed dispersion coefficient values (from the original

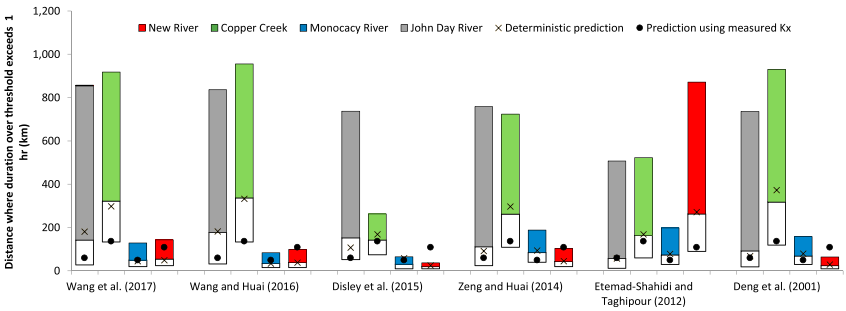

Figure 5. The 75th quantile boxplots of distance where the duration over threshold exceeded the allowed time as stated in the water quality regulation for the studied dispersion coefficient equations.

database), (ii) dispersion values predicted using each of the deterministic equations, and (iii) median plus 75% confidence intervals derived using the method described above. The horizontal line represents the specified maximum 1-hr duration of exceedance for a 1-year return period event; when the simulated duration falls below this level of compliance, the standard is achieved. Figure 4 shows that four deterministic equations overestimate the length of river stretch where the water quality threshold is exceeded compared to observed mixing parameters in the John Day River. Overall, there is considerable uncertainty regarding the length of river section that exceeds the 1-hr/1-year return period standards for ammonia when using regression equations for the determination of the dispersion coefficient. For example, when using Wang et al. (2017) equation, the duration over threshold exceeds 1 hr at 181 km downstream of the release when the deterministic dispersion coefficient is used. However, this distance varies between 28 and 854 km if the 75% confidence interval is considered. Figure 4 shows that when using the measured dispersion coefficient, the duration over threshold values fall within the 75% confidence interval bands in all cases. The Etemad-Shahidi and Taghipour (2012) equation results in the narrowest 75% confidence interval.

Figure 5 summarizes this information for all four of the rivers, showing the distance downstream of the release where the modeled pollutant has exceeded the 1-hr/1-year threshold using the deterministic dispersion coefficients from each of the regression equations and the 75% confidence intervals using the methodology outlined in section 3 as well as when using the measured value of dispersion coefficient for each river. The larger measured dispersion coefficient values for the John Day River and the Monocacy River (see Table 3) mean that the pollutant disperses faster and the 1-hr/1-year standard is achieved after a shorter distance. In most cases, there are considerable differences between the measured and deterministically estimated dispersion coefficients. The closest prediction is obtained when using the Etemad-Shahidi and Taghipour (2012) equation in the John Day river (2-km difference). The largest difference between predicted and measured values is found when using the Deng et al. (2001) equation in the Copper Creek (236-km difference). Considering all four rivers, the Disley et al. (2015) equation provides the closest predictions on average (43-km difference). When considering the 75% confidence intervals, considerable differences are observed between rivers and between equations; however, values using measured dispersion coefficients lie within predicted confidence intervals in almost all cases apart from the New River.

5. Discussion

This study examines six published equations for estimating the longitudinal dispersion coefficient with an independent analysis of published data (section 2). The study then proposes a method to propagate the uncertainty to concentration versus time profiles (section 3) and assesses the implications that this propagated uncertainty may have on testing compliance with water quality regulation (section 4).

The results showed that the equation by Disley et al. (2015) performed best in describing the longitudinal dispersion coefficient according to the efficiency criteria (higher percent accuracy, R^2, and NSC; least PBIAS; and lowest RSR), while the equation by Deng et al. (2001) resulted in a poorer performance (second highest accuracy but most PBIAS and lowest NSC). Previous studies used similar statistical criteria to evaluate equations for predicting the longitudinal dispersion coefficient (Disley et al., 2015; Etemad-Shahidi &

Taghipour, 2012; Sattar & Gharabaghi, 2015; Seo & Cheong, 1998). However, as far as the authors are aware, no studies have indicated which of these efficiency criteria would be most useful in terms of evaluating the performance of the equation when subsequently calculating concentration versus time profiles. This study presents a propagation methodology to analyze the effect of uncertainty inherent in the dispersion coefficient on the resulting concentrations when using the 1D ADE model. Section 3 uses this methodology to estimate confidence intervals on the concentration versus time profile of an independent tracer study measured in the Chillan River. Figure 3 shows that in the case of the river Chillan, five out of six equations tend to overpredict the dispersion coefficient. Although Deng et al. (2001) equation had a poor performance according to the efficiency criteria in Table 2, it provides the best visual resemblance to the observed concentrations. In contrast, Disley et al. (2015) is the best performing equation according to the efficiency criteria but underestimates the observed concentrations considerably even when the confidence interval is taken into consideration (Figure 3). This demonstrates that not all the efficiency criteria (percent accuracy, RSR, PBIAS, R^2, and NSC) presented in Table 2 appear equally suitable for selecting a dispersion equation to use with the 1D ADE model. For the same reason, Nash-Sutcliffe efficiency seems particularly unsuitable for selecting the best performing dispersion equation. As described by Krause et al. (2005), the largest disadvantage of the Nash-Sutcliffe efficiency is the fact that the differences between the observed and predicted values are calculated as squared values. This means that larger values in a data set are overemphasized, whereas lower values are neglected; however, in case of the k_x coefficient both high and low values are of equal importance. If the aim of using the 1D ADE model is to check environmental standards based on concentration-duration-frequency, then looking at the PBIAS may give a better indication of how "conservative" an equation is when checking environmental standards. A positive PBIAS value indicates that the equation is underpredicting the dispersion coefficient and hence more likely to fail a water quality standard.

The bulk river characteristics of the river Chillan would suggest that all the six longitudinal dispersion equations studied would be equally suitable for application. The performance of the equations and the scale of the uncertainty bands suggest that despite academic focus on regression equations to provide dispersion coefficients based on bulk river characteristics, considerable uncertainty remains when dispersion coefficients are utilized within modeling tools to describe concentration versus time dynamics within water quality modeling applications. Uncertainty within the estimation of dispersion coefficients is likely due to a number of reasons. These include the accuracy of the original data sets used in regression model calibration. For example, the practical difficulty in the measurement of bulk river characteristics (e.g., bed shear stress or river depth) over the same reach as the dispersion coefficient means that the calibration data sets are prone to error due to the averaging of these key geometrical and hydraulic parameters over the river reach. In addition, there is lack of information regarding the original tracer experiment data sets from which the dispersion coefficients are derived. The quantification of the dispersion coefficient is prone to measurement error if data processing techniques are not conducted in a robust manner and field experiments are not conducted appropriately. It should also be noted that bulk river characteristics cannot fully describe the complexity of mixing processes in river systems, which are heavily affected by conditions such as sinuosity; presence of vegetation, pools, and riffles; planform variability; and hyporheic exchange. It is therefore questionable if further statistical analysis of such existing data sets can produce regression equations with the potential to describe dispersion coefficients with sufficient accuracy such that model confidence intervals could be meaningfully reduced.

The implications of the uncertainty inherent within the longitudinal dispersion coefficient equations on water quality regulation were examined by calculating the duration that a pollutant exceeded a water quality standard. This calculation was carried out for four rivers in the data set shown in Wang and Huai (2016). It was observed that wide ranges of uncertainty are obtained for the John Day River and Copper Creek. This implies that the water quality failure can occur over a larger interval downstream of the pollutant release (over hundreds of kilometers). The opposite was found for the Monocacy River and New River. The uncertainty interval is smaller, making it more likely to obtain an accurate estimation of where the water quality failure occurs. However, the results for all four rivers indicate that even when using the most recent equations for estimating longitudinal dispersion coefficient, a considerable level of uncertainty inherent to these equations remains when determining water quality failures. To produce water quality simulations with lower uncertainties, robust calibration of a river-specific k_x, using dye-tracing studies over a range of flow rates is recommended. Further options include the use of more complex 2-D models in which dispersion

is less important (due to a lack of width averaging); however, this option is often limited to small reaches due to computational cost. Work on alternate modeling approaches which seek to quantify and describe processes such as transient storage are of value as much due to the potential for more stable and predictable parameters (relatable to measurable physical properties, Briggs et al., 2009) as their enhanced ability to describe specific properties of concentration distributions.

6. Conclusion

This paper examines uncertainty in 1D ADE model predictions of time-concentration profiles, given uncertainty inherent in using existing regression equations for estimating the longitudinal dispersion coefficient. Six recently published longitudinal dispersion equations are independently evaluated and compared. When considering dispersion coefficient prediction, this evaluation indicates that Disley et al. (2015) equation has the highest accuracy (73.2%), while the remaining equations have similar accuracies ranging between 61.2% and 64.4%. It is argued that evaluation criteria such as PBIAS may be important to include in the evaluation, due to its capability to indicate underprediction or overprediction of the dispersion coefficient, which are both important for estimating duration of concentration peaks over a threshold. It is also concluded that Nash-Sutcliffe is not a suitable criterion for evaluation of dispersion coefficient equations, as it neglects lower coefficient values, which for the purpose of estimating duration of concentration peaks over a threshold are equally important as high values.

Using Monte Carlo simulations, the uncertainty in the longitudinal dispersion coefficient given these six equations is propagated through the 1D ADE to create time-concentration profiles for an independent case study. Results from a case study site suggest that when using Deng et al. (2001) equation, the closest prediction of peak concentrations to observed values (approximately 3% difference between measured and 50th percentile predicted peak concentration) are obtained, as well as the narrowest uncertainty interval. However, the resulting uncertainty intervals were considerable for all the six studied regression equations. For the Disley et al. (2015) and studied Etemad-Shahidi and Taghipour (2012) equations, the measured peak concentration values were above the simulated 87.5th percentile; for the Deng et al. (2001) equation it was close to the 50th percentile, and for the other equations it was close to the 87.5th percentile. The simulated 12.5th percentile resulted in simulated peaks between 26% and 81% of the measured value.

Finally, the uncertainty methodology has been implemented into four rivers with different characteristics, and the interaction with concentration-duration-frequency type regulatory targets has been considered. It is shown that the resulting model confidence intervals are likely to be significant for assessment of regulatory compliance in areas with complex prescriptive concentration-based targets (e.g., the United Kingdom) as observed for the John Day River and the Copper Creek. Moreover, the effect of uncertainty is highly variable between rivers with different characteristics.

Within water quality assessments this highlights the value of using longitudinal dispersion coefficients derived specifically from field measurements for the river under study. A reduction of uncertainty in estimation of longitudinal dispersion coefficient using regression equations is likely to be dependent on further understanding and quantification of how other, more detailed river features affect mixing processes and dispersion coefficients and an incorporation of such features within regression-based models.

Acknowledgments
This research was conducted as part of the Marie Curie ITN—Quantifying Uncertainty in Integrated Catchment Studies (QUICS) project. This project has received funding from the European Union's Seventh Framework Programme for research, technological development, and demonstration under grant agreement 607000. The data used in this study are presented at the website (http://doi.org/10.5281/zenodo.1464738).

References

Abbaspour, K. C., Yang, J., Maximov, I., Siber, R., Bogner, K., Mieleitner, J., et al. (2007). Modelling hydrology and water quality in the pre-alpine/alpine Thur watershed using SWAT. *Journal of Hydrology*, *333*(2–4), 413–430. https://doi.org/10.1016/j.jhydrol.2006.09.014

Alizadeh, M. J., Ahmadyar, D., & Afghantoloee, A. (2017). Improvement on the existing equations for predicting longitudinal dispersion coefficient. *Water Resources Management*, *31*(6), 1777–1794. https://doi.org/10.1007/s11269-017-1611-z

Ani, E.-C., Wallis, S., Kraslawski, A., & Agachi, P. S. (2009). Development, calibration and evaluation of two mathematical models for pollutant transport in a small river. *Environmental Modelling & Software*, *24*(10), 1139–1152. https://doi.org/10.1016/j.envsoft.2009.03.008

Arnbjerg-Nielsen, K., & Harremoës, P. (1996). The importance of inherent uncertainties in state-of-the-art urban storm drainage modelling for ungauged small catchments. *Journal of Hydrology*, *179*(1–4), 305–319. https://doi.org/10.1016/0022-1694(95)02841-2

Benke, K. K., Lowell, K. E., & Hamilton, A. J. (2008). Parameter uncertainty, sensitivity analysis and prediction error in a water-balance hydrological model. *Mathematical and Computer Modelling*, *47*(11-12), 1134–1149. https://doi.org/10.1016/j.mcm.2007.05.017

Benke, K. K., Norng, S., Robinson, N. J., Benke, L. R., & Peterson, T. J. (2018). Error propagation in computer models: Analytic approaches, advantages, disadvantages and constraints. *Stochastic Environmental Research and Risk Assessment*, *32*(10), 2971–2985. https://doi.org/10.1007/s00477-018-1555-8

Beven, K., & Binley, A. (1992). The future of distributed models: Model calibration and uncertainty prediction. *Hydrological Processes*, *6*(3), 279–298. https://doi.org/10.1002/hyp.3360060305

Bottacin-Busolin, A., & Marion, A. (2010). Combined role of advective pumping and mechanical dispersion on time scales of bed form-induced hyporheic exchange. *Water Resources Research*, *46*, W08518. https://doi.org/10.1029/2009WR008892

Boxall, J. B., & Guymer, I. (2003). Analysis and prediction of transverse mixing coefficients in natural channels. *Journal of Hydraulic Engineering*, *129*(2), 129–139. https://doi.org/10.1061/(ASCE)0733-9429(2003)129:2(129)

Brevis, W., Vargas, J., Debels, P., & Link, O. (2001). Efecto de las zonas de almacenamiento transitorio en la modelación de la demanda bioquímica de oxígeno: Aplicación al Rio Chillan, Chile. Paper presented at the XV Congreso Chileno de Ingeniería Hidráulica, Concepción,Chile.

Briggs, M. A., Gooseff, M. N., Arp, C. D., & Baker, M. A. (2009). A method for estimating surface transient storage parameters for streams with concurrent hyporheic storage. *Water Resources Research*, *45*, W00D27. https://doi.org/10.1029/2008WR006959

Chapman, D. (1996). In D. Chapman (Ed.), *Water quality assessment, a guide to the use of biota, sediments and water in environmental monitoring* (2nd ed.). London: E. & F.N. Spon.

Cokelaer, T. (2014). Fitter (Version 1.0.8). Retrieved from http://github.com/cokelaer/fitter

De Smedt, F., Brevis, W., & Debels, P. (2005). Analytical solution for solute transport resulting from instantaneous injection in streams with transient storage. *Journal of Hydrology*, *315*(1–4), 25–39. https://doi.org/10.1016/j.jhydrol.2005.04.002

Deltares (2018). D-water quality processes library description. Retrieved from Delft, Netherlands:

Deng, Z. Q., Singh, V. P., & Bengtsson, L. (2001). Longitudinal dispersion coefficient in straight rivers. *Journal of Hydraulic Engineering*, *127*(11), 919–927. https://doi.org/10.1061/(ASCE)0733-9429(2001)127:11(919)

Disley, T., Gharabaghi, B., Mahboubi, A. A., & McBean, E. A. (2015). Predictive equation for longitudinal dispersion coefficient. *Hydrological Processes*, *29*(2), 161–172. https://doi.org/10.1002/hyp.10139

Dotto, C. B. S., Mannina, G., Kleidorfer, M., Vezzaro, L., Henrichs, M., McCarthy, D. T., et al. (2012). Comparison of different uncertainty techniques in urban stormwater quantity and quality modelling. *Water Research*, *46*(8), 2545–2558. https://doi.org/10.1016/j.watres.2012.02.009

El Kadi Abderrezzak, K., Ata, R., & Zaoui, F. (2015). One-dimensional numerical modelling of solute transport in streams: The role of longitudinal dispersion coefficient. *Journal of Hydrology*, *527*, 978–989. https://doi.org/10.1016/j.jhydrol.2015.05.061

Elder, J. W. (1959). The dispersion of marked fluid in turbulent shear flow. *Journal of Fluid Mechanics*, *5*(04), 544–560. https://doi.org/10.1017/S0022112059000374

Etemad-Shahidi, A., & Taghipour, M. (2012). Predicting longitudinal dispersion coefficient in natural streams using M5′ model tree. *Journal of Hydraulic Engineering*, *138*(6), 542–554. https://doi.org/10.1061/(ASCE)HY.1943-7900.0000550

Fernald, A. G., Wigington, P. J., & Landers, D. H. (2001). Transient storage and hyporheic flow along the Willamette River, Oregon: Field measurements and model estimates. *Water Resources Research*, *37*(6), 1681–1694. https://doi.org/10.1029/2000WR900338

Fischer, H. B. (1975). Discussion of 'Simple method for predicting dispersion in streams' byMcQuifer, R.S. and Keefer, T.N. *Journal of the Environmental Engineering Division, American Society of Civil Engineers*, *101*(4), 453–455.

Fischer, H. B. (1979). *Mixing in inland and coastal waters*. London: Academic Press.

Freni, G., & Mannina, G. (2010). Bayesian approach for uncertainty quantification in water quality modelling: The influence of prior distribution. *Journal of Hydrology*, *392*(1–2), 31–39. https://doi.org/10.1016/j.jhydrol.2010.07.043

F. W. R. (2012). Urban pollution management manual 3rd edition. Retrieved from http://www.fwr.org/UPM3

González-Pinzōn, R., Haggerty, R., & Dentz, M. (2013). Scaling and predicting solute transport processes in streams. *Water Resources Research*, *49*, 4071–4088. https://doi.org/10.1002/wrcr.20280

Gupta, H. V., Sorooshian, S., & Yapo, P. O. (1999). Status of automatic calibration for hydrologic models: Comparison with multilevel expert calibration. *Journal of Hydrologic Engineering*, *4*(2), 135–143. https://doi.org/10.1061/(ASCE)1084-0699(1999)4:2(135)

Guymer, I. (1998). Longitudinal dispersion in sinuous channel with changes in shape. *Journal of Hydraulic Engineering*, *124*(1), 33–40. https://doi.org/10.1061/(ASCE)0733-9429(1998)124:1(33)

Helton, J. C. (1993). Uncertainty and sensitivity analysis techniques for use in performance assessment for radioactive waste disposal. *Reliability Engineering & System Safety*, *42*(2-3), 327–367. https://doi.org/10.1016/0951-8320(93)90097-I

Hubbard, E. F., Kilpatrick, F. A., Martens, L. A., & Wilson, J. F. Jr. (1982). Measurement of time of travel and dispersion in streams by dye tracing (03-A9). Retrieved from http://pubs.er.usgs.gov/publication/twri03A9_1982

Innovyze. (2017). 1D Diffusion for Water Quality Simulations – Infoworks ICM Manual, 2017.

ISO 1100–2 (2010). *Hydrometry measurement of liquid flow in open channels—Part 2: Determination of the stage-discharge relationship*. Geneva, Switzerland: International Organization for Standardization.

Kashefipour, S. M., & Falconer, R. A. (2002). Longitudinal dispersion coefficients in natural channels. *Water Research*, *36*(6), 1596–1608. https://doi.org/10.1016/S0043-1354(01)00351-7

Krause, P., Boyle, D. P., & Bäse, F. (2005). Comparison of different efficiency criteria for hydrological model assessment. *Advances in Geosciences*, *5*, 89–97. https://doi.org/10.5194/adgeo-5-89-2005

Launay, M., Le Coz, J., Camenen, B., Walter, C., Angot, H., Dramais, G., et al. (2015). Calibrating pollutant dispersion in 1-D hydraulic models of river networks. *Journal of Hydro-environment Research*, *9*(1), 120–132. https://doi.org/10.1016/j.jher.2014.07.005

Lindenschmidt, K.-E., Fleischbein, K., & Baborowski, M. (2007). Structural uncertainty in a river water quality modelling system. *Ecological Modelling*, *204*(3–4), 289–300. https://doi.org/10.1016/j.ecolmodel.2007.01.004

Liu, H. (1977). Predicting dispersion coefficient of streams. *ASCE Journal of the Environmental Engineering Division*, *103*(1), 59–69.

Lopes, J. F., Dias, J. M., Cardoso, A. C., & Silva, C. I. V. (2005). The water quality of the Ria de Aveiro lagoon, Portugal: From the observations to the implementation of a numerical model. *Marine Environmental Research*, *60*(5), 594–628. https://doi.org/10.1016/j.marenvres.2005.05.001

Magazine, M. K., Pathak, S. K., & Pande, P. K. (1988). Effect of bed and side roughness on dispersion in open channels. *Journal of Hydraulic Engineering*, *114*(7), 766–782. https://doi.org/10.1061/(ASCE)0733-9429(1988)114:7(766)

Mannina, G., & Viviani, G. (2010a). A hydrodynamic water quality model for propagation of pollutants in rivers. *Water Science and Technology*, *62*(2), 288–299. https://doi.org/10.2166/wst.2010.285

Mannina, G., & Viviani, G. (2010b). An urban drainage stormwater quality model: Model development and uncertainty quantification. *Journal of Hydrology*, *381*(3–4), 248–265. https://doi.org/10.1016/j.jhydrol.2009.11.047

Marsili-Libelli, S., & Giusti, E. (2008). Water quality modelling for small river basins. *Environmental Modelling & Software*, *23*(4), 451–463. https://doi.org/10.1016/j.envsoft.2007.06.008

Milne, I., Mallet, M. J., Clarke, S. J., Flower, S. G., Holmes, D., & Chambers, R. G. (1992). Intermittent pollution—Combined sewer overflows, ecotoxicology and water quality standards (NRA R&D Note 123). Retrieved from http://www.fwr.org/UPM3/

Moghaddam, M. B., Mazaheri, M., & Samani, J. M. (2017). A comprehensive one-dimensional numerical model for solute transport in rivers. *Hydrology and Earth System Sciences*, *21*(1), 99–116. https://doi.org/10.5194/hess-21-99-2017

Moriasi, D. N., Arnold, J. G., Liew, M. W. V., Bingner, R. L., Harmel, R. D., & Veith, T. L. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE American Society of Agricultural and Biological Engineers*, *50*(3), 885–900.

Noori, R., Deng, Z., Kiaghadi, A., & Kachoosangi, F. T. (2016). How reliable are ANN, ANFIS, and SVM techniques for predicting longitudinal dispersion coefficient in natural rivers? *Journal of Hydraulic Engineering*, *142*(1). https://doi.org/10.1061/(ASCE)HY.1943-7900.0001062

Nordin, C. F., & Troutman, B. M. (1980). Longitudinal dispersion in rivers: The persistence of skewness in observed data. *Water Resources Research*, *16*(1), 123–128. https://doi.org/10.1029/WR016i001p00123

Norris, T., Squibbs, G., Saul, A., & Shucksmith, J. (2014). Evaluating the use of sewer network event mean concentrations within integrated catchment models. Paper Presented at the 13th International Conference on Urban Drainage, Sarawak, Malaysia,.

Noss, C., & Lorke, A. (2016). Roughness, resistance, and dispersion: Relationships in small streams. *Water Resources Research*, *52*, 2802–2821. https://doi.org/10.1002/2015WR017449

Oliphant, T. E. (2007). Python for scientific computing. *Computing in Science & Engineering*, *9*(3), 10–20. https://doi.org/10.1109/mcse.2007.58

Piotrowski, A. P., Rowinski, P. M., & Napiorkowski, J. J. (2010). Uncertainty study of data-based models of pollutant transport in rivers. Paper Presented at the River Flow, Braunschweig, Germany.

Radwan, M., El-Sadek, A., Willems, P., Feyen, J., & Berlamont, J. (2001). Modeling of nitrogen in river water using a detailed and a simplified model. *The ScientificWorld Journal* [electronic resource], *1*(2), 200–206.

Refsgaard, J. C., van der Sluijs, J. P., Brown, J., & van der Keur, P. (2006). A framework for dealing with uncertainty due to model structure error. *Advances in Water Resources*, *29*(11), 1586–1597. https://doi.org/10.1016/j.advwatres.2005.11.013

Refsgaard, J. C., van der Sluijs, J. P., Højberg, A. L., & Vanrolleghem, P. A. (2007). Uncertainty in the environmental modelling process—A framework and guidance. *Environmental Modelling & Software*, *22*(11), 1543–1556. https://doi.org/10.1016/j.envsoft.2007.02.004

Reichert, P., & Vanrolleghem, P. (2001). Identifiability and uncertainty analysis of the River Water Quality Model No. 1 (RWQM1). *Water Science and Technology*, *43*(7), 329–338.

Romanowicz, R. J., Osuch, M., & Wallis, S. (2013). Modelling of solute transport in rivers under different flow rates: A case study without transient storage. *Acta Geophysica*, *61*(1), 98–125. https://doi.org/10.2478/s11600-012-0050-8

Runkel, R. L. (1998). *One-dimensional transport with inflow and storage (OTIS): A solute transport model for streams and rivers*. Denver, Colo.: Denver, CO: U.S. Dept. of the Interior, U.S. Geological Survey. Information Services [distributor]

Rutherford, J. C. (1994). *River mixing*. West Sussex, UK: Wiley.

Sattar, A. M. A., & Gharabaghi, B. (2015). Gene expression models for prediction of longitudinal dispersion coefficient in streams. *Journal of Hydrology*, *524*, 587–596. https://doi.org/10.1016/j.jhydrol.2015.03.016

Schellart, A. N. A., Tait, S. J., & Ashley, R. M. (2010a). Estimation of uncertainty in long-term sewer sediment predictions using a response database. *Journal of Hydraulic Engineering*, *136*(7), 403–411. https://doi.org/10.1061/(ASCE)HY.1943-7900.0000193

Schellart, A. N. A., Tait, S. J., & Ashley, R. M. (2010b). Towards quantification of uncertainty in predicting water quality failures in integrated catchment model studies. *Water Research*, *44*(13), 3893–3904. https://doi.org/10.1016/j.watres.2010.05.001

Schmalle, G. F., & Rehmann, C. R. (2014). Analytical solution of a model of contaminant transport in the advective zone of a river. *Journal of Hydraulic Engineering*, *140*(7). https://doi.org/10.1061/(ASCE)HY.1943-7900.0000885

Segura, B. (2004). *Transient storage process relation with hydraulics and morphological characteristics of rivers*. Chile, Concepción, Chile: University of Concepción.

Seo, I. W., & Cheong, T. S. (1998). Predicting longitudinal dispersion coefficient in natural streams. *Journal of Hydraulic Engineering*, *124*(1), 25–32. https://doi.org/10.1061/(ASCE)0733-9429(1998)124:1(25)

Shucksmith, J., Boxall, J., & Guymer, I. (2007). Importance of advective zone in longitudinal mixing experiments. *Acta Geophysica*, *55*(1), 95–103. https://doi.org/10.2478/s11600-006-0042-7

Shucksmith, J. D., Boxall, J. B., & Guymer, I. (2010). Effects of emergent and submerged natural vegetation on longitudinal mixing in open channel flow. *Water Resources Research*, *46*, W04504. https://doi.org/10.1029/2008WR007657

Sriwastava, A., Tait, S., Schellart, A., Kroll, S., Dorpe Mieke, V., Assel Johan, V., & Shucksmith, J. (2018). Quantifying uncertainty in simulation of sewer overflow volume. *Journal of Environmental Engineering*, *144*(7), 04018050. https://doi.org/10.1061/(ASCE)EE.1943-7870.0001392

Tscheikner-Gratl, F., Bellos, V., Schellart, A., Moreno-Rodenas, A., Muthusamy, M., Langeveld, J., & Tait, S. (2018). Recent insights on uncertainties present in integrated catchment water quality modelling. *Water Research*, *150*, 368–379.

Van Der Perk, M. (1997). Effect of model structure on the accuracy and uncertainty of results from water quality models. *Hydrological Processes*, *11*(3), 227–239. https://doi.org/10.1002/(SICI)1099-1085(19970315)11:3<227::AID-HYP440>3.0.CO;2-#

van Griensven, A., & Meixner, T. (2006). Methods to quantify and identify the sources of uncertainty for river basin water quality models. *Water Science and Technology*, *53*, 51–59.

van Mazijk, A., & Veling, E. J. M. (2005). Tracer experiments in the Rhine Basin: Evaluation of the skewness of observed concentration distributions. *Journal of Hydrology*, *307*(1–4), 60–78. https://doi.org/10.1016/j.jhydrol.2004.09.022

Vandenberghe, V., Bauwens, W., & Vanrolleghem, P. A. (2007). Evaluation of uncertainty propagation into river water quality predictions to guide future monitoring campaigns. *Environmental Modelling & Software*, *22*(5), 725–732. https://doi.org/10.1016/j.envsoft.2005.12.019

Wang, Y., & Huai, W. (2016). Estimating the longitudinal dispersion coefficient in straight natural rivers. *Journal of Hydraulic Engineering*, *142*(11), 04016048. https://doi.org/10.1061/(ASCE)HY.1943-7900.0001196

Wang, Y., Huai, W., & Wang, W. (2017). Physically sound formula for longitudinal dispersion coefficients of natural rivers. *Journal of Hydrology*, *544*(Supplement C), 511–523. https://doi.org/10.1016/j.jhydrol.2016.11.058

White, W. R., Milli, H., & Crabbe, A. D. (1973). Sediment transport: An appraisal of available methods. Technical Report (INT 119). Retrieved from

Willems, P. (2008). Quantification and relative comparison of different types of uncertainties in sewer water quality modeling. *Water Research*, *42*(13), 3539–3551. https://doi.org/10.1016/j.watres.2008.05.006

Zaramella, M., Packman, A. I., & Marion, A. (2003). Applicastion of the transient storage model to analyze advective hyporheic exchange with deep and shallow sediment beds. *Water Resources Research*, *39*(7), 1198. https://doi.org/10.1029/2002WR001344

Zeng, Y., & Huai, W. (2014). Estimation of longitudinal dispersion coefficient in rivers. *Journal of Hydro-environment Research*, *8*(1), 2–8. https://doi.org/10.1016/j.jher.2013.02.005