

Article

A Computational Tool to Track Sewage Flow Discharge into Rivers Based on Coupled HEC-RAS and DREAM

Junbo Wen ^{1,2}, Mengdie Ju ^{1,2}, Zichen Jia ^{1,2}, Lei Su ³, Shanshan Wu ³, Yuting Su ³, Wenxiao Liufu ³
and Hailong Yin ^{1,2,*}

¹ College of Environmental Science and Engineering, Tongji University, Shanghai 200092, China; 2230583@tongji.edu.cn (J.W.); hjmengdie@163.com (M.J.); 2331301@tongji.edu.cn (Z.J.)

² State Key Laboratory of Pollution Control and Resource Reuse Research, Tongji University, Shanghai 200092, China

³ Guangzhou Resource Environmental Protection Technology Co., Ltd., Guangzhou 510075, China; sul@rcnzy.com (L.S.); wuss@rcnzy.com (S.W.); suyt@rcnzy.com (Y.S.); soso8537@126.com (W.L.)

* Correspondence: yinhailong@tongji.edu.cn

Abstract: Worldwide abatement of untreated sewage discharge into surface water is a challenging task. Sewage discharging into surface waters has a detrimental impact on water quality. This paper presents a MATLAB (R2018b) framework designed to identify sewage flow discharges into rivers from an inverse problem-solving perspective. The computational tool integrates a hydrodynamic model using the Hydrologic Engineering Center's River Analysis System (HEC-RAS 5.0.0) and an open-source toolbox for Differential Evolution Adaptive Metropolis (DREAM) as the inverse problem method. The proposed framework can effectively infer discharge sources in scenarios of highly transient flow based on hydraulic data at pre-set monitoring sites. To validate its capabilities, one hypothetical case and two real cases of sewage flow discharges entering a river were used to test the developed modeling framework. The results based on three performance metrics showed that this mathematical tool can be extended to simulate complex hydrodynamic flow patterns. This accomplishment underscores its potential as a valuable asset for environmental monitoring and water quality restoration efforts.



Citation: Wen, J.; Ju, M.; Jia, Z.; Su, L.; Wu, S.; Su, Y.; Liufu, W.; Yin, H. A Computational Tool to Track Sewage Flow Discharge into Rivers Based on Coupled HEC-RAS and DREAM. *Water* **2024**, *16*, 51. <https://doi.org/10.3390/w16010051>

Academic Editor: Rui Manuel Vitor Cortes

Received: 19 November 2023

Revised: 14 December 2023

Accepted: 18 December 2023

Published: 22 December 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: mathematic modeling; model integration; HEC-RAS; inverse problem; source flow tracking

1. Introduction

Urban rivers are an essential factor in the sustainable development of urban society and economies. Nevertheless, the persistent issue of untreated and illegal discharges poses a significant threat to the surface water quality of these vital urban waterways [1,2]. Therefore, there is a pressing need to develop an efficient tool for tracking sewage flow discharges to tackle the challenges posed by river pollution and to achieve pollution interception and water quality restoration.

Mathematical modeling serves as a critical tool in the identification of pollutant sources. Employing an inverse problem-solving approach, it uses measurements of contaminant concentration data at monitoring sites to infer the features of emissions, such as the volume, location, and time of the emissions. This tracking methodology has been widely used in groundwater and atmospheric studies [3–7]. In recent years, the technique has also found increasing application in the field of sewage source identification in surface water systems [8–13]; water quality sensors have already been employed by many countries. Basically, the combination of optimization methods and model governing equations in the field of water quality can provide deterministic solutions for unknown source parameters. For instance, Zhang and Xin [14] utilized the basic Genetic Algorithm (GA) to pinpoint the spill location and contaminant mass of sources in a small river channel. Jiang et al. [15] estimated the characteristics of the pollutant source and determined where, when, and

how much of the pollutant was released into a surface stream by developing an integration of Genetic Algorithm and Simulated Annealing Algorithm (IGSAA) as a pollutant source inverse method. However, the reliability of these optimization methods may be compromised due to high uncertainties in their deterministic processes and the data employed [16]. Alternatively, stochastic methods based on Bayesian inference have been suggested to overcome the disadvantages of deterministic optimization. These methods quantify the uncertainty in inverse modeling by estimating probability distributions of source parameters. Zhu et al. [17] applied genetic algorithms and Bayesian methods separately to identify instantaneous point sources and compared the performance of both methods. Jiang et al. [18] implemented a Bayesian framework for tracing the source of a significant nitrobenzene spill in the Songhua River, exploring the uncertainty factors affecting the inversion results. To enhance the performance of stochastic models, differential evolution has also been integrated into the Bayesian framework [19,20]. The generated Differential Evolution and Metropolis (DREAM) algorithm as well as the Differential Evolution and Metropolis–Hastings Markov Chain Monte Carlo (DEMH-MCMC) algorithm are capable of efficiently and accurately identifying sources of multi-point sudden water pollution incidents [11,12]. As a result, stochastic-based methods, such as the aforementioned DREAM algorithm, have been demonstrated to be more suitable for contaminant source identification due to their superior handling of uncertainty in identification results and unknown source parameters compared to deterministic-based methods like GA.

However, the previous studies have a disadvantage [11,12,14–18] because source tracking based on the water quality model usually assumes that the water flow velocity is relatively constant, which may not accurately reflect the real hydrodynamic conditions in rivers. Conversely, in cases where side flow discharge cannot be ignored compared to upstream inflow, the incorporation of a hydrodynamic model can inversely track side discharges by comparing simulated water level and flow discharge with the measured data. This enables accurate identification of flow discharge from sewage sources under varying flow velocities, as the pollutant discharge also brings side flow discharge that triggers water flow and water stage change at monitoring stations. Additionally, online hydraulic data monitoring proves to be more cost-effective than online water quality monitoring.

Therefore, there is a notable interest in the integration of inverse problem-solving methods with hydrodynamic models using low-cost hydraulic monitoring data. Recently, the integration of the hydrodynamic diffusion wave equation and the Bayesian–Markov Chain Monte Carlo (MCMC) algorithm has enabled a hydrodynamic-based river source tracing method. This approach utilizes hydraulic data to inversely estimate the location, flow rate, and timing of sewage discharges [21]. Nevertheless, the limitations of the diffusion wave equation, which only applies to single-direction flow, restrict its ability to simulate complex hydrodynamic flow patterns, such as backwater flow, tidal water flow, and highly transient flows during wet weather events or sudden sewage discharges. In contrast, HEC-RAS can be applied in the aforementioned situations, expanding its range of applications to better align with real-world scenarios [22–25]. Introducing complex hydrodynamic models like HEC-RAS that incorporate the software platform can provide more powerful tools for solving source-tracking problems.

Currently, comprehensive model systems that integrate HEC-RAS and inverse problem-solving techniques within an optimization framework are still scarce. Establishing such an integrated framework requires a programming platform that can effectively combine multiple software and open-source codes. A notable example of such a state-of-the-art programming platform is MATLAB. Recognized for its capabilities in technical computing, MATLAB integrates computation, visualization, and programming in a user-friendly environment [26]. This robust platform holds great promise in advancing the development of comprehensive solutions to address water quality challenges through the amalgamation of hydrodynamic modeling and inverse problem methodologies.

This paper presents a MATLAB framework for inverse tracking sewage flow discharges in river systems. The remainder of this paper is organized as follows: (1) Section 2

illustrates the core modules and the framework rationale; (2) Section 3 presents one hypothetical case and two real case studies under transient flow conditions to demonstrate the practical application of the modeling system. Finally, the main results are summarized in Section 4.

2. Framework Design

The MATLAB modeling system is comprised of a hydrodynamic model and an inverse problem method, which together form the basis for constructing a framework to address the inverse tracking of sewage flow discharges. The framework is built on two core modules: HEC-RAS, serving as the hydrodynamic model, and DREAM, functioning as the stochastic inverse problem algorithm. The integration of these two modules enables inverse tracking, allowing for the estimation of source discharges and discharge patterns in the river system.

2.1. HEC-RAS Modeling

A highly adopted hydrodynamic model for river flow simulation is HEC-RAS (v5.0), which was developed by the Hydrological Engineering Center/U.S. Army Corps of Engineers [27]. HEC-RAS software is a computer application crafted to simulate the flow of rivers through natural, open channels and to estimate the water surface profile [28,29].

Governing Equations

In the context of urban rivers, hydrodynamics can often be simplified to a one-dimensional (1D) model because the primary concern is the flow gradient along the longitudinal scale of the river. In this study, the HEC-RAS software performs hydrodynamic simulation with the complete 1D Saint-Venant equations, which are suitable for simulating complex hydrodynamic flow patterns during wet weather events or sudden sewage discharges. The relevant governing equations are as follows:

$$\frac{\partial A}{\partial t} + \frac{\partial Q}{\partial x} = q_L \quad (1)$$

$$\frac{\partial Q}{\partial t} + \frac{\partial}{\partial x} \left(\frac{Q^2}{A} \right) + gA \frac{\partial h}{\partial x} - gA(S_0 - S_f) = 0 \quad (2)$$

where A represents the cross-sectional area (m^2); t is time (s); Q denotes flow rate (m^3/s); x is the longitudinal distance in the channel (m); q_L is the lateral inflow per unit length of the river channel (m^2/s); h is the water depth of the channel (m); g is the acceleration due to gravity (m^2/s); S_0 is the channel slope; S_f is the friction slope.

To construct a hydrodynamic model and solve the 1D Saint-Venant equations, input data related to the parameters mentioned above are required. This includes topographic data of cross-sectional shapes, channel slopes, and hydraulic data such as flow rate, water level, and Manning coefficients. In this study, the flow regime was assumed to be unsteady, nonuniform flow within the river reach, considering dynamic upstream flow in addition to lateral inflow from sewage outlets. The established model generates simulated downstream hydraulic values (water level and flow rate) based on different combinations of lateral inflow caused by sewage sources. These values are then compared with observed data in the inverse problem algorithm to trace the source parameters of side flow discharge.

2.2. DREAM Algorithm

The stochastic inverse problem method employed in this study utilizes the DREAM Toolbox, integrated within the MATLAB environment [20]. DREAM solves complex optimization problems of water resources based on the Bayesian–Markov Chain Monte Carlo (MCMC) method [11,30–32]. In each iteration, the DREAM simultaneously runs multiple Markov chains to explore the global space and automatically adjusts the scale and orientation of the proposal distribution through differential evolution and Metropolis selection. The process aims to sample from the target posterior distribution, achieved as the Markov

chains, guided by an adaptive proposal distribution, converge towards the true posterior distribution.

2.2.1. Bayesian Theorem

According to Bayes' theorem, the posterior distribution of uncertain source parameters can be estimated from the time series of observed data [33]:

$$p(\theta|O_t) \propto p(\theta)p(O_t|\theta) \tag{3}$$

where $p(\theta|O_t)$ is the posterior distribution of the model parameters θ given the observed data O_t . $p(\theta)$ is the prior distribution of the parameters, while $p(O_t|\theta)$ represents the likelihood function of the observed data with given θ .

The likelihood function $p(O_t|\theta)$ characterizes the error between simulated data and the observation data. The log-form of the likelihood function is given as [34]:

$$p(O_t|\theta) = -\frac{n}{2}\text{Log}(2\pi) - \sum_{t=1}^n [\text{Log}(\sigma)] - \frac{1}{2} \left(\frac{M_t(\theta) - O_t}{\sigma} \right)^2 \tag{4}$$

where σ indicates model errors variance, n is the number of observed data, t shows the index representing time series, M_t and O_t are the simulated and observed time-series data, respectively.

2.2.2. Differential-Evolution and Metropolis Estimator

Differential evolution is applied as a genetic algorithm for the evolution of new candidates in the DREAM algorithm in Equations (5) and (6). Additionally, a Metropolis selection rule [35] decides whether candidates can replace their corresponding old values in Equation (7):

$$\theta_p^{i,j} = \theta_{t-1}^{i,j} + \zeta + (1 + \lambda) \gamma_{(\delta,d^*)} \sum_{k=1}^{\delta} (\theta_{t-1}^{i,a_k} - \theta_{t-1}^{i,b_k}) \tag{5}$$

where $\theta_p^{i,j}$ is the parameter vector at the current iteration t for the i th parameter in the j th chain, $\theta_{t-1}^{i,j}$ signifies the state variable of the j th chain at the current iteration of $t - 1$, $\gamma_{(\delta,d^*)} = \frac{2.38}{\sqrt{2\delta d^*}}$ is the jump rate, $\delta = 3$ is the default number of pairs used to generate the new candidate with $d^* = d$, a and b are vectors from $\{1, \dots, N\}$, the value of λ and ζ are sampled from uniform distribution, $U(-0.1, 0.1)$, and normal distribution, $N(0, 10^{-6})$, respectively.

$$\theta_p^{i,j} = \begin{cases} \theta_{t-1}^{i,j} & \text{if } U \leq 1 - CR, d^* = d^* - 1 \\ \theta_p^{i,j} & \text{otherwise} \end{cases} \tag{6}$$

where $U \sim (0, 1)$ follows a uniform distribution the value of crossover rate (CR) is sampled from a geometric sequence of different crossover probabilities, $CR = \{1/nCR, 2/nCR, \dots, 1\}$ with default setting of $nCR = 3$.

A random value U^* is sampled from a uniform distribution $U \sim (0, 1)$. The Metropolis acceptance probability is calculated as follows:

$$p_{acc}(\theta_{t-1}^{i,j} \rightarrow \theta_p^{i,j}) = \begin{cases} \min \left[1, \frac{p(\theta_p^{i,j})}{p(\theta_{t-1}^{i,j})} \right] & \text{if } p(\theta_{t-1}^{i,j}) > 0 \\ 1 & \text{if } p(\theta_{t-1}^{i,j}) = 0 \end{cases} \tag{7}$$

If $U^* < p_{acc}(\theta_{t-1}^{i,j} \rightarrow \theta_p^{i,j})$, the candidates are accepted, the chain is updated, $\theta_t^{i,j} = \theta_p^{i,j}$; otherwise, the old values are retained, $\theta_t^{i,j} = \theta_{t-1}^{i,j}$.

2.2.3. DREAM Algorithm

The convergence of the DREAM algorithm is assessed by the R-statistic in Equations (8)–(10) [36]. It takes into account the convergence of different components and calculates them with weighted values. This ensures that not only the entire chain converges but also different components converge when the value of the R-statistic is less than 1.2 for each parameter. Thereby enhancing the performance and quality of the DREAM algorithm, as proposed by [37]:

$$R_{Statistic}^i = \sqrt{\frac{N+1}{N} \frac{(T-2)W^i + 2B^i}{W^i T} - \frac{T-2}{NT}} \quad (8)$$

$$W^i = \frac{2}{N(T-2)} \sum_{j=1}^N \sum_{t=\lfloor \frac{T}{2} \rfloor}^T (\theta_t^{i,j} - \bar{\theta}^{i,j})^2 \quad (9)$$

$$\frac{B^i}{T} = \frac{1}{2(N-1)} \sum_{j=1}^N \left(\bar{\theta}^{i,j} - \frac{1}{N} \sum_{j=1}^N \bar{\theta}^{i,j} \right)^2 \quad (10)$$

where W^i is used to estimate the average level of parameter variation within each of the N chains; $\frac{B^i}{T}$ is used to estimate the variation between the means of the N chains, where N is the number of chains and T is the iteration number, and the remaining symbols are as previously defined.

It is important to analyze the convergence curves of the R-statistic. In many practical cases, the maximum number of iterations may be reached without convergence, evident when the R-statistic remains above 1.2. In such cases, if a decreasing trend in the R-statistic is observed, an increase in the maximum number of iterations should be considered. Alternatively, if the convergence curve lacks a consistent decreasing pattern, it may be necessary to add additional monitoring sites. This can be particularly effective in the case of high parameter dimensions, as it aids convergence by introducing additional constraints.

The marginal Probability Density Function (PDF) and Maximum A Posteriori (MAP) estimation are used to describe the statistical characteristics of the estimated source parameters, all of which are generated by the DREAM toolbox. The mathematical representation of the MAP can be expressed as follows:

$$MAP = \operatorname{argmax}_{\theta} p(\theta | O_t) \quad (11)$$

2.3. Coupling HEC-RAS with DREAM in MATLAB

The architecture of the coupled HEC-RAS and DREAM system for identifying sewage inputs and their discharge flows is depicted in Figure 1. In the initial preparation phase, a hydrodynamic model is established using HEC-RAS based on field surveys and data provided by local water authorities. Time series of observations, including upstream and downstream water levels and flow rates, are collected from an existing monitoring system. Corresponding DREAM parameters are set based on the actual problem, and prior distributions for inverse modeling of source parameters are defined.

The inverse modeling process is accomplished within the MATLAB system. Initially, the DREAM algorithm generates initial combination values for source parameters (discharge flow and discharge pattern) from prior information. These values are then utilized to update the unknown source parameters of lateral flows in the HEC-RAS unsteady flow file. Subsequently, the HEC-RAS model simulates changes in downstream water levels and flow rates at the monitoring site in response to sewage source discharges. A comparison between the simulated and observed time series values is then conducted to calculate the likelihood function necessary for the posterior distribution required by DREAM. The DREAM algorithm then generates samples from the Bayesian posterior distribution in Equation (3) to analyze unknown source parameters. Employing differential evolution and Metropolis selection, it generates new candidate source parameters for further iteration cycles. This iterative process continues until the optimization termination criterion

is satisfied. Finally, the results of the inversion process within the designed framework are expressed using marginal PDFs and MAP estimates. In summary, the described steps illustrate the systematic approach used by the coupled HEC-RAS and DREAM system to identify sewage flow inputs and their discharge patterns.

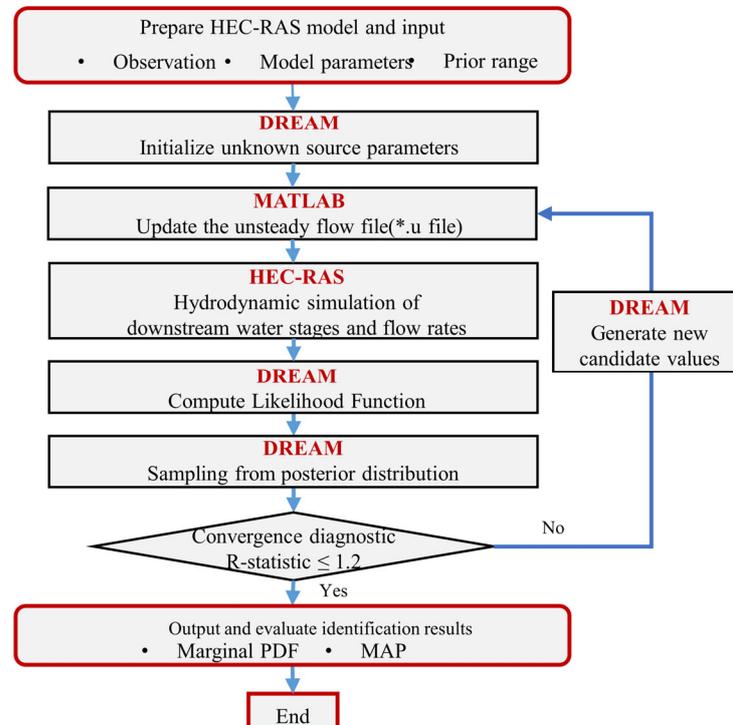


Figure 1. Schematic diagram of the inverse modeling framework developed to identify the emission source parameters.

To facilitate a clear understanding of the workflow, the scripts for running the aforementioned models and the links between these models are briefly described in Figure S1 and Listing S1–S8.

2.4. Modeling Performance Metrics

To validate the computational tool developed in this study, three error metrics were employed to assess and analyze the results of the forward computations, including the Means Relative Error (MRE), Coefficient of Determinism (R^2), the Root Mean Squared Error (RMSE), and the Nash–Sutcliffe Efficiency (NSE). These metrics allow for the reconstruction of discharge events in real cases, enabling a comparison between the hydraulic data simulated by the model and actual measurements from observed sections. If the simulated and measured values align closely, it lends credibility to the inverse problem results. The calculation formulas for these metrics are presented in Equations (12)–(14):

$$R^2 = \left(\frac{\sum_{t=1}^T (O_t - \bar{O})(M_t - \bar{M})}{\sqrt{\sum_{t=1}^T (O_t - \bar{O})^2} \sqrt{\sum_{t=1}^T (M_t - \bar{M})^2}} \right)^2 \quad (12)$$

$$RMSE = \sqrt{\frac{1}{T} \sum_{t=1}^T (O_t - M_t)^2} \quad (13)$$

$$NSE = 1 - \frac{\sum_{t=1}^T (O_t - M_t)^2}{\sum_{t=1}^T (O_t - \bar{O})^2} \quad (14)$$

where M_t represents the simulated values, O_t represents the measured values, and \bar{M} and \bar{O} represent the mean values of simulated and measured data, respectively.

3. Modeling Tool Demonstration

3.1. Hypothetical Case

The hypothetical case was a simulated scenario, which was intended to test the performance of the reliability of the proposed framework. This case simulated the sudden release of industrial wastewater in a non-steady state within an open channel. This study site is a 4000-m-long trapezoidal channel with other key parameters of water depth at 0.72 m, bottom width of 10.0 m, side slope of 0.4, channel gradient of 5×10^{-5} , and an upstream inflow of $2.0 \text{ m}^3/\text{s}$. In positive problem, this case was pulse discharges of sewage flow at $x_1 = 1200 \text{ m}$ and $x_2 = 2600 \text{ m}$. The discharge flow (q) was set at $0.3 \text{ m}^3/\text{s}$, occurring from 1:00 (T_1) to 2:59 (T_2) and at $0.2 \text{ m}^3/\text{s}$, occurring from 6:00 (T_1) to 8:59 (T_2). Key parameters for the sewage discharge included average flow, initiation time, and duration. The pulse discharge model is represented by the following equation:

$$q(t) = \begin{cases} q, & T_1 < t < T_2 \\ 0, & t < T_1 \text{ or } t > T_2 \end{cases} \quad (15)$$

Therefore, three source parameters for each sewage flow source needed to be determined in the DREAM-based framework, specifically denoted as $\{Q_i, t_i, T_i\}$. Q_i represented the average emission rate of the source, t_i represented the starting time of emission, and T_i represented the duration of the emission. Through HEC-RAS modeling with a spatial grid size of 200 m, time-variable water levels and flow rates at the downstream boundary ($x = 4000 \text{ m}$) were simulated. In the context of the inverse problem, using the previously obtained values as continuous observed values of water levels at the downstream boundary, the source discharge (flow rate and temporal pattern) can be inferred using the proposed framework.

Regarding the prior distributions for each source parameter, an even distribution was assumed within the range of $[0, 1]$ for the flow rate, $[0, 1440]$ for the starting time of emission and the duration (recorded in minutes from 0:00 to 24:00). Hence, the dimension of source parameters in this case was set at 6 ($d = 6$), for the three parameters associated with the two sewage sources. In the DREAM computation, 7 Markov chains ($N = 7$) were employed, and the pre-set number of iterations was 5000 ($T = 5000$). The model error variance of the likelihood function was set at 0.001 ($\sigma = 0.001$), while for the GA, the number of generations, the population size, the initial crossover rate, and the mutation rate for this case were set to 2000, 200, 0.8, and 0.01, respectively. For each method, simulations were conducted to reach a fit between the simulated hydrodynamic data M_t and the observed data O_t at the downstream monitoring site.

After the optimization termination criterion had been satisfied or a predefined maximum number of generations had been reached, the plot of a typical converging process of the DREAM-based optimization was automatically generated. Figure 2 illustrates the R-statistic, which is used to measure the convergence of the sampled chains for each of the unknown source parameters. The maximum R-statistic briefly reached nearly 7. However, with the correction of the outlier chains using the DREAM algorithm, this value rapidly decreased. Eventually, after approximately 2000 iterations, the R-statistic values for each parameter were found to be less than 1.2, indicating that the Markov chains had converged to the limit distribution with a high degree of accuracy.

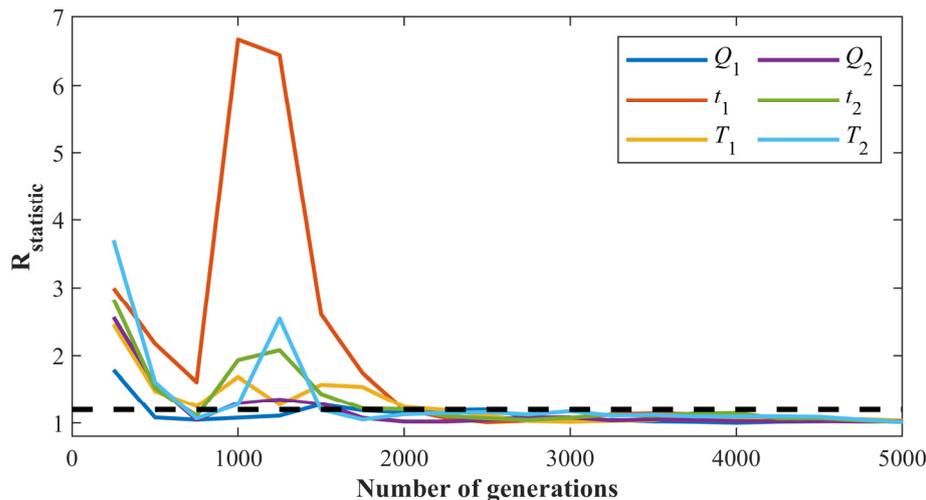


Figure 2. Convergence of R-statistic of the source parameters for hypothetical case.

Figure 3 provides the performance enhancement by the DREAM compared to GA. Specifically, GA methods exhibited higher uncertainty and lower accuracy. Even after averaging the results from 10 calculations, the mean relative errors in GA simulations stood at 2.1%, with the highest relative error for one simulation reaching 11.3% for T_1 . Conversely, the MAP values of DREAM closely matched the true values, displaying a mean relative error of merely 0.3% across the six parameters. This remarkable difference in performance signified an 85.7% improvement in the tracking accuracy of source parameters by DREAM compared to GA, thereby validating the effectiveness of the inverse algorithm. Misleading outcomes arose as GA got trapped in local optima, failing to explore the full parameter space, unlike the comprehensive exploration ensured by DREAM.

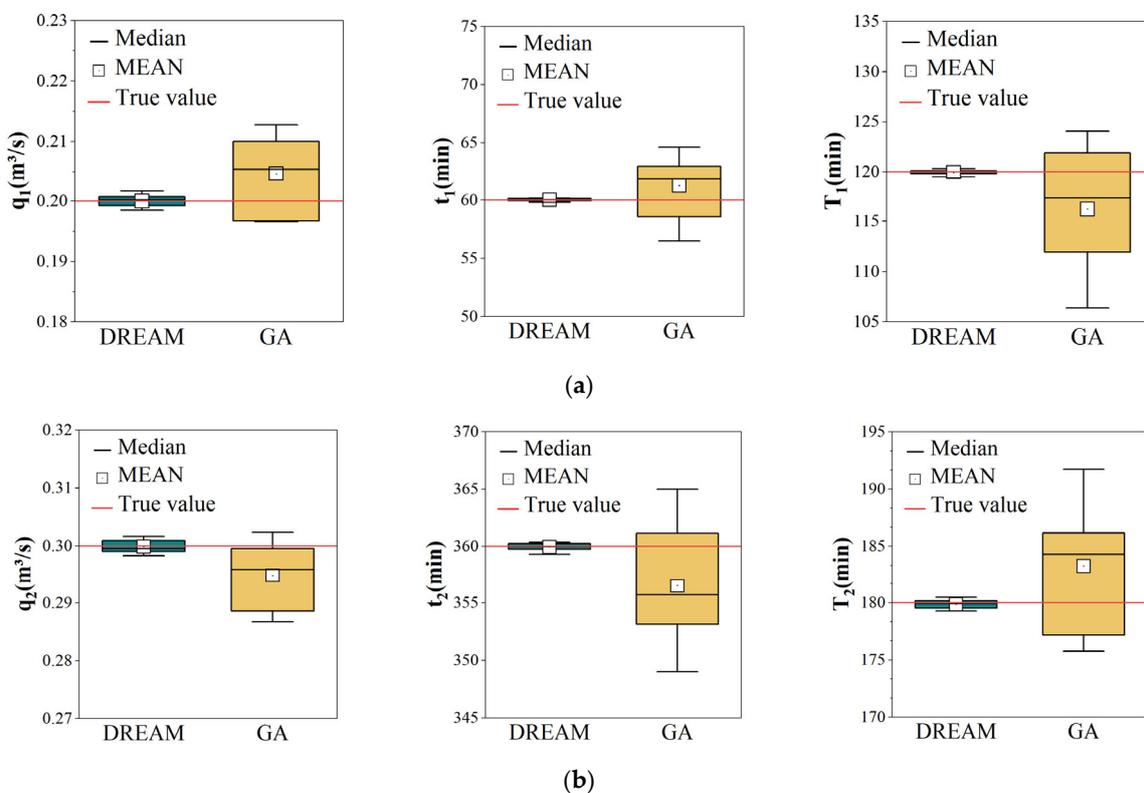


Figure 3. Comparison of the estimated source parameters obtained using the GA and DREAM: (a) Flow source at $x = 1200$ m (b) Flow source at $x = 2600$ m.

As for the grid independence solution, it was crucial to determine the spatial grid size required for a stable numerical solution that minimizes the computational load while maintaining errors in the simulated outcome at a nearly constant level. To achieve this, inverse simulations were conducted with spatial grid sizes of 100 m and 200 m. Table 1 shows the downstream water levels achieved an R^2 of 0.985, an RMSE of 0.001, and an NSE of 0.985 with a 200-m grid, indicating a satisfactory tracking accuracy. If the spatial grid size was further refined to 100 m, the R^2 , RMSE, and NSE remained almost unchanged compared to the 200-m grid solution, but the computational time increased by about 17%. Therefore, for source tracing scenarios at known discharge points, further grid refinement on the basis of 200 m did not notably enhance computational accuracy but significantly elevated computational load. Balancing the computational load and the tracking accuracy, a 200-m grid was the acceptable spatial dividing for this numerical solution.

Table 1. Modeling performances for the simulated water levels for hypothetical case.

Grid Size (m)	Computation Time (s)	R^2	RMSE	NSE
100 m	18,020	0.986	0.001	0.985
200 m	15,350	0.985	0.001	0.985

3.2. Real Case

Two real cases, including one case of sewage source tracking of hourly variable sudden discharge and the other case for multiple sewage sources tracking of discharge flows and periods during both dry and wet weather, were employed to demonstrate and verify the robustness of the above framework.

3.2.1. Real Case 1: Source Tracking of a Time-Variable Industrial Discharge into the River

The developed computation tool was demonstrated in a 1.6 km river reach within the Ci Hu River Basin of Ma'anshan City, situated in Anhui Province. Ma'anshan is an industrial city located along the Yangtze River Economic Belt. The Cihu River, which is the longest river entering the city of Ma'anshan, flows through the Cihu Economic Development Zone in its downstream region, with industrial land dominating on both sides. A schematic representation of the generalized model for a demonstration river reach is depicted in Figure 4, with two monitoring devices located upstream and downstream, as well as one time-varying sudden discharge source within the river reach.

The establishment and calibration of the HEC-RAS model relied on topographic data, including river channel elevations, channel slopes, cross-sectional areas, and time-series hydraulic data like upstream and downstream water levels and flow rates. The topographic data were obtained directly from local water authorities to build the geometric model. The accuracy of these cross-sectional profiles was verified by conducting measurements across the river, using a river vessel equipped with an Acoustic Doppler Current Profiler, from the left bank to the right bank. Based on the survey results, the geometry and topography for this river included bottom width from 4.23 to 10.07 m, bottom slope from -0.0007 to 0.0025 , bottom elevation from 2.298 to 2.503 m, and the cross-sectional areas from 4.43 to 9.55 square meters. The hydraulic data, involving time series of upstream and downstream water levels and flow rates, were measured at 1-h intervals using ultrasonic time-difference flow meters to set unsteady flow simulations. The upstream boundary and initial conditions were configured as flow hydrographs, with their values matching the measured data. Downstream boundary conditions were initially set as normal water depth to generate simulated downstream values. These simulated values were subsequently refined during the optimization process to align with observed data.

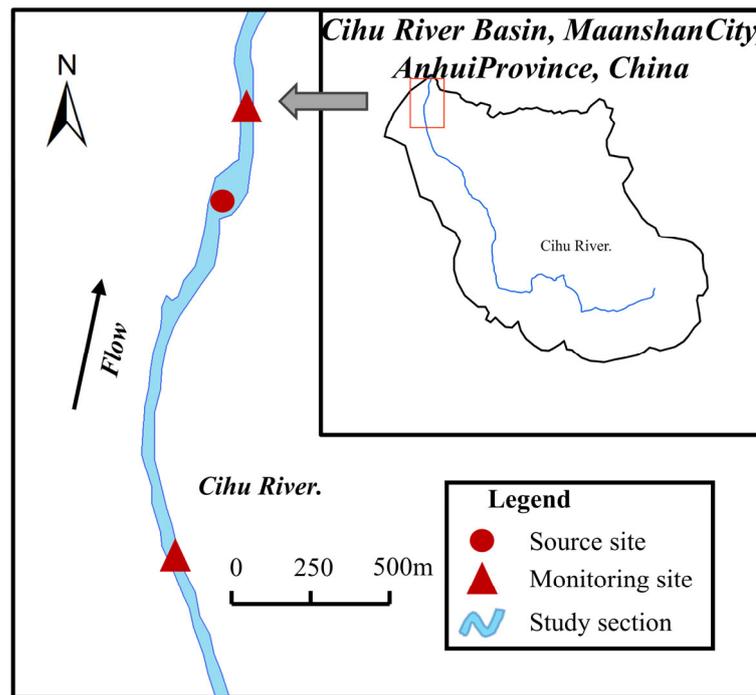


Figure 4. Description of study site and modeling test scenario for Case 1.

The Manning’s roughness coefficient (n) for the HEC-RAS model needed to be calibrated. Figure 5 shows an example of a sensitivity analysis for the calibration of roughness for the Cihu River. It can be observed that the simulated water level converges to the observed water level with a systematic selection of the n -value, showing that model calibration of roughness can improve the accuracy of model simulation significantly. Based on such sensitivity analysis, Manning’s roughness coefficient was calibrated to $n = 0.028$ for the demonstrated river reach.

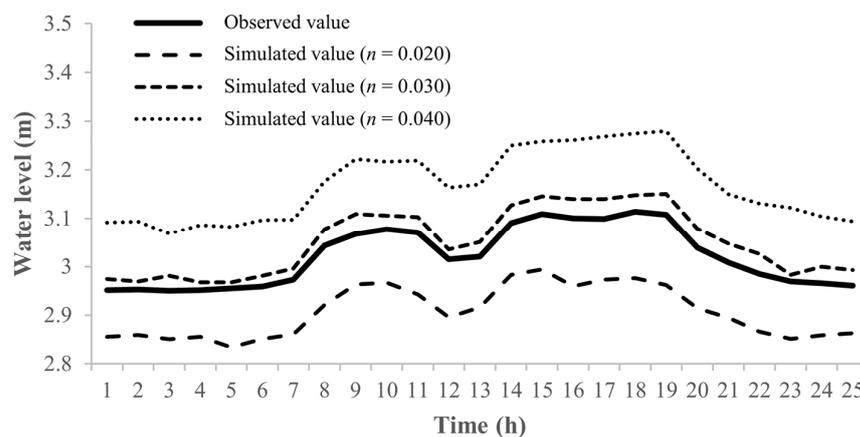


Figure 5. Sensitivity analysis of n -value at Cihu River with respect to observed and simulated water level data.

The unknown side discharge parameter was considered as a time-varying lateral flow whose boundary conditions needed to be optimized in the subsequent DREAM-based framework (i.e., discharge pattern and discharge flow). Based on Equations (3)–(7), the DREAM algorithm was configured as follows. The inverse problem aimed to determine the daily discharge flow and hourly varying coefficients for a known industrial outlet over a 24-h period, resulting in a 25-dimensional ($d = 25$) source parameter matrix $\{Q, f_1, f_2 \dots f_{24}\}$. Here, Q represented the average daily flow, and $f_1, f_2 \dots f_{24}$ represented the hourly time

coefficients of variation. The prior distribution for each source parameter was assumed to follow a uniform distribution within the range of [0, 20,000] for the flow rate and [0, 1] for the time coefficients of variation. Similarly, 7 Markov chains ($N = 7$) were employed, and the pre-set number of iterations was 10,000 ($T = 10,000$). The model error variance of the likelihood function was set at 0.001 ($\sigma = 0.001$), and the likelihood function observations O_t were determined based on the online water levels and flow rates at the downstream boundary of the channel.

Once the HEC-RAS model was constructed and validated, and the setup of the DREAM algorithm was configured, the inverse problem framework was ready for execution. During the iterative process, the previously defined 25-dimensional source parameters were continuously updated with each iteration. This iteration process was repeated until convergence was achieved. All simulations involving HEC-RAS were executed in parallel, making use of the 7 available processors in the Intel® Xeon® W-2245 CPU @ 3.90 GHz (Intel, Santa Clara, CA, USA), which corresponds to running 7 parallel Markov chains in the DREAM algorithm.

Figure S2 presents the R-statistics used to measure the convergence of the sampled chains during the iterations. The R-statistics of each source parameter was less than 1.2 after 7000 iterations, representing the convergence precision of the Markov chains or the officially declared convergence to the limit distribution. In addition, Figure 6 shows the marginal PDFs of the source parameters from the DREAM-based inverse model for this case, including 24 parameters of hourly varying coefficients within a day and 1 parameter of average daily flow. As shown in Figure 5, the solid brown line represents the density of marginal PDFs, and the blue cross marks the MAP estimate. The marginal PDFs of all source parameters followed unimodal distributions with narrow posterior ranges, indicating the high reliability of the integrated pollution tracking framework.

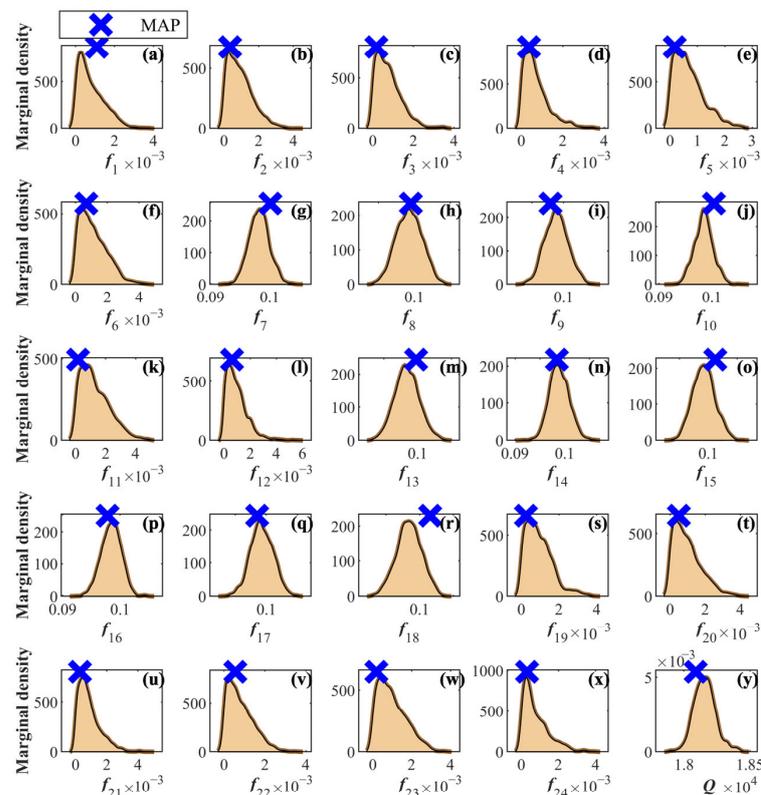


Figure 6. Marginal PDFs of source parameters for the industrial outlet in Cihu River. (a–x) Hourly varying coefficients; (y) Flow rate Abbreviations: the unit of flow rate (Q) is in cubic meters per day (m^3/d).

Utilizing the MAP values of each source parameter as the model identification results and sending these values back into the HEC-RAS model, the outcomes of the forward hydrodynamic model were compared with observed data. As shown in Figure 7, it highlights that the inverse model developed in this study can accurately reproduce the changes in water levels and flow rates caused by source flow discharges. Additionally, as illustrated in Table 2, the calculation of three error indices reveals an R^2 range of 0.950 to 0.960, an RMSE ranging only between 0.012 and 0.072, and an NSE spanning from 0.949 to 0.959. These results indicated a close agreement between the water level data and flow rate data generated through forward simulation using the inverse problem results and the actual measured data. The impressive level of accuracy demonstrated the feasibility of the developed computational tool to identify hourly variable sudden discharge flows under transient flow conditions.

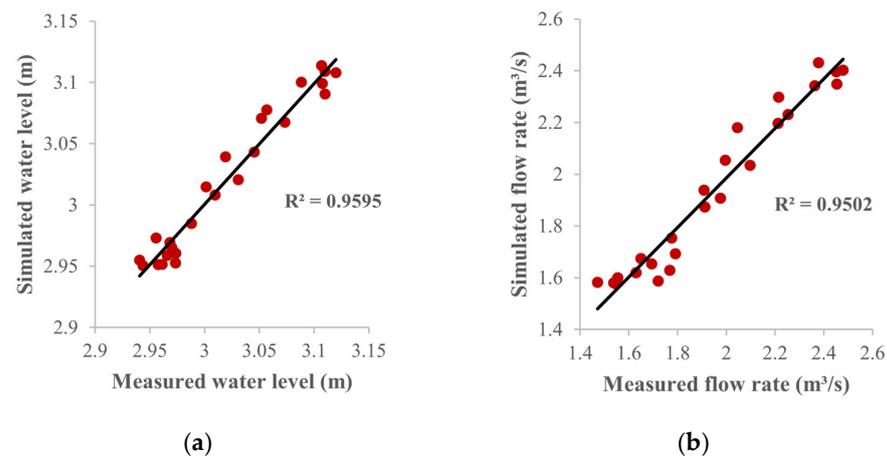


Figure 7. Performance of inversely estimated results from the DREAM-based model: (a) Water level (b) Flow rate.

Table 2. Modeling performances for the simulated downstream values.

	R^2	RMSE	NSE
Water level	0.960	0.012	0.959
Flow Rate	0.950	0.072	0.949

3.2.2. Real Case 2: Source Tracking of Multiple Sewage Discharges into the River

Real case 2 involves the inverse tracking of multiple sewage flow discharges along a 4 km river reach in Shiwan Town of Huizhou City, Guangdong Province. Huizhou City is situated in the southeastern part of Guangdong, within the northeastern region of the Pearl River Delta. Based on the on-site investigation, it was found that this river reach exhibited black water and a noticeable odor. The model domain was confined to the main river channel. A schematic representation of the generalized model for a demonstration river reach is depicted in Figure 8, with two monitoring devices located upstream and downstream, as well as three discharge sources within the river reach. This case served to validate the effectiveness of the computation tool in tracking multiple sewage sources under both dry and wet weather conditions.

To build the HEC-RAS model in this case, river topographic data, meteorological data, and hydraulic data were required. The topographic data and meteorological data were obtained from local water authorities and on-site surveys. The specific topographic data and meteorological data within the studied reach included bottom width from 12.47 to 29.88 m, bottom slope from -0.0004 to 0.0005 , bottom elevation from 2.298 to 2.503 m, and the cross-sectional areas from 6.72 to 24.87 square meters. Hydraulic data, including water level and flow rate, were available from the existing monitoring system in the study

area. The input of upstream and downstream monitoring data for dry and wet weather conditions served as boundary conditions and observation data targets for the subsequent inverse modeling process. Similarly, as shown in Figure 9, Manning’s roughness coefficient was calibrated to $n = 0.026$ for the demonstrated river reach based on sensitivity analysis.

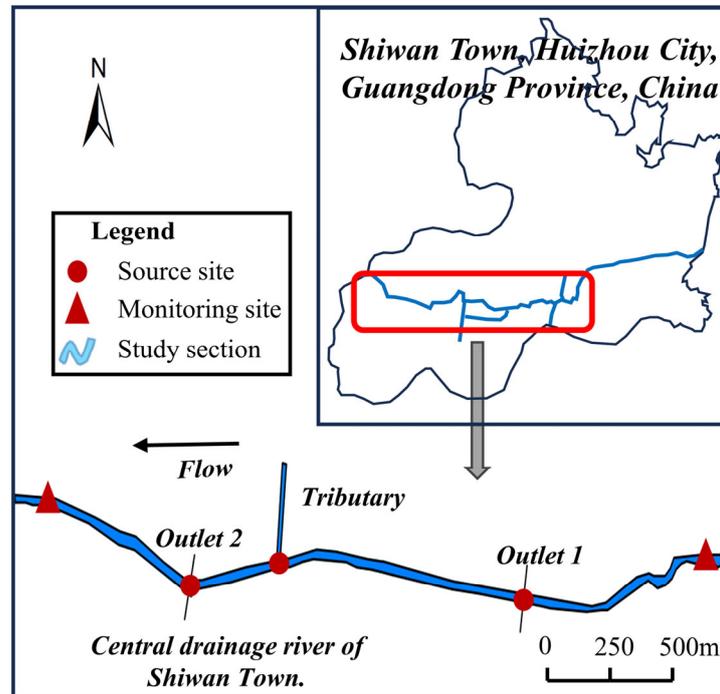


Figure 8. Description of study site and modeling test scenario for Case 2.

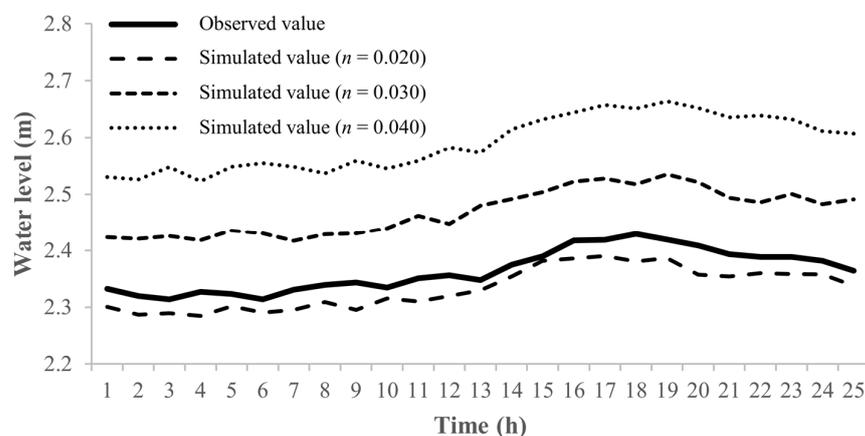


Figure 9. Sensitivity analysis of n -value at central drainage river of Shivan with respect to observed and simulated water level data.

In this case of multiple sewage discharges, the unknown side discharge was considered as a pulse discharge. Similarly, three source parameters had to be determined for each wastewater flow source, specifically denoted as $\{Q_i, t_i, T_i\}$. The prior distribution of the three source parameters was set as a uniform distribution, with $[0, 1]$ for the flow rate, $[0, 23]$ for the emission start time, and $[0, 24]$ for the duration. Therefore, the dimension of the source parameters in this case was set to 9 ($d = 9$) for the three parameters associated with the three sewage sources. Otherwise, the number of Markov chains was 7 ($N = 7$), and the pre-set number of iterations was 10,000 ($T = 10,000$). The observed water levels and flow

rates at the downstream monitoring sites were used as observation data O_t to inversely estimate the unknown source parameters.

Validation was performed under both dry and wet weather conditions to test the effectiveness of the proposed framework. In both scenarios, three lateral flows (two known outlets and one tributary flow) underwent variations at each generation throughout the optimization process. The R-statistics used to measure the convergence of the sampled chains during the iterations are presented in Figure S2, and the evolutions of the seven parallel Markov chains, as well as the resulting marginal PDFs, are presented in Figure S3.

Table 3 shows the MAP estimations of the source parameters from the DREAM-based inverse model for this case, including estimated flow rates, starting time, and duration of the two outlets and tributary. It is evident from the table that during rainfall events (based on data from the local water authority, with a daily rainfall of 19.7 mm), the flow values at each discharge outlet significantly increased compared to dry weather conditions, with an increase ranging from 45.8% to 106.4%.

Table 3. Identification results of source parameters from the DREAM-based model.

(a) Dry weather			
Dry weather	Q (m ³ /s)	t (h)	T (h)
Outlet 1	0.123	7	11
Outlet 2	0.219	12	4
Tributary	0.292	0	24
(b) Wet weather			
Wet weather	Q (m ³ /s)	t (h)	T (h)
Outlet 1	0.254	6	12
Outlet 2	0.330	7	10
Tributary	0.426	0	24

Similarly, through the utilization of MAP values obtained from the inverse model and their integration into the HEC-RAS model, downstream values were simulated. As depicted in Figures 10 and 11, a linear regression analysis comparing observed values to simulated values revealed a robust fit. Additionally, as outlined in Table 4, the analysis of performance metrics revealed R^2 values ranging from 0.915 to 0.935, RMSE values spanning from 0.010 to 0.333, and NSE values between 0.901 and 0.933. Notably, the proposed model demonstrated high reliability and accuracy in inversely estimating multiple discharge sources based on both dry and wet weather observations.

Table 4. Modeling performances for the simulated downstream values for Case 2.

(a) Dry weather			
Dry weather	R2	RMSE	NSE
Water level	0.935	0.010	0.933
Flow Rate	0.930	0.043	0.924
(b) Wet weather			
Wet weather	R2	RMSE	NSE
Water level	0.920	0.057	0.919
Flow Rate	0.915	0.333	0.901

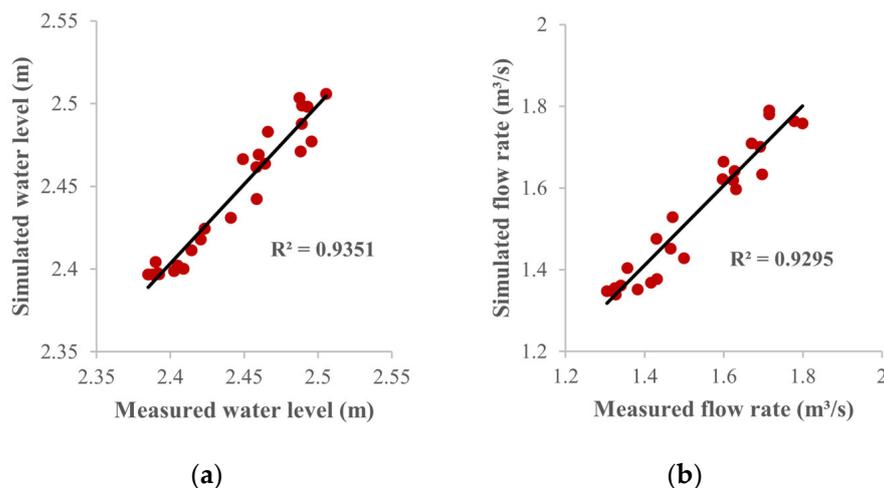


Figure 10. Performance of inversely estimated results from the DREAM-based model during dry weather: (a) Water level (b) Flow rate.

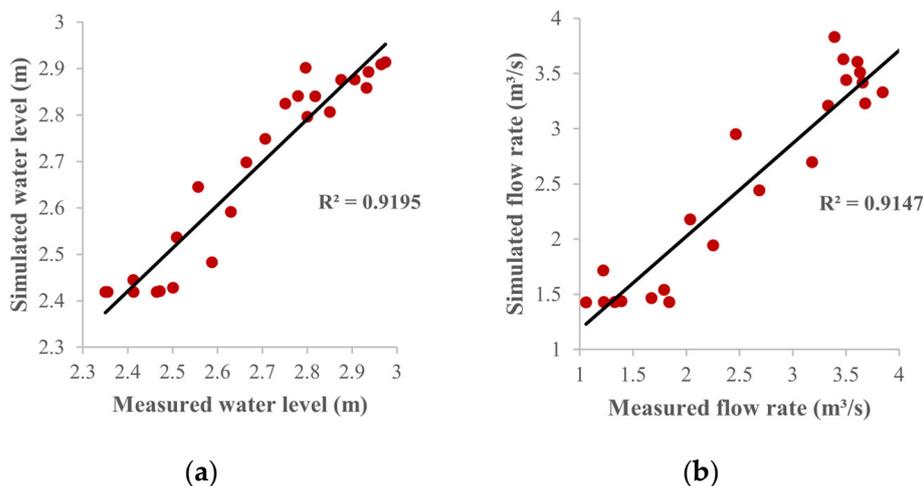


Figure 11. Performance of inversely estimated results from the DREAM-based model during wet weather: (a) Water level (b) Flow rate.

4. Conclusions

This paper presents an innovative open-source framework designed for the inverse tracking of emission patterns from known outlets in rivers under transient flow conditions based on observed river water levels and flow rates at pre-set monitoring sites. This framework is characterized by the integration of two key modules: HEC-RAS, serving as the hydrodynamic model, and DREAM, serving as the inverse problem method. Compared to diffusion wave equations that are only suitable for single-direction flow, the coupled HEC-RAS and DREAM algorithm within MATLAB can be extended to simulate source tracking under highly transient flow conditions, such as sudden discharges or in complex hydraulic conditions caused by rainfall, as demonstrated in this study. The paper also provides relevant code and scripts, serving as resources for reference and practical application in the field.

The proposed DREAM-based framework was first checked by the hypothetical case with pre-known source parameters. The MAP values of DREAM closely matched the true values, with an average relative error of only 0.3% across the six parameters, indicating an 85.7% enhancement in tracking performance compared to GA. Further, by employing real cases, the inverse framework successfully estimated source discharges, including a real scenario involving the tracking of hourly variable sudden discharge from a sewage source

and another scenario involving the tracking of discharge flows and periods from multiple sewage sources under both dry and wet weather conditions. With R^2 values ranging from 0.915 to 0.960, RMSE values from 0.010 to 0.333, and NSE values between 0.901 and 0.959, a strong agreement was observed between simulated values and actual measurements. These case study results revealed that the proposed approach served as an effective tool for inverse tracking sewage flow discharge in real-world scenarios of highly transient flow, showcasing its practical utility in pollution interception and water quality restoration.

The limitation of this study is that the presented cases only focus on the main river channel, but it is possible to extend the source tracking to the entire river network, including its tributaries, by increasing the number of monitoring points. To address this concern, future work will conduct additional cases within the river network by deploying more monitoring stations to expand the application scope of the proposed framework.

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/w16010051/s1>, Figure S1: Schematic Diagram of the MATLAB HEC-RAS framework operation; Figure S2: Convergence of R-statistic of the source parameters for Case 1; Figure S3: Convergence of R-statistic of the source parameters for Case 2. (a) Dry weather (b) Wet weather; Figure S4: Posterior marginal distribution of source parameters for the sewage flow discharge in Case 2 (a) Dry weather (b) Wet weather.; Listing S1: Script to execute designed Framework in MATLAB; Listing S2: Script to rename and relocate user-supplied files; Listing S3: Script for setting optimization options of DREAM: title; Listing S4: Script for automating the update and simulation of HEC-RAS from MATLAB; Listing S5: Script to update the unsteady flow file of HEC-RAS; Listing S6: Script to format the unsteady flow data to fit HEC-RAS; Listing S7: Script to run HEC-RAS and extract simulated water levels; Listing S8: Script for calculating the likelihood function.

Author Contributions: J.W.: conceptualization, formal analysis, visualization, writing—original draft; M.J.: conceptualization, formal analysis; Z.J.: formal analysis; L.S.: methodology; S.W.: conceptualization; Y.S.: formal analysis; W.L.: formal analysis; H.Y.: conceptualization, methodology, writing—original draft, review and editing, supervision. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (Grant 51979195), the National Key Research and Development Project (Grant 2020YFC1808201), the Key Area Research and Development Program of Guangdong Province (Grant 2020B1111350001), and the Shanghai Science and Technology Commission (Grant 20XD1430600).

Data Availability Statement: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest: Authors Lei Su, Shanshan Wu, Yuting Su, Wenxiao Liufu were employed by the company Guangzhou Resource Environmental Protection Technology Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Gonzalez, S.; Lopez-Roldan, R.; Cortina, J.L. Presence and biological effects of emerging contaminants in Llobregat River basin: A review. *Environ. Pollut.* **2012**, *161*, 83–92. [[CrossRef](#)] [[PubMed](#)]
2. Liu, J.; Wang, P.; Jiang, D.; Nan, J.; Zhu, W. An integrated data-driven framework for surface water quality anomaly detection and early warning. *J. Clean. Prod.* **2020**, *251*, 119145. [[CrossRef](#)]
3. Atmadja, J.; Bagtzoglou, A.C. State of the Art Report on Mathematical Methods for Groundwater Pollution Source Identification. *Environ. Forensics* **2001**, *2*, 205–214. [[CrossRef](#)]
4. Mirghani, B.Y.; Mahinthakumar, K.G.; Tryby, M.E.; Ranjithan, R.S.; Zechman, E.M. A parallel evolutionary strategy based simulation–optimization approach for solving groundwater source identification problems. *Adv. Water Resour.* **2009**, *32*, 1373–1385. [[CrossRef](#)]
5. Singh, S.K.; Rani, R. A least-squares inversion technique for identification of a point release: Application to Fusion Field Trials 2007. *Atmos. Environ.* **2014**, *92*, 104–117. [[CrossRef](#)]
6. Ayvaz, M.T. A hybrid simulation–optimization approach for solving the areal groundwater pollution source identification problems. *J. Hydrol.* **2016**, *538*, 161–176. [[CrossRef](#)]

7. Jha, M.; Datta, B. Three-Dimensional Groundwater Contamination Source Identification Using Adaptive Simulated Annealing. *J. Hydrol. Eng.* **2013**, *18*, 307–317. [[CrossRef](#)]
8. Cheng, W.P.; Jia, Y. Identification of contaminant point source in surface waters based on backward location probability density function method. *Adv. Water Resour.* **2010**, *33*, 397–410. [[CrossRef](#)]
9. Jing, P.; Yang, Z.; Zhou, W.; Huai, W.; Lu, X. Inversion of multiple parameters for river pollution accidents using emergency monitoring data. *Water Env. Res.* **2019**, *91*, 731–738. [[CrossRef](#)]
10. Ghane, A.; Mazaheri, M.; Mohammad Vali Samani, J. Location and release time identification of pollution point source in river networks based on the Backward Probability Method. *J. Environ. Manag.* **2016**, *180*, 164–171. [[CrossRef](#)]
11. Zhu, Y.; Chen, Z. Development of a DREAM-based inverse model for multi-point source identification in river pollution incidents: Model testing and uncertainty analysis. *J. Environ. Manag.* **2022**, *324*, 116375. [[CrossRef](#)] [[PubMed](#)]
12. Yang, H.; Shao, D.; Liu, B.; Huang, J.; Ye, X. Multi-point source identification of sudden water pollution accidents in surface waters based on differential evolution and Metropolis–Hastings–Markov Chain Monte Carlo. *Stoch. Environ. Res. Risk Assess.* **2016**, *30*, 507–522. [[CrossRef](#)]
13. Kwon, S.; Noh, H.; Seo, I.; Jung, S.H.; Baek, D. Identification Framework of Contaminant Spill in Rivers Using Machine Learning with Breakthrough Curve Analysis. *Int. J. Environ. Res. Public Health* **2021**, *18*, 1023. [[CrossRef](#)] [[PubMed](#)]
14. Zhang, S.-p.; Xin, X.-k. Pollutant source identification model for water pollution incidents in small straight rivers based on genetic algorithm. *Appl. Water Sci.* **2017**, *7*, 1955–1963. [[CrossRef](#)]
15. Jiang, D.; Zhu, H.; Wang, P.; Liu, J.; Zhang, F.; Chen, Y. Inverse identification of pollution source release information for surface river chemical spills using a hybrid optimization model. *J. Environ. Manag.* **2021**, *294*, 113022. [[CrossRef](#)] [[PubMed](#)]
16. Amiri, S.; Mazaheri, M.; Samani, J.M.V. Introducing a general framework for pollution source identification in surface water resources (theory and application). *J. Environ. Manag.* **2019**, *248*, 109281. [[CrossRef](#)] [[PubMed](#)]
17. Zhu, Y.Y.; Chen, Z.; Asif, Z. Identification of point source emission in river pollution incidents based on Bayesian inference and genetic algorithm: Inverse modeling, sensitivity, and uncertainty analysis. *Environ. Pollut.* **2021**, *285*, 117497. [[CrossRef](#)]
18. Jiang, J.P.; Chen, Y.S.; Wang, B.Y. Pollution Source Identification for River Chemical Spills by Modular-Bayesian Approach: A Retrospective Study on the ‘Landmark’ Spill Incident in China. *Hydrology* **2019**, *6*, 74. [[CrossRef](#)]
19. Vrugt, J.A.; ter Braak, C.J.F.; Diks, C.G.H.; Robinson, B.A.; Hyman, J.M.; Higdun, D. Accelerating Markov Chain Monte Carlo Simulation by Differential Evolution with Self-Adaptive Randomized Subspace Sampling. *Int. J. Nonlinear Sci. Numer. Simul.* **2009**, *10*, 273–290. [[CrossRef](#)]
20. Vrugt, J.A. Markov chain Monte Carlo simulation using the DREAM software package: Theory, concepts, and MATLAB implementation. *Environ. Model. Softw.* **2016**, *75*, 273–316. [[CrossRef](#)]
21. Yin, H.L.; Lin, Y.Y.; Zhang, H.J.; Wu, R.B.; Xu, Z.X. Identification of pollution sources in rivers using a hydrodynamic diffusion wave model and improved Bayesian-Markov chain Monte Carlo algorithm. *Front. Environ. Sci. Eng.* **2023**, *17*, 85. [[CrossRef](#)]
22. Tamiru, H.; Wagari, M. Machine-learning and HEC-RAS integrated models for flood inundation mapping in Baro River Basin, Ethiopia. *Model. Earth Syst. Environ.* **2021**, *8*, 2291–2303. [[CrossRef](#)]
23. Deshays, R.; Segovia, P.; Duviella, E. Design of a MATLAB HEC-RAS Interface to Test Advanced Control Strategies on Water Systems. *Water* **2021**, *13*, 763. [[CrossRef](#)]
24. Leon, A.S.; Tang, Y.; Qin, L.; Chen, D. A MATLAB framework for forecasting optimal flow releases in a multi-storage system for flood control. *Environ. Model. Softw.* **2020**, *125*, 104618. [[CrossRef](#)]
25. Leon, A.S.; Goodell, C. Controlling HEC-RAS using MATLAB. *Environ. Model. Softw.* **2016**, *84*, 339–348. [[CrossRef](#)]
26. MATLAB; Version 9.5.0.944444 (R2018b); The Mathworks, Inc.: Natick, MA, USA, 2018.
27. HEC-RAS. Version 5.0, Hydrologic Engineering Center, US Army Corps of Engineers. Available online: <https://www.hec.usace.army.mil/software/hec-ras/> (accessed on 27 August 2023).
28. Lamichhane, N.; Sharma, S. Development of Flood Warning System and Flood Inundation Mapping Using Field Survey and LiDAR Data for the Grand River near the City of Painesville, Ohio. *Hydrology* **2017**, *4*, 24. [[CrossRef](#)]
29. Lamichhane, N.; Sharma, S. Effect of input data in hydraulic modeling for flood warning systems. *Hydrol. Sci. J.* **2018**, *63*, 938–956. [[CrossRef](#)]
30. Vrugt, J.A.; ter Braak, C.J.F.; Gupta, H.V.; Robinson, B.A. Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modeling? *Stoch. Environ. Res. Risk Assess.* **2008**, *23*, 1011–1026. [[CrossRef](#)]
31. Laloy, E.; Vrugt, J.A. High-dimensional posterior exploration of hydrologic models using multiple-try DREAM(ZS) and high-performance computing. *Water Resour. Res.* **2012**, *48*, W01526. [[CrossRef](#)]
32. Wu, W.; Ren, J.; Zhou, X.; Wang, J.; Guo, M. Identification of source information for sudden water pollution incidents in rivers and lakes based on variable-fidelity surrogate-DREAM optimization. *Environ. Model. Softw.* **2020**, *133*, 104811. [[CrossRef](#)]
33. Gill, J. *Bayesian Methods: A Social and Behavioral Sciences Approach*, 3rd ed.; Chapman Hall/CRC: Boca Raton, FL, USA, 2015.
34. Shen, J.; Zhao, Y. Combined Bayesian statistics and load duration curve method for bacteria nonpoint source loading estimation. *Water Res.* **2010**, *44*, 77–84. [[CrossRef](#)] [[PubMed](#)]
35. Braak, C.J.F.T. A Markov Chain Monte Carlo version of the genetic algorithm Differential Evolution: Easy Bayesian computing for real parameter spaces. *Stat. Comput.* **2006**, *16*, 239–249. [[CrossRef](#)]

-
36. Gelman, A.; Rubin, D.B. Inference from Iterative Simulation Using Multiple Sequences. *Stat. Sci.* **1992**, *7*, 457–472. [[CrossRef](#)]
 37. Brooks, S.P.; Gelman, A. General methods for monitoring convergence of iterative simulations. *J. Comput. Graph. Stat.* **1998**, *7*, 434–455.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.