

An innovative framework for real-time monitoring of pollutant point sources in river networks

Original

An innovative framework for real-time monitoring of pollutant point sources in river networks / Barati Moghaddam, M., Mazaheri, M., Mohammad Vali Samani, J., Boano, F.. - In: STOCHASTIC ENVIRONMENTAL RESEARCH AND RISK ASSESSMENT. - ISSN 1436-3240. - ELETTRONICO. - (2022), pp. 1-28. [10.1007/s00477-022-02233-y]

Availability:

This version is available at: 11583/2964951 since: 2022-05-29T11:36:12Z

Publisher:

Springer Science and Business Media Deutschland GmbH

Published

DOI:10.1007/s00477-022-02233-y

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

An Innovative Framework for Real Time Monitoring of Pollutant Point Sources in River Networks

M. BaratiMoghaddam

Tarbiat Modares University <https://orcid.org/0000-0001-8346-1707>

Mehdi Mazaheri (✉ m.mazaheri@modares.ac.ir)

Tarbiat Modares University <https://orcid.org/0000-0001-8670-1710>

J. M. V. Samani

Tarbiat Modares University

Fulvio Boano

Politecnico di Torino

Research Article

Keywords: Geostatistical approach, Inverse problem, Multiple pollutant point sources, Source identification, Unsteady flow, River network

Posted Date: June 18th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-597673/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

1 **An Innovative Framework for Real Time Monitoring of Pollutant**
2 **Point Sources in River Networks**

3 **M. BaratiMoghaddam^{1,2}, M. Mazaheri^{1*}, J. M. V. Samani¹ and F. Boano²**

4 ¹ Department of Water Engineering and Management, Tarbiat Modares University, Tehran, Iran.

5 ²Department of Environment, Land and Infrastructure Engineering (DIATI), Politecnico di Torino,
6 Turin, Italy.

7

8 * Corresponding author: m.mazaheri@modares.ac.ir

9

10

11

12

13

14

15

16

17

18

19

20

21

22 **An Innovative Framework for Real Time Monitoring of Pollutant**
23 **Point Sources in River Networks**

24 **Abstract:**

25 The simultaneous identification of location and source release history in complex river
26 networks is a very complicated ill-posed problem, particularly in a case of multiple unknown
27 pollutant sources with time-varying release pattern. This study presents an innovative method
28 for simultaneous identification of the number, locations and release histories of multiple
29 pollutant point sources in a river network using minimum observation data. Considering
30 two different type of monitoring stations with an adaptive arrangement as well as real-time
31 data collection at those stations and using a reliable numerical flow and transport model, at
32 first the number and suspected reach of presence of pollutant sources are determined. Then
33 the source location and its intensity function is calculated by solving inverse source problem
34 using a geostatistical approach. A case study with three different scenarios in terms of the
35 number, release time and location of pollutant sources are discussed, concerning a river
36 network with unsteady and non-uniform flow. Results showed the capability of the
37 proposed method in identifying of sought source characteristics even in complicated cases
38 with simultaneous activity of multiple pollutant sources.

39 **Keywords:** Geostatistical approach, Inverse problem, Multiple pollutant point sources,
40 Source identification, Unsteady flow, River network.

41

42

43

44 **1. Introduction**

45 Water resources are essential to life on the Earth planet, but these limited and valuable
46 resources are increasingly under threat. Rivers in particular due to proximity to big cities and
47 extensive usage in industrial and agricultural activities, are extremely exposed to accidental or
48 intentional spills. Regarding to this issue, in recent years, a great attention has been drawn to
49 simulate fate and transport of contaminants in rivers as well as to identify pollutant sources
50 characteristics. Recovering release history of pollutant sources is essential in planning effective
51 remediation strategies. Moreover, determining the number and location of pollutant sources is
52 of great importance in order to identify responsible parties for observed pollution cloud in
53 downstream and divide remediation measure expenses among those parties (Skaggs and
54 Kabala, 1994, Liu and Ball, 1999, Atmadja and Bagtzoglou, 2001, Michalak, 2002).

55 Given known concentration data at limited downstream observation points, the pollutant
56 source identification problem is categorized as an inverse problem. Like most of the inverse
57 problems, the inverse source problem does not fulfill the well-posedness criteria of Hadamard
58 (1923). Based on Hadamard's definition, a problem is well-posed if its solution is existent,
59 unique and stable. A problem which lacks any of these features called an ill-posed problem.
60 However, since the observed pollution cloud at the downstream point, must be originated from
61 somewhere at the upstream, pollutant source identification problem always has a solution and
62 nonexistence would not raise an issue. Hence, there are two main challenges in solving an
63 inverse source problem, namely nonuniqueness and instability of the solution. The
64 nonuniqueness means that different combinations of intensity functions of the pollution sources
65 at the upstream can create a single concentration-time curve at a given observation point
66 downstream. Since time discretization of governing equation to pollution transport results in a
67 system of equations which has fewer equations (observations) than unknowns (source values),
68 multiple combinations for source characteristics might be detected which are consistent with

69 observed concentration data. To address the nonuniqueness issue researchers often assume that
70 some prior information about the unknown source is available (e.g. possible location, activity
71 duration or known intensity function, as well as, consideration a particular form of the source
72 term function). The instability issue implies to large errors in the solution following small errors
73 in measured data. It is mainly a consequence of irreversibility of dispersion phenomena, which
74 gradually smooth the pollution plume and decrease the amount of obtainable information from
75 observational data (Skaggs and Kabala, 1998). Hence, considering uncertainties in observed
76 data regarding to measurement errors and sparsity of data increases the reliability of the
77 identification results.

78 In the last 30 years, various methods have been proposed to solve pollutant source
79 identification problem in surface and groundwater which can be broadly categorized into three
80 classes: optimization-based approaches, stochastic-based approaches and mathematics-based
81 approaches. A review of those research can be found in (Atmadja and Bagtzoglou, 2001,
82 Michalak and Kitanidis, 2004b, Morrison, 2000b, Morrison, 2000a, Neupauer et al., 2000).
83 Among these methods, stochastic-based methods are becoming a trend in solving inverse
84 source problem in recent years. The most significant feature of stochastic-based approaches is
85 to treat unknown pollution source parameters as random variables and use of probability
86 distribution functions to predict those parameters. This feature provides the possibility of
87 estimation of the source characteristics in greater number of instants than available observation
88 data as well as consideration of uncertainty due to error in those data (Woodbury et al. 1998).

89 One of the stochastic-based methods which is extensively used in solving of inverse source
90 problem in groundwater is Geostatistical (GS) method. The main assumption of GS method is
91 that the unknown source function is random with a known correlation structure but unknown
92 correlation structural parameters. The optimal values of these structural parameters are
93 obtained using the geostatistical inversion theory presented by Kitanidis (1996) and the source

94 function recovered by minimizing a likelihood function while retaining the assumed correlation
95 structure. More details can be found in Kitanidis (1995, 1996) and Snodgrass and Kitanidis
96 (1997). The GS approach has been widely tested and improved in groundwater source
97 identification through hypothetical cases (Snodgrass and Kitanidis, 1997, Michalak and
98 Kitanidis, 2003, Michalak and Kitanidis, 2004b, Butera et al., 2013) and using field data
99 (Michalak and Kitanidis, 2002, Michalak and Kitanidis, 2004a, Gzyl et al., 2014). This
100 approach also had been applied only once in pollutant source identification in single-branched
101 rivers considering the effects of transient storage zone as well as linear decay processes (Boano
102 et al., 2005). Snodgrass and Kitanidis (1997) applied the GS approach for estimating the release
103 history of a conservative solute in a 1D homogeneous aquifer. Instead of using the usual
104 iterative techniques to obtain the best estimation of parameters, they combined GS techniques
105 with Bayesian theory, which provides the possibility to quantify the estimation error. Michalak
106 and Kitanidis (2002) applied the proposed approach by Snodgrass and Kitanidis (1997) for the
107 reconstruction of the contaminant release history for a 3D plume at Gloucester landfill site in
108 Ontario, Canada. Michalak and Kitanidis (2004b) combined the adjoint model with
109 geostatistical techniques in order to reduce the computational cost as well as providing the
110 possibility to use the approach in heterogeneous fields. In addition, using an adjoint model
111 provides the feasibility of application of the existing groundwater flow and transport
112 commercial codes in the framework of the proposed inverse method. Butera et al. (2013) based
113 on GS proposed a framework for simultaneous identification of location and release history of
114 a single pollutant source in 2D confined aquifers with strongly non-uniform flow field. Gzyl et
115 al. (2014) presented a multi-step method based on performing an integral pumping test and GS
116 approach to identify location and release history of a pollutant source in groundwater. The
117 results of applying this method to a complicated contamination case at the adjacent reach to an
118 abandoned chemical plant in southern Poland, indicated that it is able to successfully detect

119 suspected areas. However, the proposed methods by Butera et al. (2013) and Gzyl et al. (2014)
120 need the prior knowledge of the approximate location of pollutant source at the beginning of a
121 simulation, which is a challenge in practical applications, especially in complicated cases that
122 such information may not be available.

123 Compared to numerous studies on pollutant source identification of in groundwater, only
124 relatively few studies on solving an inverse source problem in surface waters can be found in
125 literature (El Badia and Hamdi, 2007, Hamdi, 2009, Hamdi, 2016, Andrie and El Badia, 2012,
126 Cheng and Jia, 2010, Mazaheri et al., 2015, Yang et al., 2016, Wang et al., 2018). While, the
127 pollutant transport in rivers tends to be more advection-dominated than groundwater and,
128 subsequently pollutant substance transported faster and further, which may lead to partial
129 capturing of the pollution plume at the observation points. Thus, fast and accurate identification
130 of illegal spills is more important in these environments to provide scientific support for
131 planning mitigation and adaptation strategies. Furthermore, the research of pollutant source
132 identification problem in surface waters was mainly confined to single-branch rivers and rarely
133 involved river networks. This is mainly due to hydrodynamical complexity of such systems
134 which along with inherent ill- posedness of corresponding source identification problem, form
135 a problem that is very difficult to solve. However, regarding that tributaries in a river network
136 usually are less monitored, those areas might be considered as potential places for illegal
137 discharge of pollutants. Therefore, to prevent further damage, it is necessary to pay more
138 attention to identifying pollutant sources characteristics in such environments.

139 Focusing on pollutant source identification in river networks, Telci and Aral (2011) by
140 using an adaptive sequential feature selection algorithm (Jiang, 2008), determined the location
141 of a single instantaneous source among several candidate locations. However, their proposed
142 method requires a significant amount of simulation time for training monitoring stations with
143 a large number of spill scenarios. Ghane et al. (2016) applied the backward probability method

144 to identify the source location and the released time of a single spill in a river system. Lee et
145 al. (2018) dealt with the problem of identifying the location of a single instantaneous source
146 via analyzing changes in concentration levels that observed by a sensor network in a river
147 system. By constructing random forest models, they determined the possibility that each
148 candidate location be the correct one as a number between zero and one. However, all of
149 mentioned studies considered a single pollutant source with a simple form of release (i.e. the
150 spill), while in many practical application, there are more than one active source and the release
151 functions varies with time.

152 Apart from the issue of insufficient studies on pollution source identification in rivers, most
153 of previous studies considered the location of the pollutant source to be known priori. This
154 assumption is not compatible with real-world condition, since in most cases the location of the
155 pollution source is also unknown as its intensity function. Introducing the source location as an
156 unknown, will have a significant effect on source identification process due to interaction
157 between a release at a variable source location and observational data. In other words, different
158 potential source location sets may result in significantly different solutions. Moreover, the
159 simultaneous identification of location and source release history is a very complicated ill-
160 posed problem, particularly in a case of multiple unknown pollution sources with time-varying
161 release pattern. The main motivations behind this study is to provide an innovative method for
162 simultaneous identification of the number, locations and release histories of multiple point
163 sources in a river network using minimum observational data and considering near real world
164 conditions namely unsteady and non-uniform flow as well as reactive pollutants. The proposed
165 method includes two main steps that are given below:

166 Step1: determining the number and suspected reaches to presence of sources by placement of
167 observation points in a specific manner and management of data collection at those stations.

168 Step 2: identification of exact location and intensity function of the source by solving the
 169 inverse source problem using a geostatistical approach.

170 The method is effective and easy to apply in complex river networks as well as single-branch
 171 ones. Moreover, it provides the possibility of simulators identification of all active pollutant
 172 sources. Hence the required computational time is significantly lower than common iterative
 173 methods such as simulation-optimization approach.

174 2. Material and Methods

175 2.1. Governing Equations and Statement of the problem

176 The main governing equation of solute transport in surface waters is advection-dispersion
 177 equation (ADE) (Taylor, 1954), which is a parabolic partial differential equation derived from
 178 a combination of continuity equation and Fick's first law. The one-dimensional ADE equation
 179 is as follows (Fischer et al., 1979):

$$\frac{\partial(AC)}{\partial t} + \frac{\partial(CQ)}{\partial x} - \frac{\partial}{\partial x} \left(AD \frac{\partial C}{\partial x} \right) + A\lambda C - \sum_{i=1}^m f_i(t) \delta(x - x_i) = 0 \quad (1)$$

180 where, A is the flow area, C is the solute concentration, Q is the volumetric flow rate, D is the
 181 dispersion coefficient, λ is the first-order decay coefficient, m is the number of pollution
 182 sources, $f_i(t)$ is correspondent release history of i th pollution source, $\delta(x)$ is the Dirac delta
 183 function, x_i is the i th point source release location, t and x are the time and distance,
 184 respectively. It also should be mentioned that, hydrodynamic parameters (i.e., A , Q , D) in
 185 Equation (1) are obtained from the hydrodynamics model which is based on well-established
 186 Saint-Venant equations (Wu, 2007):

$$\frac{\partial A}{\partial t} + \frac{\partial Q}{\partial x} = 0 \quad (2)$$

$$\frac{\partial Q}{\partial t} + \frac{\partial}{\partial x} \left(\frac{Q^2}{A} \right) + gA \frac{\partial z_s}{\partial x} + gAS_f = 0 \quad (3)$$

187 in which z_s and S_f are water level and energy slope, respectively.

188 The general expression of the considered problem is that there are multiple pollutant point
189 sources S_1, S_2, \dots, S_m in a river network, which the number, locations (x_1, x_2, \dots, x_m) and
190 intensity functions $(f_1(t), f_2(t), \dots, f_m(t))$ of those sources are unknown. The main
191 objectives are to present a methodology for simultaneous identification of these sources
192 characteristics (i.e. their number, locations, and intensity functions), and obtaining a unique
193 response for the considered inverse source problem with a minimum measured concentration
194 data at observation points. The proposed method consists of two main steps. The method starts
195 with the determination of a spatial range in which the source of pollution is likely to be present.
196 Then the location and approximate release history of pollution sources are recovered by means
197 of a geostatistical approach, that considered simultaneously all the possible candidates. The
198 method is effective and easy to apply in complex river networks as well as single-branch ones.
199 Moreover, since in each simulation all active pollutant sources are identified, the required
200 computational time is significantly lower than common iterative methods such as simulation-
201 optimization approach. More details are given in following sections.

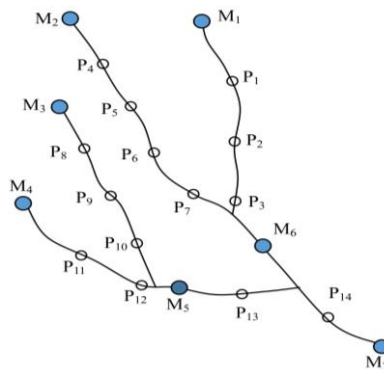
202 **2.2. Step1: Determination of the Number and Approximate Location of Pollution** 203 **Sources**

204 In order to determine the approximate location of pollutant sources, some observation points
205 are considered with a specific arrangement and data collection at those observation points are
206 managed based on specified condition of each problem. In order to provide the concentration
207 data and proceed with the identification process, two types of observation points are defined,
208 main (M_1, M_2, \dots, M_n) and secondary stations (P_1, P_2, \dots, P_k) (Figure 1). The main
209 stations collect concentration-time data continuously, but the secondary ones collect data
210 occasionally and on-demand. The placement of main and secondary stations is based on some
211 priori information including desired activity time for retrieval and accuracy of spatial range for

212 pollution source localization. The main stations are placed in a way that the travel time between
 213 two successive main stations always is less than or equal to the expected activity time for the
 214 sources. The travel time between successive main stations for each branch of the river network
 215 and is calculated using following equation (Chapra, 2008):

$$\bar{T} = \frac{\sum_{i=1}^{n-1} (C_i t_i + C_{i+1} t_{i+1})(t_{i+1} - t_i)}{\sum_{i=1}^{n-1} (C_i + C_{i+1})(t_{i+1} - t_i)} \quad (4)$$

216 in which \bar{T} is travel time, C_i is the concentration at temporal instant t_i . The secondary stations
 217 are arranged in a way that the distance between two successive stations be equal to the accuracy
 218 which expected for approximate location of the sources. This configuration of monitoring
 219 station makes it possible to identify all active pollutant sources with minimum measurement
 220 data and avoid additional data collection as well as related costs.



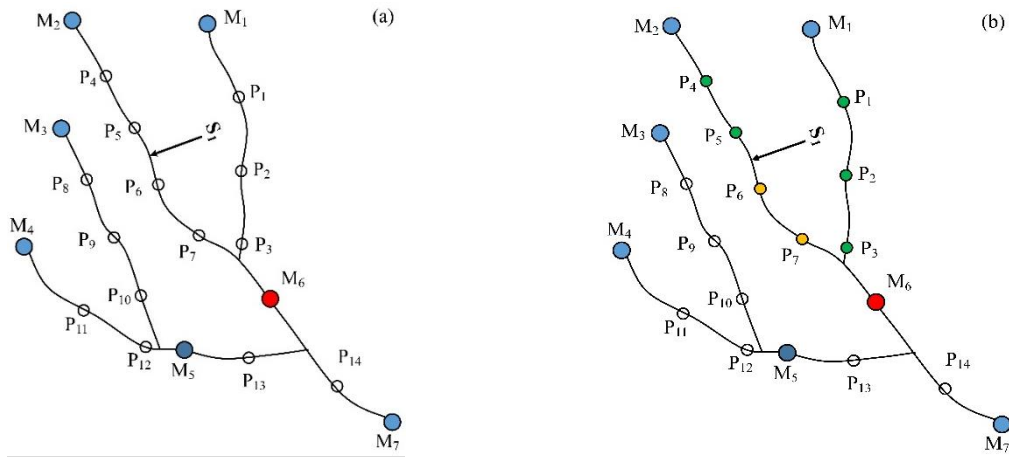
221 **Figure 1-** A hypothetical river network and arrangement of main and secondary stations
 222 The key step in the algorithm is comparison of observed and simulated concentration data
 223 in the main stations, so that any difference between these two sets of data is a sign of existence
 224 of a pollutant source at the upstream of that particular station. The simulated data are taken
 225 from an integrated flow and transport model, which solves equation (1) - (3) in a river network
 226 for a case of no active pollutant source. It is a real-time simulation model which continuously

227 executed and its outputs namely concentration-time data $C(x,t)$ are used in solving the inverse
228 source problem by proposed algorithm.

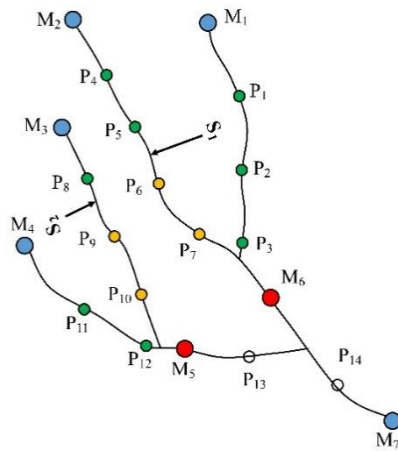
229 Once a main station detects a difference between observed and simulated concentration
230 data, depending on the type of communication topology of monitoring system a command will
231 be send from a control center or directly from that main station to all secondary stations which
232 have been located between that main station and the first main station at upstream of it, to
233 collect a concentration data at the instant of difference detection. The first secondary station
234 from upstream which shows a difference between observed and simulated data, guide us to the
235 approximate location of the source. In other words, the pollution source must be located in the
236 reach between that secondary station and the first secondary station at its upstream (Figure 2).
237 After determining the approximate location of the pollution source, following actions should
238 be done:

- 239 1. The secondary station which detected the difference as well as the secondary station
240 located at upstream of that, should start to continuous data collecting, to assure that
241 other active sources at the upstream and/or downstream of that suspected reach, will be
242 detected as well,
- 243 2. The source location should be determined more accurately, to proceed to find the
244 intensity function of detected source,
- 245 3. The forward model should be revised to include the characteristics of the identified
246 source.

247 It also should be noted that the continuous data collecting at secondary stations, which frame
248 the source location, will be stopped after the full passage of pollution cloud from secondary
249 station that located at the downstream bound of suspected reach. The identification process of
250 approximate location of a case with multiple pollutant sources, are quite similar to what was
251 described for the case with one active source (Figure 3).



252 **Figure 2-** detection of approximate location in the case of one active source



253 **Figure 3-** A case with two active sources

254 **2.3. Step 2: Recovering the Characteristics of Pollutant Sources by**
 255 **Means of a Geostatistical Method**

256 After determining the number and suspected reaches to presence of pollution sources, the exact
 257 location and approximate intensity function of pollution sources should be determined. Hence,
 258 at first the mentioned reaches are divided to some sub-reaches and the potential location of
 259 pollutant sources are considered at the center of those sub-reaches. Then, by solving an inverse
 260 source problem, the true location of the pollutant sources (i.e., where the pollutant injection has
 261 most likely originated) is determined as a location that the highest contaminant release history
 262 is obtained. In order to solve the inverse source problem a Geostatistical method (GS) has been
 263 used in this study. Regarding the linearity of equation (1) the solution of these equation subject

264 to initial and boundary conditions (i.e., $C(x,0) = C_0(x)$, $C(0,t) = C_{in}(t)$, $C(L,t) = 0$) is
 265 (Skaggs and Kabala, 1994):

$$C(\mathbf{x},t) = \int_0^t f(\tau)K(\mathbf{x},t-\tau)d\tau \quad (5)$$

266 where $K(\mathbf{x},t-\tau)$ is the transfer function (TF), that describes the effect in time at a certain
 267 location \mathbf{x} by a unitary impulse source which is released at x_0 and time τ . If M observational
 268 data be available and the time domain is discretized in N instants, a general expression of the
 269 relation between the observations and the source can be written as follows:

$$\mathbf{z} = \mathbf{h}(\mathbf{f}) + \mathbf{v} \quad (6)$$

270 where \mathbf{z} is a $[M \times 1]$ random vector of observations, \mathbf{f} is a $[N \times 1]$ random vector of discretized
 271 release history, \mathbf{h} is the model function and \mathbf{v} is a $[M \times 1]$ random vector that represents the
 272 measurement errors. The error vector \mathbf{v} is Gaussian with a zero mean and a covariance matrix
 273 as $\mathbf{R} = \sigma_R^2 \mathbf{I}$ in which \mathbf{I} is the $[M \times M]$ identity matrix. It also should be noted that $N > M$,
 274 which means that there are more unknowns than measurements. By comparing equation (6)
 275 and (5) it can be concluded that the function $\mathbf{h}(\mathbf{f})$ is linear and therefore equation (6) can be
 276 rewritten as follows:

$$\mathbf{z} = \mathbf{H}\mathbf{f} + \mathbf{v} \quad (7)$$

277 where \mathbf{H} is a $[M \times N]$ matrix known as transfer matrix and its generic element is:

$$H_{i,j} = \Delta\tau \begin{cases} K(t_i - \tau_j) & t_i > \tau_j \\ 0 & t_i < \tau_j \end{cases} \quad (8)$$

278 in which $\Delta\tau$ is time step between two successive discretization of intensity function, t_i and τ_j
 279 are observation instants and release time, respectively.

280 The $H_{i,j}$ element of transfer matrix represents the effect of a release at τ_j on observation
 281 data z_i at which collected at t_i . As shown in equation (8), to construct the \mathbf{H} matrix, it is
 282 necessary to calculate TFs at different time instants. TFs describe the response of the system

283 to a unit impulse injection. Therefore, to calculate them, the ADE equation (equation (1)) need
284 to be solved for a unitary release function at the source location and for different time instants.
285 In case of simple problems with steady flow, regular cross-sections and constant parameters,
286 TFs can be determined using analytical procedures. However, in many practical applications,
287 with unsteady flow, irregular cross-sections and variable parameters using analytical formulas
288 in evaluation of TFs values, is only possible by considering a series of simplifying assumptions.
289 As a consequence, a rough approximation in the solution of inverse problem expected, that is
290 not desirable. Due to the complex conditions that considered in this study, the transfer
291 functions have been calculated using the finite volume numerical method. To calculate the
292 value of $H_{i,j}$ terms, several runs of the numerical model were performed. In case of unsteady
293 flow, the numerical model has to be performed for all the τ_j instants that are desired to recover
294 the intensity function, i.e., N times. For each run the unit release is modelled as Dirac delta
295 function $\delta(\tau_j)$ at potential source location and breakthrough curves at observation points were
296 calculated. In other words:

$$H_{ij} = C(x, t_i) = \int_0^{t_i} \delta(\tau_j) K(x, t_i - \tau_j) d\tau \quad (9)$$

297 The equation (7) is a system of ill-posed equations that cannot be solved by conventional
298 methods. In order to overcome this difficulty, it is assumed that \mathbf{f} has a normal distribution with
299 mean and covariance as follows:

$$E[\mathbf{f}] = \mathbf{X}\beta \quad (10)$$

$$E[(\mathbf{f} - \mathbf{X}\beta)(\mathbf{f} - \mathbf{X}\beta)^T] = \mathbf{Q}(\boldsymbol{\theta}) \quad (11)$$

300 where \mathbf{X} is a $[N \times 1]$ unit vector, β is the unknown mean, $\boldsymbol{\theta}$ is a vector of unknown structural
301 parameters of the covariance function, and \mathbf{Q} is the covariance matrix of the release $f(\tau)$. In
302 this research, a Gaussian covariance matrix has been considered, whose formulation is as
303 follows:

$$\mathbf{Q}(\tau_i - \tau_j | \boldsymbol{\theta}) = \sigma^2 \exp \left[-\frac{(\tau_i - \tau_j)^2}{I_f^2} \right] \quad (12)$$

304 where $\boldsymbol{\theta}^T = [\sigma^2, I_f]$ are structural parameters.

305 The reconstruction of pollutant source intensity function in the geostatistical method consist
 306 of two steps. In the first step, known as structural analysis, the structural parameters of the
 307 covariance function θ are determined, and in the second step, the contaminant source intensity
 308 function (\mathbf{f}) is estimated using the kriging method. Structural parameters are determined by
 309 minimizing the following objective function (Snodgrass and Kitanidis, 1997):

$$L(\boldsymbol{\theta}) = -\ln [p(\mathbf{z} | \boldsymbol{\theta})] \propto \frac{1}{2} \ln (|\Sigma| \cdot |\mathbf{X}^T \mathbf{H}^T \Sigma^{-1} \mathbf{H} \mathbf{X}|) + \frac{1}{2} \mathbf{z}^T \boldsymbol{\Xi} \mathbf{z} \quad (13)$$

310 in which:

$$\Sigma = \mathbf{H} \mathbf{Q} \mathbf{H}^T + \mathbf{R} \quad (14)$$

$$\boldsymbol{\Xi} = \Sigma^{-1} - \Sigma^{-1} \mathbf{H} \mathbf{X} (\mathbf{X}^T \mathbf{H}^T \Sigma^{-1} \mathbf{H} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{H}^T \Sigma^{-1} \quad (15)$$

311 the minimization of Equation (13) is a well-posed problem, since the number of observation \mathbf{z}
 312 is greater than the number of structural parameters $\boldsymbol{\theta}$. In equations (14) and (15), \mathbf{R} is the
 313 covariance matrix of error in the observational data (\mathbf{v}). It should be noticed that the value of
 314 the unknown mean β is not relevant as it does not appear in the Equations (13)-(15). The β
 315 coefficients are eliminated from Equation (13) by averaging over all possible values of it
 316 (Hoeksema and Kitanidis, 1985, Kitanidis, 1995).

317 Once the structural parameters $\boldsymbol{\theta}$ are calculated, the intensity function is estimated through
 318 a kriging system (De Marsily, 1986):

$$\hat{\mathbf{f}} = \boldsymbol{\Lambda} \mathbf{z} \quad (16)$$

319 Equation (16) is a linear estimator. It is unbiased and minimizes the estimate error variance
 320 (Boano et al., 2005, Butera et al., 2013), in other words:

$$E[\hat{\mathbf{f}} - \mathbf{f}] = 0 \quad (17)$$

$$\min_{\hat{\mathbf{f}}} E \left[(\hat{\mathbf{f}} - \mathbf{f}) - (\hat{\mathbf{f}} - \mathbf{f})^T \right]. \quad (18)$$

321 Λ is a $[N \times M]$ matrix of Kriging weights that obtained from solving the following system of
 322 equation:

$$\begin{bmatrix} \Sigma & \mathbf{HX} \\ (\mathbf{HX})^T & 0 \end{bmatrix} \begin{bmatrix} \Lambda^T \\ \mathbf{M} \end{bmatrix} = \begin{bmatrix} \mathbf{HQ} \\ \mathbf{X}^T \end{bmatrix} \quad (19)$$

323 where \mathbf{M} is a $[1 \times N]$ matrix of Lagrange multipliers (De Marsily, 1986). The mean of the
 324 release history is then estimated by equation (16), while its covariance matrix \mathbf{V} can be
 325 evaluated as:

$$\mathbf{V} = -\mathbf{XM} + \mathbf{Q} - \mathbf{QH}^T \Lambda^T \quad (20)$$

326 Using Equation (20) the confidence interval of 95% can also be determined, so that for every
 327 instant of time t_i , the confidence interval can be calculated as $\hat{f}_i \pm 2\sqrt{V_{ii}}$ in which V_{ii} is the
 328 estimation error variance of \hat{f}_i .

329 The GS method is a practical and efficient method, but sometimes it obtains non-physical
 330 results, including negative concentrations. Usually, this problem is alleviated by introducing
 331 additional constraints to the unknown variable (Box and Cox, 1964, Snodgrass and Kitanidis,
 332 1997, Michalak and Kitanidis, 2003, 2004a). This constrain is imposed by using a power
 333 transformation of the unknown variables. The new unknown function is written as follows:

$$f^{\alpha} = \alpha (f^{1/\alpha} - 1) \quad (21)$$

334 where α is a small positive parameter, the value of which is chosen in a way that ensure $f^{\alpha} > -\alpha$
 335 . Kitanidis and Shen (1996) presented a method for choosing the optimal value of parameter α
 336 .

337 Then, the transformed variable f^{α} should be substituted to original variable f in equation (6),
 338 so equation (6) is replaced by the following one:

$$\mathbf{z} = \mathbf{h}(f^{\alpha}) + \mathbf{v} \quad (22)$$

339 in which:

$$\mathbf{h}(f^{\alpha}) = \mathbf{h} \left[\left((f^{\alpha} + \alpha) / \alpha \right)^{\alpha} \right] \quad (23)$$

340 since the model is no longer linear with respect to the transformed variable \mathbf{f}^t , the solution
 341 must be evaluated using successive iterations. More details could be found in Kitanidis (1995).
 342 The method can easily be extended to the case of m multiple independent point sources located
 343 at $\mathbf{x} = [x_1, x_2, \dots, x_m]$ and p distinct measurement points located at $\mathbf{x}_{obs} = [x_1, x_2, \dots, x_p]$.
 344 Regarding to the linearity of the ADE with respect to the concentration $C(\mathbf{x}, t)$, it can be
 345 written:

$$C(\mathbf{x}_{obs}, t) = \sum_{j=1}^m f_j(\tau) K(\mathbf{x}_{obs} - \mathbf{x}_j, t - \tau) \Delta\tau \quad (24)$$

346 In which n (is the number of observation points. The matrix form of equation $i=1,2,\dots,p$
 347 is as follows:

$$\mathbf{z} = \mathbf{H}\mathbf{f} + \mathbf{v} \quad (25)$$

348 where:

$$\mathbf{z}^T = [\mathbf{z}_1 \ \mathbf{z}_2 \ \dots \ \mathbf{z}_p] \quad (26)$$

$$\mathbf{f}^T = [\mathbf{f}_1 \ \mathbf{f}_2 \ \dots \ \mathbf{f}_m] \quad (27)$$

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_{11} & \mathbf{L} & \mathbf{H}_{1m} \\ \mathbf{M} & \mathbf{O} & \mathbf{M} \\ \mathbf{H}_{i1} & \mathbf{L} & \mathbf{H}_{pm} \end{bmatrix} \quad (28)$$

349 Equation (25) is a system of equation in which, \mathbf{H}_{ij} , $i=1,2,\dots,p$, $j=1,2,\dots,m$, is the transfer
 350 matrix corresponding with the effect of the pollutant source release at x_i on the measured
 351 concentration data at x_j . Since pollutant sources are independent, the covariance matrix is a
 352 block matrix as follows:

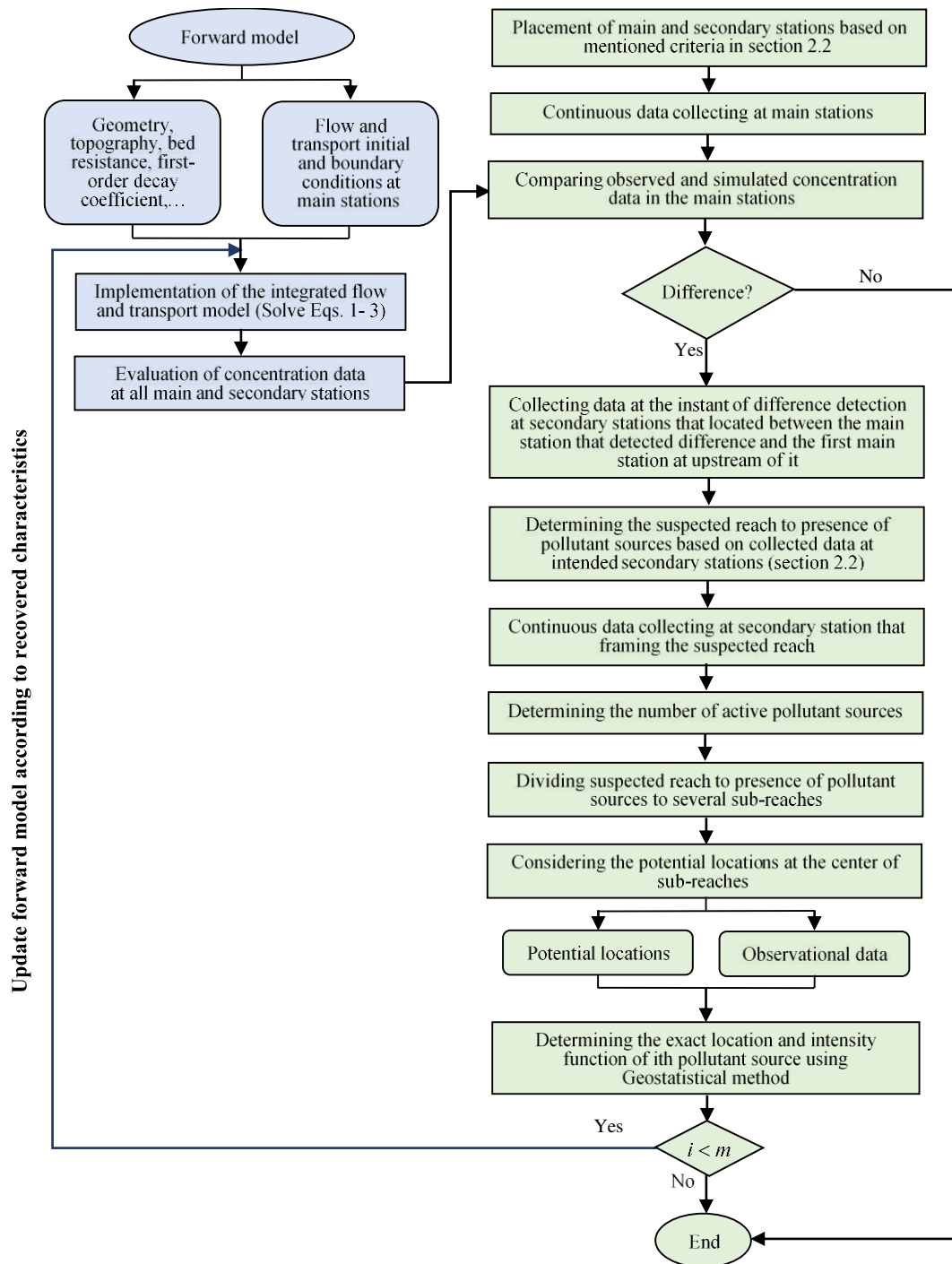
$$\mathbf{Q} = \begin{bmatrix} \mathbf{Q}_1 & \mathbf{L} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_2 & \mathbf{L} & \mathbf{0} \\ \mathbf{M} & \mathbf{M} & \mathbf{O} & \mathbf{M} \\ \mathbf{0} & \mathbf{0} & \mathbf{K} & \mathbf{Q}_m \end{bmatrix} \quad (29)$$

353 The rest of steps for solving the system of equations (22) are similar to solving for a single
 354 pollutant source, described in the previous sections. Figure 4 represents a flowchart of overall
 355 identification process.

356 **3. Results and Discussion**

357 In this section an application of the proposed method for simultaneous identification of
358 pollution source characteristics in a river network is presented. For this purpose, a hypothetical
359 river network consisting of a main stream (B1) and two tributaries (B2 and B3) with unsteady
360 flow conditions and irregular cross-sections has been considered. The general outline of the
361 considered river network along with the arrangement of main and secondary stations is shown
362 in Figure 5. The main and secondary stations were placed based on the criteria mentioned in
363 section 2.2. First, by assuming the activity duration of 10 hours and more for retrieval and
364 based on the calculated travel time from Equation (4), the location of main stations in all
365 branches, was determined. After that, based on the desired accuracy for spatial range,
366 secondary stations were located in the intervals of 8, 7 and 9 km for the main, second and third
367 branches, respectively.

368 A complete list of main and secondary stations of each branch, along with its distance from
369 the upstream and the travel time between two successive main stations, is given in Table 1. It
370 can be seen from Table 1 that the travel time between two successive main stations is always
371 less than or equal to the expected activity time for retrieval (10 hours). It also should be
372 mentioned that main stations which located at the beginning of branches, namely M1, M5 and
373 M7, are only used to record upstream boundary conditions of the forward flow and transport
374 model and are not used in the identification process.



375

376

Figure 4- Flowchart of identification process

377

378

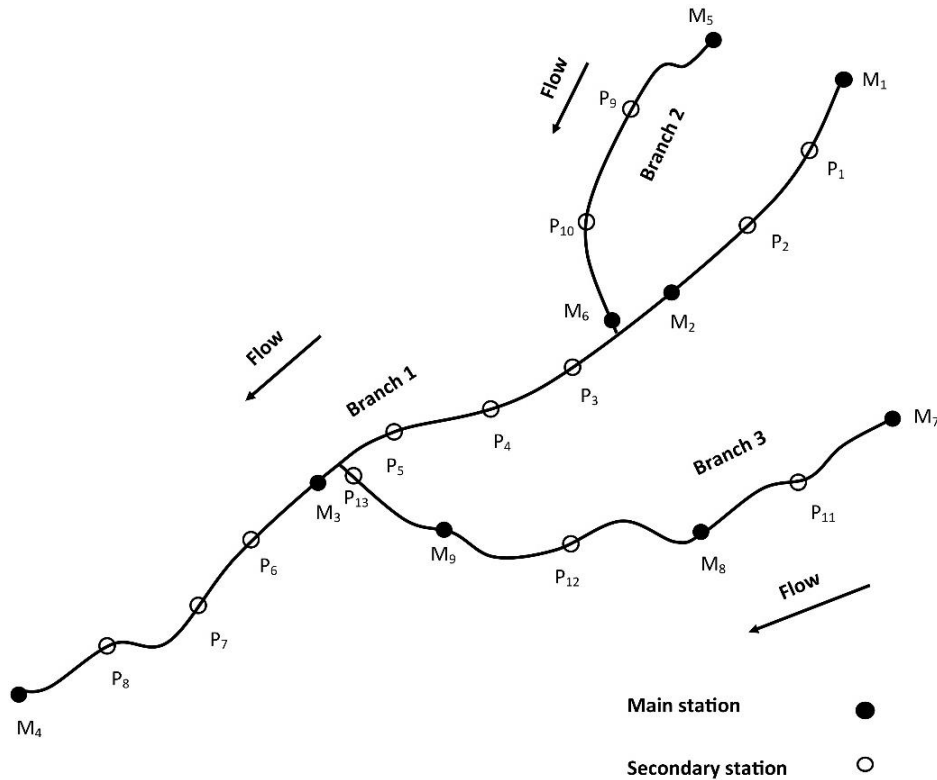
379

380

381

After placement of main and secondary stations, in order to calculate the spatial and temporal distribution of pollutant concentration in all stations, the forward flow and transport model are performed twice with give boundary condition (Figure 6). First without considering pollutant sources and then with considering them. The first set of results are used as simulated data and the second ones are used as observed data. In order to evaluate the performance of

382 proposed method three different scenarios in terms of number, release location and activity
 383 duration has been considered. The main characteristics of those scenarios were listed in Table
 384 2. Complementary explanations for each scenario are given below.

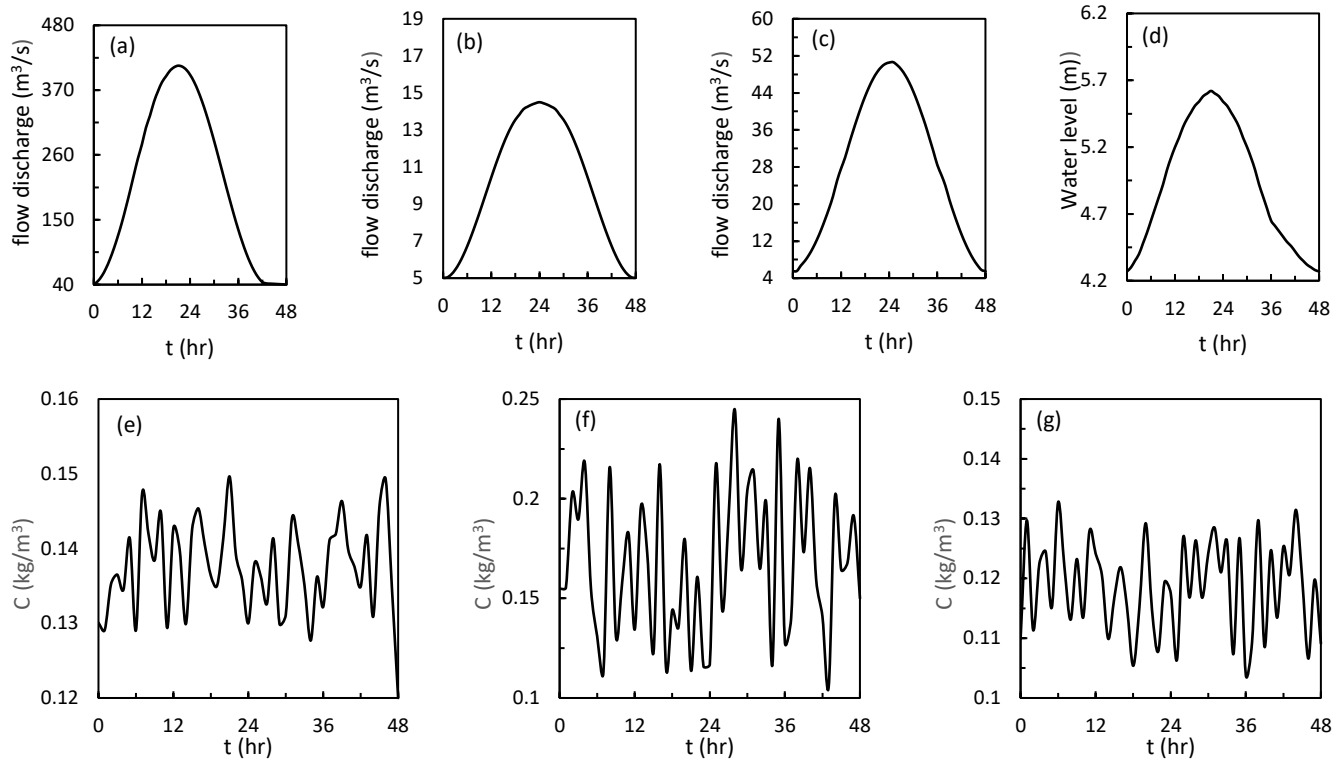


385 **Figure 5-** The schematic of hypothetical river network along with the arrangement of the main and
 386 secondary stations

387 **Table 1-** Monitoring stations

First Branch (B1)												
station	M ₁	P ₁	P ₂	M ₂	P ₃	P ₄	P ₅	M ₃	P ₆	P ₇	P ₈	M ₄
Distance from upstream (km)	0	8	16	24	32	40	48	57	65	73	81	90
Travel time (hr.)			10				9.95				9.82	
Second Branch (B2)												
station	M ₅						P ₉				P ₁₀	M ₆
Distance from upstream (km)	0						7				14	21
Travel time (hr.)							9.85					
Third Branch (B3)												
station	M ₇			P ₁₁			M ₈			P ₁₂	M ₉	P ₁₃
Distance from upstream (km)	0			9			18			27	36	45
Travel time (hr.)				9.98						10	6.12	

388



389 **Figure 6-** Flow and transport boundary conditions, (a)-(c) upstream boundary conditions of flow
 390 model at M1, M5 and M7, respectively, (d) upstream boundary conditions of flow model at M4 and
 391 (e)-(g) upstream boundary conditions of transport model at M1, M5 and M7, respectively.

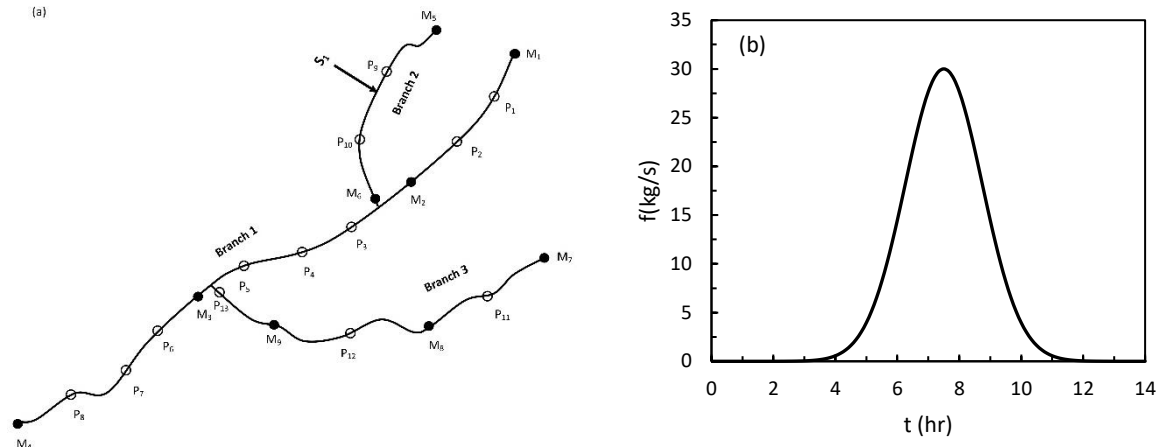
392 **Table 2-** The main characteristics of different considered scenarios

Scenario	Number of Active Sources	Release location (Branch/Distance from Upstream)	Activity Duration (hr.)	Simultaneous Active Sources
1	1	B2-8.75	11	-
2	2	B2-8.75	11	No
3- Test 1	3	B1-46	16	Yes
		B2-8.75	11	
		B1-10	13.5	
3-Test 2	3	B2-8.75	11	Yes
		B1-10	13.5	
		B1-46	16	

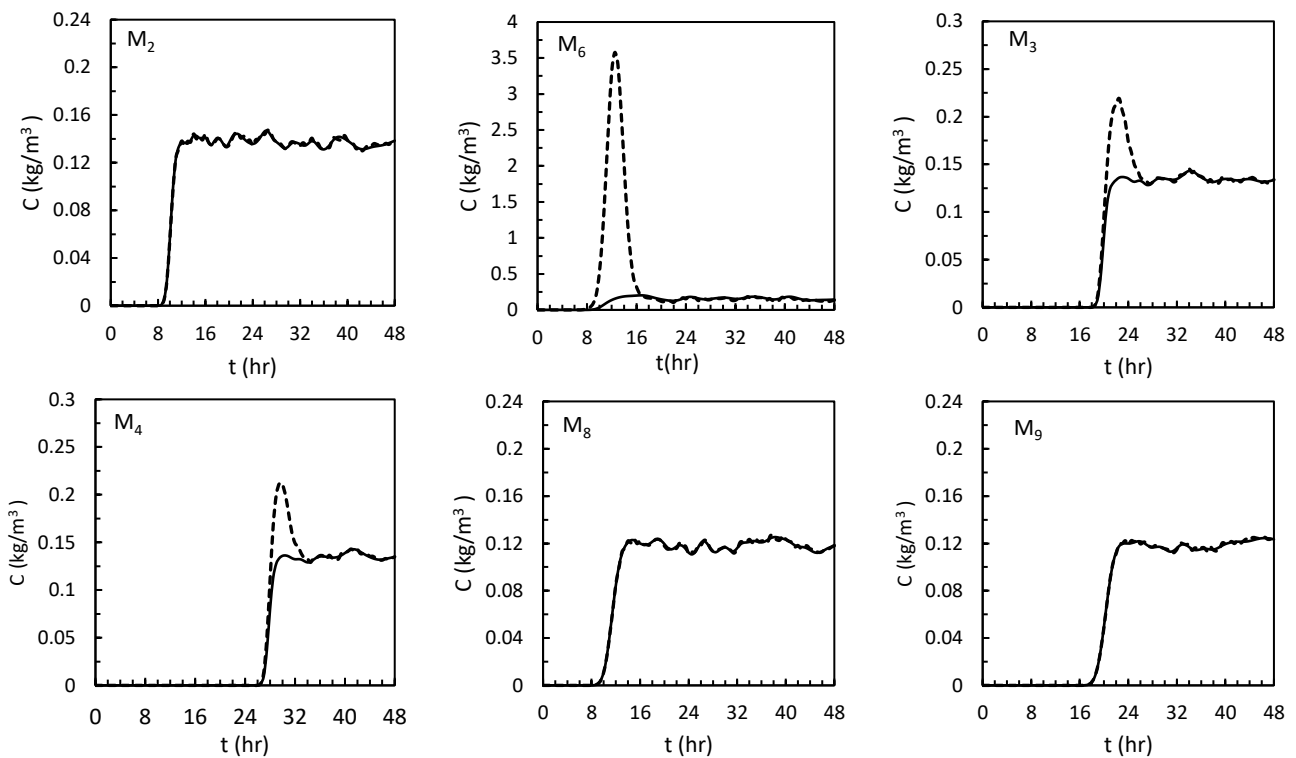
393 **3.1. Scenario 1: one active source**

394 In this example, it is assumed that there is only one active pollution source at 8.75 km of the
 395 upstream end of the second branch (B2) with activity duration of 11 hours (Figure 7). As
 396 mentioned in Section 2.2, it is necessary to first determine the suspected reach to presence of
 397 the source by comparing the observed and simulated concentration data at all main stations.
 398 Figure 8 shows a comparison of observed and simulated concentration data at main stations for
 399 this example. It should be noted that the initial period with zero concentration is due to the

400 initial condition that was chosen for the sake of simplicity, and it does not affect the results in
 401 other ways. It can be seen from Figure 8 that the first main station which recorded the difference
 402 between the observed and simulated concentration data is M6, located at 21 km of the upstream
 403 end of the second branch of the hypothetical river network.



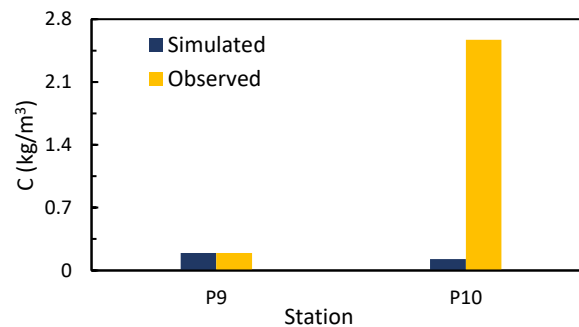
404 **Figure 7-** (a) location and (b) intensity function of pollutant source (scenario 1)



405 **Figure 8-** Comparison of simulated and observed data in main stations (scenario 1) (solid line:
 406 simulated and dashed line: observed data)

407 After detection the difference in data at M6, a concentration observation data should be
 408 collected at all secondary stations that located at upstream of M6 (namely, P9 and P10) at the
 409 instant of difference detection. Figure 9 shows the collected data at these secondary stations

410 and their comparison with the simulated data. As it can be seen from Figure 9, there is a
411 significant difference between the observed and simulated concentration data at P10, while the
412 observed and simulated data at P9 are exactly the same. Therefore, it can be concluded that the
413 release point of the pollutant source is in the reach between P9 and P10, i.e. at the range of 7
414 to 14 km of the upstream end of the B2.

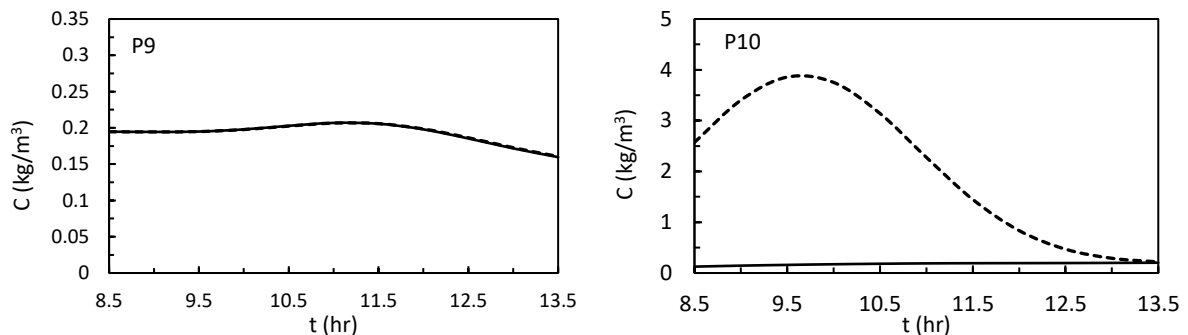


415 **Figure 9-** Comparison of simulated and observed data in secondary stations at the instant of
416 difference detection between the simulated and observed data in M6 (scenario 1)

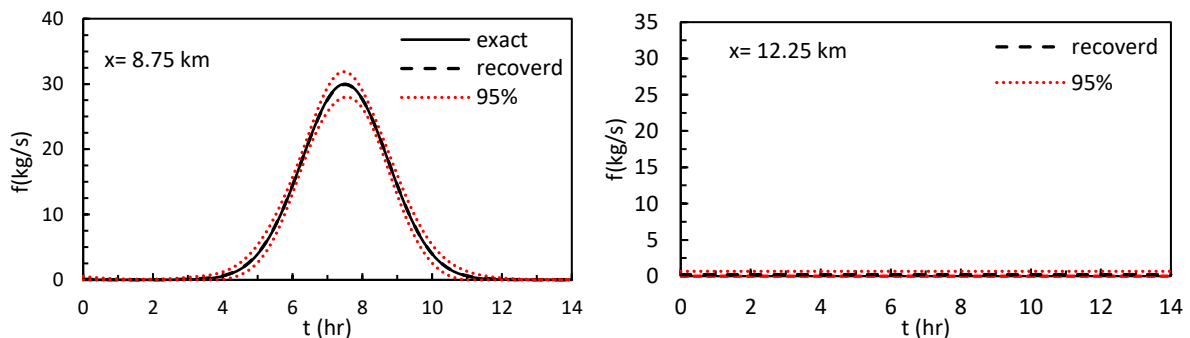
417 Subsequently, secondary stations that located at the upstream and downstream of suspected
418 reach, should begin to permanent data collection to ensure that there is no other active source.
419 Figure 10 shows a comparison of observed and simulated data at P9 and P10. As it can be seen
420 from Figure 10, there is no difference between the observed and simulated data at P9, which
421 means that there is no active source at the upstream of suspected reach, during the period of
422 activity of the discovered source. Comparison of these two series of data in P10 shows the
423 difference. According to the general form and peak concentration of observed concentration-
424 time curve and comparing it with the concentration-time curve at M6 (Figure 8), it can be
425 deduced that this difference caused by the source that just has been discovered and there are no
426 other active sources.

427 In the next step, the suspected reach to presence of pollutant source (i.e., 7 to 14 km of the
428 upstream end of the B2) is divided into two sub-reaches (namely from 7 to 10.5 km and 10.5
429 to 14 km) and the potential locations of the pollutant source are considered at the center of
430 these sub-reaches (i.e., 8.75 and 12.25 km of the upstream end of the B2). Then, by
431 implementing the inverse model and using the spatial distribution of concentration data at all

432 stations located downstream of suspected reach and at the instant of full passage of pollution
 433 cloud from P10, exact location and approximate intensity function of pollutant source are
 434 determined. Figure 11 shows the exact and recovered intensity function of the pollutant source
 435 with 95 percent confidence interval for both potential locations. As it can be seen from the
 436 Figure 11, for the case where the potential location is equal to the exact location of the assumed
 437 pollutant source (i.e. 8.75 km of the upstream end of the B2), there is a good match between
 438 the exact and recovered intensity function. While, in the case where the potential location is
 439 considered at 12.25 km of the upstream end of the B2, a close to zero amount for intensity
 440 function has been obtained.



441 **Figure 10-** Comparison of observed and simulated data in secondary stations located at the upstream
 442 and downstream of suspected reach during the period of permanent data collection by those stations
 443 (scenario 1) (solid line: simulated and dashed line: observed data)



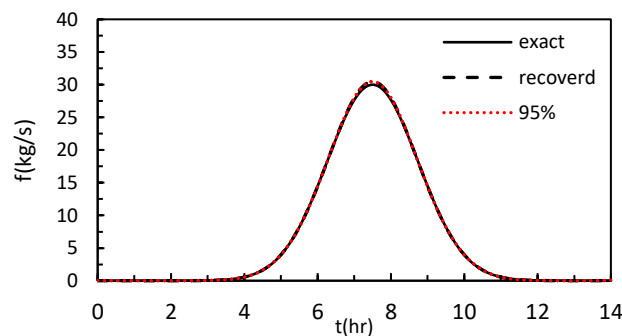
444 **Figure 11-** Recovered intensity function at two potential locations using observed data at all main and
 445 secondary stations that located downstream of the suspected reach (scenario 1)

446 After determining the exact location of the source, it is necessary to identify its intensity
 447 function more accurately, assuming that the source location is known. So, the intensity function
 448 is recovered using observed time-concentration data at the first main station at downstream
 449 (M6). The results are shown in Figure 12. The error indices for both approximate recovery

450 using the spatial distribution of concentration data and the exact recovery using time-
 451 concentration data are given in Table 3. It should be mentioned that, the indices that used to
 452 evaluate the performance of proposed method include square of correlation coefficient (R^2),
 453 root mean square error (RMSE), mean absolute error (MAE) and Euclidean distance (de). The
 454 last one, indicates the distance between the upper (σu) and lower (σl) bound of the 95%
 455 confidence interval and it is used to evaluate the uncertainty of recovered release history based
 456 on the observation data (Equation (30))

$$de = \sqrt{\sum_{i=1}^n (\sigma u_i - \sigma l_i)^2} \quad (30)$$

457 Figure 12 and the error indices in Table 3 indicate that in both cases the proposed model
 458 has been retrieved the intensity function with almost a same accuracy. The only difference is
 459 concerned with the width of 95% confidence interval, which is wider in the case of retrieval
 460 using spatial distribution of concentration data. This means that there are more release histories
 461 that consistent with the observations. This is also an indication of the increased uncertainty in
 462 estimations. The main reason for this results is sparsity of spatial distribution of concentration
 463 data compared to time-concentration data, which makes the ill-posedness issue more sever and
 464 causes more uncertainty in identification process.



465 **Figure 12-** Recovered intensity function by considering the exact location of the source and using
 466 observed concentration-time data at the first main station at downstream (M6)

467
 468
 469
 470

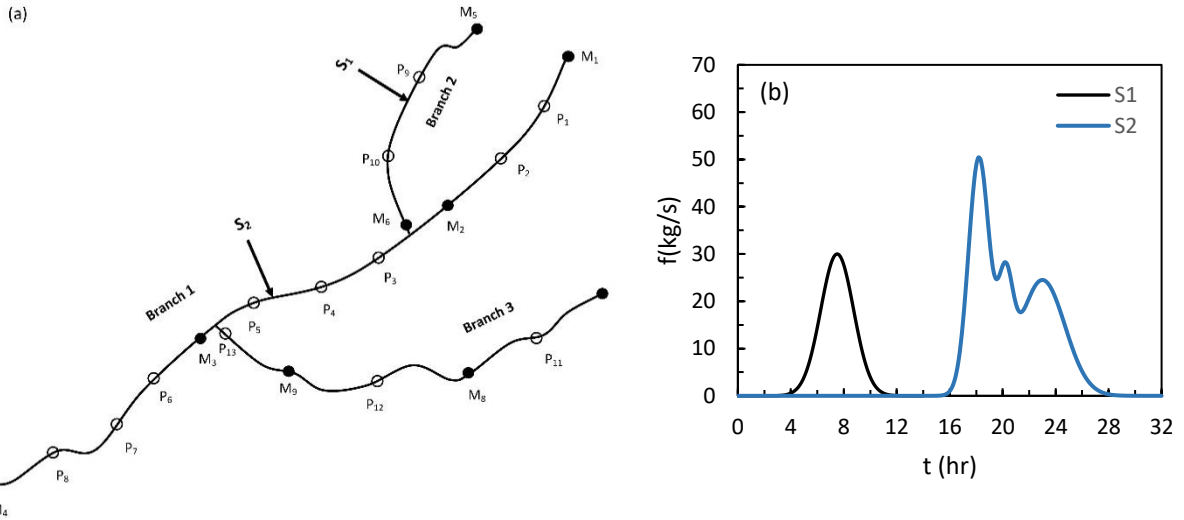
471

Table 3- Error indices of scenario 1

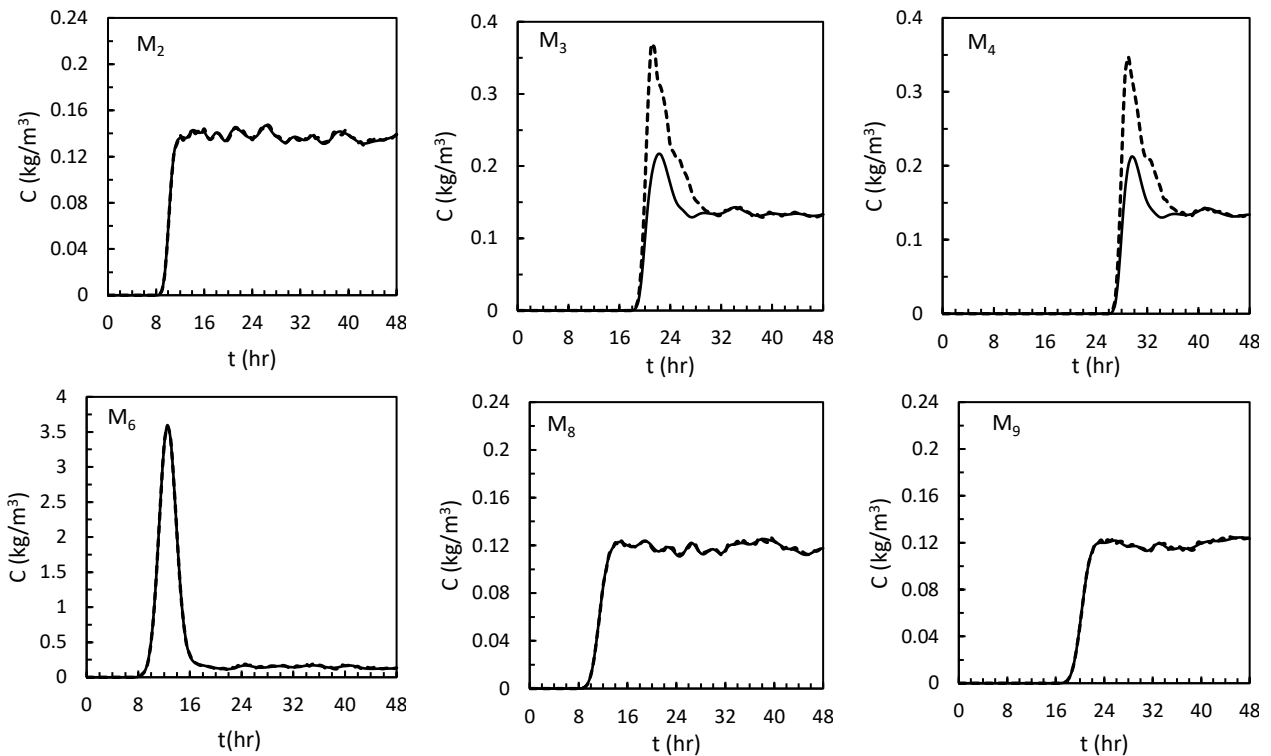
Index	Recovery using spatial distribution of concentration data	Recovery using observed concentration-time data at the first main station at downstream (M6)
R ² (%)	99.99	99.99
RMSE (kg/s)	0.05	0.04
MAE (kg/s)	0.045	0.027
de (kg/s)	9.12	0.24

472 **3.2. Scenario 2: two asynchronous active sources**

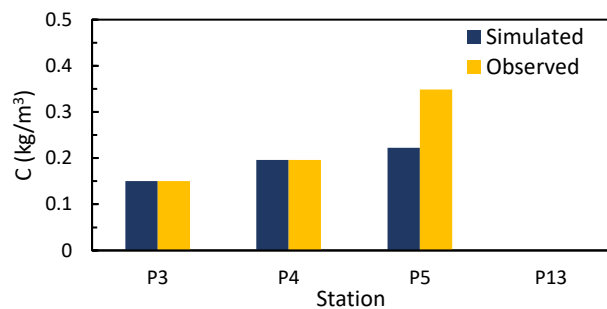
473 In this example, it was assumed that there are two active pollutant sources in the river network
474 during the simulation time, so that the start time of activity of the second source is after the end
475 of activity of the first pollutant source. The first source was considered similar to the scenario
476 one, at 8.75 km of the upstream end of the B2 and the second source assumed at 46 km of the
477 upstream end of the B1 (Figure 13). After identification of the first source, similar to the
478 scenario 1, the forward model is modified considering the identified location and release
479 history of the first source. After revising the forward model, a comparison of the observed and
480 simulated data at the main stations (Figure 14) shows that a difference between these series of
481 data at the M3 located at 57 km of the upstream end of the B1. This indicates the presence of
482 an active source at the upstream of that station. So, it is necessary to collect a concentration
483 data at the instant of recording the difference at all secondary stations located between station
484 M3 and the first main stations upstream (i.e. M2, M6 and M8). Figure 15 depicts the collected
485 data at these secondary stations and their comparison with the simulated data. As can be seen
486 from Figure 15 the only secondary station that recorded the difference between the observed
487 and simulated data is the P5 , located 48 km of the upstream end of the B1.Hence, it can be
488 said that the suspected reach to presence the second source is between P5 and the upstream
489 secondary station (P4).



490 **Figure 13-** (a) location and (b) intensity function of pollutant sources (scenario 2)

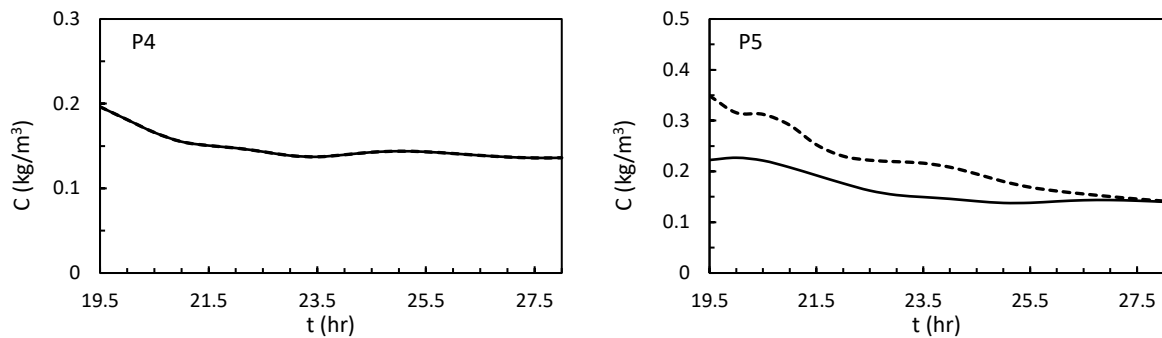


491 **Figure 14-** Comparison of simulated and observed data in main stations (scenario 2) (solid line:
492 simulated and dashed line: observed data)



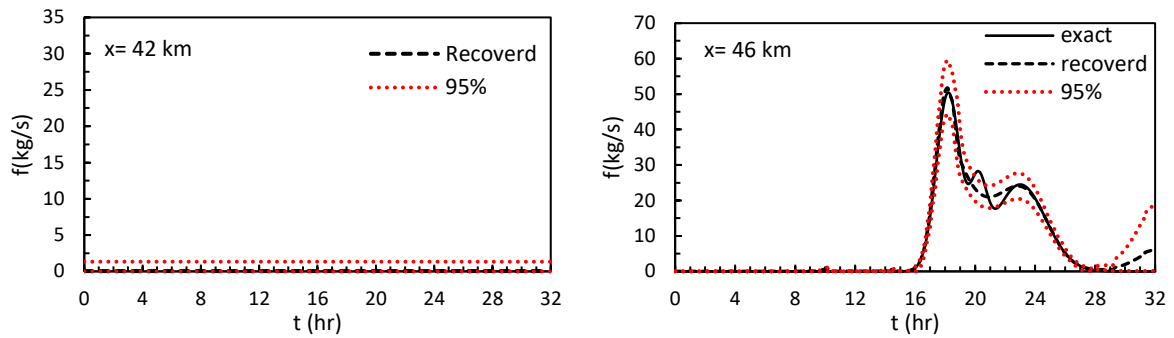
493 **Figure 15-** Comparison of simulated and observed data in secondary stations at the instant of
494 difference detection between the simulated and observed data in M3 (scenario 2)

495 Subsequently, secondary stations that are located immediately upstream and downstream
 496 of suspected reach, should begin to permanent data collection to ensure that there is no other
 497 active source. A comparison of observed and simulated data at P4 and P5 are shown in Figure
 498 16. As can be seen from Figure 16, there is no difference between the observed and simulated
 499 data at P4, which means there is no active source at the upstream during the activity time of
 500 the discovered source. A comparison of these two sets of data in the P5 represents a difference.
 501 By comparing the general form and peak concentration of observed concentration-time curve
 502 and concentration-time curve at M3 (Figure 14), it can be concluded that this difference is due
 503 to the discovered source and there are no other active pollutant sources.



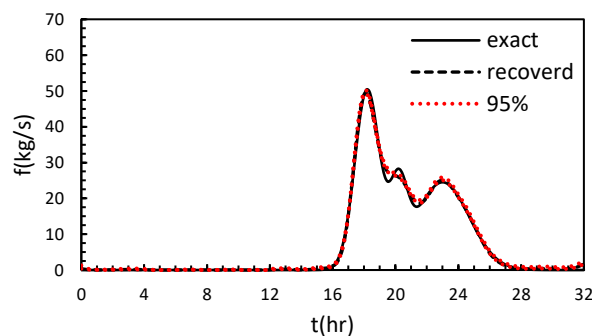
504 **Figure 16-** Comparison of observed and simulated data in secondary stations located at the upstream
 505 and downstream of suspected reach during the period of permanent data collection by those stations
 506 (scenario 2) (solid line: simulated and dashed line: observed data)

507 In the next step, the suspected reach is divided into two sub-reaches with equal length and
 508 the potential locations of the source are considered in the center of each of these sub-reaches,
 509 namely 42 and 46 km of the upstream end of the B1. Then, the exact location of S2 and its
 510 approximate intensity function are determined by implementing inverse model and using the
 511 spatial distribution of concentration data in all stations located at the downstream of suspected
 512 reach. The results are presented in Figure 17. According to these results, it can be concluded
 513 that the second source is located 46 km of the upstream end of the B1, which is corresponded
 514 to the assumed location.



515 **Figure 17-** Recovered intensity function at two potential locations using observed data at all main and
 516 secondary stations that located downstream of the suspected reach (scenario 2)

517 Finally, by assuming the known source location and using concentration-time data at M3,
 518 the intensity function is determined more accurately. The results are shown in Figure 18. The
 519 error indices for both approximate recovery using the spatial distribution of concentration data
 520 and the exact recovery using time-concentration data are given in Table 4. Figure 18 and the
 521 error indices in Table 4, suggested that the accuracy of the results obtained using the
 522 concentration-time data is slightly better than the accuracy of the results obtained using the
 523 spatial distribution of concentration data. In addition, the 95% confidence interval opening is
 524 wider at the case of recovery with spatial distribution of concentration data, which is interpreted
 525 as more uncertainty in results. Given that the spatial distribution of concentration data are
 526 usually sparse and the number of available data is much less than the desired temporal instants
 527 for retrieval of intensity function, the existence of a higher degree of uncertainty in the results
 528 is inevitable.



529 **Figure 18-** Recovered intensity function by considering the exact location of the source and using
 530 observed concentration-time data at the first main station at downstream (M3)

531
 532

533

Table 4- Error indices of scenario 2

Index	Recovery using spatial distribution of concentration data	Recovery using observed concentration-time data at the first main station at downstream (M3)
R ² (%)	98.32	99.68
RMSE (kg/s)	1.64	0.6975
MAE (kg/s)	0.7996	0.3823
de (kg/s)	81.2366	15.2694

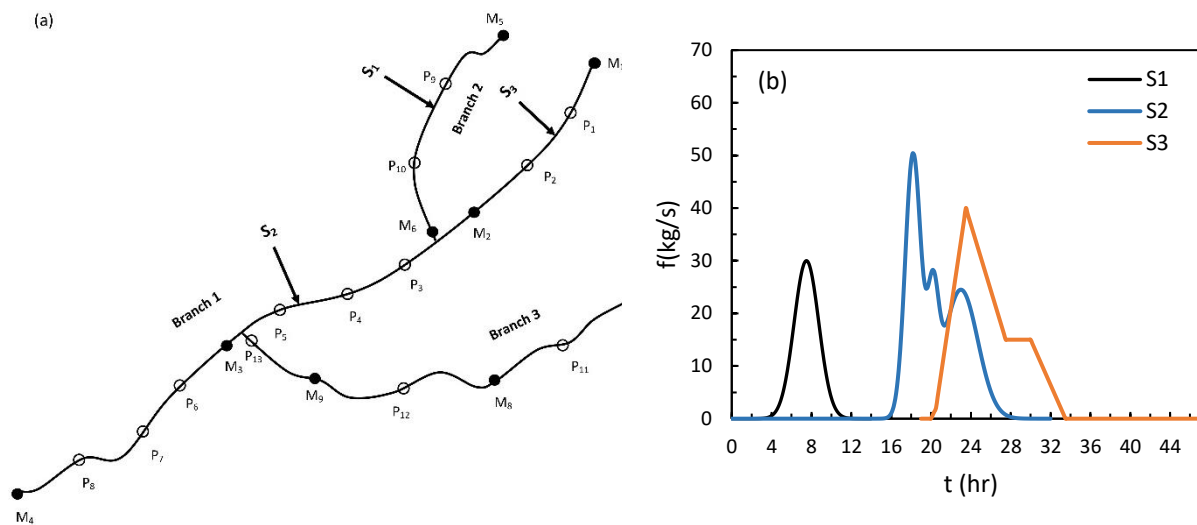
534 **3.3. Scenario 3: three active sources, with at least two simultaneously active** 535 **sources**

536 In order to show the capabilities of the proposed model in the case where several sources are
537 simultaneously active, this example considered the identification of three sources that a part of
538 the activity time of two of those sources coincide. The first source similar to the scenario 1 has
539 been considered at 8.75 km of the upstream end of the B2 and the other two sources considered
540 at 10 and 46 km of the upstream end of the B1. It is also assumed that the activity time of the
541 last two sources is after the end of the activity of the first source. In addition, it assumed that
542 part of the activity time of the sources that located at 10 and 46 km of the upstream end of the
543 B1 is simultaneous. This example is presented for two different cases in terms of the start
544 activity time of pollutant sources. Complementary explanations for each case are given below.

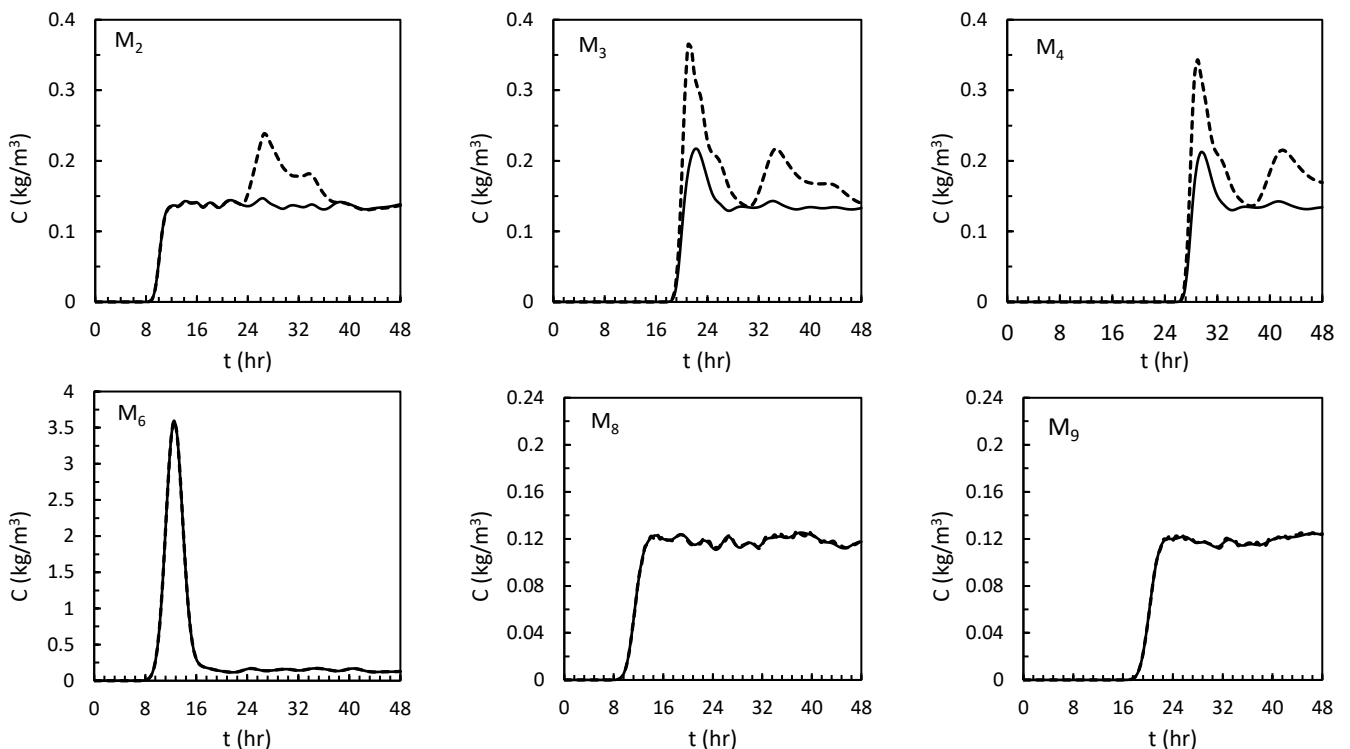
545 **a) Test 1**

546 In the first case, it is assumed that the source at 46 km of the upstream end of the B1 starts its
547 activity earlier than the source at 10 km of the upstream end of the B1 (Figure 19). After
548 identification of the first source, similar to what described in the scenario 1, the forward model
549 is modified according to recovered source characteristics. After revising the forward model, a
550 comparison of observed and simulated data at the main stations (Figure 20), first shows a
551 difference between these two set of data at the M3 located at 57 km of the upstream end of the
552 B1. A few hours later, while the pollution cloud has not yet completely passed the M3, a
553 difference between the observed and simulated data at the M2 at 24 km of the upstream end of
554 the B1, is recognized. This means two sources are simultaneously active at upstream of these

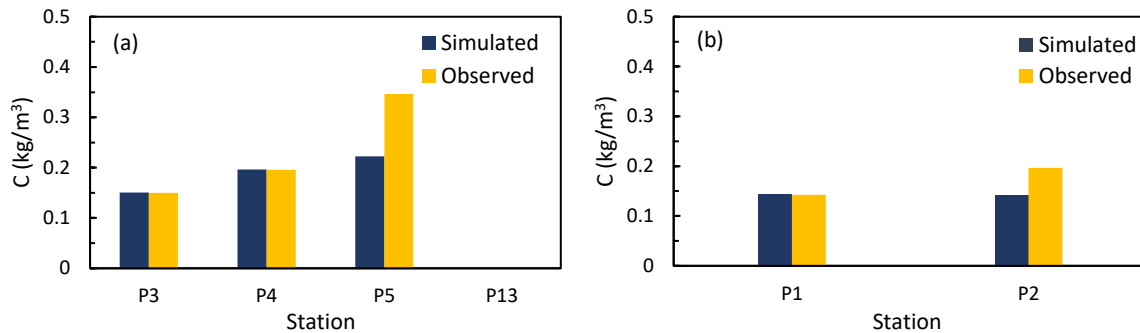
555 two main stations. In order to correctly identify the suspect reaches to presence of these two
 556 sources, it is necessary to collect a concentration data at the instant of difference detection at
 557 all secondary stations between station M3 and M2 and the first main stations that located at
 558 upstream of them. Figure 21 (a) and (b) represent a comparison of observed and simulated data
 559 at sought secondary stations and at the instant of difference detection in M3 and M2,
 560 respectively.



561 **Figure 19-** (a) location and (b) intensity function of pollutant sources (scenario 3-test 1)



562 **Figure 20-** Comparison of simulated and observed data in main stations (scenario 3-test1) (solid line:
 563 simulated and dashed line: observed data)

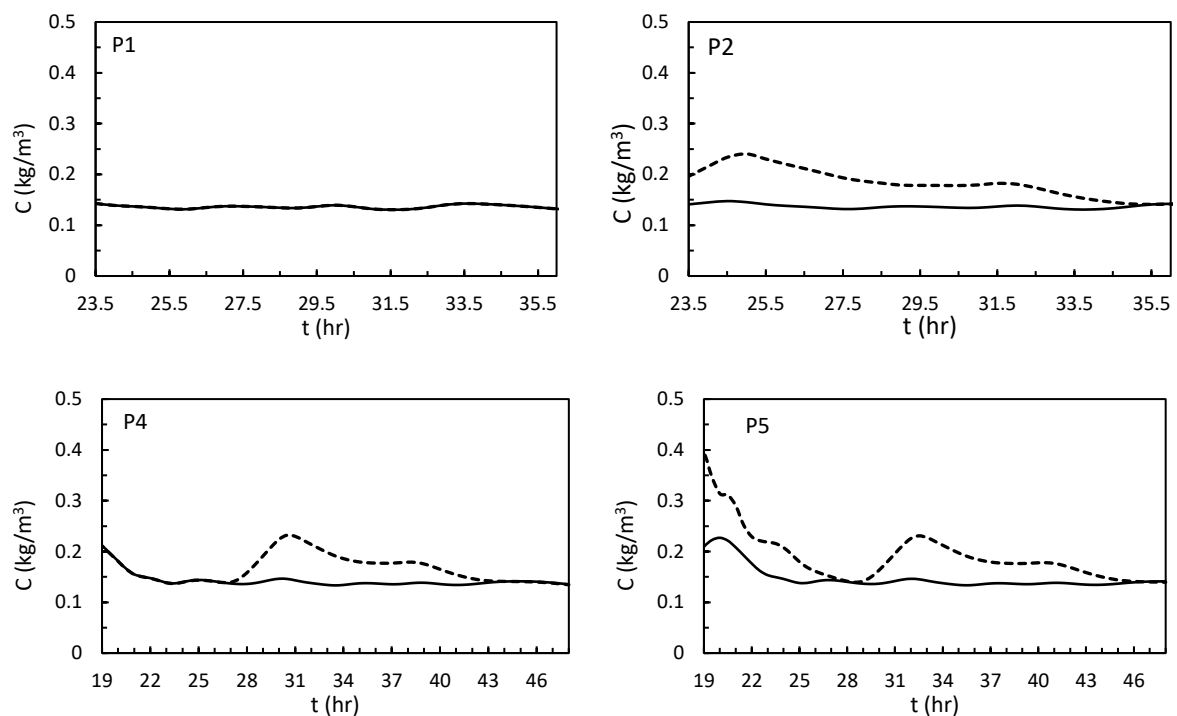


564 **Figure 21-** Comparison of simulated and observed data in secondary stations at the instant of
 565 difference detection between the simulated and observed data in (a) M3 and (b) M2 (scenario 3-test1)

566 In addition, to ensure that there are no other active sources, the concentration-time data that
 567 has been collected at secondary stations at the upstream and downstream of suspected reaches
 568 are compared with simulated data. It should be noted that the beginning instant of data
 569 collection is from the instant of difference detection at M3 and M2. Figure 22 shows a
 570 comparison of observed and simulated data at secondary stations P1, P2, P4 and P5. As can be
 571 seen from Figure 22, there is no difference between observed and simulated data at P1, which
 572 means that there is no active source at upstream of that station during the activity of detected
 573 source. A comparison of these two sets of data in P2 shows a difference. By comparing the
 574 general form and peak concentration of C-t curve with C-t curve at M2 (Figure 20), it can be
 575 deduced that this difference is due to the discovered source and there is no other active source.
 576 Observed and simulated data at P4 and P5 also show difference. By a similar argument, it can
 577 be concluded that this difference is due to the discovered sources and that there is no other
 578 active source at upstream of these stations.

579 After determining the suspected reaches to presence of two sources, their exact location and
 580 approximate intensity function are recovered using the spatial distribution of concentration data
 581 in all downstream stations. Given that the source which located at 40 to 46 km of the upstream
 582 end of the B1 has started its activity earlier, its exact location must be determined first. It
 583 should be noted that this case is fundamentally different from the two previous two scenarios.
 584 In the two previous scenarios, the spatial distribution of concentration data which used to
 585 determine the exact location and approximate intensity function had been collected at the

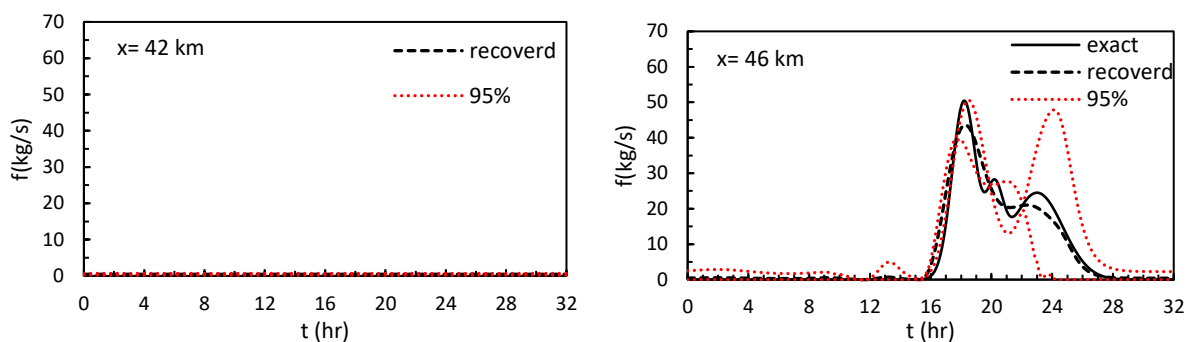
586 instant of full passage of the pollution cloud from downstream secondary station. However, in
 587 this test, due to the simultaneous activity of two pollutant sources, the exact location and
 588 intensity function of the second pollutant source are determined using the spatial distribution
 589 of concentration data at the instant of discovering the effect of third source. This is because the
 590 observed data at the instant the full passage of pollutant cloud from the downstream secondary
 591 station represented the combined effects of two sources, and therefore using of them may lead
 592 to incorrect identification results. While at the instant of detection third source, its effect has
 593 not yet reached the downstream, and the data that has been recorded at downstream main and
 594 secondary stations shows only the effect of second source.



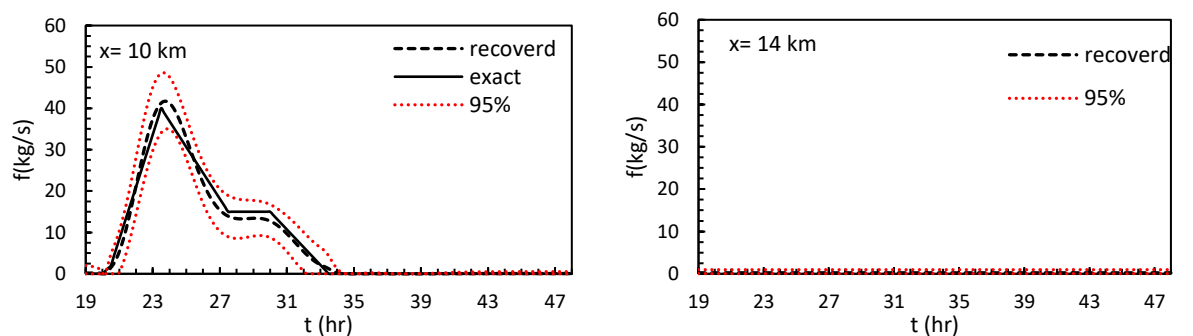
595 **Figure 22-** Comparison of observed and simulated data at secondary stations located at the upstream
 596 and downstream of suspected reaches during the period of permanent data collection by those
 597 stations (scenario 3-test 1) (solid line: simulated and dashed line: observed data)

598 The identification process is started by dividing the suspected reaches to presence of second
 599 and third sources into two equal length sub-reaches. Then, potential locations of the pollutant
 600 sources are considered in the center of those sub-reaches and by implementing the inverse
 601 model the exact location and approximate intensity function of each source is determined.
 602 Figure 23 shows the results of inverse model implementation for two potential locations for

603 second source, i.e. 42 and 46 km of the upstream end of the B1. As can be seen from it, a close
 604 to zero and a non-zero intensity functions have been obtained for 42 and 46 km potential
 605 locations, respectively. Therefore, it can be concluded that the second source of is located at
 606 46 km of the upstream end of the B1, which corresponds to the assumed location. Subsequently,
 607 the location of third source is also determined using the spatial distribution of concentration
 608 data at the instant that pollutant cloud fully passes from P2. The results of the inverse model
 609 implementation for the two potential locations, i.e. 10 and 14 km of the upstream end of the
 610 B1, are shown in Figure 24. As indicated in this figure, a non-zero intensity function is obtained
 611 for the potential location of 10 km. So, it can be concluded that the third source is released at
 612 10 km of the upstream end of the B1, which corresponds to the assumed location.



613 **Figure 23-** Recovered intensity function of S_2 at two potential locations using observed data at all
 614 main and secondary stations that located downstream of the suspected reach at the instant of recording
 615 the difference between the simulated and observed data in M2 (scenario 3-test1)

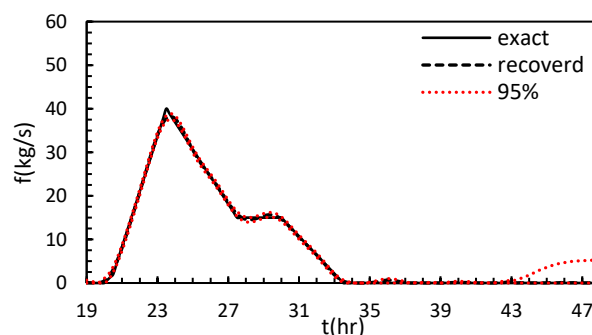


616 **Figure 24-** Recovered intensity function of S_3 at two potential locations using observed data at all
 617 main and secondary stations that located downstream of the suspected reach at the instant that the
 618 pollution cloud completely passes the P2 (scenario 3-test1)

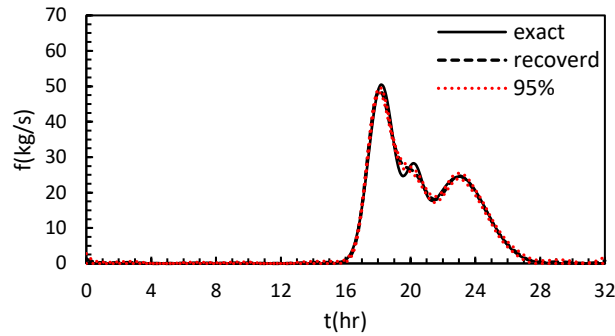
619 Once the location of pollutant sources has been determined, their intensity functions should
 620 be recovered more accurately, assuming known source locations. Due to the simultaneous
 621 activity of two pollutant sources in this case, by starting from upstream, at first the exact

622 intensity function of the third pollutant source (located 10 km from upstream of B1) is retrieved
623 using the concentration time data at M2. Then the forward model is modified, considering the
624 obtained characteristics of this source. Thus, the C-t observed data at M3 will only include the
625 effect of the second pollutant source ((located 46 km from upstream of B1)), and the exact
626 intensity function of this source can also be calculated.

627 The results of the recovery of the third source intensity function using the C-t observed data
628 at M2 are shown in Figure 25. Figure 26 shows the results of exact recovery of the intensity
629 function of second source using the C-t observed data at M3 after deducting the effect of third
630 source. The error indices for both approximate and exact recovery of the third source intensity
631 function are given in Table 5. As can be seen from Figure 25 and Figure 26 and the error indices
632 of Table 5, the accuracy of the results obtained using the c-t data is slightly better than the
633 accuracy of the results obtained using the spatial distribution of concentration data. In addition,
634 the 95% confidence interval width is narrower for the case of exact recovery, which indicates
635 less uncertainty in obtained results in this case. The main reason for this is the difference in the
636 number of observational data in these two cases. Since the spatial distribution of concentration
637 data is usually sparse and the number of available data is much less than the number of desired
638 instant for recovery of intensity function, the degree of uncertainty in retrieved results
639 increases.



640 **Figure 25-** Recovered intensity function of S_3 by considering the exact location of the source and
641 using observed concentration-time data at the first main station at downstream (M2)



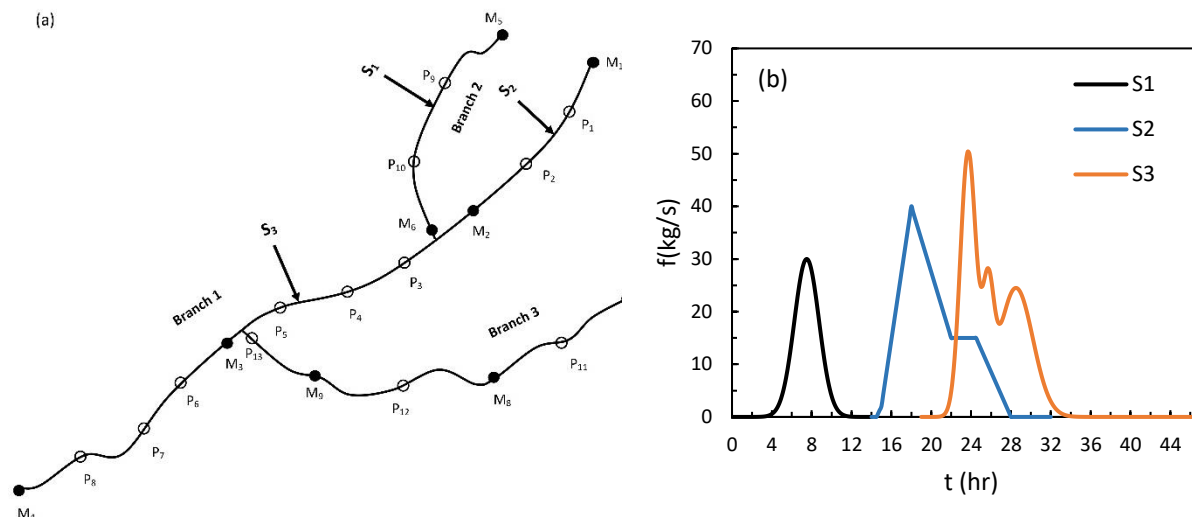
642 **Figure 26-** Recovered intensity function of S_2 by considering the exact location of the source and
 643 using observed concentration-time data at the first main station at downstream (M3) and after revising
 644 the forward model

645 **Table 5-** Error indices of scenario 3- test 1

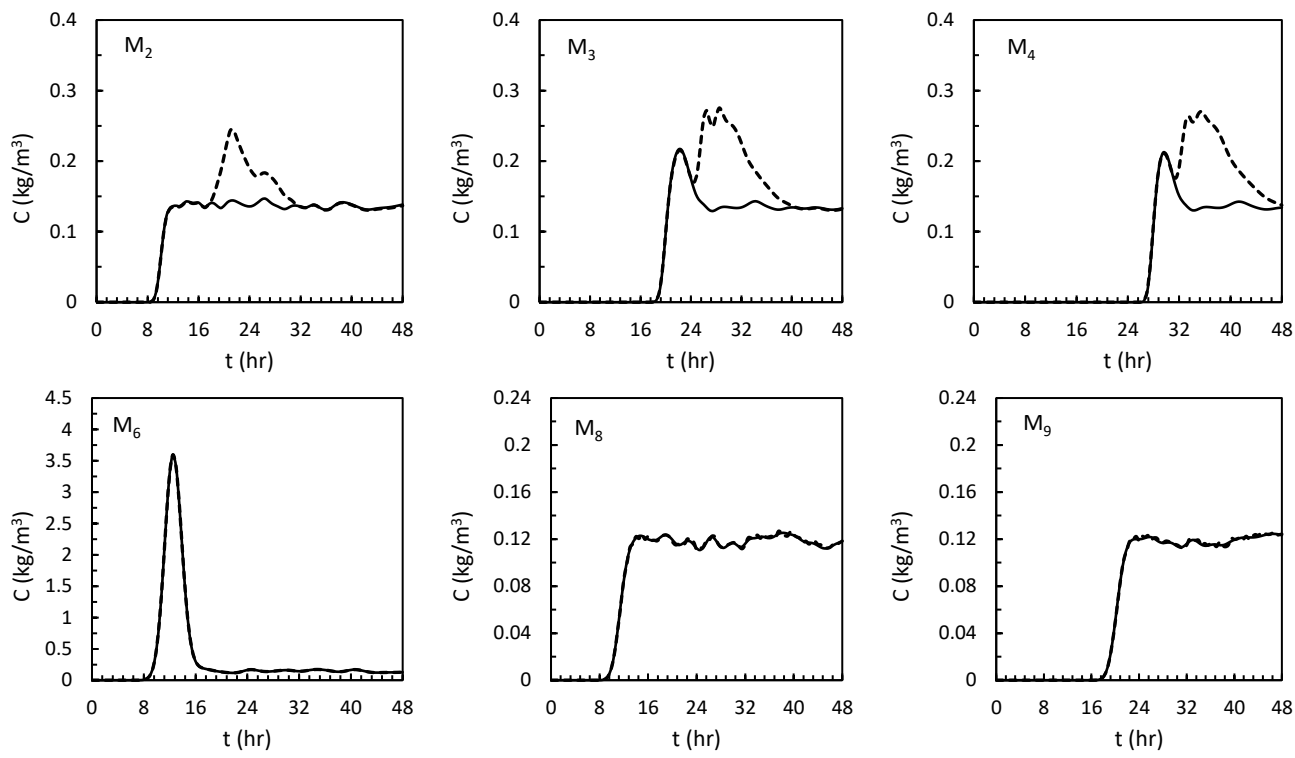
Index	S_2 (46B1)		S_3 (10B1)	
	Recovery using spatial distribution of concentration data	Recovery using observed concentration-time data at the first main station at downstream (M3)	Recovery using spatial distribution of concentration data	Recovery using observed concentration-time data at the first main station at downstream (M2)
R^2 (%)	95.99	99.64	98.55	99.96
RMSE (kg/s)	2.4674	0.7477	1.5034	0.3585
MAE (kg/s)	1.4293	0.3975	0.9168	0.2196
de (kg/s)	184.7864	14.9453	92.1065	24.5408

646 **b) Test 2**

647 In the second case, it is assumed that the source at 10 km of the upstream end of the B1 starts
 648 its activity earlier than the source at 46 km o the upstream end of the B1(Figure 27), which
 649 creates a different condition in identification process than the first test. After identification of
 650 the first source, similar to what described in the scenario 1, the forward model is modified
 651 according to the recovered source characteristics. After revising the forward model, a
 652 comparison of observed and simulated data at the main stations shows a difference between
 653 these two set of data at the M2 located at 24 km of the upstream end of the B1 (Figure 28).



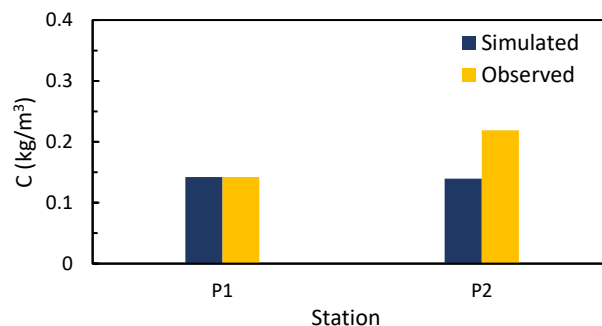
654 **Figure 27-** (a) location and (b) intensity function of pollutant sources (scenario 3-test 2)



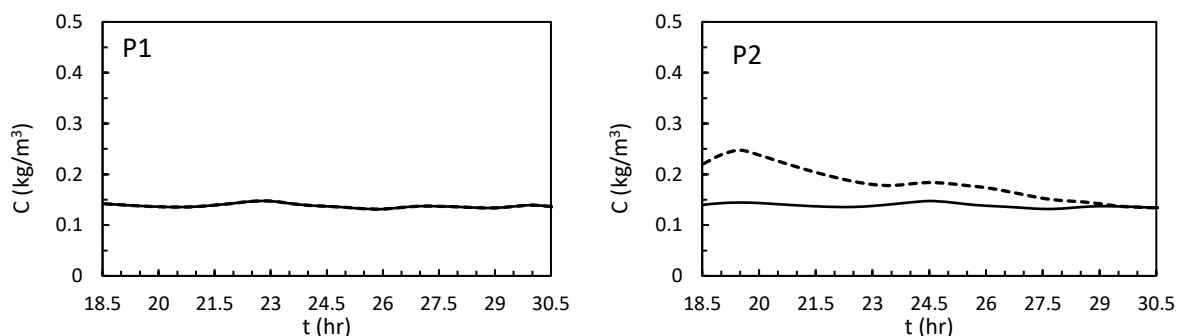
655 **Figure 28-** Comparison of simulated and observed data in main stations (scenario 3-test2) (solid line:
 656 simulated and dashed line: observed data)

657 Once the difference between observed and simulated data sets was detected , it is necessary
 658 to collect a concentration data at the instant of difference detection at all secondary stations
 659 located between station M2 and the first main station at upstream (namely P1 and P2), and
 660 compare those data with corresponding simulated data (Figure 29). According the Figure 29 It
 661 can be deduced that the suspected reach to presence the second source is in between P1 and P2
 662 (i.e. at 8 to 16 km of the upstream end of the B1). Also, in order to ensure that there are no

663 other simultaneously active sources upstream and downstream of the suspected reach, the
 664 observed and simulated data are compared at P1 and P2 during the permanent data recording
 665 period by these stations (Figure 30). As shown in Figure 30, there is no difference between the
 666 observed and simulated data at P1, which means that there is no active source upstream of
 667 suspected reach during the detection period. However, a comparison of these two sets of data
 668 in P2 shows a difference. Regarding the general form and peak concentration of observed C-t
 669 curve with the corresponding one at M2 (Figure 28), it can be inferred that this difference is
 670 due to the discovered pollutant source and there is no other active pollutant source.



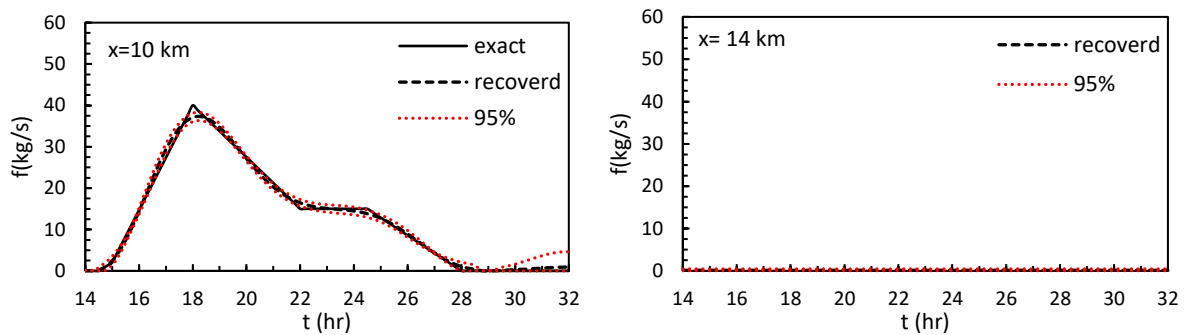
671 **Figure 29-** Comparison of simulated and observed data in secondary stations at the moment of
 672 difference detection between the simulated and observed data in M2 (scenario 3-test2)



673 **Figure 30-** Comparison of observed and simulated data in secondary stations located at the upstream
 674 and downstream of suspected reach during the period of permanent data collection by those
 675 stations (scenario 3-test 2) (solid line: simulated and dashed line: observed data)

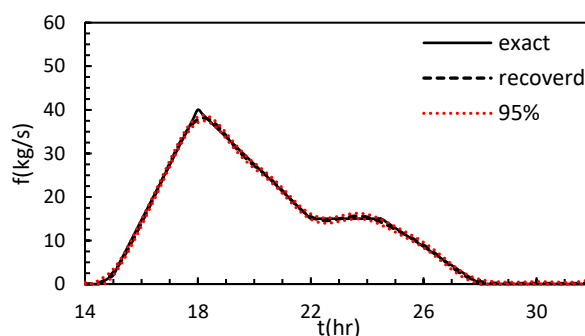
676 In the next step, the determined suspected reach is divided into two equal length sub-reaches
 677 of and the potential locations of the pollutant source is considered in the center of each of these
 678 sub-reaches, i.e. 10 and 14 km of the upstream end of the B1. Then, the inverse model is
 679 implemented using spatial distribution of concentration data at all station located at
 680 downstream of the suspected reach. The results of the inverse model implementation for both

681 potential locations are presented in Figure 31 .As indicated in figure, for a potential location of
 682 14 km the intensity function is obtained close to zero, while for a potential location of 10 km a
 683 non-zero intensity is obtained. Therefore, it can be concluded that the second pollutant source
 684 is located at 10 km from upstream of the B1, which corresponds to the assumed location.



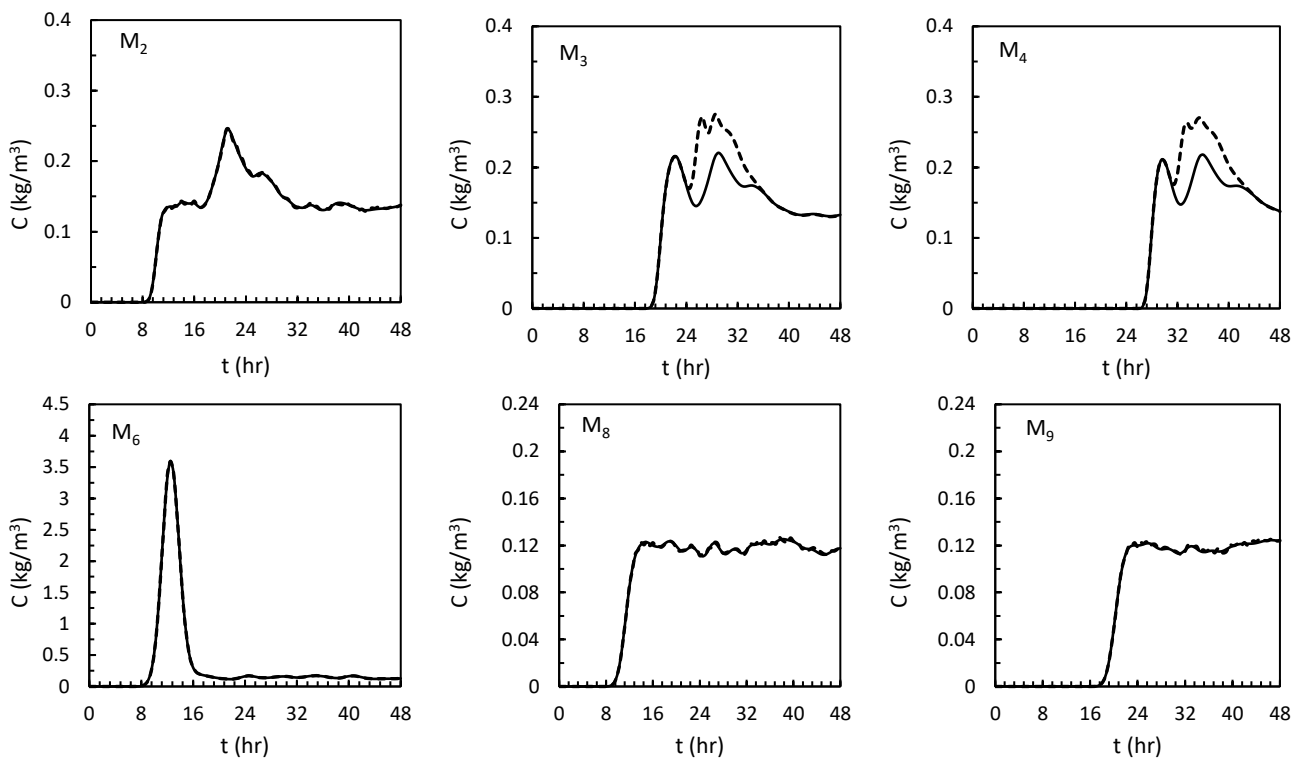
685 **Figure 31-** Recovered intensity function of S_2 at two potential locations using observed data at all
 686 main and secondary stations that located downstream of the suspected reach at the instant that the
 687 pollution cloud completely passes the P2 (scenario 3-test2)

688 After determining the exact location of the source, its intensity function are recovered more
 689 accurately, assuming the source location is known and using observed C-t data at M2 (Figure
 690 32). The error indices for both approximate and exact recovery of the intensity function for
 691 second source are given in Table 6. As shown in Figure 32 and the error indices in Table 6, the
 692 accuracy of the results obtained using the C-t data is slightly better than the accuracy of the
 693 results obtained using the concentration spatial series data and the 95% confidence interval
 694 width is narrower as well. So, uncertainty associated with retrieved results are less in this case.
 695 The main reason for this is availability of more observation data compare to the case of
 696 recovery with spatial distribution of concentration data.

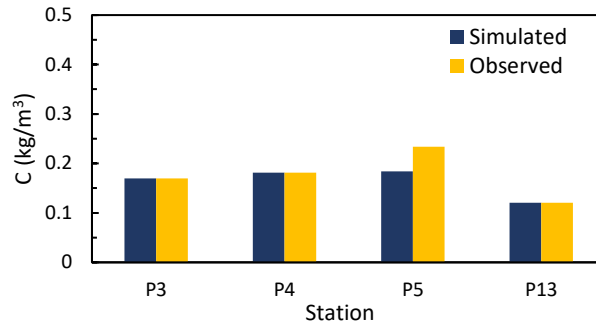


697 **Figure 32-** Recovered intensity function of S_2 by considering the exact location of the source and
 698 using observed concentration-time data at the first main station at downstream (M2)

699 After identification the characteristics of the second pollutant source, the forward model is
 700 modified according to determined characteristics and the observed and simulated data that
 701 obtained by modified forward model are compared. Comparison of these two sets of data
 702 indicates the existence of difference at the M3 (Figure 33). Therefore, it can be concluded that
 703 a pollution source is active upstream of this station. By comparing the concentration data at the
 704 instant of difference detection in all secondary stations that located between the M3 and the
 705 first main station at upstream (M2) (Figure 34), the suspected reach to presence the third
 706 pollutant source is determined between 40 to 48 km of the upstream end of the B1.

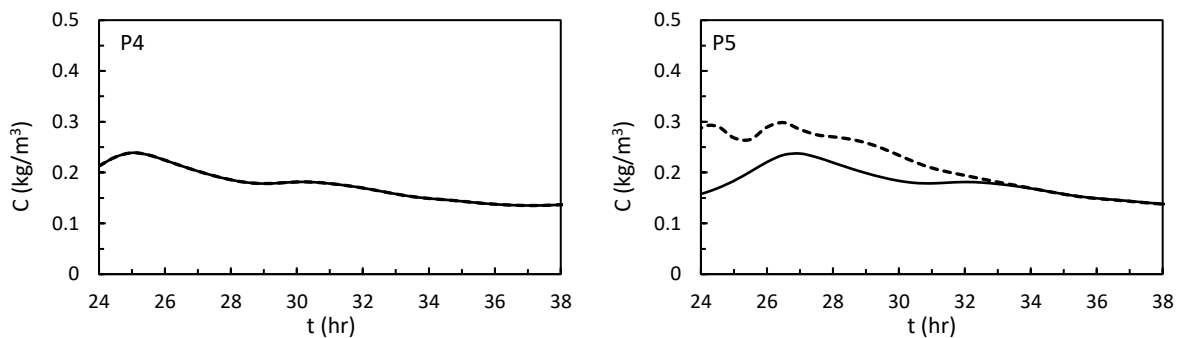


707 **Figure 33-** Comparison of simulated and observed data in main stations (scenario 3-test2) after
 708 identification of S_2 and revising the forward model (solid line: simulated and dashed line: observed
 709 data)
 710
 711



712 **Figure 34-** Comparison of simulated and observed data in secondary stations after identification of S_2
 713 and at the instant of difference detection between the simulated and observed data in M3 (scenario 3-
 714 test2)

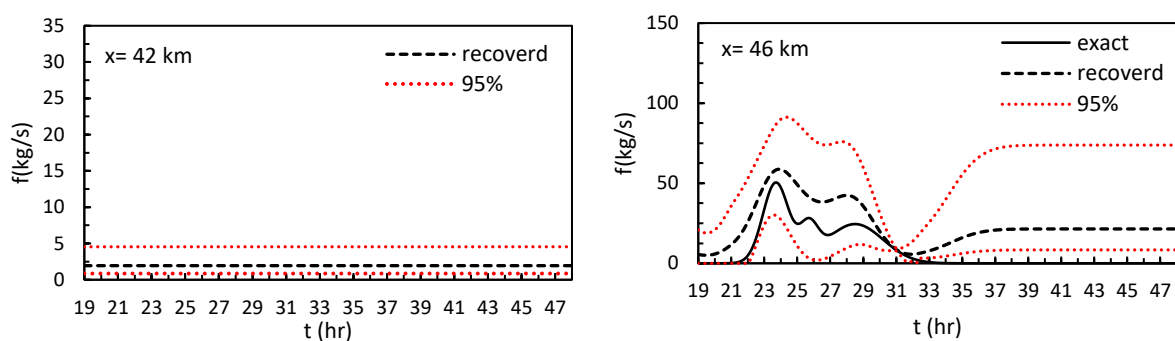
715 In order to ensure that there are no other simultaneously active sources at upstream and
 716 downstream of the suspected reach, the observed and simulated data at P4 and P5 secondary
 717 stations are compared during the permanent data recording period by these stations(Figure 35).
 718 As can be seen from Figure 35 there is no difference between the observed and simulated data
 719 at P4, which means that there are no other active sources during the identification period. A
 720 comparison of these two sets of data in the P5 shows the difference. By comparing the general
 721 form and peak concentrations of Observed C-t curve with the associated one at M3 (Figure 33),
 722 it can be argued that this difference is due to the discovered contaminant source and there are
 723 no other active sources.



724 **Figure 35-** Comparison of observed and simulated data in secondary stations located at the upstream
 725 and downstream of suspected reach to presence of S_3 during the period of permanent data collection
 726 by those stations (scenario 3-test 2) (solid line: simulated and dashed line: observed data)

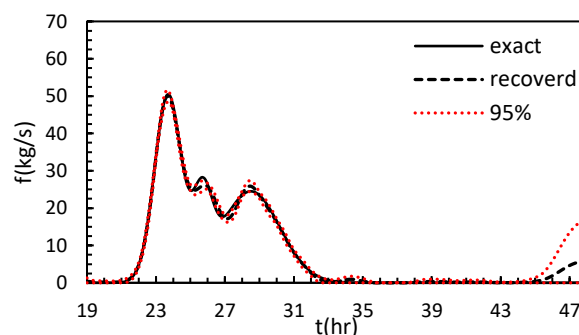
727 In the next step, the exact location and approximate intensity function of the third source
 728 are recovered by implementation inverse model for two potential source locations which are in
 729 the center of two equal length sub-reaches, i.e. 42 and 46 km of the upstream end of the B1.
 730 The results have been presented in Figure 36 that shows that the location of the third source

731 must be at 46 km of the upstream end of the B1, for which a non-zero intensity function had
 732 been obtained. However, as shown in this figure, there is no good match between the recovered
 733 and the exact intensity function. The reason for this is the time delay in identifying the effect
 734 of the third pollutant source at M3, due to the synchronization of its activity with the second
 735 pollutant source. As a result, some part of information about the third source of the pollutant is
 736 lost and consequently retrieval accuracy had been reduced and associated uncertainty
 737 increased.



738 **Figure 36-** Recovered intensity function of S_3 at two potential locations using observed data at all
 739 main and secondary stations that located downstream of the suspected reach at the instant of recording
 740 the difference between the simulated and observed data in M3 (scenario 3-test2)

741 After determining the exact location of the third pollutant source, its intensity function is
 742 retrieved more accurately using the C-t data at M3 (Figure 37). As it is clear from the figure,
 743 the model has succeeded in recovering the intensity function of the mentioned pollutant source
 744 with good and acceptable accuracy. The error indices presented in Table 6 also confirm this.



745 **Figure 37-** Recovered intensity function of S_3 by considering the exact location of the source and
 746 using observed concentration-time data at the first main station at downstream (M3) and after revising
 747 the forward model

748
 749
 750

751

Table 6- Error indices of scenario 3- test 2

Index	S ₂ (10B1)		S ₃ (46B1)	
	Recovery using spatial distribution of concentration data	Recovery using observed concentration-time data at the first main station at downstream (M2)	Recovery using spatial distribution of concentration data	Recovery using observed concentration-time data at the first main station at downstream (M3)
R ² (%)	99.61	99.88	74.0750	98.92
RMSE (kg/s)	0.7702	0.4162	17.4011	1.3823
MAE (kg/s)	0.5797	0.3057	15.9637	0.7126
de (kg/s)	20.8116	11.716	741.1620	52.0625

752 **4. Conclusion**

753 This study has been presented an innovative multistep method for simultaneous identification
754 of the number, location and release history of pollutant source in a river network considering
755 unsteady and non-uniform flow. The only priori information that the method needs are the
756 expected activity period for recovery, accuracy of spatial range for retrieval the source location
757 and the travel time of each branch. Based on those priori information, at first an adaptive
758 arrangement of observation points is proposed. Then suspect reaches to presence of pollutant
759 sources are delineate by comparing the simulated and observed breakthrough curves at
760 considered stations. In this step, the number of all simultaneous active pollution sources is also
761 determined. Then, the suspected reaches are divided to some sub-reaches and it is assumed that
762 the origin of possible sources is in the center of those sub- reaches. At the second step the
763 location and approximate release history of pollution sources are recovered by means of a
764 geostatistical approach, that considered simultaneously all the possible candidates. The source
765 location is considered as the location where the highest amount of released pollutant is
766 estimated. Finally, the exact release history is determined using the temporal distribution of
767 observed concentration data at the first downstream main station.

768 The proposed method is suitable for practical applications, since it is based on one-
769 dimensional flow and transport models and considers the complicated real-world conditions.
770 The method is effective and easy to apply in complex river networks as well as single-branch

771 ones. Moreover, since in each simulation it is possible to identify all active pollutant sources,
772 the required computational time is significantly lower than common iterative methods such as
773 simulation-optimization approach. Another significant advantage of the proposed method is
774 that it provides unique results for sought characteristics, using minimum observational data. In
775 fact, if the observation points placed based on suggested pattern, obtaining the unique results
776 is guaranteed. The results of application of method to a hypothetical river network for different
777 scenarios in terms of the number, release time and location of pollutant sources, showed
778 that the methodology performs very well in case of large-scale river networks. The given
779 results were acceptable regarding to a limited requirement inputs. Of course, the quality of the
780 recovery is dependent on the accuracy of the observation data. So, the uncertainty associated
781 with results due to using erroneous observational data, was considered also through 95 percent
782 confidence interval. This paper is one of the first attempts to solve the complicated and ill-
783 posed problem of simultaneous identification of all characteristics of multiple pollutant sources
784 in a complex river network. There are several aspects that need further investigation. Currently,
785 the application of proposed method is limited to cases in which the activity time of pollutant
786 sources are equal to or greater than expected activity time for recovery. Some measures such
787 as considering random data collecting in secondary station might alleviate this problem. This
788 is a subject for our future study.

789 **References**

- 790 Andrle, M. and El Badia, A. 2012. Identification of multiple moving pollution sources in
791 surface waters or atmospheric media with boundary observations. *Inverse problems*,
792 28, 075009.
- 793 Atmadja, J. and Bagtzoglou, A. C. 2001. State of the art report on mathematical methods for
794 groundwater pollution source identification. *Environmental forensics*, 2, 205-214.
- 795 Boano, F., Revelli, R. and Ridolfi, L. 2005. Source identification in river pollution problems:
796 A geostatistical approach. *Water resources research*, 41.
- 797 Box, G. E. and Cox, D. R. 1964. An analysis of transformations. *Journal of the Royal Statistical*
798 *Society: Series B (Methodological)*, 26, 211-243.

799 Butera, I., Tanda, M. G. and Zanini, A. 2013. Simultaneous identification of the pollutant
800 release history and the source location in groundwater by means of a geostatistical
801 approach. *Stochastic Environmental Research and Risk Assessment*, 27, 1269-1280.

802 Chapra, S. C. (2008). Surface water-quality modeling. Illinois: Waveland press.

803 Cheng, W. P. and Jia, Y. 2010. Identification of contaminant point source in surface waters
804 based on backward location probability density function method. *Advances in Water*
805 *Resources*, 33, 397-410.

806 De Marsily, G. 1986. Quantitative Hydrogeology: Groundwater Hydrology for Engineers
807 Academic Press. Inc., Orlando, Florida.

808 El Badia, A. and Hamdi, A. 2007. Inverse source problem in an advection-dispersion-reaction
809 system: application to water pollution. *Inverse Problems*, 23, 2103.

810 Fischer, H. B., Koh, R. C., Brooks, N. H., list, E. J. and Imberger, J. 1979. Mixing in Inland
811 and Coastal Waters. Academic Press.

812 Ghane, A., Mazaheri, M. and Samani, J. M. V. 2016. Location and release time identification
813 of pollution point source in river networks based on the Backward Probability Method.
814 *Journal of environmental management*, 180, 164-171.

815 Gzyl, G., Zanini, A., Frączek, R. and Kura, K. 2014. Contaminant source and release history
816 identification in groundwater: a multi-step approach. *Journal of contaminant*
817 *hydrology*, 157, 59-72.

818 Hadamard, J. 1923. Lectures on Cauchy's Problem in Linear Partial Differential Equations,
819 Yale University Press.

820 Hamdi, A. 2009. The recovery of a time-dependent point source in a linear transport equation:
821 application to surface water pollution. *Inverse Problems*, 25, 075006.

822 Hamdi, A. 2016. Detection-Identification of multiple unknown time-dependent point sources
823 in a 2 D transport equation: application to accidental pollution. *Inverse Problems in*
824 *Science and Engineering*, 1-25.

825 Hoeksema, R. J. and Kitanidis, P. K. 1985. Comparison of Gaussian conditional mean and
826 kriging estimation in the geostatistical solution of the inverse problem. *Water*
827 *Resources Research*, 21, 825-836.

828 Jiang, H. 2008. Adaptive feature selection in pattern recognition and ultra-wideband radar
829 signal analysis. California Institute of Technology.

830 Kitanidis, P. K. 1995. Quasi-linear geostatistical theory for inversing. *Water resources*
831 *research*, 31, 2411-2419.

832 Kitanidis, P. K. 1996. On the geostatistical approach to the inverse problem. *Advances in Water*
833 *Resources*, 19, 333-342.

834 Kitanidis, P. K. and Shen, K.-F. 1996. Geostatistical interpolation of chemical concentration.
835 *Advances in Water Resources*, 19, 369-378.

836 Lee, Y. J., Park, C. and Lee, M. L. 2018. Identification of a Contaminant Source Location in a
837 River System Using Random Forest Models. *Water*, 10, 391.

838 Liu, C. and Ball, W. P. (1999). Application of inverse methods to contaminant source
839 identification from aquitard diffusion profiles at Dover AFB, Delaware. *Water*
840 *Resources Research*, 35, 1975-1985.

841 Mazaheri, M., Mohammad Vali Samani, J. and Samani, H. M. V. 2015. Mathematical Model
842 for Pollution Source Identification in Rivers. *Environmental Forensics*, 16, 310-321.

843 Michalak, A. M. 2002. Environmental Contamination with Multiple Potential Sources and the
844 Common Law: Current Approaches and Emerging Opportunities. *Fordham*
845 *Environmental Law Journal*, 14, 147-206.

846 Michalak, A. M. and Kitanidis, P. K. 2002. Application of Bayesian inference methods to
847 inverse modelling for contaminants source identification at Gloucester Landfill,
848 Canada. *DEVELOPMENTS IN WATER SCIENCE*, 47, 1259-1266.

849 Michalak, A. M. and Kitanidis, P. K. 2003. A method for enforcing parameter nonnegativity
850 in Bayesian inverse problems with an application to contaminant source identification.
851 *Water Resources Research*, 39.

852 Michalak, A. M. and Kitanidis, P. K. 2004a. Application of geostatistical inverse modeling to
853 contaminant source identification at Dover AFB, Delaware. *Journal of hydraulic
854 research*, 42, 9-18.

855 Michalak, A. M. and Kitanidis, P. K. 2004b. Estimation of historical groundwater contaminant
856 distribution using the adjoint state method applied to geostatistical inverse modeling.
857 *Water Resources Research*, 40.

858 Morrison, R. D. 2000a. Critical Review of Environmental Forensic Techniques: Part I.
859 *Environmental Forensics*, 1, 157-173.

860 Morrison, R. D. 2000b. Critical review of environmental forensic techniques: Part II.
861 *Environmental Forensics*, 1, 175-195.

862 Neupauer, R. M., Borchers, B. and Wilson, J. L. 2000. Comparison of inverse methods for
863 reconstructing the release history of a groundwater contamination source. *Water
864 Resources Research*, 36, 2469-2475.

865 Skaggs, T. H., and Z. J. Kabala. (1994). Recovering the release history of a groundwater
866 contaminant. *Water Resour. Res.*, 30(1), 71– 79.

867 Skaggs, T. H. and Kabala, Z. 1998. Limitations in recovering the history of a groundwater
868 contaminant plume. *Journal of Contaminant Hydrology*, 33, 347-359.

869 Snodgrass, M. F. and Kitanidis, P. K. 1997. A geostatistical approach to contaminant source
870 identification. *Water Resources Research*, 33, 537-546.

871 Taylor, G. 1954. The dispersion of matter in turbulent flow through a pipe. *Proceedings of the
872 Royal Society of London. Series A. Mathematical and Physical Sciences*, 223, 446-468.

873 Telci, I. T. and Aral, M. M. 2011. Contaminant source location identification in river networks
874 using water quality monitoring systems for exposure analysis. *Water Quality, Exposure
875 and Health*, 2, 205-218.

876 Wang, J., Zhao, J., Lei, X. and Wang, H. 2018. New approach for point pollution source
877 identification in rivers based on the backward probability method. *Environmental
878 Pollution*, 241, 759-774.

879 Woodbury, A., Sudicky, E., Ulrych, T. J. and Ludwig, R. 1998. Three-dimensional plume
880 source reconstruction using minimum relative entropy inversion. *Journal of
881 Contaminant Hydrology*, 32, 131-158.

882 Wu, W. 2007. *Computational river dynamics*, CRC Press.

883 Yang, H., Shao, D., Liu, B., Huang, J. and Ye, X. 2016. Multi-point source identification of
884 sudden water pollution accidents in surface waters based on differential evolution and
885 Metropolis–Hastings–Markov Chain Monte Carlo. *Stochastic environmental research
886 and risk assessment*, 30, 507-522.

887
888
889
890
891
892
893

894 **Declarations**

895

896 **Funding**

897 No funding was received for conducting this study.

898 **Conflicts of interest/Competing interests**

899 The authors declare that they have no conflicts of interest.

900 **Availability of data and material**

901 Not applicable.

902 **Code availability**

903 Not applicable.

904