

Foul sewer model development using geotagged information and smart water meter data

Jia, Yueyi; Zheng, Feifei; Zhang, Qingzhou; Duan, Huan Feng; Savic, Dragan; Kapelan, Zoran

DOI

[10.1016/j.watres.2021.117594](https://doi.org/10.1016/j.watres.2021.117594)

Publication date

2021

Document Version

Accepted author manuscript

Published in

Water Research

Citation (APA)

Jia, Y., Zheng, F., Zhang, Q., Duan, H. F., Savic, D., & Kapelan, Z. (2021). Foul sewer model development using geotagged information and smart water meter data. *Water Research*, 204, Article 117594. <https://doi.org/10.1016/j.watres.2021.117594>

Important note

To cite this publication, please use the final published version (if applicable). Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights. We will remove access to the work immediately and investigate your claim.

1 **Real-time foul sewer hydraulic modelling driven by water consumption data from water**
2 **distribution systems**

3 Qingzhou Zhang¹, Feifei Zheng², Yueyi Jia³, Dragan Savic⁴ and Zoran Kapelan⁵

4 ¹**Qingzhou Zhang:** Postdoctoral Research Fellow, College of Civil Engineering and
5 Architecture, Zhejiang University, China. wdswater@gmail.com.

6 ²**Feifei Zheng:** Corresponding author, Professor, College of Civil Engineering and
7 Architecture, Zhejiang University, China. feifeizheng@zju.edu.cn. Tel: +86-571-8820-6757.
8 Postal address: A501, Anzhong Building, Zijingang Campus, Zhejiang University, 866
9 Yuhangtang Rd, Hangzhou, China 310058.

10 ³**Yueyi Jia:** PhD candidate, College of Civil Engineering and Architecture, Zhejiang
11 University, China. yueyi@zju.edu.cn.

12 ⁴**Dragan Savic:** Chief Executive Officer, KWR Water Research Institute,
13 Dragan.Savic@kwrwater.nl, Professor, Centre for Water Systems, University of Exeter, North
14 Park Road, Exeter, EX4 4QF, United Kingdom.

15 ⁵**Zoran Kapelan:** Professor, Department of Water Management, Delft University of
16 Technology, The Netherlands, z.kapelan@tudelft.nl

17

18 **Abstract:** One way to address many issues (e.g., illicit inflows) within foul sewer systems
19 (FSSs) is via real-time hydraulic models. However, a bottleneck within real-time FSS
20 modelling is the lack of spatio-temporal manhole inflow data. To address this problem, this
21 paper proposes a new method to develop real-time FSS models driven by water consumption
22 data from associated water distribution systems (WDSs) that often have a proportionally
23 larger number of sensors. Within the proposed method, the FSS manholes are integrated with
24 the WDS water consumption nodes based on their underlying physical connections. An
25 optimization approach is subsequently proposed to identify the transfer factor k between
26 nodal water consumption and FSS manhole inflows based on historical observations. These
27 identified k values combined with the acquired real-time nodal water consumption data from
28 the WDS equipped with a dense network of sensors drive the FSS real-time modelling. The
29 proposed method is applied to two real FSSs and results show that it can produce sewer flow
30 and manhole water depth simulations matching well with observations at the monitoring
31 locations with averaged R^2 , NSE and KGE (Kling-Gupta efficiency) around 0.99, 0.88 and
32 0.92, respectively. It is anticipated that real-time models developed by the proposed method
33 can be useful for the efficient FSS management and operation.

34 **Key words:** foul sewer system; water consumption data; real-time models; water distribution
35 system

36 **1 Introduction**

37 Sewer networks are traditionally designed to collect wastewater from residential, commercial
38 and/or industrial clients or possible stormwater from urban surfaces due to rainfall events.
39 Collected wastewater is transported then downstream to wastewater treatment plants
40 (WWTPs) or released directly into rivers (Bailey et al., 2019). These sewer networks are
41 often called combined sewer systems (CSSs), which have been widely used in large cities
42 around the world (Li et al., 2014). In recent years, there is a growing trend in separating CSSs
43 into independent storm drainage systems and foul sewer systems (FSSs, Schilperoort et al.
44 2013). The former are used to convey urban runoff solely to surface water bodies (e.g., rivers)
45 and the latter deliver sewerage collected from houses and commercial buildings before being
46 conveyed to treatment facilities. Such a separation is mainly driven by the purpose to
47 improve urban water environments as combined sewer overflows (CSOs, Black and
48 Endreny, 2006) would inevitably threaten the ecological health of the receiving water
49 (Joseph-Duran et al., 2015).

50 Over the past decade, many FSSs around the world have experienced significant changes due
51 to population growth and quick urbanization, which is especially the case in many developing
52 countries such as China (Sweetapple et al., 2018). These changes are mainly represented by
53 the expanded spatial scales of FSSs, the increased complexity in their topology structures and
54 the aged systems (Rokstad and Ugarelli, 2015, Huang et al., 2018). This, consequently,
55 results in significant challenges/difficulties for effective FSS management and operation, and
56 hence many issues exist (Garda et al., 2016). A typical issue is the deposits in the FSSs,

57 including sediments (Seco et al., 2018), fat, oil and grease (FOG, Liu et al., 2016) and toilet
58 papers (Eren and Karadagli, 2012). All those deposits can directly affect flow capacity of the
59 sewers, causing overflows from CSOs and manholes as well as potential water quality issues
60 (e.g, odor issues, Liu et al., 2016; Talaiekhosani et al., 2016). Another common issue is the
61 illicit discharges from local factories (Irvine et al. 2011; Banik et al. 2017), where these
62 discharges often contain toxic substances (e.g., heavy metals) that are often beyond the
63 processing capacity of the downstream WWTPs. This, therefore, can result in functional
64 failures of WWTPs and consequently significant contamination of the receiving water body
65 (McCall et al. 2016). In addition to issues of deposits and illicit discharges, leaks of the
66 sewers, groundwater infiltration and illicit connections between FSSs and stormwater pipes
67 are frequently reported, inducing serious contamination to the surrounding water
68 environments (Lepot et al., 2017; Beheshti and Saegrov, 2019).

69 The issues mentioned above have appreciably affected the urban water environments due to
70 the resultant overflows from CSOs/manholes and leaks of FSSs. One way to address these
71 issues is the placement of sensors within the FSS to monitor the water depths and sewer flows,
72 thereby triggering a warning when the observations are significantly higher or lower than the
73 historical data (Ahm et al. 2016). However, due to the high purchase cost and intensive
74 maintenance efforts associated with these sensors, the monitoring network is often sparse for
75 the majority of the FSSs (Kleidorfer et al., 2012). Consequently, a warning associated with
76 the potential issues (e.g., overflows or leaks) can be only available for the very limited
77 number of FSS locations in the proximity of sensors. In addition, the abnormal observations
78 at the monitoring locations may be caused by sudden discharges increases caused by the

79 water users, rather than the illicit discharges, resulting in a potentially high likelihood of false
80 warning (Koch et al. 2011). More importantly, relying solely on the observations from the
81 sewer sensors cannot offer predictions on the hydraulic status of the entire FSS in future
82 (Bruen and Yang, 2006). To this end, real-time sewer hydraulic modelling can be promising
83 in addressing the issues associated with the FSSs as mentioned above, where the hydraulic
84 variables such as water depths and sewer flows across the entire FSS are simulated in
85 real-time. These simulations, combined with observations at the monitoring locations, can be
86 used to deduce whether leaks, illicit discharges, deposits and illicit connections exist in the
87 FSS, as well as facilitate the localization of such events.

88 Manhole inflow data (i.e., sewer discharges of the water users) at a high time resolution (say
89 every 30 minutes) is the key to enable the development of a real-time FSS hydraulic model.
90 However, discharge data of such a high temporal and spatial resolution are typically
91 unavailable in engineering practice, resulting in a large challenge for real-time modeling of
92 FSS (Breinholt et al., 2013). To deal with this problem, a widely used approach is to calibrate
93 a model to estimate manhole inflows with the aid of limited in-sewer observations (Korving
94 and Clemens, 2005). While some calibration methods are available in the literature (e.g., di
95 Pierro et al., 2005, Khu et al., 2006, Broekhuizen et al., 2020), they mainly focus on
96 calibrating the underlying rainfall-runoff relationship for the combined sewer systems in an
97 off-line manner, thereby predicting the floods or sewer overflows caused by rainfall. These
98 previously published methods, therefore, cannot be used or at least are difficult to estimate
99 FSS manhole inflows in real-time.

100 The real-time management of the FSS has received great attention over the past few decades,
101 with the main focus on system real-time control based on observations (Schütze et al., 2002;
102 Sara et al., 2020). More specifically, real-time control is defined as a timely operation of an
103 FSS based on continuously monitored process data. Those data are water levels and sewer
104 flows in the system, with operations including the activation of pumps, sluice gates and weirs
105 used to improve system performance (e.g., reduce the overflows, Schütze et al., 2003).
106 However, these real-time control studies operated the hydraulic facilities (e.g., pumps) with
107 the aid of system observations rather than FSS simulations, and hence they differ
108 significantly from the real-time FSS hydraulic modelling, which is the aim of the present
109 study.

110 The main difficulty associated with the calibration of FSS manhole inflows based on the
111 limited number of monitoring sites is the “equifinality” (Khu et al., 2006). More specifically,
112 a large number of manhole inflow combinations can produce similar agreements between
113 simulated and observed water levels at monitoring locations. As a result, it is very difficult, if
114 not impossible, to identify a particular parameter set (i.e., manhole inflow combination) that
115 can represent the true underlying spatial distribution of the discharges from water users into
116 the FSS.

117 To address the “equifinality” issue, this paper proposes a new method to enable the
118 development of real-time FSS hydraulic model. Within the proposed method, the FSS model
119 is integrated with its corresponding water distribution system (WDS) hydraulic model for the
120 same area being considered. Such a model integration approach is possible as the WDS

121 models have already been widely used (Walski et al., 2003). In addition, the number of
122 sensors (e.g., smart demand meters, pressure sensors and flow meters) deployed in the WDSs
123 can be large, which is, at least partly, driven by the quick developments of the Internet of
124 Things in recent years (Zheng et al., 2018). Such a dense sensor network can greatly facilitate
125 the estimation of real-time nodal water consumption for the WDS models as demonstrated in
126 previous studies (Creaco et al., 2019). This is especially the case in recent years as smart
127 demand meters have been increasingly used in many WDSs, providing water consumption
128 data for many users (not only large users but also residential users) in a real-time manner
129 (typically every 15 or 30 minutes, Creaco et al., 2018). Such near real-time and high-density
130 spatial water consumption data can be assimilated with the limited in-sewer observations to
131 develop a real-time FSS hydraulic model. This is the key feature and novelty of the method
132 presented in this paper.

133 The concept of incorporating water consumption data into FSS modeling can be dated back to
134 Bruke et al. (1986), where an FSS model was calibrated using monthly water use records.
135 More recently, Bailey et al. (2019) presented a new FSS model, where the stochastically
136 simulated water demands were imported into the sewer network model. While these limited
137 previous studies have made great contributions in assimilating water use records into FSS
138 modelling (mainly used for FSS design purpose), the water consumption data used are either
139 collected manually at a very low time resolution (e.g., monthly, Bruke et al., 1986) or
140 provided by a stochastic simulator (Bailey et al., 2019). Consequently, these data cannot
141 represent the true underlying temporal and spatial variations of the manhole inflows.
142 Therefore, they cannot be used to develop real-time FSS models, which is the focus of this

143 study.

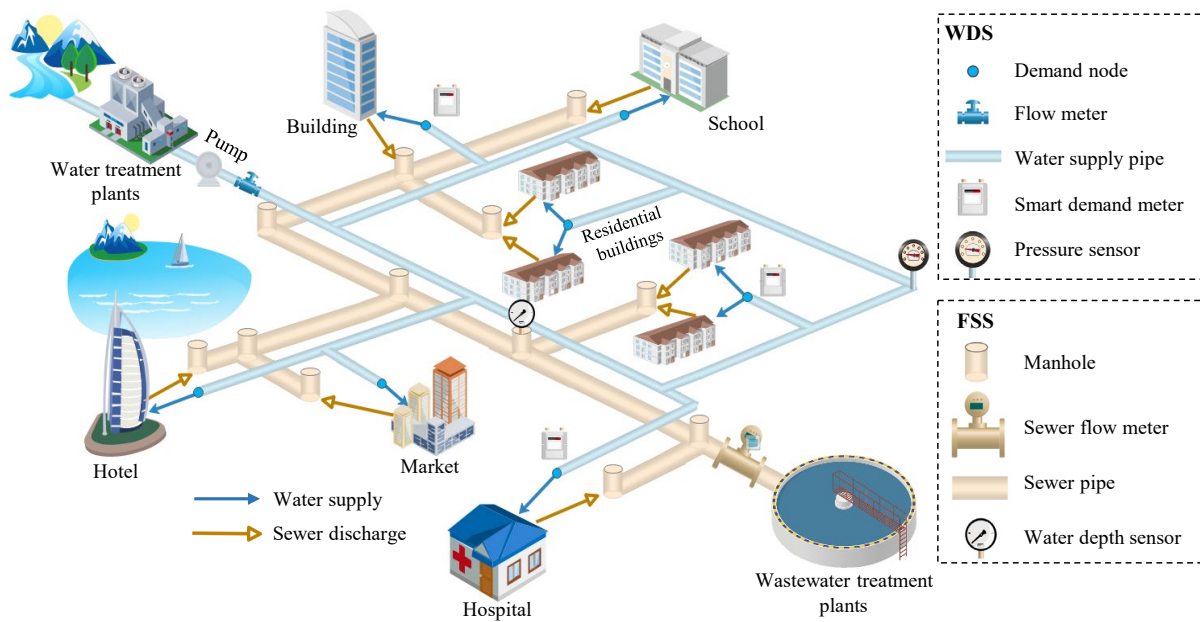
144 The key feature of the proposed method is that the real-time FSS model is developed using a
145 large number of existing sensors within the WDSs. This implies that it is not necessary to
146 deploy a large number of sewer sensors (which is often very expensive in terms of both
147 sensor purchases and maintenance) to enable real-time sewer modelling, making the proposed
148 method attractive for practical applications. This paper is organized as follows. The proposed
149 methodology is described in Section 2, followed by the descriptions of the case studies
150 considered in Section 3. Results and discussions are given in Section 4. Finally, the conclusion
151 section (Section 5) shows the main observations and implications of this paper.

152 **2. Methodology**

153 **2.1 The overall modelling concept**

154 Fig.1 illustrates the overall concept of the proposed method, where a foul sewer system (FSS)
155 and a water distribution system (WDS) for a small area are presented. Typically, raw water
156 from reservoirs or rivers is pumped into the water treatment plants in order to improve water
157 quality to a required standard (Wu et al., 2011). Subsequently, the treated water is conveyed
158 to the WDS, satisfying demands for various users including residents, schools, hospitals,
159 industrial and commercial buildings, as shown in Figure 1. To ensure water supply safety,
160 sensors are often deployed in the WDS (Figure 1), including pressure sensors, flow meters
161 and smart demand meters. The latter have been increasingly being deployed in recent years to
162 monitor water consumptions for the users in a near real-time manner (Creaco et al., 2018).

163 Consequently, such a dense sensor network enables the development of real-time WDS
 164 modelling, which has been an important trend within the water supply domain (both research
 165 and industry) due to its great merits in facilitating effective system management as
 166 highlighted in Creaco et al. (2019).



167
 168 **Fig. 1 An illustration of the concept for the proposed modelling method, where a water**
 169 **distribution system and a foul sewer system are presented**

170 Inherently, local residents or commercial/industrial users discharge sewage after water
 171 consumption as illustrated in Figure 1. Sewer pipes collect and convey the sewage to
 172 downstream wastewater treatment plants, with a limited number of water depth or sewer flow
 173 sensors installed to monitor hydraulic state of the system. Consequently, the following
 174 equation can be used to represent the underlying relationship between water consumption and
 175 sewage discharge for user i :

176
$$d_i = F(q_i, k_i, t_i) \tag{1}$$

177 where d_i is the sewer discharge rate of user i (i.e., manhole inflow rate) resulting from its
178 water consumption q_i taken from the WDS, t_i represents the time delay, i.e., the time
179 between the clean water entering the user property and the time it reaches the local sewer
180 network, k_i is the transfer factor for user i , representing the proportion of supplied water
181 that ends up in the local sewer network; k_i typically has a value between 0.7 and 1.0
182 (Behzadian and Kapelan 2015). Equation (1) represents the fundamental rule/assumption in
183 the proposed method used to build the connections between the WDS water consumption
184 data and the FSS manhole inflows.

185 Fig.2 presents the overall methodology of the proposed method, with two modules involved.
186 The first module consists of three phases, which are carried only once in an offline manner,
187 and the second module only has the fourth phase (Phase 4) of the proposed method, which
188 runs in real-time. The details are given below.

189 *Phase 1: Integrate the WDS and FSS models (carry out once).* Within this phase, the
190 FSS and WDS models are developed with hydraulic facility information (e.g., water supply
191 pipes, tanks, sewer pipes) taken from external sources such as the GIS or asset management
192 system. This is followed by the building of the connections between each WDS demand node
193 and the FSS manhole based on the spatial distance with details given in Section 2.2.

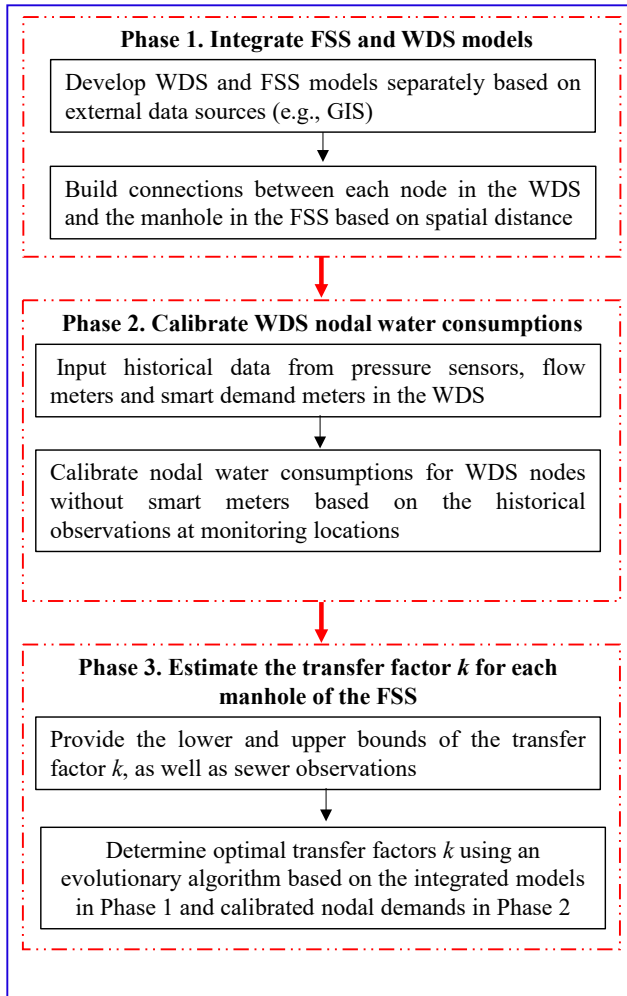
194 *Phase 2: Calibrate nodal water consumptions of the WDS.* It should be highlighted that
195 the calibration of the nodal water consumptions in Phase 2 is conducted offline, which is used
196 to provide data for Phase 3. More specifically, Based on a particular time period of historical

197 data from pressures sensors, flow meters and smart demand meters deployed in the WDS, the
198 nodal water consumption without smart demand meters are estimated for a given time
199 resolution (often equals the time resolution of the flow or pressure sensors) with details given
200 in Section 2.3

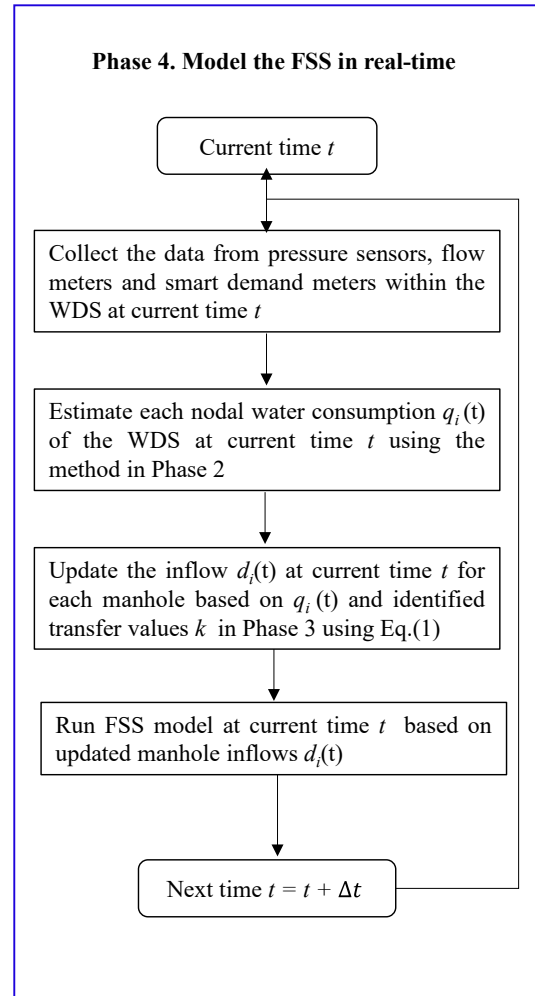
201 *Phase 3: Estimate the transfer factor k for each manhole of the FSS.* According to the
202 identified relationship between WDS nodes and the FSS manholes in Phase 1, as well as the
203 calibrated nodal water consumptions in Phase 2, the transfer factor k is determined. For this,
204 an evolutionary algorithm (EA) is applied with the objective function defined in Eq. (9-13)
205 and using sewer observations, with details given in Section 2.4.

206 *Phase 4: Model the FSS in a real-time manner.* Data from pressure sensors, flow meters
207 and the available smart demand meters in the WDS are acquired at the current time t . These
208 data are used as the inputs for the real-time WDS modelling to estimate water consumption
209 for each node (q_i) within the WDS (Section 2.3). Eq. (1) is subsequently used to update the
210 manhole inflows d_i based on the known q_i and identified k values (Phase 3). Finally, the FSS
211 is modelled by updating manhole inflows d_i in real-time. This offers short-term hydraulic
212 predictions (water depths at manholes and flow rates in sewer pipes) of the entire FSS with a
213 particular time resolution (if say every 30 minutes used in this paper).

Offline module (carry out once)



Real-time module



214

215

Fig. 2 The overall methodology of the proposed method

216 It is noted that a few assumptions are made in the proposed method, with the justification
217 given below.

218 (i) Given that the proposed FSS real-time modelling method is driven by water
219 consumption data from the WDS, the number of available smart demand meters in the WDS
220 is important to ensure the high accuracy of the FSS simulations. For the two case studies
221 considered in this paper, the number of smart demand meters is reasonably high, making
222 them perfectly suited for the demonstration of the proposed method. However, some WDSs

223 may have relatively low coverage of the smart demand meters (e.g., only installed for large
224 demand users). While such a case would not affect the application of the proposed method,
225 the accuracy of the WDS nodal water consumption values and the FSS real-time simulations
226 can be affected. However, it is anticipated that smart demand meters are increasingly used by
227 water utilities as a result of the quick developments in the Internet of Things (Zheng et al.,
228 2018; Creaco et al., 2019), and hence the applicability of the proposed method is only going
229 to grow.

230 (ii) The proposed method assumes that observations from the WDS and FSS sensors
231 (including smart demand meters) are accurate within the applications in this study. However,
232 in reality, observation errors can exist due to the sensor malfunctions or signal transmission
233 issues. Therefore, it is necessary to incorporate the potential observation errors into the
234 modelling framework. Although that is an important study direction, it is beyond the scope of
235 the paper and will be the focus of future work.

236 (iii) As shown in Equation (1), the time t_i implies that the nodal water
237 consumption q_i estimated at time t using the smart demand meter or the real-time
238 calibration method (section 2.3) should correspond to the manhole inflow at time $t + t_i$. The
239 value of t_i can be dependent on the particular user properties, including the characteristics
240 of the water supply area associated with the demand node, as well as the physical
241 characteristics (e.g., length and slopes) of the connecting sewer pipes between users and the
242 corresponding manholes. In this study, t_i is ignored as this value is typically small, ranging
243 from several minutes up to 15 minutes for many cases (Wu et al., 2011). This assumption is
244 considered valid in our study as the time resolution used for real-time FSS modelling in this

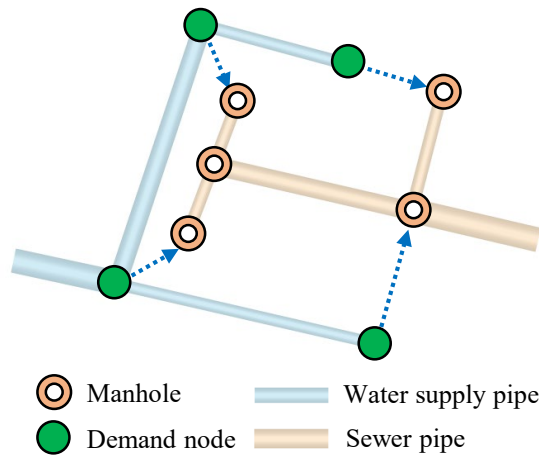
245 paper is lower (i.e., 30 minutes).

246 (iv) In this study, a linear transfer function with a constant factor of k is proposed to
247 describe the underlying relationship between the nodal water consumption and manhole
248 inflows. While being simple in practical implementation, the transfer function as well as the k
249 factor can be affected by not only the time delay t_i in Equation (1), but also the infiltration
250 inflows and the properties of water users. More specifically, the transfer function may be
251 different between the water users with or without smart demand sensors, and the k factor may
252 temporally vary, or even change as a function of different water users. These influences need
253 further consideration in future study along this research line.

254 **2.2 Integrate the FSS and WDS models**

255 Typically, FSS and WDS models are developed with the aid of the geographic information
256 system (GIS) or the asset management system for the analyzed area (Behzadian and Kapelan,
257 2015; Huang et al., 2018). The details (e.g., locations and length) of various system
258 components including pipes, tanks, valves and pumps can be taken from the GIS. This is
259 often followed by system skeletonization in which many facilities (mainly small pipes) are
260 removed or simplified without significantly affecting the hydraulic properties of the original
261 full system (Huang et al., 2020). To enable the practical application of Eq. (1), it is important
262 to build the connection between each demand node i , representing the water consumption in
263 the WDS model, and the manhole, representing the facility to collect the sewages in the FSS
264 model. Such a connection indicates that the consumption at the demand node i is received by
265 its associated manhole. To this end, the WDS and FSS models are integrated within their
266 development processes in this study.

267 Figure 3 illustrates the proposed integration method, where a water demand node in the WDS
268 model is assigned to the manhole of the FSS model within a shortest distance from each other.
269 The rationale behind this is that manholes are often built near the water users (nodes in the
270 WDS model) to collect their sewerage discharges. Consequently, two cases are available as
271 shown in Figure 3, which are (i) one demand node is assigned to a manhole, and (ii) multiple
272 demand nodes are assigned to a single manhole. In addition to these two relatively simple
273 cases, in practice, one demand node can be associated with multiple manholes, which is
274 possible when this demand node represents many users. However, it is difficult to know the
275 proportion of the total discharge associated with each relevant manhole. For the sake of
276 simplicity, a single manhole with the minimum spatial distance to this demand node is
277 selected to deliver the total discharge. While such a simplification can cause an unrealistic
278 hydraulic status in a very small area relative to the original full system, its impacts on the
279 overall results can be negligible. Since each demand node (say node i) in the WDS is
280 assigned to a particular manhole in the FSS model, the water consumption of this node (q_i) is
281 considered as the approximate manhole inflows (d_i). Their underlying flow relationship
282 needs to be further accurately determined with the incorporation of the transfer factor k_i as
283 shown in Equation (1).



284

285

Fig. 3 Illustration of the proposed integration method for FSS and WDS model

286

developments

287

It is noted that when a higher level of accuracy is needed for a practical application,

288

individual water consumption and sewer connections could be identified if required. This will

289

lead to a slight modification on the WDS and FSS model topologies, which can better reflect

290

the flows of supplied drinking water and generated wastewater by different users.

291

2.3 Calibrate nodal water consumption based on historical observations

292

Based on the built connections between each WDS node and its corresponding FSS manhole

293

as described in Section 2.2, nodal water consumption is the driver for triggering the real-time

294

FSS modelling. In the WDSs, many smart demand meters can be available, providing near

295

real-time water consumption data (if say every 15 minutes or 30 minutes) for WDS nodes

296

(users), especially for water users with large demand. However, in practice, it may not be

297

necessary to have smart meters installed at each demand node, and hence this study adopts a

298

calibration method to enable the estimation of water consumption at the nodes without smart

299 demand meters.

300 In this study, the numerical approach described in Zhang et al. (2018) is selected to calibrate
301 the nodal water consumptions due to its demonstrated efficiency. The objective function of
302 this adopted numerical method is formulated as the weighted sum of squared differences
303 between the field-observed and model-simulated responses (pressures and flows) at
304 monitoring points in the WDS within a particular time period (i.e., the time resolution of the
305 monitoring data), i.e. as follows:

$$306 \quad \text{Min} : f(\mathbf{q}) = \sum_{i=1}^{NH} w_h^i [H_i^o - H_i(\mathbf{q})]^2 + \sum_{j=1}^{NF} w_q^j [Q_j^o - Q_j(\mathbf{q})]^2 = \begin{bmatrix} \mathbf{H}^o - \mathbf{H}(\mathbf{q}) \\ \mathbf{Q}^o - \mathbf{Q}(\mathbf{q}) \end{bmatrix}^T \mathbf{W} \begin{bmatrix} \mathbf{H}^o - \mathbf{H}(\mathbf{q}) \\ \mathbf{Q}^o - \mathbf{Q}(\mathbf{q}) \end{bmatrix} \quad (2)$$

307 where \mathbf{q} is the vector of nodal water consumptions in the WDS, including known water
308 consumption data at nodes with smart demand meters and unknown nodal water consumption
309 data; NH and NF are the numbers of observed nodal pressures and pipe flows, respectively;
310 w_h^i and w_q^j are the weighting factors for observed pressures at the i^{th} node and observed
311 flows at the j^{th} pipe, respectively, where $w_h^i = 1/(H_i^o)^2$ and $w_q^j = 1/(Q_j^o)^2$ are used in this
312 study following Kun et al. (2015) and Zhang et al. (2018). H_i^o and $H_i(\mathbf{q})$ are the observed
313 and simulated pressures at the i^{th} node respectively; Q_j^o and $Q_j(\mathbf{q})$ are the observed and
314 simulated flows at the j^{th} pipe respectively. Equation (2) can be expressed in the matrix
315 form (see above) by using $\mathbf{H}^o = [H_1^o, H_2^o, \dots, H_{NH}^o]^T$, $\mathbf{Q}^o = [Q_1^o, Q_2^o, \dots, Q_{NF}^o]^T$,
316 $\mathbf{H}(\mathbf{q}) = [H_1(\mathbf{q}), H_2(\mathbf{q}), \dots, H_{NH}(\mathbf{q})]^T$, $\mathbf{Q}(\mathbf{q}) = [Q_1(\mathbf{q}), Q_2(\mathbf{q}), \dots, Q_{NF}(\mathbf{q})]^T$, and
317 $\mathbf{W} = \text{diag}([w_h^1, w_h^2, \dots, w_h^{NH}, w_q^1, w_q^2, \dots, w_q^{NF}])$.

318 Expanding Eq. (2) through the first-order Taylor series:

319
$$f(\mathbf{q} + \Delta\mathbf{q}) \approx \begin{bmatrix} \Delta\mathbf{H} - \mathbf{J}_H \Delta\mathbf{q} \\ \Delta\mathbf{Q} - \mathbf{J}_Q \Delta\mathbf{q} \end{bmatrix}^T \mathbf{W} \begin{bmatrix} \Delta\mathbf{H} - \mathbf{J}_H \Delta\mathbf{q} \\ \Delta\mathbf{Q} - \mathbf{J}_Q \Delta\mathbf{q} \end{bmatrix} \quad (3)$$

320 where $\Delta\mathbf{H} = \mathbf{H}^o - \mathbf{H}(\mathbf{q})$ and $\Delta\mathbf{Q} = \mathbf{Q}^o - \mathbf{Q}(\mathbf{q})$ are the differences between the observed and

321 simulated values of nodal pressures and pipe flows, respectively; $\mathbf{J}_H = \frac{\partial \mathbf{H}(\mathbf{q})}{\partial \mathbf{q}}$ and

322 $\mathbf{J}_Q = \frac{\partial \mathbf{Q}(\mathbf{q})}{\partial \mathbf{q}}$ are the Jacobian matrix with details given in Zhang et al. (2018). Since Eq. (2)

323 is a convex function (Kun et al., 2015), the minimum objective value of Eq. (2) can be

324 obtained when its first-order derivative (Eq. (3)) equals to zero, that is:

325
$$\frac{df(\mathbf{q} + \Delta\mathbf{q})}{d\Delta\mathbf{q}} = -2 \begin{bmatrix} \mathbf{J}_H \\ \mathbf{J}_Q \end{bmatrix}^T \mathbf{W} \begin{bmatrix} \Delta\mathbf{H} - \mathbf{J}_H \Delta\mathbf{q} \\ \Delta\mathbf{Q} - \mathbf{J}_Q \Delta\mathbf{q} \end{bmatrix} = 0 \quad (4)$$

326 By solving Eq. (4), $\Delta\mathbf{q}$ can be obtained as follows:

327
$$\Delta\mathbf{q} = \left(\begin{bmatrix} \mathbf{J}_H \\ \mathbf{J}_Q \end{bmatrix}^T \mathbf{W} \begin{bmatrix} \mathbf{J}_H \\ \mathbf{J}_Q \end{bmatrix} \right)^{-1} \begin{bmatrix} \mathbf{J}_H \\ \mathbf{J}_Q \end{bmatrix}^T \mathbf{W} \begin{bmatrix} \Delta\mathbf{H} \\ \Delta\mathbf{Q} \end{bmatrix} \quad (5)$$

328
$$\mathbf{q}^{s+1} = \mathbf{q}^s + \Delta\mathbf{q}^s \quad (6)$$

329 where $s = 0, 1, \dots, S$ is the iteration number (S is the maximum allowed number of

330 iterations). It is highlighted that the water consumption at nodes with smart demand meters

331 are known within the entire calibration process and hence $\Delta\mathbf{q}$ is only considered for the

332 nodes without smart meters. To ensure the estimated nodal water consumption values are

333 practically meaningful, the domain knowledge has been incorporated within the calibration

334 process in this study as shown below (Wu et al., 2010):

335
$$q_i^{s+1} = \begin{cases} q_i^{\min}, & \text{if } q_i^{s+1} < q_i^{\min} \\ q_i^{\max}, & \text{if } q_i^{s+1} > q_i^{\max} \\ q_i^{s+1}, & \text{others} \end{cases} \quad (7)$$

336 where $q_i^{\min} = (1-p) \times q_i^{\text{initial}}$ and $q_i^{\max} = (1+p) \times q_i^{\text{initial}}$ are the minimum and maximum
 337 allowed water consumptions at node i respectively; p is the percentage generally within
 338 10%~20% in practice (Zhang et al., 2018); q_i^{initial} is estimated using

339
$$q_i^{\text{initial}} = \frac{l_i}{L_T - L_M} (Q_T - Q_M) \quad (8)$$

340 where l_i is the length of the pipe associated with node i ; L_T and L_M is the total pipe
 341 length of all nodes and the length of pipes associated with smart demand meters respectively;
 342 Q_T is the total water consumption of the WDS at a given time period (e.g., 30 minutes),
 343 which is estimated based on the flow meters installed at the outlet of the water treatment
 344 plants and volume changes in the tanks if available; Q_M is the sum of the water
 345 consumption values measured by the available smart demand meters within the WDS at a
 346 given time period.

347 The calibration process at each time period (i.e., the time resolution of the monitoring data,
 348 e.g. 30 minutes) is executed by iteratively updating $\Delta \mathbf{q}$ in Eq. (6) until the maximum value
 349 of vector $\|\Delta \mathbf{q}\|$ is smaller than a given threshold value ε (e.g. $\varepsilon = 0.1$). The entire
 350 calibration process is executed again once the monitoring data from sensors are updated,
 351 representing a real-time hydraulic calibration for the WDS. It is noted that the pipe resistance
 352 coefficients are not calibrated in a real-time manner as these values are not likely to change

353 over a short time period (Kun et al., 2016).

354 **2.4 Estimate the transfer factor k for each FSS manhole**

355 As stated in Eq. (1), the nodal consumption data determined in Section 2.3 (q_i) cannot be
356 directly taken as the manhole inflows (d_i) due to the inevitable loss during the transporting
357 process within the facilities of the users (Behzadian and Kapelan, 2015). In this study, a
358 transfer factor k_i is used to represent the proportion of water consumption used by node i
359 that has been collected by its corresponding manhole. Such a factor can vary as a function of
360 the properties of the water users, such as user types (commercial users or common resident
361 users) and habits of water usages (Bailey et al., 2019). Therefore, the transfer factor needs to
362 be calibrated for each demand node based on the nodal water consumption data and the sewer
363 observations (e.g., sewer flow rates or water depth in the manholes) in the FSS. In this study,
364 the transfer factor k_i associated with each demand node is considered to be approximately
365 constant over time because the user properties are overall constant over a short time period
366 (Bailey et al., 2019).

367 To calibrate the transfer factor $\mathbf{K} = [k_1, k_2, \dots, k_n]^T$ of the entire FSS with a total of n
368 manholes with external inflows, the following objective function is defined,

$$\text{Min} : F(\mathbf{K}) = \sum_{t=T_w}^T \left(\sum_{i=1}^M [g(h_i^o(t)) - g(h_i^s(t))]^2 + \sum_{j=1}^N [g(f_j^o(t)) - g(f_j^s(t))]^2 \right) \quad (9)$$

$$[\mathbf{h}_i^s, \mathbf{f}_j^s] = [h_i^s(t_1), h_i^s(t_2), \dots, h_i^s(T); f_j^s(t_1), f_j^s(t_2), \dots, f_j^s(T)] = F_s(\mathbf{D}(T)) \quad (10)$$

385 modelling as described in Section 2.3. k_i^{\min} and k_i^{\max} are the minimum and maximum
386 allowable values of k_i , which can be determined by engineering experience. In this study,
387 $k_i^{\min} = 0.7$ is used for each demand node of the WDS, and $k_i^{\max} = 1.0$ is used for each WDS
388 node with smart demand meters, but $k_i^{\max} = 1.3$ is used for WDS nodes without smart meters.
389 This is because water consumptions of nodes without smart meters are calibrated using the
390 method described in Section 2.3, and hence the identified values can inevitably deviate from
391 the true water consumption values at a certain extent. To mitigate this potential impact, the
392 maximum value of the transfer factor for these nodes is increased to 1.3. In this paper, an
393 evolutionary algorithm (EA, Zheng et al., 2017) combined with the SWMM package is
394 employed to solve the optimization problem defined in Eq. (9-13). While different EAs are
395 available in literature, Borg (Hadka and Reed, 2013) is used in this study due to its
396 well-demonstrated performance in dealing with complex water resources optimization
397 problems, with more algorithm details in Section 3.2.

398 In the proposed FSS calibration method, manhole inflows are considered as the only
399 calibration parameters due to their large temporal and spatial variations, with which the
400 transfer factor k for each manhole can be estimated. It should be noted that Manning's
401 roughness coefficients of the sewer pipes can also affect the hydraulics of the FSS. However,
402 previous studies have shown that the impacts of the small to moderate variation in Manning's
403 roughness coefficients of sewer pipes are limited (Rossman and Huber, 2017). In addition, the
404 physical pipe properties (e.g., pipe ages and materials) that affect the Manning's roughness
405 coefficient are unlikely to vary in a short time period (Zhang et al., 2018) and hence it is not
406 considered within the real-time FSS modelling. It is highlighted that the values of

407 $\mathbf{K} = [k_1, k_2, \dots, k_n]^T$ are calibrated using a particular time period of historical water
408 consumption data and in-sewer observations in an offline manner (carried out once as shown
409 in Figure 2).

410 **2.5 Model the FSS in real-time**

411 It is noted that Phases 1-3 in Sections 2.2-2.4 are carried offline (in the offline module as
412 shown in Figure 2), aimed to identify the transfer factors between the WDS nodal water
413 consumptions and the FSS manhole inflows. This is followed by the real-time FSS modelling
414 (real-time module of the proposed method in Figure 2) with the following steps.

415 **Step 1:** Collect the data from pressure sensors, flow meters and the available smart demand
416 meters in the WDS at current time t ,

417 **Step 2:** Estimate the water consumption for each WDS node without smart demand meters,
418 $q_i(t)$ in Equation (1), using the method described in Section 2.3 (Phase 2) based on the
419 observations from Step 1.

420 **Step 3:** Update the manhole inflow $d_i(t)$ based on $q_i(t)$ and the identified transfer factor k in
421 Phase 3 of the offline module using Equation (1).

422 **Step 4:** Run the FSS hydraulic model based on the manhole inflow $d_i(t)$, producing the water
423 depths and sewer flows for the entire FSS within the time resolutions (30 minutes in this
424 study). This is followed by moving to Step 1 at $t = t + \Delta t$ where Δt is the time resolution
425 of the FSS modelling ($\Delta t = 30$ minutes in this study).

426 **2.6 Metrics used for performance evaluation**

427 Five statistical metrics are used in this paper to evaluate the performance of the proposed
 428 method in simulating the FSS hydraulic variables. They are the absolute percentage error
 429 (*APE*), the mean absolute percentage error (*MAPE*), the coefficient of determination (R^2), the
 430 Nash-Sutcliffe model efficiency (*NSE*), and the Kling-Gupta Efficiency (*KGE*). These five
 431 metrics are selected due to their wide applications in assessing the model performance within
 432 the water resources domain (Mu et al., 2020). The *APE* between the i^{th} observation Y_i and
 433 its corresponding simulation \hat{Y}_i is defined as

$$434 \quad APE = \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right| \times 100\% \quad (14)$$

$$435 \quad MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right| \times 100\% \quad (15)$$

436 where n is the total number of data points. As shown in Eq. (14) and (15), a lower value of
 437 *APE* or *MAPE* indicate an overall better model performance. The R^2 is a goodness-of-fit
 438 measure for linear regression models, which can be mathematically described as (Gujarati et
 439 al., 2009):

$$440 \quad R^2 = \frac{\left(\sum_{i=1}^n (Y_i - \tilde{Y})(Y_i - \bar{Y}) \right)^2}{\sum_{i=1}^n (Y_i - \tilde{Y})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2} \quad (16)$$

441 where \bar{Y} represents the mean of the observations and \tilde{Y} is the mean of the simulations. A
 442 large value of R^2 represents a better model performance. The *NSE* is defined as follows
 443 (Nash and Sutcliffe, 1970), with a larger value implying a better model performance:

444

$$NSE = 1 - \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \quad (17)$$

445 The *KGE* metric is mathematically described as follows (Knoben et al., 2019):

446

$$KGE = 1 - \sqrt{(r-1)^2 + \left(\frac{\sigma_{sim}}{\sigma_{obs}} - 1\right)^2 + \left(\frac{\mu_{sim}}{\mu_{obs}} - 1\right)^2} \quad (18)$$

447 where r is the linear correlation between observations and simulations; σ_{sim} and σ_{obs} are
 448 the standard deviation in simulations and observations, respectively; μ_{sim} and μ_{obs} are the
 449 mean of simulations and observations, respectively. A large value of *KGE* means that the
 450 simulations can match observations better, with $KGE=1$ representing the best model
 451 performance.

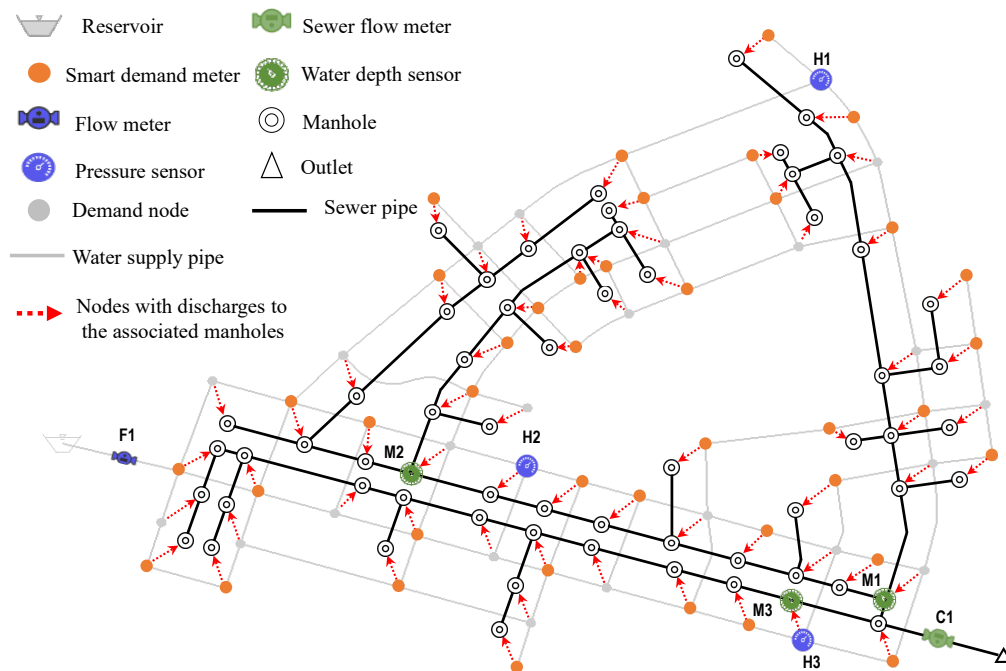
452 3. Case studies

453 3.1 Case study description

454 Two real-world FSSs in China, the Benk network (BKN) and the Xiuzhou network (XZN),
 455 are selected as case studies to demonstrate the utility of the proposed method. These two
 456 FSSs are selected as their associated WDSs have good coverage of monitoring sensors,
 457 especially the smart demand meters. In addition, BKN and XZN respectively represent scales
 458 of a relatively small region and a town, aimed to demonstrate the utility of the proposed
 459 method in handling the FSSs with different complexity levels.

460 BKN consists of one outlet, 64 manholes and 64 sewer pipes (Figure 4), delivering the

461 wastewater for the users with water supplied by a WDS (referred to as WDS-BNK).
 462 WDS-BNK is composed of one reservoir, 65 nodes and 93 pipes, as well as one flow meter,
 463 three pressure sensors and 40 smart water demand meters (Figure 4), providing
 464 approximately 4,800 m³ of water per day. As shown in Figure 4, one in-sewer flow meter and
 465 three water depth sensors with a 30-minute time resolution have been installed in BKN, with
 466 an average discharge of about 4,100 m³/day. The dotted arrow lines in Figure 4 represent the
 467 receiving manhole for each demand node determined based on the spatial distances.
 468 Observations from the WDS-BNK and BNK sensors are recorded for consecutive 31 days
 469 without rainfall or snowfall events in winter with a 30-minute time resolution.



470

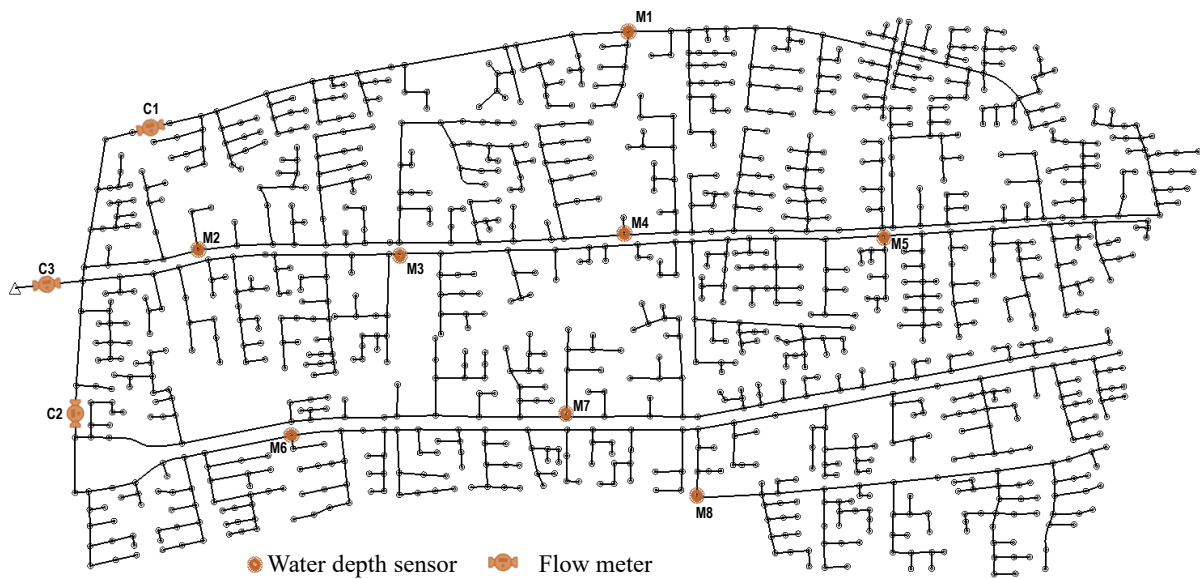
471 **Fig. 4 The layout and sensor locations of the BKN case study and its corresponding**

472

WDS-BNK

473 The XZN system is a large-scale complex FSS in Jiaying City, with a total length of
 474 approximately 86 km and an average discharge of about 21,500 m³/day. The layout of the

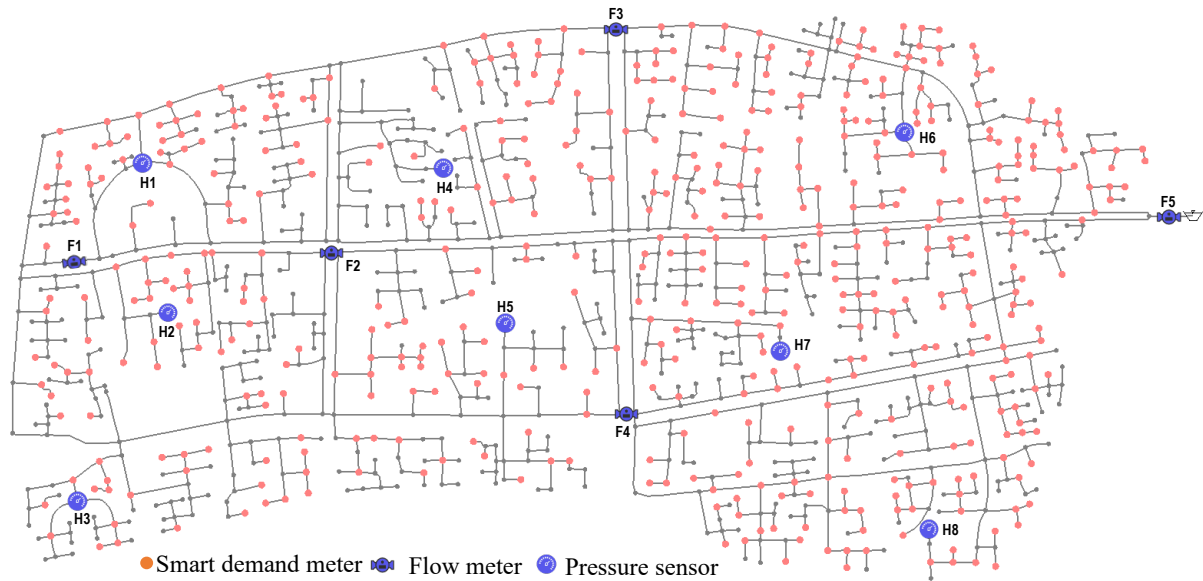
475 XZN network is shown in Fig. 5, consisting of one outlet, 1,214 manholes and 1,214 sewer
476 pipes. As shown in Fig. 5, three flow meters and eight water depth sensors have been
477 installed in this FSS. The WDS that supplies water demands for this area (referred as
478 WDS-XZN) has one reservoir, one pump station, 1,119 nodes and 1,137 water consumption
479 pipes as shown in Fig. 6. In the WDS-XZN network, five flow meters, eight pressure sensors
480 and 525 smart demand meters are deployed as illustrated in Fig. 6. The WDS-XZN network
481 supplies approximately 23,150 m³ per day for a population about 107,500 living in this area
482 within the Jiaxing City. As the same for the BKN network, the data from the WDS-XZN and
483 XZN sensors are recorded for consecutive 31 days without rainfall or snowfall events in
484 winter with a 30-minute time resolution.



485

486

Fig. 5 The layout and sensor locations of the XZN case study



487

488

Fig. 6 The layout and sensor locations of the WDS-XZN

489

3.2 Implementation of the proposed method

490

The EPANET2.0 and SWMM5.1 (Rossman, 2000, 2010) were used as WDS and FSS

491

hydraulic simulation model respectively in this study. For both case studies, historical data of

492

the first 17 consecutive days from WDS sensors with a 30-minute time resolution were used

493

to estimate the water consumptions of nodes without smart meters. This led to a total of 816

494

($17 \times 24 \times 2$) time periods with nodal water consumptions to be calibrated for each WDS.

495

These estimated nodal water consumption data were subsequently used to identify the

496

transfer factors k of the FSS based on sewer observations at the first 17 days.

497

The WDS and FSS sewer observations of the remaining 14 days ($14 \times 24 \times 2$ data points used

498

for model validation) were used to run the real-time FSS models with a 30-minute time

499

resolution. In other words, the first set of WDS observations at the validation period (the last

500

14 days) was considered as the observations at time t in the real-time module of Figure 2 (Δt

501 =30 minutes), followed by the execution of the four steps in Section 2.5.

502 For the nodal water consumption calibration, the termination error was set as
503 $\max(\|\Delta\mathbf{q}\|) \leq 0.1$ (Eq. 6), the maximum allowed iterations was $S=100$ (Eq. 6), and the
504 adjustment range of nodal water consumptions was $p=20\%$ for each WDS (Eq. 7). For the
505 WDS-BNK (Figure 4), observations of the first 17 days from two pressure sensors (H1 and
506 H3) and the flow meter F1 were used for calibration, and the records of pressure sensor H2
507 were used for validation. For the WDS-XZN (Figure 6), observations of the first 17 days
508 from H1, H3, H4, H6 and H8 pressure sensors, as well as F2, F3, F4 and F5 flow meters were
509 utilized for model calibration, and the records of H2, H5, H7 and F1 were used for validation.
510 The first three days were considered as the warming-up time period for the FSS model setting
511 up as stated in Eq. (9), i.e., $T_w=3$ days. The observations of the next 14 days were used for
512 FSS model calibration, and the remaining observations of 14 days were utilized for validating
513 the performance of the real-time FSS models. The linear scale function $g()$ in Eq. (9) for
514 each case study is defined as

$$515 \quad g(x) = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (17)$$

516 where x represents the observed or simulated values at monitoring points; x_{\min} and x_{\max}
517 are the lower and upper bounds, respectively. These two parameters for each monitoring point
518 are determined by analyzing historical observation data over 14 days (i.e., the calibration time
519 period) in this paper.

520 The evolutionary algorithm Borg (Hadka and Reed, 2013) was selected to solve the proposed

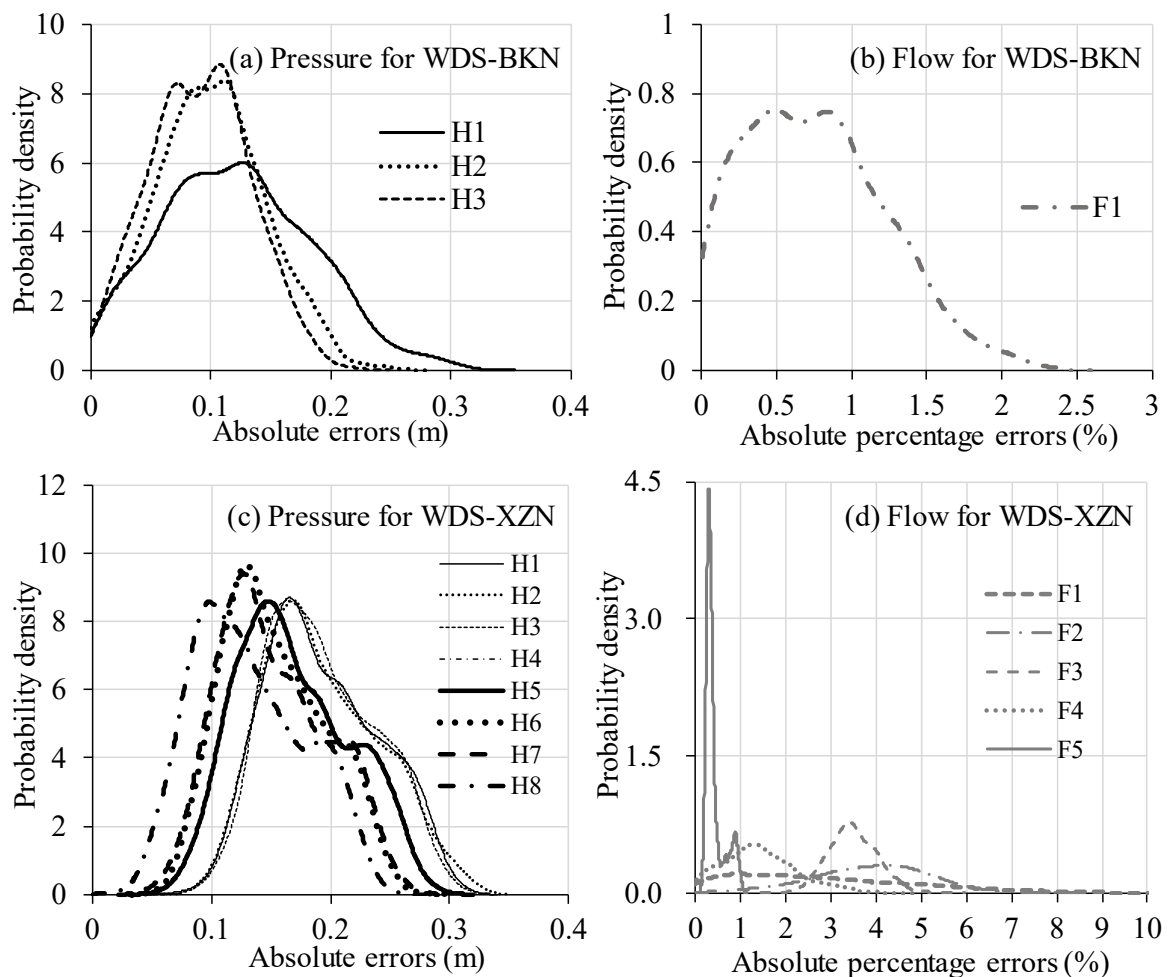
521 calibration problem defined in Eq. (9-13) due to its great performance in handling complex
522 urban water resources and engineering optimization problems (Reed et al., 2013, Zheng et al.,
523 2016). The initial population size of Borg applied to BKN and XZN case studies were 500
524 and 1,000 respectively, and the maximum allowable iterations are 50,000 for both case
525 studies. The default values of the other parameters of Borg were used in this study as they
526 have been validated through various applications (Wang et al., 2014). Five Borg runs with
527 different random number seeds were applied to each case study, and the results showed that
528 the final optimization results were overall similar across different runs. Therefore, the results
529 of a typical Borg run were presented to enable discussions for each of the two FSS case
530 studies.

531 **4. Results and discussions**

532 **4.1 Calibration results of WDS nodal water consumptions**

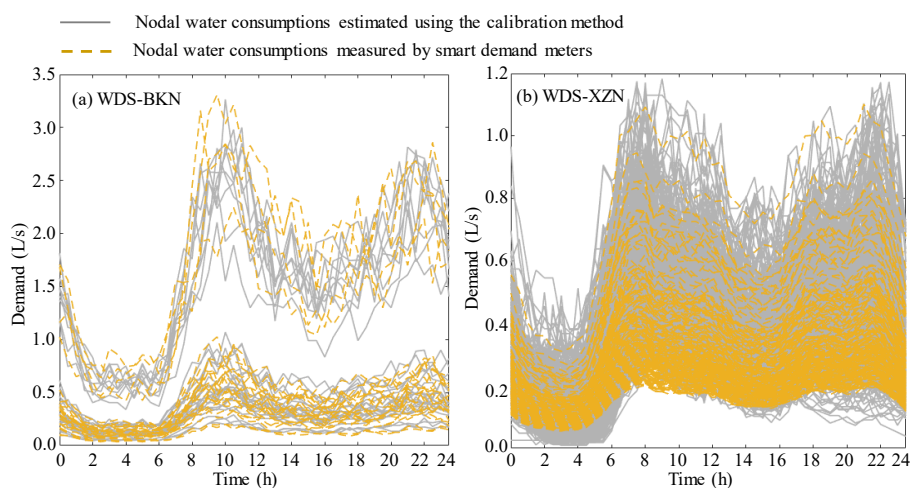
533 For each FSS case study, nodal water consumptions of its associated WDS need to be
534 calibrated at each time period, resulting in a total of 816 calibration runs using the calibration
535 method as described in Section 2.3. The resultant time consumption was approximately 25
536 seconds and 10 minutes for WDS-BKN and WDS-XZN systems, respectively, on a PC with a
537 2.60-GHz Intel Core i9-7980XE and 2 GB of RAM. Fig. 7 shows the density plot of the
538 errors between observations and simulations at the monitoring locations for both case studies. It
539 is seen that, for the WDS-BKN case, more than 90% of the absolute errors (*AEs*) is less than
540 0.30m for each pressure monitoring point (including the H2 sensor used for model validation),
541 with the maximum *AE* being 0.32m across the three pressure monitoring points. In terms of

542 flow, more about 93% of absolute percentage errors (*APEs*) are smaller than 1.5%, with the
 543 maximum *APE* being 2.40% as shown in Fig. 7(a, b). For the WDS-XZN (Fig. 7(c, d)), the
 544 differences between the simulated and observed pressure values at the eight monitoring
 545 locations are negligible (including H2, H5 and H7 used for validation), with all *AEs* being
 546 lower than 0.4m. Relative to pressure, the deviations between the flow simulations and
 547 observations are slightly larger (Figure 7(d)), with the majority of *APEs* smaller than 5% and
 548 the maximum *APE* being 9.8% (F1 used for validation).



549
 550 **Fig. 7 The probability density distribution of the errors between observations and**
 551 **simulations for all monitoring locations of the WDS-BKN and WDS-XZN**

552 To further demonstrate the quality of calibration results, the criteria defined in Walski et al.
553 (2003) were used to verify the simulation accuracies. As stated in Walski et al. (2003), a
554 satisfactory WDS model calibration should ensure 85% of pressure errors within $\pm 0.2\text{m}$, 100%
555 of pressure errors within $\pm 0.5\text{ m}$, trunk main flow errors (flows more than 10% of the total
556 demands) within $\pm 5\%$, and the other flow errors within $\pm 10\%$. The calibration results of the
557 two WDSs satisfied these criteria, implying that the calibration was successful as the resultant
558 nodal water consumptions can reproduce the overall hydraulics of the WDS. Figure 8 presents
559 the nodal water consumptions over the 31 days with a resolution of 30-minute for the two
560 WDSs ((a) for the WDS-BKN and (b) for WDS-XZN), where the grey solid lines represent the
561 calibrated nodal water consumptions and the orange dotted lines indicate the nodal water
562 consumptions measured by smart demand meters. Despite some variations, all the nodal water
563 consumptions exhibited an overall similar trend for both WDSs, with two peak demand periods
564 occurring at each demand node as shown in Fig. 8, which matches well with the typical water
565 use properties (Zhang et al., 2018).



566

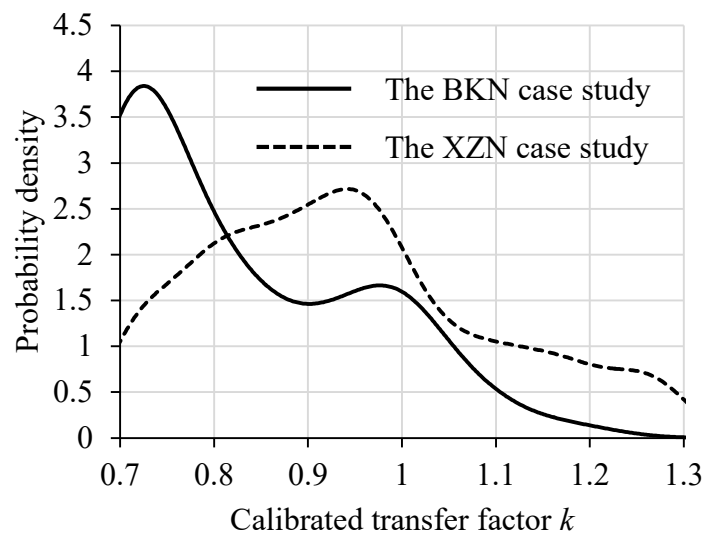
567 **Fig. 8 Nodal water consumptions of the two WDSs at a typical day with a 30-min time**

568

resolution

569 **4.2 Estimated transfer factor values**

570 Figure 9 shows the distribution of the probability density of the identified transfer factor k
571 values for all manholes of the BKN and XZN based on the historical data over the first 17
572 calibration days (observations of the first three days were used as model setting-up). Such an
573 optimization (Section 2.4) took 4.86 and 56 hours respectively based on the same computing
574 platform as mentioned above. It can be seen that the majority of k values is within the range
575 of 0.7~1.0 for the BKN and XZN, with a mean value of 0.83 and 0.92 respectively, meaning
576 that around 83% and 92% of the total water consumptions have been collected by the FSS of
577 BKN and XZN in this area, respectively. This demonstrates that the calibrated k values for all
578 manholes were overall practically meaningful (Behzadian and Kapelan, 2015).



579

580 **Fig. 9 The density probability distribution of the identified transfer factor k for the BKN**
581 **and XZN case study**

582 It is noted that around 10% and 28% calibrated k values were greater than 1 as shown in
583 Figure 1. Such values were only allowed for the WDS nodes without smart demand sensors,
584 and hence their nodal water consumptions were estimated using the calibration method

585 described in Section 2.3. While Fig. 7 showed that the calibration results can reproduce the
586 overall hydraulics of the WDS at the monitoring locations, the calibrated nodal water
587 consumptions might inevitably deviate from the true values at a certain extent (Zhang et al.,
588 2018). To mitigate this potential impact, the value of k for the FSS manholes associated with
589 WDS nodes without smart demand meters was allowed to have a range between 0.7 and 1.3,
590 as previously stated. This led to that a proportion of k values were greater than 1 as shown in
591 Fig. 9.

592 Fig. 10 shows the FSS calibrated results (the first 17 days) corresponding to the transfer factor
593 values presented in Figure 9. It is seen that the simulated flows in C1 in the small BKN case
594 study matched well with the observations (Fig. 10(a)), where all APE values were lower than
595 5.0% and the mean APE value was 1.16%. For the XZN case study (Fig. 10(e)), the maximum
596 and the mean APE values between simulations and observations within the calibration period
597 at the C3 monitoring location were 13.68% and 3.02% respectively. Therefore, it can be
598 deduced that the simulations matched well with the observations for such a large XZN case
599 study. While the APE values at the period with relatively low sewer flows were relatively large,
600 their corresponding absolute errors (AEs) were overall low as shown in Figure 10 (c,f). For
601 example, the maximum AE value was 1.88 L/s for the BKN case study with an average flow of
602 48.47 L/s in C1 (Fig. 10(c)). Similarly, the maximum AE value was 30.51 L/s for the XZN case
603 study with an average flow of 255.10 L/s in C3 (Fig.10(f)).

604 Tables 1 and 2 present the values of the performance metrics applied to the simulations and
605 observations at monitoring locations for both case studies. As shown in this table, for the BKN

606 case study, the averaged values of *MAPE*, R^2 , *NSE* and *KGE* over four different monitoring
607 locations within the calibration period are 3.61%, 0.99, 0.94 and 0.94 respectively. For the large
608 XZN case study, the averaged values of *MAPE*, R^2 , *NSE* and *KGE* over 11 different monitoring
609 locations within the calibration period are 4.98%, 0.98, 0.89 and 0.93 respectively. This implied
610 that FSS calibration (aimed to estimate the transfer factor) was overall successful.

616 **Table 1 Values of the performance metrics applied to the simulations and observations**

617 **within the validation period for the BKN case study**

Sensor ID	Calibration period				Validation period			
	<i>MAPE</i> (%)	<i>R</i> ²	<i>NSE</i>	<i>KGE</i>	<i>MAPE</i>	<i>R</i> ²	<i>NSE</i>	<i>KGE</i>
M1	3.18	0.99	0.97	0.93	3.20	0.99	0.96	0.93
M2	2.06	0.99	0.97	0.93	2.12	0.99	0.97	0.92
M3	8.05	0.99	0.84	0.89	8.04	0.99	0.84	0.88
C1	1.16	0.99	0.99	0.99	1.15	0.99	0.99	0.98
Average	3.61	0.99	0.94	0.94	3.63	0.99	0.94	0.93

618 **Table 2 Values of the performance metrics applied to the simulations and observations**

619 **within the validation period for the XZN case study**

Sensor ID	Calibration period				Validation period			
	<i>MAPE</i> (%)	<i>R</i> ²	<i>NSE</i>	<i>KGE</i>	<i>MAPE</i> (%)	<i>R</i> ²	<i>NSE</i>	<i>KGE</i>
M1	7.85	0.98	0.78	0.87	7.79	0.97	0.77	0.87
M2	6.49	0.97	0.83	0.90	6.81	0.96	0.80	0.88
M3	6.82	0.98	0.81	0.89	6.43	0.96	0.81	0.89
M4	8.09	0.98	0.77	0.87	8.08	0.97	0.77	0.87
M5	2.83	0.98	0.95	0.96	3.33	0.96	0.94	0.96
M6	7.09	0.97	0.79	0.89	7.07	0.96	0.79	0.88
M7	3.56	0.98	0.93	0.91	4.23	0.96	0.91	0.91
M8	3.10	0.97	0.95	0.92	3.45	0.96	0.94	0.91
C1	2.89	0.99	0.99	0.99	3.20	0.99	0.99	0.99
C2	3.00	0.99	0.99	0.99	3.57	0.98	0.98	0.99
C3	3.02	0.99	0.99	0.99	3.62	0.98	0.98	0.99
Average	4.98	0.98	0.89	0.93	5.23	0.97	0.88	0.92

620 **4.3 Performance of the real-time FSS modelling**

621 Results in the calibration period demonstrated that the proposed method was capable of

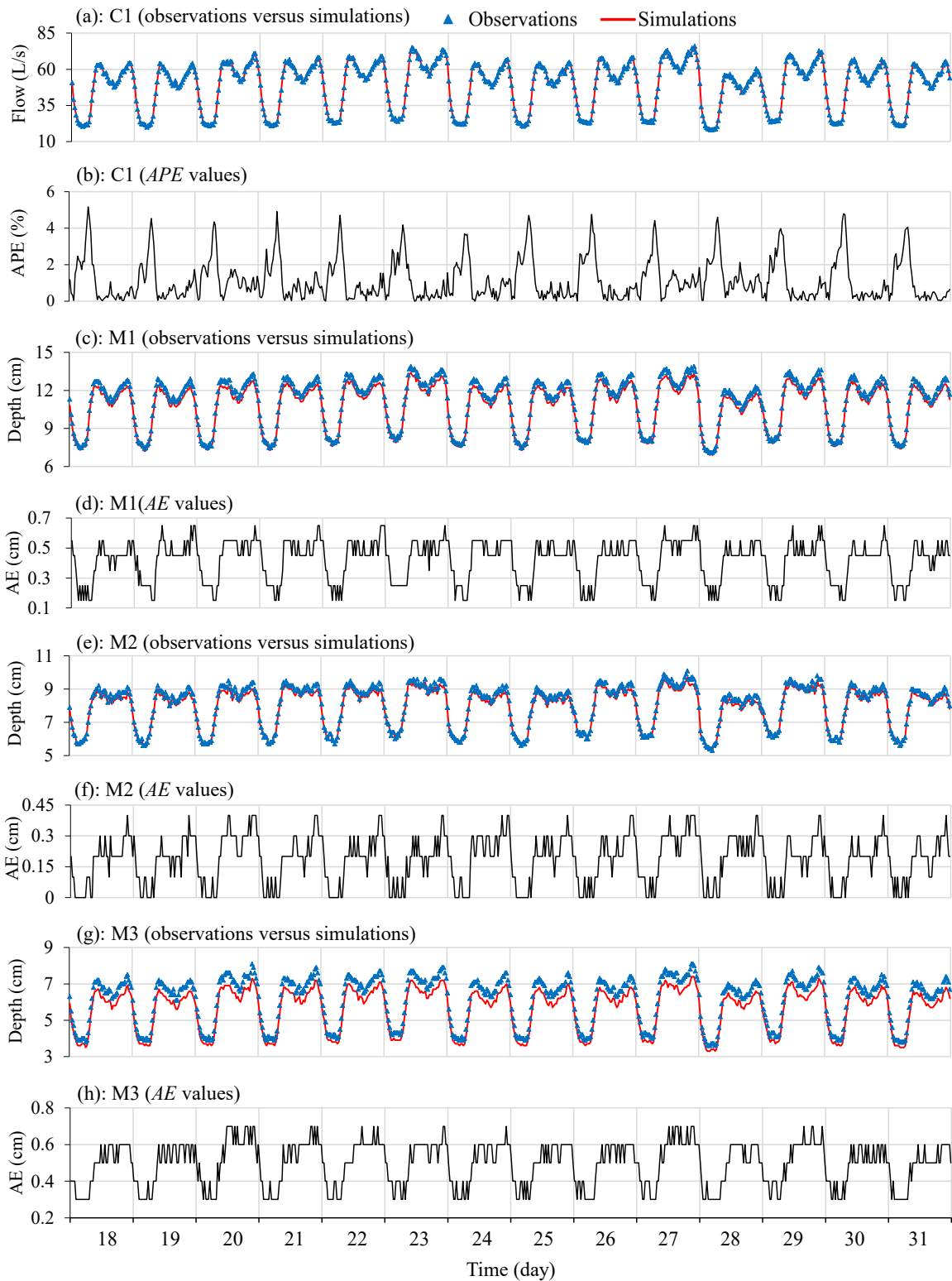
622 identifying suitable transfer factors that can match well simulations and observations at the

623 monitoring locations. This section validated the performance of the real-time FSS models

624 driven by the WDS consumption data in modelling the sewer hydraulics and such a

625 performance evaluation was conducted using the observations from the 17th to the 31st days (i.e.,
626 validation period). The steps of the real-time FSS modelling were presented in Figure 2
627 (real-time module). Figures 11 and 12 show the observations versus observations of the
628 monitoring locations every 30 minutes within the validation period for both case studies.

629 It is seen from Figure 11 that the sewer flow and the water depth simulations matched well with
630 the observations within the validation period at the four monitoring locations (C1, M1, M2 and
631 M3) in the BKN case study. More specifically, the maximum flow *APE* value was 4.91%, and
632 the maximum absolute error of water depth was 0.7 cm across M1, M2 and M3 locations.
633 Similarly, the differences between the simulation and observations for C1, C2, M1 and M5
634 monitoring locations were also matched very well for the XZN case study as shown in Fig.
635 12. For this large FSS, the maximum flow *APE* value was 13.45% and the maximum absolute
636 error of water depth was 1.4 cm (similar observations can be made for other monitoring
637 locations).



638

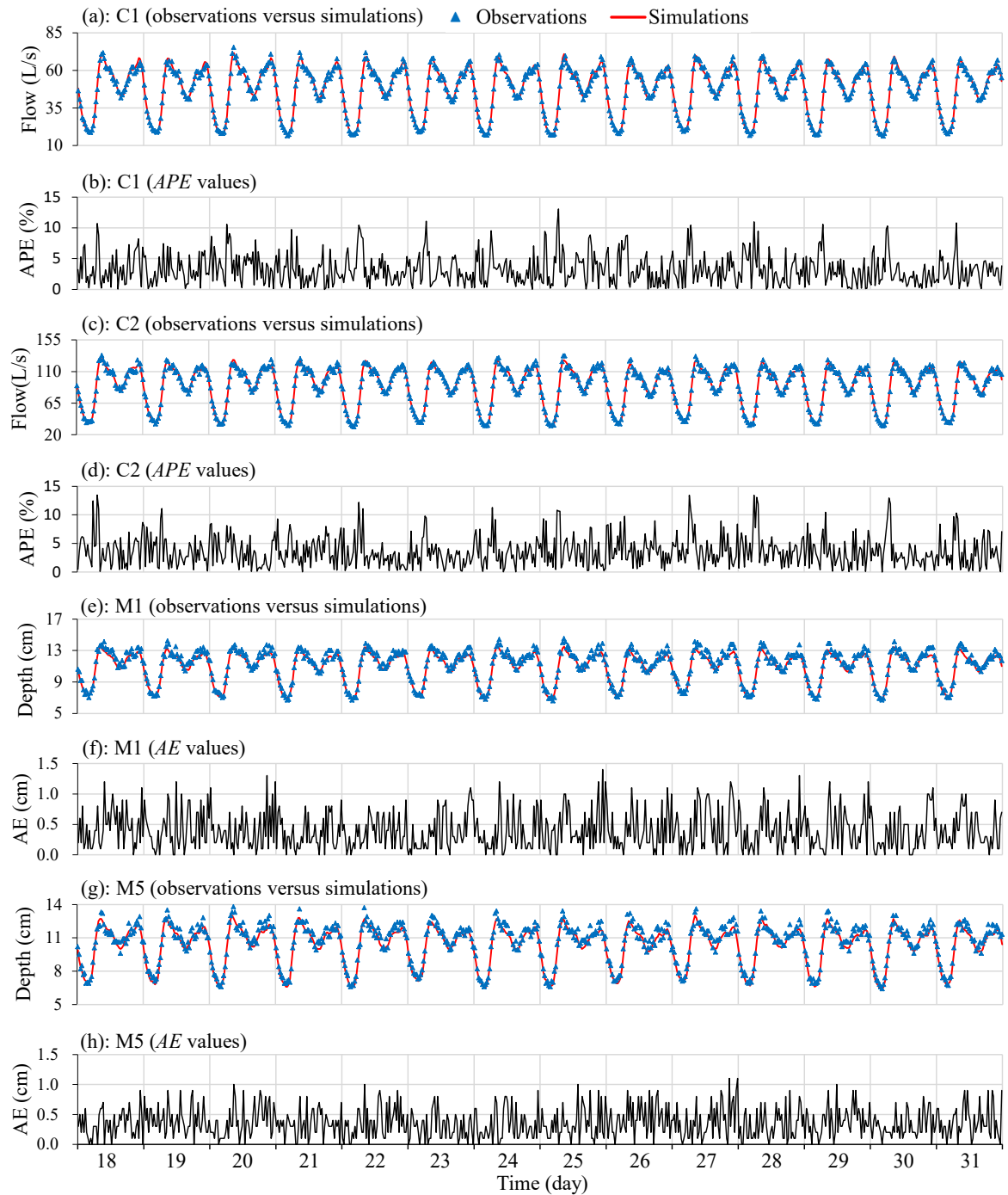
639

Fig. 11 Observations versus simulations, as well as the APEs or AEs for the four

640

monitoring locations (shown in Fig. 4) within the validation period of the BKN case study

641



642

643 **Fig. 12 Observations versus simulations, as well as the APEs or AEs for the four**
 644 **monitoring locations (shown in Fig. 5) within the validation period of the XZN case study**

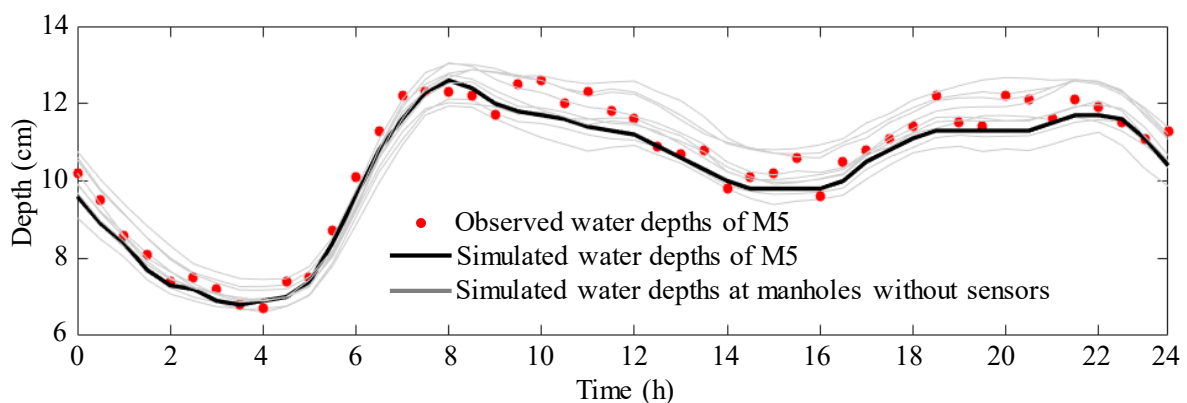
645 The values of performance metrics applied to the observations and simulations within the

646 validation period for both case studies are also presented in Tables 1 and 2 respectively. As

647 shown in these two tables, the averaged values of *MAPE*, R^2 , *NSE* and *KGE* over four different
648 monitoring locations within the validation period are 3.63%, 0.99, 0.94 and 0.93 respectively
649 for the BKN case study. The averaged values of *MAPE*, R^2 , *NSE* and *KGE* over four different
650 monitoring locations within the validation period are 5.23%, 0.97, 0.88 and 0.92 respectively
651 for the XZN case study. Overall, the performance of the FSS models within the validation
652 period was similar or slightly worse than the calibration period for both case studies (see Tables
653 1 and 2). This indicated that (i) there was a low likelihood of over-fitting within the calibration
654 process due to the similar performance between the calibration and validation period, and (ii)
655 the real-time FSS models driven by WDS water consumption data were effective in accurately
656 simulating the sewer hydraulics at a high time resolution (very 30 minutes).

657 The real-time model was able to offer a great opportunity to enable the comparison between the
658 simulations and observations at monitoring locations at a very high time resolution (every 30
659 minutes in this paper), followed by a warning trigger if large deviations between the
660 simulations and observations were observed. More specifically, a threshold can be determined
661 by long-term historical data for each monitoring location as did in Qi et al. (2018). If the
662 deviations between the simulations and observations at a particular monitoring location go
663 beyond the specified range, a warning can be triggered efficiently. It should be highlighted that
664 since the real-time FSS model developed using the proposed method has already accounted for
665 the inflow variation caused by the change in water consumption, the false warning rate is
666 expected to be significantly reduced. Therefore, the proposed real-time FSS model can be a
667 useful tool for the development of an efficient warning system, aimed to detect the potential
668 hydraulic issues (e.g., leaks and illicit inflows) for the FSSs.

669 In addition to providing accurate simulations at the monitoring locations, the proposed method
670 was also able to produce real-time simulations for the manholes and sewer pipes without
671 monitoring sensors. While the accuracies of these simulations cannot be directly evaluated due
672 to the unavailability of observations, it can be anticipated that they can reasonably represent the
673 true hydraulics of the manholes and sewer pipes without monitoring sensors. This was because
674 the real-time FSS model was driven by the water consumption data from the water distribution
675 system, where nodal water consumptions were either measured by smart demand meters or
676 estimated with the aid of an intensive sensor (pressure and flow sensors) coverage. As shown in
677 Figure 13, water depths of 10 manholes near M5 sensor of the XZN case study over a typical
678 day within the validation period exhibited a similar and reasonable trend. These accurate
679 hydraulic simulations at the manholes and pipes without monitoring sensors can be useful to
680 enable the efficient localization of leaks, deposits or illicit inflows, through comparing the
681 simulations with the sampled observations from the field survey.



682

683 **Fig. 13 Water depth simulations and observations of M5, as well as the water depth**
684 **simulations of 10 manholes near M5 without sensors in the XZN case study in 18th day**
685 **within the validation period**

686 5. Conclusions

687 This paper proposes a novel method to develop a real-time foul sewer system (FSS) model
688 driven by water consumption data from its associated water distribution system (WDS) that
689 often has a large number of sensors such as pressure sensors, flow meters and smart demand
690 meters. Within the proposed method, the FSS and the WDS models are integrated to build
691 physical connections between water consumption nodes and their corresponding manholes
692 based on spatial distances. This is followed by a proposal of an optimization approach to
693 identify the transfer factor k between nodal water consumptions and FSS manhole inflows
694 according to historical observations. Subsequently, real-time nodal water consumption data
695 are acquired using an efficient calibration approach based on the dense sensors in the WDS.
696 Finally, these nodal water consumption data combined with the identified k values drive the
697 FSS real-time modelling.

698 Two real FSS case studies, the smaller BKN with 64 sewer pipes and 64 manholes and the large
699 ZXN case study with 1214 sewer pipes and 1214 manholes have been used to test/validate and
700 demonstrate the proposed method. The results obtained demonstrate that the proposed method
701 can produce real-time predictions of water depths and flows that are in good agreements with
702 the corresponding observations at monitoring locations. The evidence for this can be found in
703 the high mean values of R^2 , NSE and KGE metrics obtained across different monitoring
704 locations, which are 0.99, 0.94 and 0.93 of the small BNK case study, and 0.97, 0.88 and 0.92
705 for the large XZN case study, respectively. In addition to providing accurate simulations at the
706 monitoring locations, the proposed method is expected to produce reasonable real-time

707 simulations for the manholes and sewer pipes without monitoring sensors. This deduction is
708 based on that the real-time FSS model is driven by the WDS water consumption data that are
709 either measured by smart demand meters or estimated based on a large number of sensors
710 (pressure and flow sensors). This implies that the “equifinality” problem can be successfully
711 addressed by using the proposed method. Therefore, the developed real-time FSS model offers
712 an important tool to facilitate effective and efficient foul sewer system management and
713 operation.

714 Finally, it is acknowledged that the proposed method is developed ignoring a number of
715 uncertainties that exist in reality. These include potential inaccuracies of WDS and FSS sensor
716 measurements (e.g., smart demand meters, water depth sensors, flow sensors), the potential
717 impacts of the ignorance of the water travelling time within the user property and the influence
718 of the variation in Manning’s coefficients of the sewer pipes. These uncertainties need to be
719 more systematically investigated in a future study.

720 **Acknowledgement**

721 This work is funded by the National Natural Science Foundation of China (Grant No.
722 51922096), and the Excellent Youth Natural Science Foundation of Zhejiang Province, China
723 (LR19E080003).

724 **References**

725 Ahm, M., Thorndahl, S., Nielsen, J.E. and Rasmussen, M.R. (2016) Estimation of combined
726 sewer overflow discharge: a software sensor approach based on local water level
727 measurements. *Water Science and Technology* 74(11), 2683-2696.

728 Bailey, O., Arnot, T.C., Blokker, E.J.M., Kapelan, Z., Vreeburg, J. and Hofman, J. (2019)
729 Developing a stochastic sewer model to support sewer design under water conservation
730 measures. *Journal of Hydrology* 573, 908-917.

731 Banik, B.K., Di Cristo, C., Leopardi, A. and de Marinis, G. (2017) Illicit intrusion
732 characterization in sewer systems. *Urban Water Journal* 14(4), 416-426.

733 Beheshti, M. and Saegrov, S. (2019) Detection of extraneous water ingress into the sewer
734 system using tandem methods - a case study in Trondheim city. *Water Science and*
735 *Technology* 79(2), 231-239.

736 Behzadian, K. and Kapelan, Z. (2015) Modelling metabolism based performance of an urban
737 water system using WaterMet(2). *Resources Conservation and Recycling* 99, 84-99.

738 Behzadian, K., Kapelan, Z., Venkatesh, G., Brattebø, H. and Sægrov, S. (2014) WaterMet2: A
739 tool for integrated analysis of sustainability-based performance of urban water systems.
740 *Drinking Water Engineering Sciences* 7, 63-72.

741 Black, J. and Endreny, T. (2006) Increasing stormwater outfall duration, magnitude, and
742 volume through combined sewer separation. *Journal of Hydrologic Engineering* 11(5),
743 472-481.

744 Breinholt, A., Grum, M., Madsen, H., Thordarson, F.O. and Mikkelsen, P.S. (2013) Informal
745 uncertainty analysis (GLUE) of continuous flow simulation in a hybrid sewer system
746 with infiltration inflow - consistency of containment ratios in calibration and validation?
747 *Hydrology and Earth System Sciences* 17(10), 4159-4176.

748 Broekhuizen, I., Leonhardt, G., Marsalek, J. and Viklander, M. (2020) Event selection and
749 two-stage approach for calibrating models of green urban drainage systems. *Hydrology*

750 and Earth System Sciences 24(2), 869-885.

751 Bruen, M. and Yang, J.Q. (2006) Combined hydraulic and black-box models for flood
752 forecasting in urban drainage systems. Journal of Hydrologic Engineering 11(6),
753 589-596.

754 Burke, C., Molzahn, R., Pherson, P. and Coutts, P. (1986) An evaluation of a sanitary sewer
755 system using a computer model. Tunnelling and Underground Space Technology -
756 TUNN UNDERGR SPACE TECHNOL 1, 153-161.

757 Creaco, E., Campisano, A., Fontana, N., Marini, G., Page, P.R. and Walski, T. (2019) Real
758 time control of water distribution networks: A state-of-the-art review. Water Research
759 161, 517-530.

760 Creaco, E., Signori, P., Papiri, S. and Ciaponi, C. (2018) Peak Demand Assessment and
761 Hydraulic Analysis in WDN Design. Journal of Water Resources Planning and
762 Management 144(6).

763 Di Pierro, A., Hankin, C. and Wiklicky, H. (2005) Quantitative static analysis of distributed
764 systems. Journal of Functional Programming 15, 703-749.

765 Du, K., Long, T.-Y., Wang, J.-H. and Guo, J.-S. (2015) Inversion Model of Water Distribution
766 Systems for Nodal Demand Calibration. Journal of Water Resources Planning and
767 Management 141(9).

768 Eren, B. and Karadagli, F. (2012) Physical Disintegration of Toilet Papers in Wastewater
769 Systems: Experimental Analysis and Mathematical Modeling. Environmental Science &
770 Technology 46(5), 2870-2876.

771 Garda, A., Castillo, F., Binet, G., Litrico, X. and Gil, A. (2016) Needs and potential of

772 unmanned vehicles in sewers. *Houille Blanche-Revue Internationale De L Eau* (1),
773 24-29.

774 Gujarati, D.N. (2009) *Basic econometrics*. Tata McGraw-Hill Education.

775 Guo, D., Zheng, F., Gupta, H. and Maier, H.R. (2020) On the Robustness of Conceptual
776 Rainfall-Runoff Models to Calibration and Evaluation Data Set Splits Selection: A Large
777 Sample Investigation. *Water Resources Research* 56(3).

778 Hadka, D. and Reed, P. (2013) Borg: An Auto-Adaptive Many-Objective Evolutionary
779 Computing Framework. *Evolutionary Computation* 21(2), 231-259.

780 He, G., Zhang, T., Zheng, F. and Zhang, Q. (2018) An efficient multi-objective optimization
781 method for water quality sensor placement within water distribution systems considering
782 contamination probability variations. *Water Research* 143, 165-175.

783 Huang, D., Liu, X.H., Jiang, S.Z., Wang, H.C., Wang, J.Y. and Zhang, Y.K. (2018) Current
784 state and future perspectives of sewer networks in urban China. *Frontiers of*
785 *Environmental Science & Engineering* 12(3), 16.

786 Huang, Y., Zheng, F., Duan, H.-F., Zhang, T., Guo, X. and Zhang, Q. (2019) Skeletonizing
787 Pipes in Series within Urban Water Distribution Systems Using a Transient-Based
788 Method. *Journal of Hydraulic Engineering* 145(2).

789 Irvine, K., Maryc, R., Vermette, S., Bakert, J. and Kleinfelder, K. (2011) Illicit discharge
790 detection and elimination: Low cost options for source identification and trackdown in
791 stormwater systems. *Urban Water Journal* 8(6), 379-395.

792 Joseph-Duran, B., Ocampo-Martinez, C. and Cembrano, G. (2014) Hybrid modeling and
793 receding horizon control of sewer networks. *Water Resources Research* 50(11),

794 8497-8514.

795 Kapelan, Z. (2002) Calibration of water distribution system hydraulic models.

796 Khu, S.T., di Pierro, F., Savic, D., Djordjevic, S. and Walters, G.A. (2006) Incorporating
797 spatial and temporal information for urban drainage model calibration: An approach
798 using preference ordering genetic algorithm. *Advances in Water Resources* 29(8),
799 1168-1181.

800 Kleidorfer, M., Leonhardt, G. and Rauch, W. (2012) Identifiability analysis in conceptual
801 sewer modelling. *Water Science and Technology* 66(7), 1467-1474.

802 Knobens, W.J.M., Freer, J.E. and Woods, R.A. (2019) Technical note: Inherent benchmark or
803 not? Comparing Nash-Sutcliffe and Kling-Gupta efficiency scores. *Hydrology and Earth
804 System Sciences* 23(10), 4323-4331.

805 Koch, M.W. and McKenna, S.A. (2011) Distributed Sensor Fusion in Water Quality Event
806 Detection. *Journal of Water Resources Planning and Management* 137(1), 10-19.

807 Korving, H. and Clemens, F. (2005) Impact of dimension uncertainty and model calibration
808 on sewer system assessment. *Water Science and Technology* 52(5), 35-42.

809 Lepot, M., Makris, K.F. and Clemens, F. (2017) Detection and quantification of lateral, illicit
810 connections and infiltration in sewers with Infra-Red camera: Conclusions after a wide
811 experimental plan. *Water Research* 122, 678-691.

812 Li, T., Zhang, W., Feng, C. and Shen, J. (2014) Performance assessment of separate and
813 combined sewer systems in metropolitan areas in southern China. *Water Science and
814 Technology* 69(2), 422-429.

815 Lim, J.S., Kim, J., Friedman, J., Lee, U., Vieira, L., Rosso, D., Gerla, M. and Srivastava, M.B.

816 (2013) SewerSnort: A drifting sensor for in situ Wastewater Collection System gas
817 monitoring. *Ad Hoc Networks* 11(4), 1456-1471.

818 Liu, Y., Tugtas, A.E., Sharma, K.R., Ni, B.-J. and Yuan, Z. (2016) Sulfide and methane
819 production in sewer sediments: Field survey and model evaluation. *Water Research* 89,
820 142-150.

821 McCall, A.-K., Bade, R., Kinyua, J., Lai, F.Y., Thai, P.K., Covaci, A., Bijlsma, L., van Nuijs,
822 A.L.N. and Ort, C. (2016) Critical review on the stability of illicit drugs in sewers and
823 wastewater samples. *Water Research* 88, 933-947.

824 Mu, L., Zheng, F., Tao, R., Zhang, Q. and Kapelan, Z. (2020) Hourly and Daily Urban Water
825 Demand Predictions Using a Long Short-Term Memory Based Model. *Journal of Water
826 Resources Planning and Management* 146(9).

827 Nash, J.E. and Sutcliffe, J. (1970) River Flow Forecasting Through Conceptual Models: Part
828 1. — A Discussion of Principles. *Journal of Hydrology* 10, 282.

829 Qi, Z., Zheng, F., Guo, D., Maier, H.R., Zhang, T., Yu, T. and Shao, Y. (2018) Better
830 Understanding of the Capacity of Pressure Sensor Systems to Detect Pipe Burst within
831 Water Distribution Networks. *Journal of Water Resources Planning and Management*
832 144(7).

833 Rauch, W., Bertrand-Krajewski, J. L., Krebs, P., Mark, O., Schilling, W., Schütze, M., and
834 Vanrolleghem, P. A. (2001). Mathematical modelling of integrated urban drainage
835 systems. Keynote paper. In *Second International Conference on Interactions between
836 sewers, treatment plants and receiving waters-INTERURBA II, Lisbon/Portugal*, 89-106.

837 Reed, P.M., Hadka, D., Herman, J.D., Kasprzyk, J.R. and Kollat, J.B. (2013) Evolutionary

838 multiobjective optimization in water resources: The past, present, and future. *Advances*
839 *in Water Resources* 51, 438-456.

840 Rokstad, M.M. and Ugarelli, R.M. (2015) Evaluating the role of deterioration models for
841 condition assessment of sewers. *Journal of Hydroinformatics* 17(5), 789-804.

842 Rossman, L.A. (2000) users manual. National Risk Management Research Laboratory.

843 Rossman, L.A. (2010) Storm Water Management Model User's Manual.

844 Rossman, L.A. and Huber, W. (2017) Storm water management model reference manual
845 volume II–hydraulics. US Environmental Protection Agency, 190.

846 Sambito, M., Di Cristo, C., Freni, G. and Leopardi, A. (2020) Optimal water quality sensor
847 positioning in urban drainage systems for illicit intrusion identification. *Journal of*
848 *Hydroinformatics* 22(1), 46-60.

849 Sara C. Troutman, Nancy G. Love, Branko Kerkez. (2020) Balancing water quality and flows
850 in combined sewer systems using real-time control. *Environmental Science: Water*
851 *Research & Technology* 6 (5), pages 1357-1369.

852 Savic, D.A., Kapelan, Z.S. and Jonkergouw, P.M.R. (2009) Quo vadis water distribution
853 model calibration? *Urban Water Journal* 6(1), 3-22.

854 Schilperoort, R., Hoppe, H., de Haan, C. and Langeveld, J. (2013) Searching for storm water
855 inflows in foul sewers using fibre-optic distributed temperature sensing. *Water Science*
856 *and Technology* 68(8), 1723-1730.

857 Schütze M, Campisano A, Colas H, Schilling W and Vanrolleghem, P. A. (2002). Real-time
858 control of urban wastewater systems-where do we stand today? *Global Solutions for*
859 *Urban Drainage*, 1-17.

860 Schütze, M., Campisano, A., Colas, H., Vanrolleghem, P., and Schilling, W. (2003). Real-time
861 control of urban water systems. In International Conference on Pumps,
862 Electromechanical Devices and Systems Applied to Urban Water Management PEDS,
863 22-25.

864 Seco, I., Schellart, A., Gomez-Valentin, M. and Tait, S. (2018) Prediction of Organic
865 Combined Sewer Sediment Release and Transport. Journal of Hydraulic Engineering
866 144(3).

867 Sweetapple, C., Astaraie-Imani, M. and Butler, D. (2018) Design and operation of urban
868 wastewater systems considering reliability, risk and resilience. Water Research 147,
869 1-12.

870 Talaiekhosani, A., Bagheri, M., Goli, A. and Khoozani, M.R.T. (2016) An overview of
871 principles of odor production, emission, and control methods in wastewater collection
872 and treatment systems. Journal of Environmental Management 170, 186-206.

873 Walski, T.M., Chase, D.V., Savic, D.A., Grayman, W., Beckwith, S. and Koelle, E. (2003)
874 Advanced water distribution modeling and management.

875 Wang, Q., Guidolin, M., Savic, D. and Kapelan, Z. (2015) Two-Objective Design of
876 Benchmark Problems of a Water Distribution System via MOEAs: Towards the
877 Best-Known Approximation of the True Pareto Front. Journal of Water Resources
878 Planning and Management 141(3).

879 Wu, Z.Y., Farley, M., Turtle, D., Kapelan, Z., Boxall, J., Mounce, S., Dahasahasra, S., Mulay,
880 M. and Kleiner, Y. (2011) Water loss reduction, Bentley Institute Press.

881 Wu, Z.Y., Sage, P. and Turtle, D. (2010) Pressure-Dependent Leak Detection Model and Its

882 Application to a District Water System. *Journal of Water Resources Planning and*
883 *Management* 136(1), 116-128.

884 Xie, X., Zhang, H. and Hou, D. (2017) Bayesian Approach for Joint Estimation of Demand
885 and Roughness in Water Distribution Systems. *Journal of Water Resources Planning and*
886 *Management* 143(8).

887 Zhang, Q., Zheng, F., Duan, H.-F., Jia, Y., Zhang, T. and Guo, X. (2018) Efficient Numerical
888 Approach for Simultaneous Calibration of Pipe Roughness Coefficients and Nodal
889 Demands for Water Distribution Systems. *Journal of Water Resources Planning and*
890 *Management* 144(10).

891 Zheng, F., Tao, R., Maier, H.R., See, L., Savic, D., Zhang, T., Chen, Q., Assumpcao, T.H.,
892 Yang, P., Heidari, B., Rieckermann, J., Minsker, B., Bi, W., Cai, X., Solomatine, D. and
893 Popescu, I. (2018) Crowdsourcing Methods for Data Collection in Geophysics: State of
894 the Art, Issues, and Future Directions. *Reviews of Geophysics* 56(4), 698-740.

895 Zheng, F., Zecchin, A.C., Maier, H.R. and Simpson, A.R. (2016) Comparison of the
896 Searching Behavior of NSGA-II, SAMODE, and Borg MOEAs Applied to Water
897 Distribution System Design Problems. *Journal of Water Resources Planning and*
898 *Management* 142(7).

899 Zheng, F., Zecchin, A.C., Newman, J.P., Maier, H.R. and Dandy, G.C. (2017) An Adaptive
900 Convergence-Trajectory Controlled Ant Colony Optimization Algorithm With
901 Application to Water Distribution System Design Problems. *Ieee Transactions on*
902 *Evolutionary Computation* 21(5), 773-791.