



## Automatic mapping of urban wastewater networks based on manhole cover locations

Nanée Chahinian, Carole Delenne, Benjamin Commandre, Mustapha Derras, Laurent Deruelle, Jean-Stéphane Bailly

### ► To cite this version:

Nanée Chahinian, Carole Delenne, Benjamin Commandre, Mustapha Derras, Laurent Deruelle, et al.. Automatic mapping of urban wastewater networks based on manhole cover locations. Computers, Environment and Urban Systems, 2019, 78, pp.101370. 10.1016/j.compenvurbsys.2019.101370 . hal-02275903

**HAL Id: hal-02275903**

**<https://hal.science/hal-02275903>**

Submitted on 11 Sep 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Automatic mapping of urban wastewater networks based on manhole cover locations

Nanée Chahinian<sup>a</sup>, Carole Delenne<sup>a,b</sup>, Benjamin Commandré<sup>a,b</sup>, Mustapha Derras<sup>c</sup>, Laurent Deruelle<sup>c</sup>, Jean-Stéphane Bailly<sup>d,e</sup>

<sup>a</sup>*HSM, Univ. Montpellier, CNRS, IRD, Montpellier, France*

<sup>b</sup>*Inria Lemon, CRISAM - Inria Sophia Antipolis - Méditerranée, France*

<sup>c</sup>*Berger-Levrault, Montpellier, France*

<sup>d</sup>*LISAH, Univ Montpellier, INRA, IRD, Montpellier SupAgro, Montpellier, France*

<sup>e</sup>*AgroParisTech, Paris, France*

---

## Abstract

Accurate maps of wastewater networks in cities are mandatory for an integrated management of water resources. However, in many countries around the world this information is unavailable or inaccurate. A new mapping method is put forward to create wastewater network maps using manhole cover locations as a prime information source. These locations could be available via ground surveys, remote sensing techniques or stakeholder's databases. A new algorithm is developed which considers manhole covers as the nodes of the network and connects them automatically. It minimizes cost functions defined by industry rules thus generating an optimized network. The various input data and the rules used to build the deterministic tree-shaped graph being uncertain, a stochastic version of the algorithm is also put forward to generate a set of probable networks in addition to the optimized one. The method is tested on the wastewater networks of Prades-le-Lez and validated on the town of Ramonville Saint Agne. Both towns are located in Southern France, are part of the two most dynamic metropolitan areas of France and are under constant urban pressure due to their proximity with the cities of Montpellier and Toulouse. The shape and topology of the mapped networks are compared to the actual ones. The results indicate an overall good agreement between the layouts of the real and generated

networks. The proposed algorithm may thus be used to map wastewater networks from sampled georeferenced manhole covers, elevation and street network databases. Additional sources of information are however necessary to recreate the networks full geometry and insure proper conveyance. The low error values and high scores for completeness, correctness and quality indicate that the method is robust and may be adapted and tested on other study zones.

11 *Keywords:* Wastewater, Urban hydrology, Multigraph, Network connectivity,  
12 Mapping

---

## 13 **1. Introduction**

14 Urban expansion is an ongoing process both in developing and developed  
15 countries. According to the latest global statistics published by the UN 55%  
16 of the world's population is currently residing in urban areas as opposed to  
17 30% in 1950. By 2050 this percentage is estimated to reach 68% [35]. The  
18 increase in population often leads to urban sprawl and city managers have to  
19 constantly extend water access and sanitation services to new peripheral areas.  
20 Getting accurate and updated information on the underground wastewater and  
21 stormwater networks is a cumbersome task, especially in cities undergoing urban  
22 expansion [24, 30]. With the development of smart city technologies, a growing  
23 number of towns are getting equipped with electronic sensors which are able  
24 to log and transmit data continuously. It thus becomes easier to monitor the  
25 evolution of environmental variables in quasi real time and to make adequate  
26 decisions based on the prevailing conditions. However, data on the location and  
27 the geometric features of water networks are still incomplete or missing [10].  
28 This is especially due to the fact that many countries only recently passed bills  
29 on the localization accuracy and precision that contractors need to report back  
30 when they undertake works in the vicinity of underground utility networks. In  
31 France for instance, this bill was passed in 2012 [22].

32 Attempts have been made to find the optimal design configuration for wastew-  
 33 ater and stormwater networks based on topographic and hydraulic constraints  
 34 or economic costs [40, 2, 21, 27, 31]. Network layout optimisation problems have  
 35 been thoroughly investigated in operational research since the 1950s [16, 7]. In  
 36 combinartorial optimisation this equates to finding the minimum spanning tree  
 37 of an undirected, connected graph and is known as such as put forward by  
 38 Borvka in 1926 [32] and later by Kruskal and Prim [23, 36] (see a review in  
 39 [17, 7]) or finding the shortest path to connect N-points commonly known as  
 40 the Steiner tree problem [19, 13].  
 41 Some authors have generated virtual urban networks [26, 39]. However, in ap-  
 42 plications where the actual network has to be mapped, exogenous factors such  
 43 as existing network branches may render these solutions quite inefficient. Fewer  
 44 attempts have been made in the literature to reproduce the layout of actual  
 45 drainage networks, on urban or farmed catchments [6]. In [4], [8] and [11] the  
 46 node to link information was provided by urban databases or local operators. In  
 47 the case of [10], who have developed a Bayesian Mapping model for buried util-  
 48 ity networks using utility records coupled with MTU sensor data and manhole  
 49 surveys, the node to link connection was established through hypothesis testing  
 50 using the Expectation Maximization Algorithm. Their results were validated  
 51 against 2 site specific data.

52 When working on an urban catchment where very little information about  
 53 the network configuration is available, one solution may be to infer it based on  
 54 visible features such as manhole covers. Recent works have shown that these  
 55 could be localized by using new processing techniques and high resolution aerial  
 56 imagery [29, 33, 34]. Although the precise location of all the manhole covers is  
 57 not mandatory to build accurate hydraulic models of the wastewater network,  
 58 this is far from being the only piece of information required by modelling soft-

ware products. The conveyance should be assessed via parameters such as pipe size, shape and roughness. Pipe slope is also a parameter of great importance as classical hydraulic models cannot compute gravity fed flow on counter-slopes, unless pumping stations are installed.

In this paper, we put forward a methodology to generate a map of a wastewater or a stormwater network based on manhole cover or sewer inlet locations. The paper is structured as follows: the optimization algorithm and the cost functions are presented in section 2.1. Section 2.2 presents the stochastic version of this algorithm, allowing input data and rules uncertainty propagation up to the generated network design. A test using real world data from two French towns is presented in section 3 followed by a discussion and conclusions in section 4.

## 2. Network cartography from georeferenced points

Starting from a set of georeferenced points (manhole covers) and assuming the position of the network's general outfall, a methodology is put forward to retrieve the links between the points, i.e. the network pipes or edges. In a first, purely deterministic approach, the algorithm starts from the outfall point and chooses to link points with edges that minimize different cost functions. Secondly, this method is generalized with stochasticity introducing a probability based on the edges' costs allowing uncertainty propagation in the network generation.

We first assume that manhole covers are located accurately and are given as a set  $S$  of georeferenced points  $M_i(x_i, y_i)$ , that constitute the nodes of the network to be mapped. Manhole localization may be done manually or automatically by using Very High Resolution aerial images as presented in [15, 34]. The general chart of the algorithm is presented in Figure 1 and the different steps

85 of the method are detailed in the following subsections. The full procedure is  
 86 automated and written in Python language. Using the input points in ESRI  
 87 shapefile format, it returns a new shapefile with an associated attribute table  
 for the created segments.

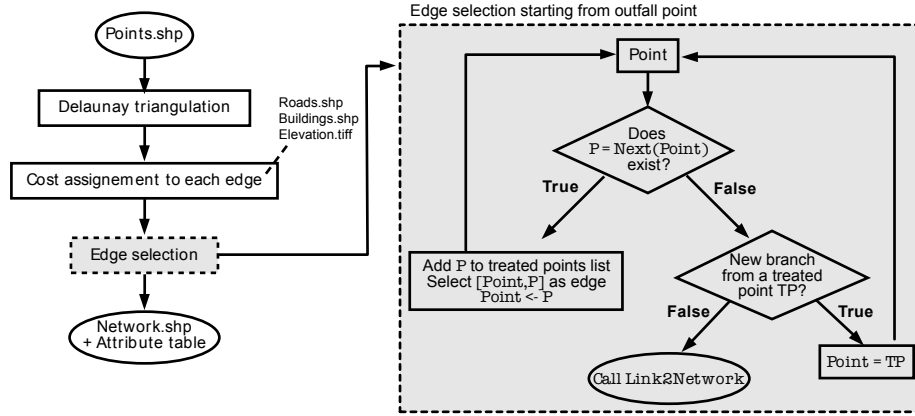


Figure 1: General chart (left) and edge selection algorithm (right). At the end of the process the `Link2Network` function is called for all remaining untreated points.

88

### 89 2.1. Creation of a valued directed graph

90 In a first step, a Delaunay triangulation is computed from the set  $S$ . This  
 91 operation consists in connecting the points of set  $S$  to form triangles such as  
 92 no point  $M_i$  is inside the circumcircle of any triangle. In Euclidian distance, a  
 93 Delaunay triangulation can be proven to be a spanning graph of set  $S$ . However,  
 94 as this rule has no set rationale for wastewater networks, points located within  
 95 a radius set by the user are also considered as possible neighbouring points. The  
 96 edges thus formed give a consistent subset of all the possible connections between  
 97 points of  $S$ , called initial graph in the following. Then, a value  $c(M_i M_j)$  is  
 98 assigned to each edge  $M_i M_j$  which defines the "cost" to make wastewater flowing  
 99 from  $M_i$  to  $M_j$ . Of course, the edge is directed so that  $c(M_i M_j) \neq c(M_j M_i)$ .

100 At this stage, the cost function is based on two criteria: *i*) length: favouring

101 smallest edges, and *ii*) slope: favouring gravity fed flow within current industry  
 102 rules used in France. If no other information than manhole cover position is  
 103 considered available, the points elevation is assessed from a Digital Elevation  
 104 Model (DEM) and assumed related to that of the pipes. This is a strong hy-  
 105 pothesis, often used in hydrological modelling [9] but that is not entirely true  
 106 for stormwater networks where ground slopes may be countered by adjusting  
 107 the trench dimensions or adding pumping devices. Thus, water may flow in the  
 108 opposite direction to surface flow/slope. However, in data scarce situations this  
 109 approximation is retained as it is the only possibility. This assumption was ver-  
 110 ified by calculating the coefficient of determination between the DEM elevation  
 111 and the pipe's inlet depths for the data we had access to. The results show  
 112 that  $R^2=0.97$  for Prades-Le-Lez. In the following, a third criterion based on  
 113 the angle between two adjacent pipes will be added. When information on land  
 114 use is available, which is the case in France via IGN BD-TOPO<sup>©</sup>, penalties are  
 115 assigned to edges that cross roads ( $P_r$ ) or buildings ( $P_b$ ) using a buffer created  
 116 around the roads' polylines. Note that similar criteria have also been used by  
 117 [8] to generate a synthetic stormwater map.

118 The cost function is then defined as:

$$c(M_i M_j) = \alpha_L C_{L_{ij}} + \alpha_S C_{S_{ij}} + P_r + P_b \quad (1)$$

119 Where:

- 120 •  $C_{L_{ij}}$  is the cost associated to the length  $L_{ij}$  of edge  $M_i M_j$ . In Southern  
 121 France, the maximal distance recommended between two manhole covers  
 122 is about 80 m [5]. In the framework of network cartography with partial  
 123 information on the manhole covers' position *e.g.* obtained through remote  
 124 sensing data, the cost associated with the length increases linearly from 0  
 125 to 1 between 0 m and  $L_{\max}$  (see Figure 2a), where  $L_{\max}$  is chosen equal

126 to 160 m, in the following:

$$C_{L_{ij}} = \begin{cases} \frac{L_{ij}}{L_{\max}} & \text{if } L_{ij} < L_{\max} \\ 1 & \text{else} \end{cases} \quad (2)$$

127  $L_{\max}$  is set to twice the recommended value as a tolerance to the older  
 128 portions of networks which were designed before the fairly recent recom-  
 129 mendation was made and to take into account non-reported manhole cover  
 130 positions.

- 131 •  $C_{S_{ij}}$  is the cost associated to the slope  $S_{ij}$  of edge  $M_i M_j$ . The recom-  
 132 mended slope to ensure gravity fed flow is between 0.3% and 0.7% and  
 133 should not be greater than 5% nor negative (counterslope); the cost asso-  
 134 ciated to the slope is thus defined as (see Figure 2b):

$$C_{S_{ij}} = \begin{cases} 0 & \text{if } 0.3\% < S_{ij} < 0.7\% \\ \frac{|S_{ij}-0.7\%|}{10\%-0.7\%} & \text{if } 0.7\% < S_{ij} < 10\% \\ -\frac{|S_{ij}-0.3\%|}{1.3\%} & \text{if } -1\% < S_{ij} < 0.3\% \\ 1 & \text{if } -1\% > S_{ij} \text{ OR } S_{ij} > 10\% \end{cases} \quad (3)$$

135 Note that the cost associated to the slope may be relaxed if the slope  
 136 is estimated using ground elevation that may not be representative of  
 137 underground elevation.

- 138 •  $P_r$  is defined as the length of edge that is outside a road divided by a  
 139 distance  $d$ . In this application  $d=20$  m so that  $P_r$  is greater than one if  
 140 more than 20 m of an edge are outside the road buffer.
- 141 •  $P_b$  is defined as the percentage of edge that crosses a building multiplied  
 142 by a number  $N$ . In this application  $N=4$ , so that  $P_b$  is greater than one



143 if more than 25% of the edge is crossing a building.

144 • The alpha coefficients are mere weighing coefficients. The user may set  
145 the alpha values without resorting to calibration simply by choosing the  
146 relative weight of each criterion. Alpha values reflect also the quality of  
147 the used datasets. For instance, if the slopes used to remap the network  
148 are known to be more accurate than the manhole cover locations which  
149 determine the pipe lengths, a higher alpha value may be given to slope  
150 alpha  $\alpha_S$  and a lower value to the length alpha ( $\alpha_L$ ) and vice versa. The  
151 alpha values may also be determined by referring to the local rules of  
152 practice and their evolution through time, or based on the development  
153 history of a given town. In France the rules of practice are based on  
154 national technical guidelines. At regional level, good practice rules may  
155 be established as well. Hence, there is little room for improvisation or  
156 modification for a given town. However, if for other countries these rules  
157 are set at regional or district level than the user may change them. If these  
158 guidelines change over time, it is also possible to modify the weighing of the  
159 parameters. If partial field data is available, they may also be evaluated by  
160 calibration or trial and error, provided the thresholds of the cost functions  
161 are modified accordingly.

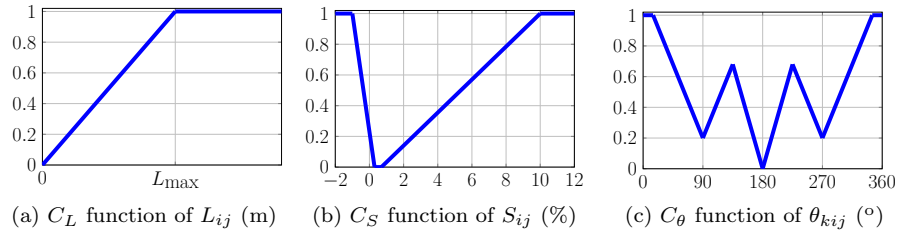


Figure 2: Cost functions, related to edge length (left), slope (middle) and angle (right). Note that the angle cost function cannot be computed directly on the initial graph but is used during the pipe selection procedure, to chose between candidate edges.

162 This first step results in an directed graph with valued edges among which  
 163 the network pipes will be selected. The attribute table of the generated shapefile  
 164 contains useful information for each edge, such as point ID and elevation, edge  
 165 length and slope, as well as cost values and road and building penalty values  
 166 (see Table 1).

Symbol	Definition
$M_u$ (resp. $M_d$ )	Upstream (resp. downstream) point ID
$z_u$ (resp. $z_d$ )	Estimate of upstream (resp. downstream) point elevation
$L_{ud}$ and $C_{L_{ud}}$	Edge's length and associated cost
$S_{ud}$ and $C_{S_{ud}}$	Edge's slope and associated cost
$P_{rud}$	Road penalty
$P_{bud}$	Building penalty

Table 1: Attributes associated to the initial graph shapefile for each edge  $(M_u M_d)$ .

167 The pipe selection algorithm (Figure 1 right) from the possible edges is  
 168 applied to link the points of set  $S$  starting from the outfall.

169 Cost function given by equation (1) is used to select the edge  $ij$  starting  
 170 from the current point  $i$ . At this step the following new criterion  $C_{\theta_{kij}}$  is added.  
 171 The coefficients 0.8 and 0.4 have been chosen to get the specific shape of the  
 172 cost function plotted in Figure 2c, favouring  $180^\circ$  angles between two adjacent  
 173 edges  $ki$  and  $ij$ , and  $90^\circ$  angles to a lesser extent, and giving the highest penalty  
 174 to acute angles:

$$C_{\theta_{kij}} = \begin{cases} 1 & \text{if } |360 - \theta| < 30 \\ 0.8 \frac{|90 - \theta|}{60} + 0.2 & \text{if } |90 - \theta| < |180 - \theta| \\ 0.4 \frac{|180 - \theta|}{90} & \text{if } |180 - \theta| < |90 - \alpha| \text{ and } |180 - \theta| \leq |270 - \theta| \\ 0.8 \frac{|270 - \theta|}{60} + 0.2 & \text{if } |270 - \theta| < |180 - \theta| \end{cases} \quad (4)$$

175 The "angle cost"  $C_{\theta_{ij}}$  of a given edge  $ij$  is computed as the sum of the costs  
 176  $C_{\theta_{kij}}$  for all points  $k$  already linked to point  $i$ . The total cost function  $C(M_i M_j)$

177 given in equation (1) is thus modified as follows:

$$c(M_i M_j) = \alpha_L C_{L_{ij}} + \alpha_S C_{S_{ij}} + \alpha_\theta C_{\theta_{ij}} + P_r + P_b \quad (5)$$

178 Starting from the outfall, the edge with the lowest cost is chosen as pipe  
179 network provided that it is lower than a maximal admissible cost **MaxCost**.  
180 In the following **MaxCost** is set equal to one, so that, given the penalties on  
181 the road or building crossings, the points located on the roads are first linked  
182 together. All the linked points are labeled as **Treated Point** and stored in a  
183 list. When it is not possible to connect a new pipe to point  $i$  with an admissible  
184 cost, the algorithm browses the list of already treated points to find a new  
185 path, called **Branch**, still favouring points located inside the road buffer. When  
186 all the possible branches have been explored, the algorithm browses the list of  
187 non treated points to connect them to the network using the **Link2Network**  
188 function with a relaxed constraint on road crossing. The pseudo code of the  
189 main treatment and of the Next function are presented in algorithms 1 and 2.

190 If some points are still unlinked at the end of this process, the point with the  
191 lowest elevation value is automatically defined as a new outfall, as stormwater  
192 networks might have more than one outlet on a given catchment. The linking  
193 operation starts over for all the remaining points (if any).

194 The entire process thus yields a multigraph with possible unlinked points  
195 corresponding to the false positives of a previous detection step. Although some  
196 authors consider these as "infeasible solutions" when optimizing network design  
197 [1], they do correspond to real-world situations where, due to urban growth, the  
198 network layout may be altered over time.

---

**Algorithm 1** Pseudo code of main treatment (gen stands for the NetworkGenerator class)

---

```

1: PtBranch  $\leftarrow$  [outfall]
2: while MainTREATMENT do
3:   BRANCH  $\leftarrow$  False
4:   idPt  $\leftarrow$  PtBranch[-1] # last point added to the list
5:   Next  $\leftarrow$  pointNext(gen, TreatedPt, idPt)
6:   if Next then # there is a Next on the same branch
7:     gen.NetWk.append(gen.edges[(Next, idPt)])
8:     gen.edges[(Next, idPt)].branchID = branchID
9:     TreatedPt.append(Next)
10:    PtBranch.append(Next)
11:   else
12:     # if no more point can be added, try to start a new branch from
    already treated points
13:      $i = 0$ 
14:     while (not BRANCH and  $i < \text{len}(\text{TreatedPt})$ ) do
15:       idPt  $\leftarrow$  TreatedPt[i]
16:        $i \leftarrow i + 1$ 
17:       Next  $\leftarrow$  pointNext(gen, TreatedPt, idPt)
18:       if Next then
19:         BRANCH  $\leftarrow$  True # New branch
20:         gen.NetWk.append(gen.edges[(Next, idPt)])
21:         TreatedPt.append(Next)
22:       if not BRANCH then # no new branch found
23:         # browse untreated points to link them to network
24:         couple  $\leftarrow$  Link2Network(gen, TreatedPt)
25:         if couple then
26:           TreatedPt.append(couple[0])
27:           gen.NetWk.append(gen.edges[(couple[0], couple[1])])
28:         else
29:           MainTREATMENT = False

```

---

## 199 2.2. Stochastic approach

200 Each step of the previously developed process is subject to uncertainty: the  
201 cost function was defined from best practice recommendations given in the form  
202 of intervals that may not always be strictly followed; pipe slopes might not be  
203 reported accurately or could be determined using ground elevation data; maps  
204 of manhole positions might be erroneous or incomplete. Validation maps pro-  
205 duced at the district or city scale may also have cartographic errors due to edge

---

**Algorithm 2** Pseudo code of function Next

---

```

1: NextPt  $\leftarrow$  NONE
2: Weight  $\leftarrow$  MaxCost
3: for Neigh in gen.neighbour[idPt] and not in TreatedPt do
4:   # Find DownPt, the downstream point of idPt and compute angle
5:   angleCost = gen.angleCost(DownPt, idPt, Neigh)
6:   # Find Neigh2, a possible upstream point of idPt
7:   for Neigh2 in gen.neighbour[idPt] do
8:     if idPt is the downstream point of Neigh2 then
9:       # add angle cost for all already present edges
10:      angleCost  $\leftarrow$  angleCost + gen.angleCost(Neigh2, idPt, Neigh)
11:   # Compute total cost
12:   gen.computeCost(gen.edges[(Neigh, idPt)])
13:   # keep the minimum cost (or use the probability in the stochastic ver-
   sion)
14:   if gen.edges[(Neigh, idPt)].cost < Weight then
15:     Weight  $\leftarrow$  gen.edges[(Neigh, idPt)].cost
16:   NextPt  $\leftarrow$  Neigh
   return NextPt

```

---

206 matching and rubber-sheeting. It is thus admitted that the layout of the ac-  
207 tual wastewater network cannot be retrieved perfectly. The aim of the process is  
208 then to propagate input data and rules' uncertainty and consequently map a set  
209 of probable network that explicitly represent the uncertainty in the wastewater  
210 networks accordingly. This enables to further propagate this mapping uncer-  
211 tainty into the hydraulic software to reproduce the main discharge uncertainty  
212 at outfall.

213 In this stochastic approach, for a given downstream node  $M_i$  we define a  
214 probability distribution function  $P_{M_i M_j}$  for each edge  $M_i M_j$  of the initial graph,  
215 based on the edge's cost value:

$$P_{M_i M_j} = \frac{\frac{1}{c_{M_i M_j}}}{\sum_j \frac{1}{c_{M_i M_j}}} \quad (6)$$

216 Lower costs are assigned to the edges that meet the constraints imposed on  
217 length, slope and angle, hence their selection as pipe has a higher occurrence

218 probability. These correspond to a small-set of edges, which would have had a  
 219 lower selection probability if the process was purely random.

220 For a given point  $M_i$ , instead of choosing the edge with the minimum cost,  
 221 the upstream point  $M_j$  is randomly sampled from possible neighbours in the  
 222 initial graph using multinomial laws with parameters from equation (6). Several  
 223 runs of the algorithm are performed so that, at the end of the process, a set of  
 224 probable networks is thus obtained.

### 225 2.3. Validation procedure

226 The mapped network is validated against the actual map via positional errors  
 227 and changes in network hierarchy.

228 Positional errors are calculated using the criteria of completeness, correct-  
 229 ness and quality, put forward in [18] to evaluate automatically extracted road  
 230 networks, and total error is defined using [28]: [25]

$$\text{Completeness} = \frac{\text{TP}}{\text{TP}+\text{FN}} \quad (7a)$$

$$\text{Correctness} = \frac{\text{TP}}{\text{TP}+\text{FP}} \quad (7b)$$

$$\text{Quality} = \frac{\text{TP}}{\text{TP}+\text{FP}+\text{FN}} \quad (7c)$$

$$\text{Error} = \frac{\text{FN}+\text{FP}}{\text{RL}} \quad (7d)$$

231 where TP represents true positives (*i.e.* the length of correctly mapped pipes),  
 232 FN false negatives (length of pipes that are not mapped by the algorithm but  
 233 that do exist in the validation database), FP false positives length of pipes  
 234 that are mapped by the algorithm but that are not reported in the validation

235 database) and where RL stands for Real Length *i.e.* the total length of pipes  
 236 as reported in validation database. Completeness represents the percentage of  
 237 the reference network which lies within the buffer around the mapped pipe.  
 238 Correctness represents the percentage of correctly mapped pipes, *i.e.*, the per-  
 239 centage of the mapped pipes which lie within the buffer around the reference  
 240 network. Quality is a measure of the overall goodness of the final result. It  
 241 takes into account completeness and correctness.

242 The optimum value is 1 for the three first criteria and is 0 for the error. In  
 243 this application the buffer is flat and double sided. Its width is set to  $w = 5$  m,  
 244 *i.e.*  $\pm 2.5$  m. This means that mapped pipes are considered to be identical to  
 245 the real pipes if they fall within a 5 m strip. This distance may not meet the  
 246 precision limits imposed by current legislation in developed countries (such as  
 247 the U.S. Clean Water Act, or the French environmental legislation [22] revised  
 248 in 2018). However, these maps may be used to plan database updating field  
 249 work. In addition, in many developing countries, no maps exist at all and even  
 250 a 5 m tolerance may be considered as an improvement.

251 Network hierarchy is based on Shreve's magnitude [38], which is a numer-  
 252 ical measure of its branching complexity and implicitly accounts for topology.  
 253 The network is assumed to be a mathematical tree and channels or branches,  
 254 which have no upstream junctions or tributaries, have an order of 1. When two  
 255 branches of order  $i$  and  $j$  join, the resulting downstream branch has an order  
 256  $i + j$ . Changes in network hierarchy will be directly reflected by the outlets  
 257 order.

### 258 **3. Application and results**

259 The methodology is first tested on the town of Prades-le-Lez in South East-  
 260 ern France ( $34^{\circ}41'51''\text{N}$ ;  $3^{\circ}41'51''\text{E}$ ) and validated on the town of Ramonville-

261 Saint-Agne in the South West of France ( $43^{\circ}34'04''\text{N}$ ;  $3^{\circ}54'07''\text{E}$ ).

262 The Prades-le-Lez database for the existing manhole locations is provided  
263 by Montpellier Méditerranée Métropole, the local operator. The town limits the  
264 manhole cover positions and the initial graph are presented in Figure 3. The  
265 database includes 792 manhole covers and 23.45 km of pipes with a mean length  
266 of 28 m and standard deviation of 15 m. All the other geographical data used in  
267 this study are available through the French National Institute of Geography's  
268 databases (BD-TOPO<sup>®</sup> and ALTI-RGE<sup>®</sup>). The network is not fully connected  
269 as can be seen on Figure 3 in the eastern part of the town.

270 The Ramonville-Saint-Agne database has been obtained through the French  
271 government's open access gateway (<https://www.data.gouv.fr/fr/>). Ramonville-  
272 Saint-Agne is a small town located in the urban area of Toulouse, the 4th biggest  
273 city in France. Despite its small size, it has a high population density, 2140,7  
274 inhabitants.km<sup>-2</sup> in 2015 i.e. 13829 inhabitants over 6km<sup>2</sup> and is under con-  
275 stant urban pressure due to its proximity with the city of Toulouse. The network  
276 consists of 1878 nodes and 58.9 km of pipes, with a mean length of 30 m and  
277 standard deviation of 16 m.

278 The distance between the two towns is of 249 km. They are both part of the  
279 two most dynamic metropolitan areas of France, with a mean annual population  
280 growth of 2.9 % between 2010 and 2015 for Ramonville-Saint-Agne and 2.8%  
281 for Prades-le-Lez. In comparison, the city of Paris scores -0.3% over the same  
282 period and its metropolitan area 0.5%. Both catchments are larger and their  
283 networks are bigger than most of the study cases reported in the literature  
284 [3, 8, 31, 37].

### 285 3.1. Prades-le-Lez

286 In both applications, the road buffer width for the road penalty  $P_r$  is set  
287 to 8 m ( $\pm 4$  m). Using the 792 manhole cover locations, 9946 oriented edges



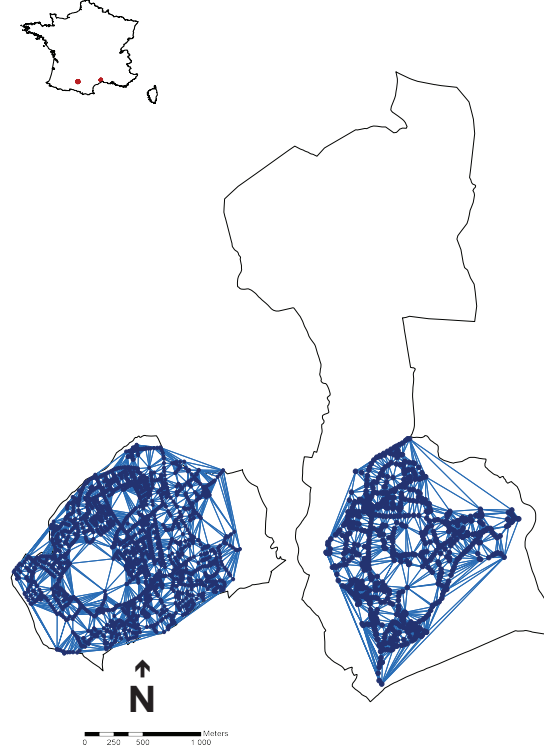


Figure 3: The two study catchments, Ramonville-Saint-Agne (Left) and Prades-Le-Lez (Right). Limits of the municipality, manhole cover positions as reported in the stakeholder's database and corresponding initial graphs.

are defined. The best ones, having the optimal cost values, will be selected as  
 pies for the mapping process. In the following subsections, first the optimal  $\alpha$   
 values are identified, then the "optimal" networks with regards to the various  
 criteria are presented. Finally the set of probable networks obtained using the  
 stochastic procedure is analyzed.

### 3.2. Selection of optimal $\alpha$ values

A sensitivity analysis of the results is performed with respect to the three  
 parameters  $\alpha_L$ ,  $\alpha_S$  and  $\alpha_\theta$  which are varied from 0 to 1 by step of 0.1, while

296 respecting a total sum value of 1. Figure 4 represents the results in terms of  
 297 error (eq. 7d).

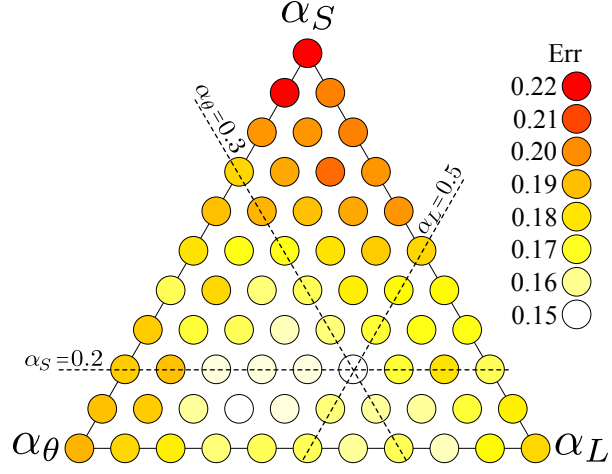


Figure 4: Results according to different values of  $\alpha$  in term of error (eq. 7d). Each parameter  $\alpha$  is equal to one at the corresponding top of the triangle and zero at the opposite base. The best results are obtained with  $\alpha_L = 0.5$ ,  $\alpha_S = 0.2$  and  $\alpha_\theta = 0.3$ .

298 The best result is obtained for  $\alpha_L = 0.5$ ,  $\alpha_S = 0.2$  and  $\alpha_\theta = 0.3$ , with an  
 299 error equal to 0.158. This corresponds also to best quality (0.85) and correctness  
 300 (0.92) and to a completeness value of 0.92. The resulting network is presented  
 301 in Figure 5a.

302 Note that all four criteria have better values with this algorithm than when  
 303 using the Kruskal algorithm with ranked error criteria on slope presented in [14].  
 304 Low  $\alpha_S$  values would suggest that the slope is less important than distance when  
 305 linking the network nodes which does not seem quite logical for a gravity fed  
 306 flow network. A possible explanation is that we used terrain slope values, via  
 307 the Digital Elevation Model, which are not representative of pipe slopes. Indeed,  
 308 pipe layers often adjust the trench dimensions to imposed slope values through  
 309 digging and filling. Thus, associating a cost function to the terrain slope values  
 310 does not impact the simulation outputs greatly in terms of position.

311 Best completeness (0.93) is obtained with  $\alpha_L = 0.2$ ,  $\alpha_S = 0.2$  and  $\alpha_\theta = 0.6$

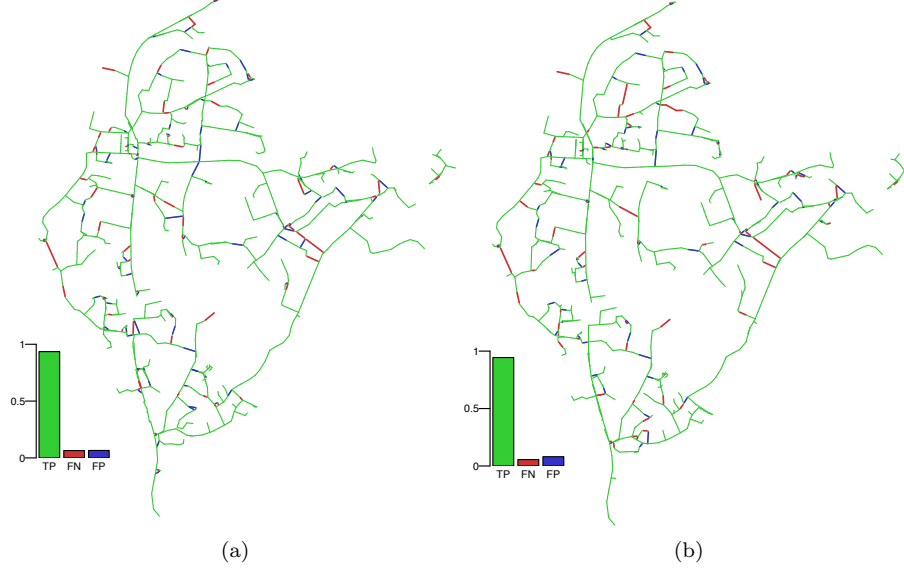


Figure 5: Best result: a) in term of error, quality and correctness, obtained with  $\alpha_L = 0.5$ ,  $\alpha_S = 0.2$  and  $\alpha_\theta = 0.3$ ; b) in terms of completeness obtained with  $\alpha_L = 0.2$ ,  $\alpha_S = 0.2$  and  $\alpha_\theta = 0.6$ .

(Figure 5b). The higher  $\alpha_\theta$  and lower  $\alpha_L$  values result in a network that has more right angle connections but edges of similar length: in comparison with the best error results, the difference between the minimal edge values are of 0.1 m, while the maximum values are 0.6 m different. Given the positional errors and the precision of the data, these differences are thought to be not significant. There are however 6 non-connected manhole covers when angles are given a higher weight compared to 2 when distance is given a higher weight.

In both instances isolated pipe segments can be found, highlighting the fact that the proposed procedure allows for disconnected graphs in order to reproduce sub-networks, such as the one located in the eastern part of the catchment. However, with lower  $\alpha_L$  values the number of disconnected parts increases from 2 to 5. These are not false connections but rather incomplete connections as can be seen when overlaying the simulated networks maps to the actual map.

325 Additional dissimilarities with the ground truth (see Figure 6 and following sub-  
326 section) also rise from the fact that the mapping algorithm considers that flow  
327 is gravity fed while the real network has a mainforce in the western part of the  
328 catchment.

### 329 *3.3. Network connectivity and layout*

330 In terms of network layout and connectivity, the maximum Shreve's magni-  
331 tude of both the "Best Network" and the "Most Complete" network are within  
332 the same order of magnitude as the real network's: Shreve's magnitude at the  
333 outlet is of 131 for the actual network, 134 for the best network and 149 for the  
334 most complete network (see Figure 6). However, while the real network has its  
335 highest orders on the western part of the catchment, both simulated networks  
336 have higher orders in their eastern branch. This is because the error criteria we  
337 have used to assess the results, rely only on position and not on flow direction.  
338 Indeed connections are allowed even for pipes with counter slopes. Thus 43%  
339 of the connected pipes have a slope cost value of 1 for the "Best Network" and  
340 39.7% for the "Most complete" network. All of these pipes are true positives  
341 and do exist in the validation database and 59% of them have reported negative  
342 slope values.

343

344 To impose correct flow directions in the network, solutions with higher alpha  
345  $\alpha_s$  values should be selected. Our results however indicate that all solutions with  
346  $\alpha_s \geq 0.7$  and  $\alpha_L < 0.2$  fail to connect the upstream and downstream parts of the  
347 network resulting in lower Shreve's magnitudes at the outlet and networks with  
348 less spatial extent. The corresponding Shreve's magnitudes vary between 33 and  
349 59. Given the catchment's topographic configuration, relaxing the condition on  
350 gravity fed flow is the only possibility of insuring connectivity throughout the  
351 network. However, none of our solutions is able to reproduce the impact of the

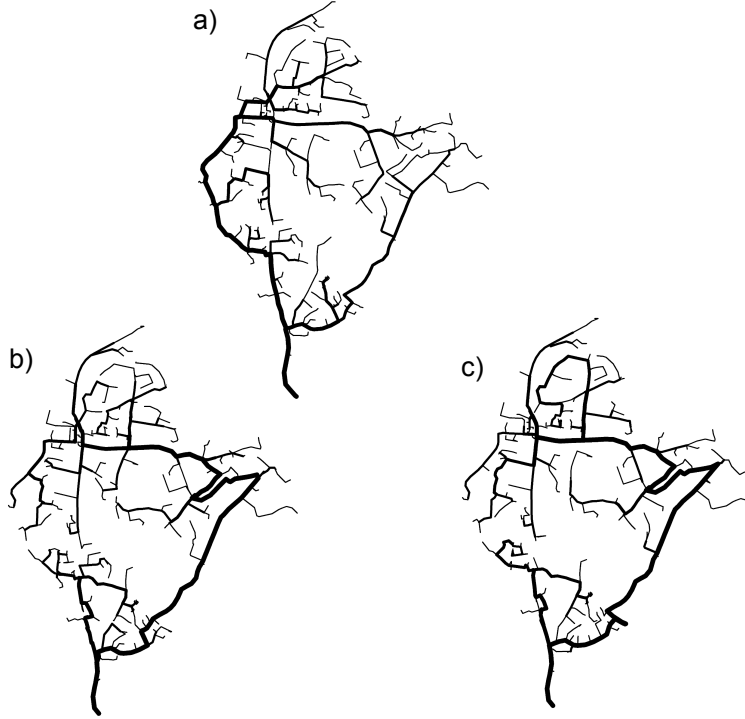


Figure 6: Comparison of Shreve's number between the actual and mapped networks. a) actual network; b) best mapped network obtained in terms of error, quality and correctness; c) best mapped network obtained in terms of completeness.

mainforce and to insure flow in the lower western part of the catchment. Such a-priori knowledge may be translated by assigning an overall null cost value to the corresponding edges thus forcing the connection algorithm to select them. The same procedure may be used if information is also available on the location of pumping stations.

#### 3.4. Uncertainty propagation

Using best  $\alpha$  parameters determined previously, the algorithm in its stochastic version is run 500 times. When some edges are selected either in one direction or in the other, the two occurrence probabilities are summed, as no information is available in the validation database on the actual network topology and flow

362 directions. Thus 1570 non oriented pipes are selected from the 9946 edges of  
 363 the initial graph. The results are plotted in Figure 7.

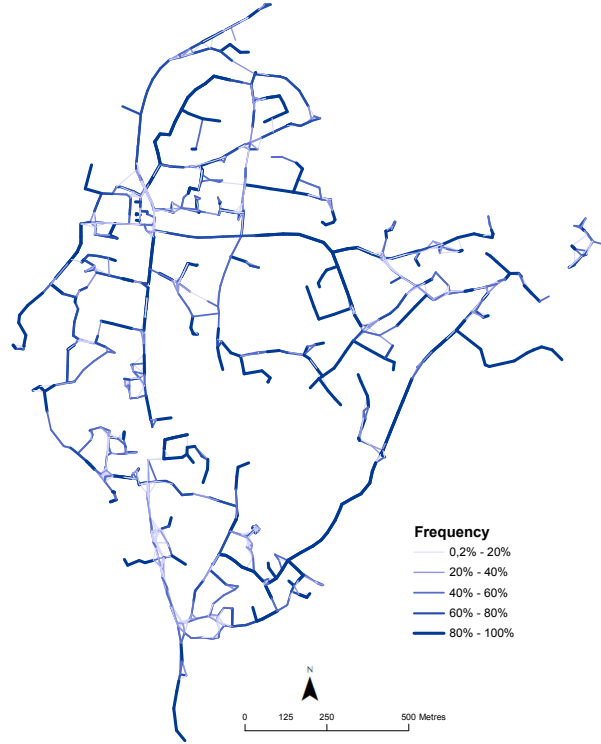


Figure 7: Results of uncertainty propagation. Darker lines correspond to pipes with higher occurrence frequency.

364 For the sake of clarity, only pipes selected at least for one network are rep-  
 365 resented, according to their occurrence frequency in the 500 runs. 708 of these  
 366 pipes are reported in the validation database and 862 are not. The occurrence  
 367 probabilities of these True and False pipes are plotted in Figure 8 in boxplot  
 368 format.

369 The results indicate that the True Positive distribution median is equal to  
 370 86%, meaning that 50% of pipes of the validation database are selected by the  
 371 stochastic process in more than 86% of the 500 simulations. More than 25% are

372 selected each time and 25% have an occurrence frequency lower than 55.95%.  
 373 Half of the false pipes have an occurrence frequency lower than 16.8%.

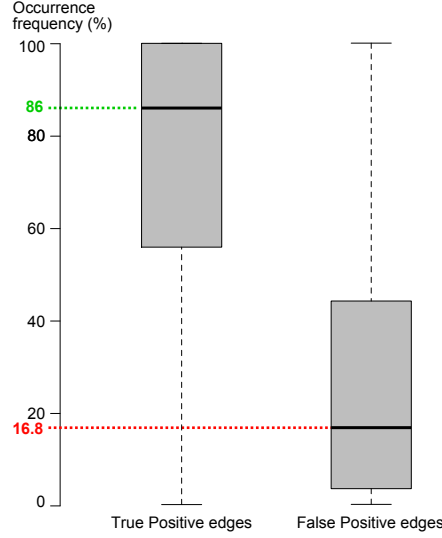


Figure 8: Results of the uncertainty propagation. Frequencies of True Positive and False Positive pipes

374 It can be seen in Figure 7 that the pipes with the highest occurrence prob-  
 375 ability (90% to 100%) correspond to the starting segment of the outlet and to  
 376 the branches with no tributaries that are the outer pipes of the network. The  
 377 median length of these pipes is 31.2 m (mean = 32.8 m;  $\sigma=15.7$ ;  $N=332$ ). The  
 378 pipes with the lowest occurrence probability (0.2% to 20%) are those linking  
 379 nodes with very similar elevation values or that have counter slope sections. On  
 380 average they are longer with a median length of 41.2 m (mean length=42.7 m,  
 381  $\sigma=18.7$ ,  $N=737$ ).

382 38 of the 8376 non selected edges are reported in the validation database  
 383 and 34 of them have a prohibitive  $C_L$  value as they fall outside the road buffer,  
 384 some run in fact parallel to the buffer (see Figure 9).

385 The results of the uncertainty propagation show that, in our case study, half  
 386 of the generated network pipes are certain. They also show that about 85 % of



Figure 9: Examples of initial graph edges that are never selected as pipes while being part of the validation database.

graph edges linking manhole cover locations are certain not to host a network  
pipe. These results are however case study and input data dependent.

#### 4. Validation

In order to test the robustness of the method we randomly sampled 75%,  
50% and 25% of the manhole positions i.e. 592, 395 and 198 points respec-  
tively. We ran the algorithm again with the parameters corresponding to the  
lowest error, best quality and correctness i.e  $\alpha_L = 0.5$ ,  $\alpha_S = 0.2$  and  $\alpha_\theta = 0.3$ .  
When omitting intermediate manhole covers, the possibility of crossing roads  
and buildings increases, as manholes are mandatory at pipe junctions and cross-



roads. The cost function value associated to length, slope and angles may also increase. The **MaxCost** parameter is thus taken respectively equal to 1.5, 2 and 4 for the 75%, 50% and 25% samples. The results are presented in figure 10 and show that:

- When using 75% of the positions, 581 pipes are created and 2 points remain unconnected. The positional errors increase with completeness dropping to 0.79 while quality and correctness reach 0.6 and 0.72 respectively. The overall error is of 0.51.
- When using 50% of the positions, 379 pipes are created and 3 points are left unconnected. The error increases to 1.05 and correctness, quality and completeness drop to 0.48, 0.37 and 0.63 respectively.
- When using 25% of the manhole cover locations 196 pipes get created and no points are left unconnected. The resulting network has very low quality criteria (error = 1.87, Correctness = 0.23, Quality = 0.17, Completeness = 0.37) but is not fragmented because of the increase of the **MaxCost** parameter.

These results show that the outcome of our method is dependent on the density of points used to generate the network. If less than half of the manhole positions are detected through field data or photo-interpretation, the algorithm will fail to reproduce a map of acceptable quality. However, as can be seen in Figure 10, the overall layout of the network is still visible and the main trunks are mapped even when using 25% of the original dataset.

A second validation test consisted in running the algorithm on the town of Ramonville-Saint-Agne. We ran the algorithm again with the same parameters i.e  $\alpha_L = 0.5$ ,  $\alpha_S = 0.2$  and  $\alpha_\theta = 0.3$ .

The corresponding error is of 0.28, quality is 0.76, correctness is 0.84 and com-

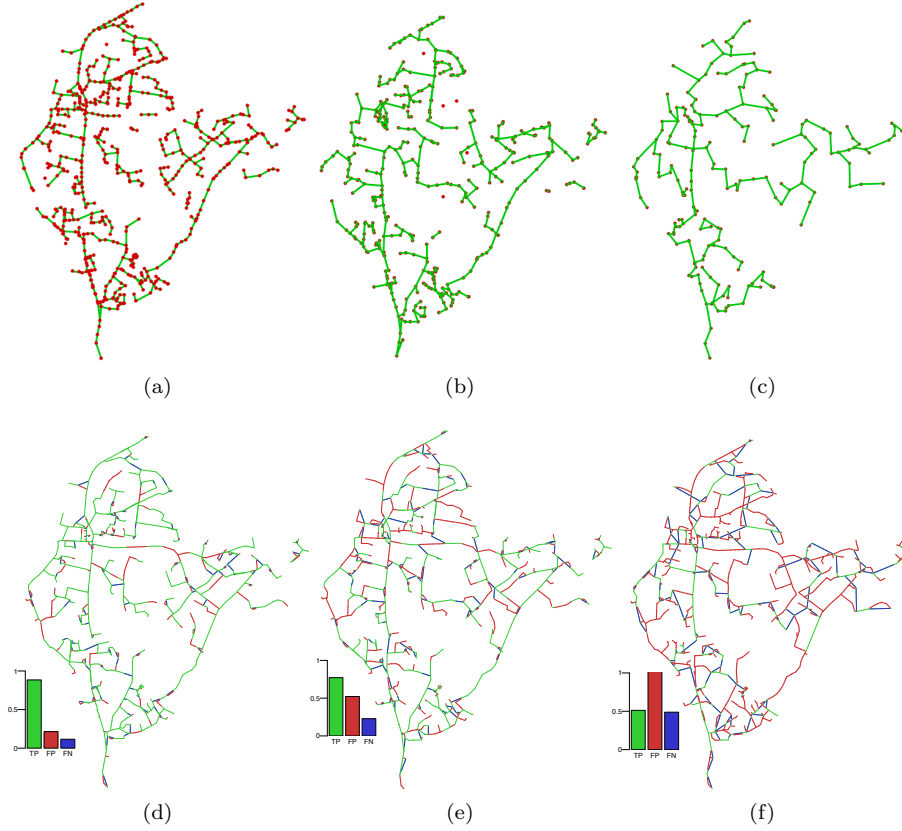


Figure 10: Results obtained with 75% (a,d), 50% (b,e) and 25% (c,f) of the database points. Up: initial points (red) and mapped network (green), down: validation against real network.

422 pleteness 0.89. 1817 pipes are mapped by the algorithm and there are only 32  
 423 un-connected manhole covers 30 of which are located outside the road buffer and  
 424 are less likely to be selected by the algorithm. This is a common problem as  
 425 50.2 % of the false negatives, i.e. pipes that do exist in the validation database  
 426 but that the algorithm did not map, are located outside the road buffer. 24%  
 427 correspond to short pipes ( $< 20$  m) and 1% to pipes  $> 80$  m. The proportions  
 428 are similar for the false positives, with 40% of short pipes and 0.9% of pipes  
 429 with a length  $> 80$  m. 30% of the false positive have acute angles and 42% have  
 430 a slope cost The results of the sensitivity test on Ramonville-Saint-Agne pre-

431 sented in Figure 11. They show that three triplets give the best results in terms  
 432 of lowest error and highest completeness (0.89), correctness (0.85) and quality  
 433 (0.77) :  $\alpha_L = 0.8$ ,  $\alpha_S = 0.0$  and  $\alpha_\theta = 0.2$ ;  $\alpha_L = 0.8$ ,  $\alpha_S = 0.1$  and  $\alpha_\theta = 0.1$   
 434 and  $\alpha_L = 0.7$ ,  $\alpha_S = 0.1$  and  $\alpha_\theta = 0.2$ . The range of error values calculated  
 435 on this catchment is higher than those calculated on Prades-Le-Lez. However  
 436 using lower coefficients for slope i.e. lower  $\alpha_S$  values, still yields better results.

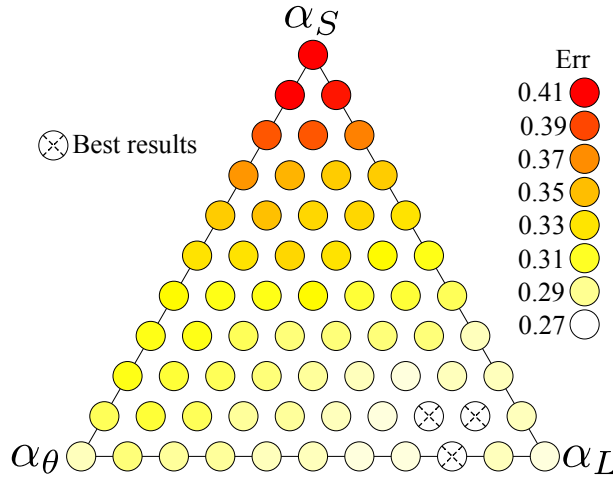


Figure 11: Error values according to different values of  $\alpha$  for Ramonville-Saint-Agne (eq. 7d).

437 Figure 12 shows the networks generated using both the parameters found  
 438 for Prades-Le-Lez and with the first triplet. The higher influence of distance vs  
 439 slope is again thought to be due to the use of surface elevation data which is  
 440 not necessarily representative of the pipe slopes : 43 % of the true positives in  
 441 figure 12b have a slope cost function value of 1. Unfortunately 92% of the slope  
 442 values are missing in the validation database. The remaining 8% have a mean  
 443 slope value of 2% ( $\sigma = 2.56\%$ ;  $N=95$ ).

444 In terms of connectivity however, the results are very poor. This is due to  
 445 the fact that, on the one hand, Ramonville-Saint-Agne's actual network is not  
 446 fully connected according to the digital maps we obtained despite the existence  
 447 of three mainforces. On the other hand, as highlighted previously, the DEM

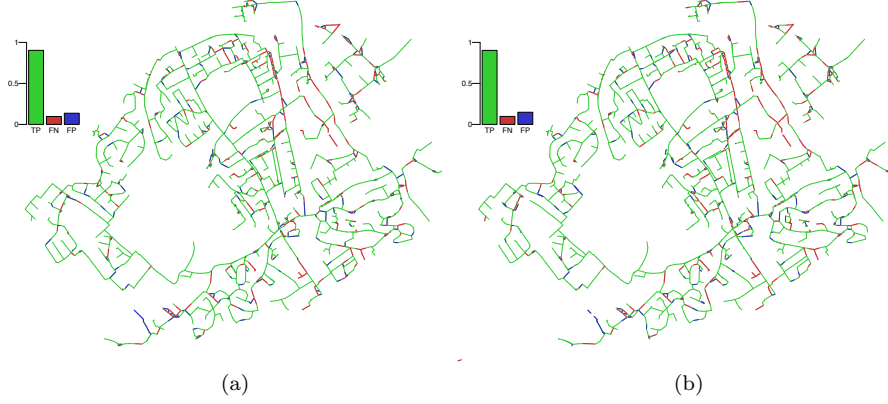


Figure 12: Results obtained on Ramonville-Saint-Agne with a)  $\alpha_L = 0.5$ ,  $\alpha_S = 0.2$  and  $\alpha_\theta = 0.3$  ; b)  $\alpha_L = 0.8$ ,  $\alpha_S = 0.0$  and  $\alpha_\theta = 0.2$ .

448 data indicates that more than 40 % of the true positives, i.e. pipes that do  
 449 exist in reality have counter slopes. Hence, our algorithm creates 308 network  
 450 branches but the highest Shreve order we calculate is 7 while the sub-network  
 451 of the validation database has a Shreve order of 387.

## 452 5. Discussion and conclusion

453 In this work we put forward a method to build a map of an underground  
 454 network using a set of georeferenced points. The method has been developed  
 455 to help contractors, managers or scientists infer the location of the wastewater  
 456 network and may be used at the scale of any administrative entity: borough,  
 457 town or water board where information is missing.

458 The method we developed requires less input data than some of the methods  
 459 described in the literature which consider the network layout as an optimiza-  
 460 tion problem based on pipe size or network cost [2, 20, 40, 31, 37]. However,  
 461 the connectivity data we have produced in this work is not enough to run a  
 462 hydraulic or hydrological model of the network. A procedure to recreate the  
 463 networks full geometry *i.e.* pipe diameter, rim and invert elevation and depth,

464 and insure proper conveyance is currently under development with a method  
465 that does not involve any calibration to determine these values and relies only  
466 on hydraulic continuity rules and industry practices. Similarly to the approach  
467 followed by [3] we use scripts to propagate pipe attribute data. The code is  
468 currently being validated. The main challenge is to translate these industry  
469 practices into cost functions as we have done in this work. Note that the func-  
470 tions are generic and the thresholds may be changed according to the practices  
471 of a given country. Thus our algorithm may be used to generate the topology  
472 of the network which may be used in hydraulic modelling or integrated in asset  
473 management databases. Though the later should ideally contain information  
474 on pipe location, geometry and condition, in many developing and developed  
475 countries they do not exist or have very fragmentary data. Pipe condition and  
476 age are not always reported for instance. It would be presumptuous on our part  
477 to claim that we would be able to generate such information using only manhole  
478 cover positions. We are currently addressing this issue using text-mining and  
479 NLP techniques [12].

480 The algorithm presented in this work is very efficient in terms of computation  
481 time as only 6 min are required to build a single network of 792 nodes and  
482 calculate the corresponding validation criteria, using a standard PC (INTEL  
483 CORE I7-5930K, 3.5GHz with 16Gb RAM). For the stochastic approach, the  
484 Delaunay graph being built once for all, the time is reduced to about 4 min per  
485 probable network.

486 Two types of information are necessary to run our algorithm: the location  
487 of the network nodes *i.e.* manhole covers and the slope of the underlying pipes.  
488 This is the geometric feature which is most difficult to determine at this stage.  
489 It is also the most sensitive parameter both to control gravity fed flow and to  
490 model it. In some examples [2, 21, 27, 40], slope has been calibrated according to

491 discharge and velocity. This procedure is not indicated in our case because the  
 492 main point of mapping the wastewater network is to use it in urban hydrological  
 493 modelling to simulate discharge values based on measured precipitation data and  
 494 not pre-determine them. Slope is often one of the least documented variables  
 495 in French urban databases available in open access. In the case of Prades-le-  
 496 Lez for instance, only 46% of the records have slope values and most of them  
 497 are incoherent. This complicates both the validation process and the attempts  
 498 made to find a rule or a method to predetermine slope values. Assuming that  
 499 terrain and underground slopes are parallel would be a suitable hypothesis as  
 500 long as there are no local constraints such as bedding planes or obstacles like  
 501 other buried networks, which could result in different layouts. Despite our best  
 502 efforts, we could not find any open access geotechnical reports, which could  
 503 give us an indication on bedding planes. One solution would be to infer slope  
 504 values based on the ground elevation and laying depths, knowing that in certain  
 505 instances the laying depth is modified to insure adequate slope values and flow  
 506 conditions. A linear relationship was established between the roadway elevation  
 507 and the pipe's upstream and downstream inlet depths ( $R^2=0.97$ ) for Prades-le-  
 508 Lez. This result is promising and the method will be tested on other catchments.  
 509 It does require accurate fine scale elevation data but a growing number of towns  
 510 and county councils are providing this type of information through their open  
 511 data web sites. In France for instance, based on the modified law n2016-1321  
 512 of October 7 2017, all local authorities employing more than 50 agents have to  
 513 put administrative documents and public data online by October 2018. The  
 514 lack of accurate data on pipe slopes may be compensated by information on  
 515 the location of pumps or mainforces. The network's topology may be corrected  
 516 by assigning an overall null cost value to the corresponding edges thus forcing  
 517 the connection algorithm to select them. The same procedure may be used

518 if the information on flow directions is available. This algorithm is mainly  
519 developed for combined sewer outflows. It may also be used to map the layout  
520 of stormwater networks provided inlet grates are used as additional location  
521 indicators. The validation dataset we have used is one of the largest used in the  
522 literature to validate network mapping algorithms. The low error values and  
523 high scores for completeness, correctness and quality indicate that the method  
524 is robust and may be adapted and tested on other study zones.

## 525 Acknowledgments

526 This study is part of the project Cart'Eaux funded by the European Regional  
527 Development Fund (ERDF). The authors would like to acknowledge the IGN  
528 (French National Geographical Institute), Montpellier Métropole and SICOVAL  
529 (Syndicat Intercommunal pour l'aménagement et le développement des coteaux  
530 et de la vallée de l'Hers) for access to their digital databases. They also would  
531 like to thank the three anonymous reviewers whose comments helped greatly  
532 improve the quality of the manuscript.

## 533 References

- 534 [1] M. Afshar. Improving the efficiency of ant algorithms using adaptive re-  
535 finement: Application to storm water network design. *Advances in Water*  
536 *Resources*, 29(9):1371–1382, sep 2006.
- 537 [2] M. Afshar. Partially constrained ant colony optimization algorithm for the  
538 solution of constrained optimization problems: Application to storm water  
539 network design. *Advances in Water Resources*, 30(4):954–965, 2007.
- 540 [3] M. Aguilar and R. Dymond. Innovative Technologies for Storm-Water Man-  
541 agement Programs in Small Urbanized Areas. *Journal of Water Resources*  
542 *Planning and Management*, 140(11):04014029, 2013.

- [4] A. Allard, K. Chancibault, and H. Andrieu. Construction of drainage network of an entire urban area - Case of Nantes East. In *Novatech 2013*, pages 1–10, Lyon, 2013.
- [5] Aqua Valley. Charte qualité des réseau d’eau potable et d’assainissement en Languedoc-Roussillon. Technical report, AquaValley, 2017.
- [6] J.S. Bailly, F. Levavasseur, and P. Lagacherie. A spatial stochastic algorithm to reconstruct artificial drainage networks from incomplete network delineations. *International Journal of Applied Earth Observation and Geoinformation*, 13(6):853–862, 2011.
- [7] F. Bazlamaçci and K. Hindi. Minimum-weight spanning tree algorithms A survey and empirical study. *Computers and Operations Research*, 28(8):767–785, 2001.
- [8] R. Bertsch, V. Glenis, and C. Kilsby. Urban flood simulation using synthetic storm drain networks. *Water (Switzerland)*, 9(12), 2017.
- [9] K. Beven. *Rainfall-Runoff Modelling: The primer*. John Wiley & Sons, Chichester, 2001.
- [10] M. Bilal, W. Khan, J. Muggleton, E. Rustighi, H. Jenks, S. Pennock, P. Atkins, and A. Cohn. Inferring the most probable maps of underground utilities using Bayesian mapping model. *Journal of Applied Geophysics*, 150:52–66, 2018.
- [11] F. Blumensaat, M. Wolfram, and P. Krebs. Sewer model development under minimum data requirements. *Environmental Earth Sciences*, 65(5):1427–1437, 2012.
- [12] T. Bonnabaud La Bruyère, N. Chahinian, C. Deleenne, L. Deruelle, M. Derras, F. Frontini, R. Panckhurst, M. Roche, L. Sautot, and M. Teis-



- seire. Mégadonnées, données liées et fouille de données pour les réseaux  
d’assainissement. In M. Brando, C., Frontini, F., Roche, editor, *Atelier  
Humaintés Numériques Spataialisées (HumaNS’2018)*, pages 15–19, 2018.
- [13] M. Brazil, R. Graham, D. Thomas, and M. Zachariasen. On the history of  
the Euclidean Steiner tree problem. *Archive for History of Exact Sciences*,  
68(3):327–354, 2014.
- [14] B. Commandré, N. Chahinian, J.S. Bailly, M. Chaumont, G. Subsol, F. Ro-  
driguez, M. Derras, L. Deruelle, and C. Delenne. Automatic reconstruction  
of urban wastewater and stormwater networks based on uncertain manhole  
cover locations. In *14th IWA/IAHR International Conference on Urban  
Drainage, ICUD 2017*, volume 2012, 2017.
- [15] B. Commandre, D. En-Nejjary, L. Pibre, M. Chaumont, C. Delenne, and  
N. Chahinian. Manhole cover localization in aerial images with a deep  
learning approach. *International Archives of the Photogrammetry, Remote  
Sensing and Spatial Information Sciences - ISPRS Archives*, 42(1W1):333–  
338, 2017.
- [16] B. Golden, M. Ball, and L. Bodin. Current and future research directions  
in network optimization. *Computers and Operations Research*, 8(2):71–81,  
1981.
- [17] R. L. Graham and P. Hell. On the History of the Minimum Spanning Tree  
Problem. *Annals of the History of Computing*, 7(1):43–57, 1985.
- [18] O. Heipke, C., Mayer, H., Wiedemann, C., Jamet. Evaluation of auto-  
matic raod extaction. *International Archives Photogrammetry and Remote  
Sensing*, 32:47–57, 1997.

- [19] F. K. Hwang and D. Richards. Steiner tree problems. *Networks*, 22(1):55–89, 1992.
- [20] S. Ivić, L. Grbčić, and S. Družeta. Cooperative Random Walk for Pipe Network Layout Optimization. *International Journal of Applied Engineering Research*, 11:2839–2847, 2016.
- [21] J. Izquierdo, I. Montalvo, R. Pérez, and V. S. Fuertes. Design optimization of wastewater collection networks by pso. *Computers and Mathematics with Applications*, 56(3):777–784, 2008.
- [22] JORF. Arrêté du 15 février 2012 pris en application du chapitre IV du titre V du livre V du code de l’environnement relatif à l’exécution de travaux à proximité de certains ouvrages souterrains, aériens ou subaquatiques de transport ou de distribution. *Journal Officiel de la République Française*, 45:2988, 2012.
- [23] J. Kruskal. On the Shortest Spanning Subtree of a Graph and the Traveling Salesman Problem. *Proceedings of the American Mathematical Society*, 7(1):48–50, 1956.
- [24] N. Metje, P. Atkins, M. Brennan, D. Champan, H. Lim, J. Machell, J. Muggleton, S. Pennock, J. Ratcliffe, M. Redfern, C. Rogers, A. Saul, Q. Shan, S. Swingler, and A.N. Thomas. Mapping the underworld: State of the art review. *Tunnelling and underground space technology*, 22:568–586, 2007.
- [25] du Tourisme et de la Mer Ministère de l’Équipement, des Transports, du logement. Fascicule n 70 Ouvrages d’assainissement Titre I : Réseaux. *Bulletin officiel*, page p 149, 2003.
- [26] M. Möderl, D. Butler, and W. Rauch. A stochastic approach for automatic generation of urban drainage systems. *Water Science and Technology*:

- 617        *a journal of the International Association on Water Pollution Research*,  
618        59(6):1137–43, 02 2009.
- 619 [27] R. Moeini and M. H. Afshar. Layout and size optimization of sanitary  
620        sewer network using intelligent ants. *Advances in Engineering Software*,  
621        51:49–62, 2012.
- 622 [28] I. Molloy and T. F. Stepinski. Automatic mapping of valley networks on  
623        Mars. *Computers and Geosciences*, 33(6):728–738, 2007.
- 624 [29] M Moy de Vitry, K Schindler, J Rieckermann, and Leit ao JP. Sewer inlet  
625        localization in uav image clouds: Improving performance with multiview  
626        detection. *Remote Sensing*, 10(7):706, 2018.
- 627 [30] J. M. Muggleton, M. J. Brennan, and Y. Gao. Determining the location of  
628        buried plastic water pipes from measurements of ground surface vibration.  
629        *Journal of Applied Geophysics*, 75(1):54–61, 2011.
- 630 [31] K. Navin and Y. Mathur. Layout and Component Size Optimization of  
631        Sewer Network Using Spanning Tree and Modified PSO Algorithm. *Water*  
632        *Resources Management*, 2016.
- 633 [32] J. Nešetil, E. Milková, and H. Nešetilová. Otakar Borvka on minimum  
634        spanning tree problem: Translation of both the 1926 papers, comments,  
635        history. *Discrete Mathematics*, 233(1-3):3–36, 2001.
- 636 [33] H. Niigaki, J. Shimamura, and M. Morimoto. Circular object detection  
637        based on separability and uniformity of feature distributions using bhat-  
638        tacharyya coefficient. In *21st Int. Conf. on Pattern Recognition (ICPR)*,  
639        2012.
- 640 [34] J. Pasquet, T. Desert, O. Bartoli, M. Chaumont, C. Delenne, G. Sub-  
641        sol, M. Derras, and N. Chahinian. Detection of manhole covers in high-

- 642 resolution aerial images of urban areas by combining two methods. *IEEE*  
643 *Journal of Selected Topics in Applied Earth Observations and Remote Sens-*  
644 *ing*, 9(5):1802–1807, 2016.
- 645 [35] Population Division of the Department of Economic and Social Affairs of  
646 the United. World Urbanization prospects: The 2018 Revision, 2018.
- 647 [36] R. C. Prim. Shortest Connection Networks And Some Generalizations. *Bell*  
648 *System Technical Journal*, 36(6):1389–1401, 1957.
- 649 [37] G. Rezaei, M. Afshar, and M. Rohani. Layout optimization of looped  
650 networks by constrained ant colony optimisation algorithm. *Advances in*  
651 *Engineering Software*, 70:123–133, 2014.
- 652 [38] R. L. Shreve. Statistical law of stream numbers. *The Journal of Geology*,  
653 74(1):17–37, 1966.
- 654 [39] R. Sitzenfrei, M. Möderl, and W. Rauch. Graph-based approach for gener-  
655 ating virtual water distribution systems in the software vibe. *Water Science*  
656 *and Technology: Water Supply*, 10(6):923–932, 2010.
- 657 [40] G. A. Walters. The design of the optimal layout for a sewer network.  
658 *Engineering Optimization*, 9(1):37–50, 1985.