Sewer Network Layout Selection and Hydraulic Design Using a Mathematical Optimization Framework

Natalia Duque [1,2,3], **Daniel Duque [4,5]**, **Andrés Aguilar [3]** and **Juan Saldarriaga [6,*]**

[1] Eawag, Swiss Federal Institute of Aquatic Science and Technology, 8600 Dübendorf, Switzerland; natalia.duquevillarreal@eawag.ch

[2] Institute of Environmental Engineering, ETH Zurich, 8093 Zurich, Switzerland

[3] Water Distribution and Sewerage Systems Research Center, University of the Andes, 111711 Bogotá, Colombia; af.aguilar691@uniandes.edu.co

[4] Department of Industrial Engineering, Center for the Optimization and Applied Probability, Universidad de los Andes, 111711 Bogotá, Colombia; d.duque25@gmail.com

[5] Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, IL 60208, USA

[6] Department of Civil and Environmental Engineering, Water Distribution and Sewerage Systems Research Center, University of the Andes, 111711 Bogotá, Colombia

* Correspondence: jsaldarr@uniandes.edu.co

Abstract: This paper proposes an iterative mathematical optimization framework to solve the layout and hydraulic design problems of sewer networks. The layout selection model determines the flow rate and direction per pipe using mixed-integer programming, which results in a tree-like structured network. This network layout parametrizes a second model that determines hydraulic features including the diameter and the upstream and downstream invert elevations of pipes using a shortest path algorithm. These models are embedded in an iterative scheme that refines a cost function approximation for the first model upon learning the actual design cost from the second model. The framework was successfully tested on two sewer network benchmarks from the literature and a real sewer network located in Bogotá, Colombia, that is proposed as a new instance. For both benchmarks, the proposed methodology found a better solution with up to 42% cost reduction compared to the best methodologies reported in the literature. These are near-optimal solutions with respect to construction cost that satisfy all hydraulic and pipe connectivity constraints of a sewer system.

1. Introduction

A sewer network is an infrastructure that fundamentally comprises pipes and manholes, and is characterized by a network layout and a hydraulic design. The layout of a sewer network defines the flow direction within the network following a tree-like structure where all manholes are connected to the outfall through a unique path. Given a layout, a hydraulic design establishes the diameter for each pipe and the invert elevation of its two endpoints. The *sewer network design* (SND) problem aims to obtain a minimum-cost design capable of transporting a specific flow rate for each pipe, while satisfying all hydraulic and operational constraints to comply with local regulations. The input information for

the SND problem includes inflows at each manhole (rain and/or wastewater), topographic information, available pipe materials and diameters, density and water viscosity, and hydraulic design constraints.

Significant literature addresses the SND problem. The predominant approach to solve the problem considers the layout selection (LS) and hydraulic design (HD) as independent (and subsequent) problems. This separation allows for tractability of the problem at the price of design optimality. In what follows, we summarize some of the heuristic and exact techniques that have been proposed in literature to address the LS and HD problems.

A group of studies solve the HD problem for a given layout using mathematical programming (MP) approaches, such as linear programming (LP) [1–4]; non-linear programming (NLP) [5,6]; and dynamic programming (DP) [7–9]. A DP approach suffers from the problem of dimensionality in its application in the design of sewer systems [10]. To improve the dimensionality of DP-like approaches, Duque et al. [9] propose a graph modelling framework for the design of a series of pipes. The arcs of a graph model different combinations of invert elevations and diameters, a path represents a feasible hydraulic design and the best hydraulic design is obtained by means of a shortest path (SP) algorithm. Since these MP-based approaches can be computationally expensive, some authors use metaheuristics that provide a good solution with less computational resources. Genetic algorithms (GA) have been applied to water and sewer engineering problems to obtain least-cost designs [11–13]. Due to the high adaptability of GAs, they have also been used in combination with other models for sewer networks design. Cozzolino et al. [14] propose a combination with hydrologic and hydraulic models, Cisty [15] couples a GA with LP, Haghighi and Bakhshipour [12] with integer programming, and Hassan et al. [16] with heuristic programming (HP). Other heuristic approaches used to solve the hydraulic design problem are ant colony optimization (ACO) [17,18], simulated annealing (SA) together with tabu search (TS) [19,20], particle swarm optimization (PSO) [21], and cellular automata (CA) [22,23]. Zaheri et al. [24] implemented a two-phase simulation-optimization method to define iteratively the HD with CA while EPA SWMM is used to simulate the network. These heuristic strategies do not have optimality guarantees, and the burden to achieve globally optimal designs might be further expanded since the problem is solved sequentially.

Some authors have addressed the LS problem independently from the HD. This sub-problem is considered a hard problem that grows exponentially with the number of pipes. Walters [8] uses DP for optimizing the layout of a sewer network by selecting the position of all manholes along the trunk sewer after determining the order in which branches connect into it. Navin et al. [25] use Kruskal's algorithm to find the minimum spanning tree (MST) over a base graph in which edge lengths serve as the weights. From the MST, the authors generate a predefined number of trial spanning trees for the HD problem. Hsie et al. [26] propose a Steiner minimal tree (SMT) problem solved with a mixed-integer programming (MIP) to select a layout. The objective function minimizes the buried depth of the pipes assuming that they represent the main construction costs in sewer systems. The construction cost for the pipes is simplified as a constant cost per unit length, which considers pipe and excavation costs. Moreover, the methodology uses fixed design parameters (e.g., Manning's roughness coefficient, filling ratio and minimum and maximum velocities) for the formulation of linear equations in the MIP. Some metaheuristics have also been used to solve the LS, such as GA [27,28] and ACO [29]. Haghighi [28] develops a loop-by-loop cutting algorithm in which he applied GAs for determining the near-optimal layout using graph theory. The loop-by-loop cutting algorithm has been successfully used to select those pipes of a base network that can be 'cut' to generate an open network layout. Later, Haghighi and Bakhshipour [20] use TS coupled with the loop-by-loop cutting algorithm for the optimization of the LS and HD. Other studies optimize both the layout and hydraulic design, yet in a sequential fashion. Moeini and Afshar [18,30] use a tree-growing algorithm (TGA) to generate layouts, ACO to determine pipe diameters and NLP to determine pipe slopes [31]. Navin et al. [32] use a combination of an MST for the layout selection and modified particle swarm optimization for the hydraulic design. Other approaches that sequentially solve both problems include [33–37]. In particular, Li and Matthew [34] are among the first authors to tackle both problems. Their method uses an SP

spanning tree to obtain a first layout, diameters are derived from hydraulic principles, and invert elevations are computed using discrete differential dynamic programming (DDDP). The layout is iteratively improved using the search direction method that identifies pipes for which the flow rate can be adjusted to decrease the costs, while keeping the remaining pipe weights constant.

In this paper we propose an iterative method to solve the SND problem in which the network layout and the hydraulic design are computed with exact methodologies. Our method relies on MIP to obtain a network layout and uses DP to obtain a hydraulic design. The latter is an extension of the methodology proposed by Duque et al. [9] originally intended for series of pipes in sewer systems. An important feature of our methodology is that the mathematical model that describes each problem approximates the cost of a complete design using different but aligned objective functions. The objective function in the hydraulic design is a better representation of the real cost of a system because it is based on the diameters and invert elevations of the pipes, which are unknown in the LS problem. To circumvent this problem, our iterative scheme arises as a mechanism to refine an approximation of the objective function in the model for the LS problem by regressing previous layout decisions with their corresponding cost of the optimal hydraulic design. Refining the cost function approximation improves network layout in terms of the overall design cost, significantly reducing the optimality burden that arises from decoupling the problems.

Our methodology is novel with respect to previous literature in several aspects. For starters, *any* procedure for the layout selection relies on a surrogate for the cost function. For example, Navin et al. [25] use pipe lengths as a surrogate for the cost when computing the MST, while Hsie et al. [26] use ground cutting cost (expressed as a linear function of the invert elevations) plus a constant value that resembles the pipe cost per unit length. Our surrogate for the cost function approximates the construction cost of each pipe as a linear function of the flow direction and the flow rate, and the parameters of such functions are updated at every iteration. The simplicity of the cost function allows for a tractable model while the function updates help to guide the global search towards a near-optimal layout. Moreover, when solving the hydraulic design for a specific layout, there is a choice between solving for diameters and slopes sequentially or jointly. For example, Li and Matthew [34] fix pipe diameters based on design principles and solve a DP to define the crown elevations at the extremes of the pipes. Moeini and Afshar [18,30] use ACO to select diameters for each pipe and derive slopes assuming a maximum flow depth and a fixed elevation at the root node. Moeini and Afshar [31] extend the procedure in [18] and derive slopes by heuristically solving an NLP via BFGS algorithm (see e.g., Nocedal and Wright [38] for BFGS optimality conditions). In contrast, our methodology jointly optimizes both diameter and invert elevations with an exact method, i.e., for a fixed layout our hydraulic design is optimal. Lastly, we argue that our iterative scheme is unique across all methods discussed so far. Most heuristics rely on randomized procedures [18,30,31] or layout enumeration [25,32] to obtain different designs across iterations. Few authors incorporate a tabu list to guide the global search [30,31]. A notable exception is Li and Matthew [34] with their search direction algorithm, which essentially leverages the gradient of the objective function to update the current layout. Our approach is fundamentally different and proves effective to addressing the sewer design problem. To summarize our contributions:

- we propose an MIP to model the LS problem over a network with general topography,
- we extend the methodology in Duque et al. [9] to generate a hydraulic design, and
- we propose a novel iterative scheme in which the objective function in the LS model approximates the true hydraulic-based cost, and this approximation is refined as the method progresses.

Hereafter, the paper is organized as follows: the next section provides a rigorous description of the problem statement and defines key concepts that we use throughout the development of our methodology, and the subsequent section describes the proposed methodology in detail. Afterwards, two benchmark instances from the literature and one case study in Bogotá (Colombia) illustrate the

performance of the methodology. Finally, the last section concludes the paper and outlines future research avenues.

2. Problem Statement and Definitions

In this section, we provide a formal problem statement of the LS and the HD, as well as the general design assumptions and terminology used in our methodology. Both problems are modelled using graph theory [39]. Each problem uses a particular graph representing different features of the network. The LS graph is connected to the HD graph through an auxiliary graph, the tree-structured graph, which translates the layout graph into an open network (Figure 1). The layout selection graph (top) represents the network topology and is based on manholes, pipes and an outfall. The tree-structured graph (middle) is the representation of the layout as a tree-like open network by using additional nodes to represent all the manholes at the extremes of the branches. Later, the hydraulic design graph (bottom) is built following the tree-structured graph's connectivity and generates a set of design nodes for each tree-node in the middle graph, which at the same time belongs to a manhole on the top graph. Design nodes are used to establish the position and dimensions of the pipes. Therefore, there are as many design nodes for one manhole as combinations of available pipe diameters and invert elevations. Possible invert elevations start at the minimum excavation limit and decrease based on a fixed elevation change (ΔZ), until the maximum excavation limit. The relationship between the graphs will be explained along with the mathematical formulation of the optimization framework.
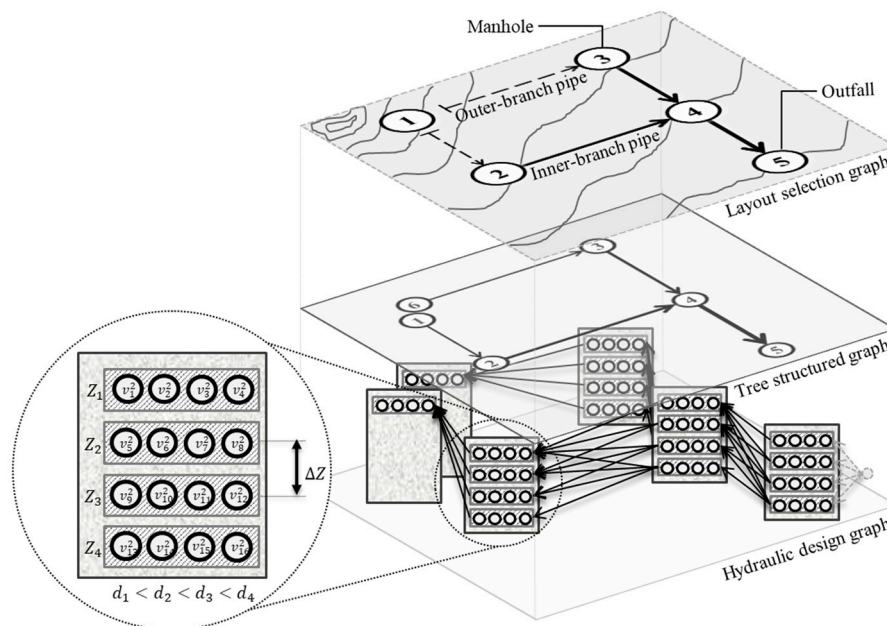

Figure 1. Relationship of the three sewer network design problem graphs. The layout selection graph (top) represents the network topology. The tree-structured graph (middle) is the representation of the layout as a tree-like open network by using additional nodes to represent all the manholes at the extremes of the branches. Finally, the hydraulic design graph (bottom) defines the hydraulic design options following the tree-structured graph's connectivity.

For both the LS and the HD, we made the following assumptions:

- The network must transport water from a discrete number of sources (manholes, sinks, etc.) to an outfall at a specific location [8].
- The flow moves by gravity over a tree-like structured network.
- There is a manhole between adjacent pipes due to changes in slope, diameter and/or flow direction [40].

- A uniform flow is assumed for the hydraulic design of each pipe [40].
- The HD methodology can use the Manning or Darcy–Weisbach and Colebrook–White resistance equations.

2.1. Layout Selection

The input of the LS problem can be described by an undirected graph $\mathcal{G} = (\mathcal{M}, \mathcal{E})$, where $\mathcal{M} = \{m_1, \dots, m_K\}$ is a set of nodes and $\mathcal{E} \subset \{(m_i, m_j) : m_i \in \mathcal{M}; m_j \in \mathcal{M}; i < j\}$ is a set of edges (undirected links between two nodes). Node $m_i \in \mathcal{M}$ represents the *i*-th manhole of the sewer network with the coordinates x_i and y_i and ground elevation Z_i. An edge $(m_i, m_j) \in \mathcal{E}$ represents the span between two manholes in which a pipe must be installed, but with no indication of flow direction. Individual inflows Q_i [m³/s] enter the system at every manhole m_i and are transported through the pipes towards the outfall m_K.

For a layout to be feasible [28], (1) no cycles are accepted, (2) all manholes (nodes) must be connected to the outfall, (3) there must be a pipe on every edge, (4) there is a single outfall, (5) several pipes can flow into a manhole, but at most one pipe can go out from every manhole, forming a unique directed path from every manhole to the outfall, (6) flow rates must satisfy mass balance equations at every manhole (no water is lost or stored).

Given the undirected base graph $\mathcal{G}$, a solution to the LS problem defines flow directions, flow rates, and connection types of every edge $(m_i, m_j) \in \mathcal{E}$ representing a sewer pipe. The flow direction of an edge determines whether the flow goes from m_i to m_j or otherwise. The flow rate of an edge resembles the flow rate of the respective pipe. The connection type of an edge describes the function of the pipes. The network is modelled as a tree-like structure with two types of pipe: inner-branch and outer-branch pipes (Figure 1). Through a series of pipes, each branch of the network follows a unique path towards the outfall. An outer-branch pipe is considered as the outmost pipe of a branch which receives, at most, the inflow from its upstream manhole. The inner-branch pipes are the rest of the pipes in the network. Outer-branch pipes do not have upstream pipes, while outer-branch and inner-branch pipes always drain into inner-branch pipes. We formally define the underlying directed graph that represents a sewer network layout later in our methodology.

2.2. Hydraulic Design

An HD defines the diameter and invert elevation of the endpoints of each pipe. This problem is represented by an additional graph as shown in Figure 1. The input data for the HD problem include the tree-like network layout, flow rates for each pipe, a cost function depending on the diameter and excavation depth, roughness of the pipe material (Manning or k_s coefficients), kinematic viscosity of the water, and some hydraulic and maintenance constraints. We consider the following hydraulic constraints: minimum pipe diameter, maximum filling ratio, minimum wall shear stress, minimum and maximum velocity, and minimum and maximum slope. For instance, upstream and downstream invert elevations of a pipe are chosen so that there is gravity-driven flow. For each manhole, there are as many design nodes as combinations of available pipe diameters and invert elevations. Possible invert elevations start at the minimum excavation limit and decrease based on a fixed elevation change (ΔZ), until the maximum excavation limit. To avoid blockages, every downstream pipe diameter is larger than or equal to its upstream pipe diameter.

3. Methodology

We propose an iterative optimization framework to solve the SND problem. At each iteration, we obtain a network layout using MIP and produce a hydraulic design using DP. The mathematical model that describes the LS problem consists of integer and continuous variables that represent layout decisions, linear constraints that define layout feasibility and a linear objective function that resembles the true cost of a system. The HD problem is framed as an SP problem on a graph that encodes both design decisions and constraints for a given network layout. A solution to the HD problem provides a

better assessment of the true cost of a system than that obtained by the LS model; hence, we perform an extra step at each iteration in which the cost coefficients of the linear objective function in the LS model are updated based on all the network layouts seen so far and their corresponding construction costs obtained from the HD. This updating scheme relies on multiple linear regression models that regress the flow rate and flow direction of a pipe against its construction cost.

3.1. Layout Selection Model

To generate a network layout, we formulate an MIP [41] that models flow direction, flow rate, and connection type for every pipe of the sewer network. The model consists of decision variables that represent the flow rate and direction; linear constraints that mathematically describe the layout feasibility; and a linear objective function that approximates the construction cost of a complete design based on layout decisions. Our model is a variant of the network design problem [42] in which we incorporate additional constraints unique to sewer networks.

Recall that the input of the LS problem is given by the undirected graph $\mathcal{G} = (\mathcal{M}, \mathcal{E})$ and the inflow parameter Q_i for every manhole $m_i \in \mathcal{M}$. Let $\mathcal{T} = \{t_1, t_2\}$ be the set of connection types, with the understanding that t_1 represents outer-branch pipes and t_2 represents inner-branch pipes. Hence, layout decisions can be described by the following variables:

- x_{ijt} is a binary variable that takes the value of one if the flow direction between manholes m_i and m_j is from m_i to m_j and the connection is of type $t \in \mathcal{T}$.
- q_{ijt} is a nonnegative real-valued variable that represents the amount of flow from m_i to m_j if the connection type is $t \in \mathcal{T}$, in the same units as the inflow Q_i in each manhole m_i.

To fully describe a feasible layout with these decision variables, we require several constraints to couple them (e.g., if $x_{ijt} = 0$ we require that $q_{ijt} = 0$) and ensure feasibility as described in the problem statement (namely, a directed path from every manhole to the outfall node, a tree-like structure and mass balance at each node). To consider every possible layout, the model includes both x_{ijt} and x_{jit} variables for every edge $(m_i, m_j) \in \mathcal{E}$ and connection type $t \in \mathcal{T}$ (as well as q_{ijt} and q_{jit}). For the sake of correctness in our model, we define a set of arcs (directed links between two nodes) induced by set $\mathcal{E}$ as $\mathcal{A}_L = \{(m_i, m_j) : (m_i, m_j) \in \mathcal{E}\} \cup \{(m_j, m_i) : (m_i, m_j) \in \mathcal{E}\}$, as well as indexed sets of outgoing pipes from manhole m_i, $\mathcal{M}_i^+ = \{j : (m_i, m_j) \in \mathcal{A}_L\}$, and incoming pipes to manhole m_i, $\mathcal{M}_i^- = \{j : (m_j, m_i) \in \mathcal{A}_L\}$. We set the parameter $Q_K = -\sum_{i=1}^{K-1} Q_i$ to represent the total flow in the system needing to reach the outfall node. To couple x_{ijt} and q_{ijt} with linear constraints, we define lower and upper bounds for variable q_{ijt} as $\ell_{ij} = Q_i/e_i$ and $u_{ij} = |Q_K|$, where e_i is the number of adjacent manholes to m_i. Finally, for every arc $(m_i, m_j) \in \mathcal{A}_L$ we define c_{ij} as an estimate of cost per flow unit that traverses from m_i to m_j and a_{ij} as a fixed cost estimate for selecting the flow direction $m_i \rightarrow m_j$. These parameters are incorporated in the objective function of the model which serves as an oracle of the true cost function of a complete design. Equation (1) defines the objective function that assesses the cost estimate of layout decisions.

$$\min\left(\sum_{t \in \mathcal{T}} \sum_{(i,j) \in \mathcal{A}_L} c_{ij} q_{ijt} + \sum_{t \in \mathcal{T}} \sum_{(i,j) \in \mathcal{A}_L} a_{ij} x_{ijt} \right) \tag{1}$$

Linear constraints (Equations (2)–(10)), ensure feasible connections throughout the network as well as proper water balance, assuming there are neither storage nor water losses. The mass balance constraint (Equation (2)) ensures that all the inflow gets transported and guarantees the existence of a unique directed path from every manhole to the outfall node.

$$\sum_{j \in \mathcal{M}_i^+} \sum_{t \in \mathcal{T}} q_{ijt} - \sum_{j \in \mathcal{M}_i^-} \sum_{t \in \mathcal{T}} q_{jit} = Q_i \qquad \forall\, m_i \in \mathcal{M} \tag{2}$$

Equations (3) and (4) establish minimum and maximum flow rates for every pipe, respectively. Note that these two equations couple the decision variables x_{ijt} and q_{ijt} in a linear fashion, avoiding the product of x_{ijt} with q_{ijt} in the flow balance constraints. This is a relevant aspect of the formulation that allows for a computationally trackable model.

$$q_{ijt} \geq \ell_{ij} x_{ijt} \qquad \forall \left(m_i, m_j\right) \in \mathcal{A}_L, \ t \in \mathcal{T} \tag{3}$$

$$q_{ijt} \leq u_{ij} x_{ijt} \qquad \forall \left(m_i, m_j\right) \in \mathcal{A}_L, \ t \in \mathcal{T} \tag{4}$$

Equation (5) enforces a single flow direction and connectivity type for every pair of manholes $\left(m_i, m_j\right) \in \mathcal{E}$.

$$\sum_{t \in \mathcal{T}} \left(x_{ijt} + x_{jit}\right) = 1 \ \forall \left(m_i, m_j\right) \in \mathcal{E} \tag{5}$$

Equation (6) ensures that at most one inner-branch pipe, $t_2 \in \mathcal{T}$, goes out of every manhole to ensure a tree-like layout.

$$\sum_{j \in \mathcal{M}_i^+} x_{ijt_2} \leq 1 \ \forall \ m_i \in \mathcal{M} \backslash \{m_K\} \tag{6}$$

Equation (7) ensures that outer-branch and inner-branch pipes always drain into inner-branch pipes. By definition, an inner-branch pipe cannot drain into an outer-branch pipe. The left-hand side of this constraint counts the number of pipes going into manhole m_i. The right-hand side checks if an inner-branch type is going out (note that this sum is at most equal to one from the previous constraint). Hence, there can only be an inner-branch pipe going out of m_i, if other incoming pipes are selected.

$$\sum_{j \in \mathcal{M}_i^-} \sum_{t \in \mathcal{T}} x_{jit} \geq \sum_{j \in \mathcal{M}_i^+} x_{ijt_2} \qquad \forall \ m_i \in \mathcal{M} \backslash \{m_K\} \tag{7}$$

Equation (8) states the maximum flow to be transported by an outer-branch pipe as the inflow coming from the upstream manhole.

$$\sum_{j \in \mathcal{M}_i^+} q_{ijt_1} \leq Q_i \qquad \forall \ m_i \in \mathcal{M} \backslash \{m_K\} \tag{8}$$

Equations (9) and (10) define the mathematical domain of the variables.

$$q_{ijt} \geq 0 \ \forall \left(m_i, m_j\right) \in \mathcal{A}_L, \ t \in \mathcal{T} \tag{9}$$

$$x_{ijt} \in \{0, 1\} \ \forall \left(m_i, m_j\right) \in \mathcal{A}_L, \ t \in \mathcal{T} \tag{10}$$

The optimization problem described above can be solved, i.e., obtain numerical values for x_{ijt} and q_{ijt} that minimize the objective function in Equation (1), via an off-the-shelf MIP solver (e.g., Xpress-MP Optimizer [43], Gurobi Optimizer [44]). The algorithm that solves such problems is beyond the scope of this paper, but we take for granted that we can retrieve from such solvers a numerical solution of x_{ijt} and q_{ijt} for every pair of manholes $\left(m_i, m_j\right) \in \mathcal{A}_L$ and connection type $t \in \mathcal{T}$ for which the variables were defined. Note that a solution to the LS problem directly tells us where the network is open (i.e., forms a tree-like structured network) by scanning the resulting outer-branch pipes. We use x and q to denote vectors that contain all the decision variables x_{ijt} and q_{ijt}, respectively; and define $\hat{x}$ and $\hat{q}$ as the numerical solution to the problem.

3.2. Layout Representation as a Tree-Like Directed Graph

Before jumping into the HD model, we describe a procedure that maps a layout described by $\hat{x}$ to a tree-like directed graph denoted by $\mathcal{G}_T = (\mathcal{N}_T, \mathcal{A}_T)$. In this graph, there is one node for every manhole

in $\mathcal{M}$, and there is one additional node for every outer-branch pipe in solution $\hat{\mathbf{x}}$. To formalize this construct, let $\mathcal{N}_T = \{v_1, \ldots, v_K, \ldots, v_T\}$ be the set of tree nodes in graph $\mathcal{G}_T$, l_k be a label that represents the index of the manhole associated to tree node v_k, and $\mathcal{N}_T^i = \{v_k : v_k \in \mathcal{N}_T, ; l_k = i\}$ be a subset of $\mathcal{N}_T$ that contains tree nodes associated with manhole $m_i \in \mathcal{M}$. Figure 1 (middle) shows an example of this at manhole $m_1 \in \mathcal{G}$ that is duplicated to obtain two nodes $v_1, v_6 \in \mathcal{N}_T^1$. These extra nodes resemble the upstream manhole of the outer-branch pipes in layout $\hat{\mathbf{x}}$. The arcs of the network connect pairs of nodes in $\mathcal{N}_T$ to resemble the layout decisions defined by $\hat{\mathbf{x}}$. To avoid confusion with the nodes of graph $\mathcal{G}$, we henceforth refer to nodes in $\mathcal{N}_T$ as tree nodes.

Tree nodes $v_1, \ldots, v_K$ have one-to-one correspondence with manholes $m_1, \ldots, m_K$ so that $l_k = k$ for $k = 1, \ldots, K$ and nodes $v_{K+1}, \ldots, v_T$ have labels $l_k \in \{1, \ldots, K\}$ for $k = K+1, \ldots, T$. To define the set of arcs $\mathcal{A}_T$, we consider connections of types t_1 (outer-branch) and t_2 (inner-branch) separately and define:

- $\mathcal{A}_T^{t_1} = \Big\{ (v_k, v_{k'}) : k = K+1, \ldots, T; \ k' = 1, \ldots, K; \ \hat{x}_{l_k, k', t_1} = 1 \Big\}$,
- $\mathcal{A}_T^{t_2} = \Big\{ (v_k, v_{k'}) : k = 1, \ldots, K; \ k' = 1, \ldots, K; \ \hat{x}_{k, k', t_2} = 1 \Big\}$, and
- $\mathcal{A}_T = \mathcal{A}_T^{t_1} \cup \mathcal{A}_T^{t_2}$

3.3. Hydraulic Design Model and Solution Approach

Duque et al. [9] propose a methodology to solve the HD problem for sewer systems consisting of a single series of pipes. The key idea in their methodology is that finding the optimal hydraulic design is framed as an SP problem. In an SP problem, a directed graph is equipped with a cost function that assigns a weight to every arc, and the goal is to find a minimum-cost path between a pair of nodes. In the graph proposed by Duque et al. [9], a feasible design corresponds to a path in the graph that encodes diameter and invert elevation decisions for every pipe of the series. Nodes in this network span all combinations of diameters and discretized invert elevations at every manhole. An arc in this graph encodes a diameter and invert elevations for a specific pipe. To obtain an optimal design, costs are assigned to the arcs of the graph and an SP algorithm retrieves the minimum-cost path.

A network layout of general topology poses several modelling and algorithmic challenges to extend the methodology in Duque et al. [9]. The foremost of these challenges is that in a tree-like layout, because there are multiple starting points (outer-branch pipes), there are multiple paths that connect every manhole to the outfall that needs to be evaluated simultaneously. This is in contrast to the case in which a series of pipes is considered because there is only one starting point. Furthermore, manholes merge multiple pipes upstream into one inner-branch pipe downstream (see manhole 5 in Figure 1 for an example). The combination of these two issues hinders the possibility of applying a one-to-one SP algorithm as in Duque et al. [9] because in our setting there are multiple starting nodes and the branch merging issue. Therefore, we found the individual shortest paths from the outfall to every outer manhole, and afterwards checked that the manholes with two or more incoming pipes would have an outgoing pipe at the lowest elevation.

The graph that represents all hydraulic designs for a given network layout requires input: a layout represented as a tree-structured graph $\mathcal{G}_T$, flow rates $\hat{q}$, a discrete set of commercially available diameters $\mathcal{D} = \{d_1, \ldots, d_B\}$, the roughness of the pipes (Manning coefficient or k_s), and the kinematic viscosity of water ν. The design flow rate $Q_{kk'}$ is obtained from the arcs in $\mathcal{G}_T$, which defines the design flow rate with magnitude [m³/s] and direction for every pipe in the network. Hence, $Q_{kk'} = \hat{q}_{ijt}$ for $i = l_k$, $j = l_{k'}$, and t is such that $x_{ijt} = 1$. To ease notation, we henceforth refer to the manhole associated with the tree node v_k by its index label l_k (as opposed to m_{l_k}).

In what follows, we first introduce an *HD graph* and then present the procedure for obtaining the optimal HD. Let $\mathcal{G}_D = (\mathcal{N}_D, \mathcal{A}_D)$ be a directed graph where nodes and arcs are defined as follows:

- $\mathcal{N}_D = \Big\{ v_1^1, \ldots, v_i^k, \ldots, v_{D^T}^T \Big\}$ is the set of nodes in the HD graph $\mathcal{G}_D$, where v_i^k is the *i*th *design node* associated to the tree node $v_k \in \mathcal{N}_T$ and $v_{D^T}^T$ is the last design node in the outfall node $v_T \in \mathcal{N}_T$.

Therefore, $\mathcal{N}_D$ is the union of disjoint subsets $\mathcal{N}_D^k = \left\{v_1^k, \ldots, v_{D^k}^k\right\}$ for every tree node $v_k \in \mathcal{N}_T$, i.e., $\mathcal{N}_D = \cup_{v_k \in \mathcal{N}_T} \mathcal{N}_D^k$.

- $\mathcal{A}_D = \left\{\left(v_i^k, v_j^{k'}\right) : v_i^k \in \mathcal{N}_D^k, \, v_j^{k'} \in \mathcal{N}_D^{k'}, (v_k, v_{k'}) \in \mathcal{A}_T\right\}$ is the set of arcs of the HD graph $\mathcal{G}_D$.

Every node v_i^k is equipped with two attributes: $Z\left(v_i^k\right)$ represents a potential invert elevation in [m] at which a pipe can meet manhole l_k and $\delta\left(v_i^k\right)$ represents a possible diameter in [m] for an upstream pipe of manhole l_k. There are as many nodes in $\mathcal{N}_D^k$ as combinations of diameters in $\mathcal{D}$ and invert elevations at manhole l_k. The latter is determined by the excavation limits at manhole l_k and a discretization parameter ΔZ [m] selected by the user (e.g., 0.1 m or 0.01 m). The invert elevations that are considered for manhole l_k are Z_{min}^k, $Z_{min}^k + \Delta Z$, $Z_{min}^k + 2\Delta Z$, $\ldots, Z_{max}^k$, where Z_{min}^k and Z_{max}^k denote the minimum and maximum invert elevations, respectively. Arc $\left(v_i^k, v_j^{k'}\right)$ represents a pipe from manhole l_k to manhole $l_{k'}$ with diameter $d\left(v_i^k, v_j^{k'}\right) = \delta\left(v_j^{k'}\right)$ and slope $s\left(v_i^k, v_j^{k'}\right) = \frac{Z\left(v_i^k\right) - Z\left(v_j^{k'}\right)}{L\left(v_i^k, v_j^{k'}\right)}$, where $L\left(v_i^k, v_j^{k'}\right)$ is the pipe length (Figure 2). Note that both diameter and invert elevations for every pipe fully characterize a solution to the HD problem and, therefore, they represent the decision variables of the problem. In our formulation of the problem, these decision variables are ultimately mapped to the arcs in $\mathcal{A}_D$ and their value is determined upon solving a shortest path problem as we explain next.
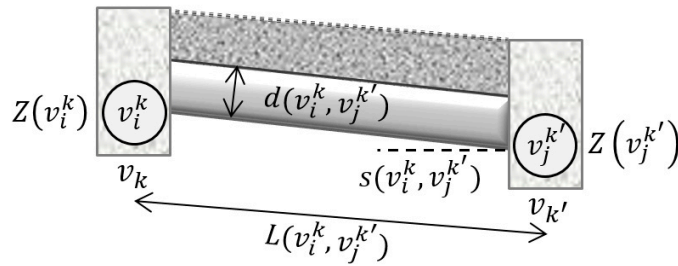

Figure 2. Hydraulic design representation of a single pipeline (diameter-slope) [9].

By construction, $\mathcal{G}_D$ ensures that basic hydraulic constraints are met. For instance, if water flows from manhole l_k towards manhole $l_{k'}$, only arcs satisfying a non-decreasing diameter and a positive slope (in favour of gravity) exist. Formally, arc $\left(v_i^k, v_j^{k'}\right)$ is part of graph $\mathcal{G}_D$ if and only if $\delta\left(v_i^k\right) \leq \delta\left(v_j^{k'}\right)$, $s\left(v_i^k, v_j^{k'}\right) > 0$, and all regulation-dependent constraints are satisfied. Note that typical hydraulic constraints such as minimum and maximum depth, flow velocity, shear stress and quasi-critical flow conditions, can be trivially evaluated for every arc $\left(v_i^k, v_j^{k'}\right)$ because all the required input for the evaluation is known, i.e., diameter, slope, and flow rate ($Q_{kk'}$).

Every arc in $\mathcal{A}_D$ is equipped with a cost $C\left(v_i^k, v_j^{k'}\right)$ that represents the corresponding construction cost (and potentially maintenance and operational cost). In our case studies we adopt the cost function introduced by Li and Matthew [34], however, we emphasize that any cost function can be employed. Similar to how our model trivializes the evaluation of hydraulic constraints, highly non-linear cost functions are reduced to a parameter of the design graph. Most cost functions are defined in terms of pipe diameter $d\left(v_i^k, v_j^{k'}\right)$, pipe length $L\left(v_i^k, v_j^{k'}\right)$, and average excavation depth $h\left(v_i^k, v_j^{k'}\right) = 0.5\left(h\left(v_i^k\right) - h\left(v_j^{k'}\right)\right)$, where $h\left(v_i^k\right)$ is the buried depth at the extreme of a pipe, i.e., $h\left(v_i^k\right) = Z^k - Z\left(v_i^k\right)$.

To find an optimal hydraulic design for the given input layout, we solve a one-to-all shortest path problem on $\mathcal{G}_D$ starting from a dummy node v_D that we connect to every node in $\mathcal{N}_D^T$ (Figure 1).

A solution to such a problem is a shortest path spanning tree $\mathcal{P}$ rooted in v_D. Formally, the underlying optimization problem is defined by Equation (11).

$$\min_{\mathcal{P}} \sum_{(v_i^k, v_j^{k'}) \in \mathcal{P}} C\left(v_i^k, v_j^{k'}\right) \tag{11}$$

subject to $\mathcal{P}$ being a shortest path spanning tree. The constraints that enforce $\mathcal{P}$ to be a shortest path spanning tree are omitted as we adopt an algorithmic approach to solve the problem instead of a mathematical programming approach (as in the LS problem). A generic formulation of the SP problem can be found in Ahuja et al. [39]. We use the Bellman–Ford algorithm [45] to optimally solve problem 11. The Bellman–Ford algorithm relies on labels at each node that save the cumulative cost from the starting node (v_D in our case). This label-correcting algorithm updates the cumulative cost for each node when a least-cost label is found and saves its predecessor node [39]. When all the nodes of the graph are evaluated, the cost labels are optimal. For more details about this algorithm, we refer the reader to Duque et al. [9].

When different series of pipes in the network have a common pipe, the solution of the one-to-all SP problem might include more than one arc between the manholes associated with the shared pipe. If that is the case, the final design is adjusted by choosing the inner-branch pipe with lower invert elevations (deeper pipe) and the largest diameter (e.g., pipe between tree layout nodes v_4 and v_5 in Figure 3). This adjustment ensures the design feasibility and does not compromise optimality since the selected pipe (deepest and largest in diameter) also belongs to the optimal solution of the one-to-all SP problem. The solution of this problem encodes the optimal hydraulic design of all the branches for a fixed input layout.
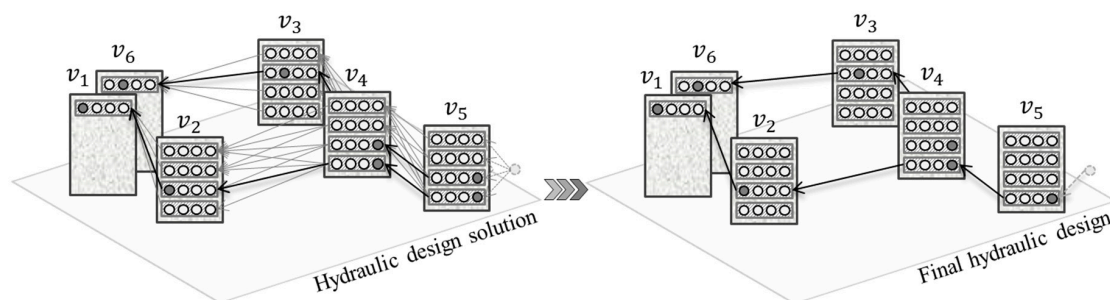

Figure 3. Selection of the definitive hydraulic design, eliminating double pipes generated from different shortest paths.

The entire procedure for the hydraulic design is summarized as follows:

1. Add a dummy design node v_D connecting every design node in the outfall $\mathcal{N}_D^T$ (see Figure 1);

2. Assign a cost value $C\left(v_i^k, v_j^{k'}\right)$ for every arc in $\mathcal{A}_D$;

3. Reverse all arcs in $\mathcal{A}_D$ and execute the Bellman–Ford algorithm [45] to obtain the one-to-all shortest paths from the outfall node v_D to every other node in $\mathcal{N}_D$;

4. Retrieve the best paths from v_D to any design node in $\mathcal{N}_D^k \ \forall \ v_k \in \mathcal{N}_T$;

5. Select the deepest and largest diameter pipe when multiple design arcs $\left(v_i^k, v_j^{k'}\right) \in \mathcal{A}_D$ that are part of different shortest paths belong to the same tree arc $(v_k, v_{k'}) \in \mathcal{A}_T$, since only one pipe should exist in each section.

3.4. Iterative Scheme

We conclude our methodological section with a description of the iterative scheme and an outline of the entire algorithm. As mentioned, during a single iteration the LS and HD problems are solved

in a sequential manner. Figure 4 summarizes the three modules and modelling frameworks for a single iteration.
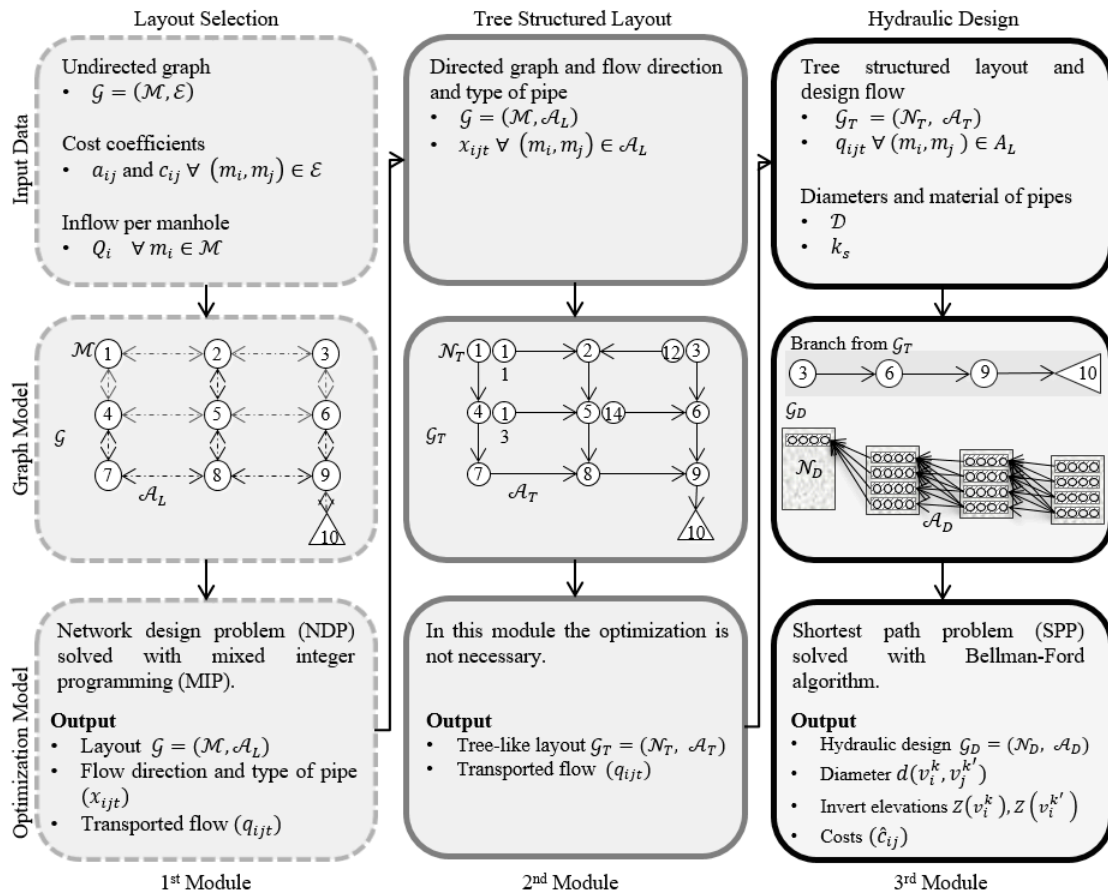

Figure 4. Sewer networks design (SND) modules and modelling frameworks for a single iteration.

Recall that the cost of a sewer network is a function of the diameter, length, and excavation volume of each pipe. However, diameters and excavation volumes are not decision variables of the LS problem and therefore the objective function (Equation (1)) approximates the hydraulic design cost as a function of the flow rates **q** and the flow direction **x**. Equation (1) implies a linear model to approximate the cost of a pipe from manhole m_i to manhole m_j, as a function of q_{ijt} and x_{ijt}. Let $\hat{c}_{ij}$ be a *prediction* of the cost of a pipe from m_i to m_j under the linear model in Equation (12).

$$\hat{c}_{ij} = c_{ij}q_{ijt} + a_{ij}x_{ijt} \tag{12}$$

Our goal is to estimate parameters c_{ij} and a_{ij} for all $\left(m_i, m_j\right) \in \mathcal{A}_L$ as a new input to the LS problem. We estimate c_{ij} and a_{ij} independently for each arc by means of linear regression. For a particular arc $\left(m_i, m_j\right)$, we regress previous values $\hat{q}_{ijt}$ and $\hat{x}_{ijt}$ against the cost obtained in the hydraulic design for this pair of manholes, i.e., $C(v, v')$ for some arc $(v, v') \in \mathcal{P}$ that belongs to the shortest path spanning tree for which nodes v and v' are associated to manholes m_i and m_j, respectively. Note that indices for v and v' were dropped to avoid cumbersome notation. As the algorithm progresses, we obtain new observations of $\hat{c}_{ij}$, q_{ijt}, and x_{ijt} that refine the estimation of the parameters. Refining the estimation ultimately improves the prediction of the cost and, therefore, the quality of subsequent layouts.

Figure 5 presents the entire framework in a flow diagram. We start with an initialization procedure that generates a hydraulic design based on a randomized layout. This step is necessary to obtain the observations for the parameter estimation in the regression step. The randomized layouts are obtained by simply assigning random values to c_{ij} and a_{ij} and solving the MIP model. In the iterative phase,

the algorithm executes the three modules described above until a maximum number of iterations is met. After each iteration, the linear regression considers all the costs obtained from the hydraulic designs, providing a better estimate of the LS cost coefficients c_{ij} and a_{ij}.
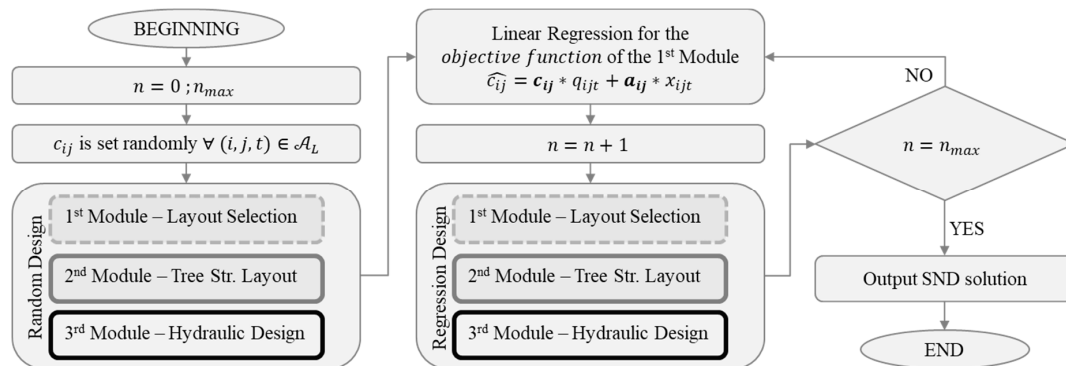

Figure 5. Sewer networks design (SND) problem flow chart.

3.5. Software and Data

The SND framework is mainly coded in Java. The layout selection module, however, is coded in Mosel, which is a mathematical modelling and programming language to model the MIP problem [43]. The SND software is available to download from https://bitbucket.org/nduquevillarreal/snd_java-xpress.git. It was designed to have few data requirements for the design of sewer systems. We provide the input data of our case studies in the Supplementary Material.

3.6. Case Studies

To compare our approach to others found in the literature, we tested our methodology on two benchmark networks. The first benchmark, proposed by Li and Matthew [34], has 79 pipes and 78 manholes and has been used over the years by several authors to test their algorithms [4,12,46,47]. The second benchmark network proposed by Moeini and Afshar [31], has 81 manholes and 144 pipes. In addition, we also tested our methodology in a network that is part of the real sewer network in Bogota, Colombia. This network, labelled *Chicó*, has 109 manholes, 160 pipes with lengths between 65 m and 204 m, steep ground and a total flow of 1.526 m^3/s. For the hydraulic design of these networks, we used the cost function, design parameters and constraints proposed by Li and Matthew [34]. Table 1 presents the hydraulic constraints and Equations (13) and (14) present the construction costs for a pipe f_p and a manhole f_m, respectively [34].

$$f_p = \begin{pmatrix} \left(4.27 + 93.59d^2 + 2.86dh + 2.39h^2\right)L & with\ d \leq 1\ m\ and\ h \leq 3\ m \\ \left(36.47 + 88.96d^2 + 8.70dh + 1.78h^2\right)L & with\ d \leq 1\ m\ and\ h > 3\ m \\ \left(20.50 + 149.27d^2 - 58.96dh + 17.75h^2\right)L & with\ d > 1\ m\ and\ h \leq 4\ m \\ \left(78.44 + 29.25d^2 + 31.80dh - 2.32h^2\right)L & with\ d > 1\ m\ and\ h > 4\ m \end{pmatrix} \tag{13}$$

$$f_m = \begin{pmatrix} 136.67 + 166.19d^2 + 3.5dh + 16.22h^2 & with\ d \leq 1\ m\ and\ h \leq 3\ m \\ 132.91 + 790.94d^2 - 280.23dh + 34.97h^2 & with\ d \leq 1\ m\ and\ h > 3\ m \\ 209.74 + 57.53d^2 + 10.93dh + 19.88h^2 & with\ d > 1\ m\ and\ h \leq 3\ m \\ 210.66 - 113.04d^2 + 126.43dh - 0.60h^2 & with\ d > 1\ m\ and\ h > 3\ m \end{pmatrix} \tag{14}$$

Table 1. Hydraulic design constraints for round pipes [34].

	Constraint	Value	Condition
1	Minimum diameter	0.2 m	Always
2	Maximum filling ratio	0.6	$d \leq 0.3$ m
		0.7	$0.35 \leq d \leq 0.45$ m
		0.75	$0.50 \leq d \leq 0.90$ m
		0.8	$d \geq 1.0$ m
3	Minimum velocity	0.7 m/s	$d \leq 0.5$ m
		0.8 m/s	$d > 0.5$ m
4	Maximum velocity	5.0 m/s	
5	Minimum gradient	0.003	Flow rate < 0.015 m^3/s

The variables f_p and f_m are calculated in terms of the pipe diameter d, the length L, and average buried depth h. The network construction cost is the sum of all costs of pipes and manholes in the network. Annual maintenance cost is 4.2% of the construction cost, considered for a 10-year period [34]. Construction and operational costs of pumping are not considered since our design considers gravity-driven systems. Additionally, we used concrete pipes (Manning coefficient $n = 0.014$ and the following list of diameters: {0.2, 0.25, 0.3, 0.35, 0.38, 0.4, 0.45, 0.5, 0.53, 0.6, 0.7, 0.8, 0.9, 1.0, 1.05, 1.20, 1.35, 1.4, 1.5, 1.6, 1.8, 2, 2.2, 2.4}.

The input data and detailed results for all networks can be found in the Supplementary Material and the Git repository. The input data includes the Id of the manholes and pipes, inflow Q_i in [m^3/s] and the x, y and z coordinates in [m] at each manhole. The results for the hydraulic design for each pipe include upstream and downstream manhole, type [inner/outer], diameter [m], upstream and downstream invert elevation [m], flow rate [m^3/s] and slope [–].

4. Results

4.1. Benchmark Li and Matthew

After 10 iterations using an elevation change of $\Delta Z = 0.1$ m, the hydraulic design for this layout costs ¥ 1.33×10^6 and has a maintenance cost of ¥ 0.55×10^6. The design has a maximum excavation depth of 9.45 m, and took a computational time of 45 min to solve both the LS and HD. An additional iteration was done to design the same layout with higher precision (i.e., a finer elevation change of $\Delta Z = 0.01$ m). Figure 6 shows the layout and hydraulic design obtained using $\Delta Z = 0.01$ m. The construction and maintenance costs were reduced to ¥ 1.29×10^6 and ¥ 0.54×10^6, respectively. For this design, we obtained a maximum excavation depth of 9.38 m and a computational time of 58 min for a single iteration of the hydraulic design. The detailed hydraulic design can be found in the Supplementary Material.

Table 2 shows a comparison of the results found in the literature for the network design proposed by Li and Matthew [34], which do not include pumping. All the authors presented in this table use the same constraints (Table 1) and cost functions comprising pipe f_p and manhole f_m costs (Equations (13) and (14)) [34]. We compare results obtained using different optimization approaches for cases where the layout is taken by Li and Matthew [34] and only the hydraulic design is optimized [4,12,46,47] and where, as for the present work, both layout and hydraulic design are optimized [20,28,31,48]. Our methodology obtained the lowest cost after 10 iterations with a modest computational time.
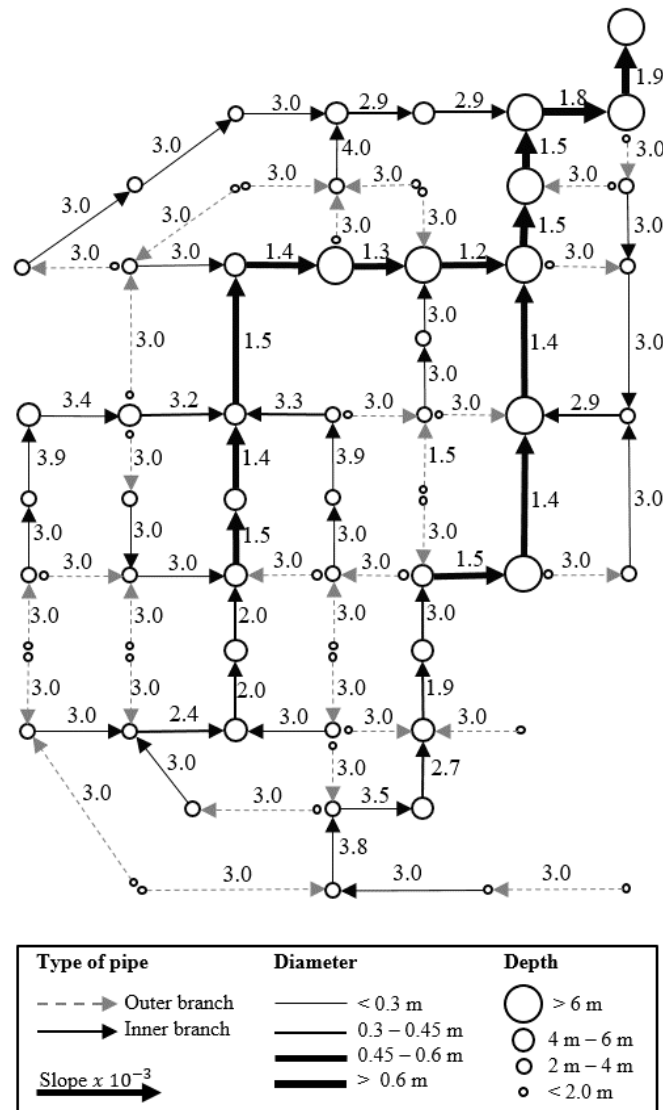
Figure 6. Layout and hydraulic design ($\Delta Z = 0.01$ m) of the benchmark proposed by Li and Matthew [34], using the proposed methodology.

Table 2. Comparison of the construction cost for the benchmark proposed by Li and Matthew [34].

Method	Researchers	Optimized Layout	Construction Cost (Yuan) $\times 10^6$
MGA	Pan and Kao [46]	No	$Y = 1.91$
Adaptive GA	Haghighi and Bakhshipour [12]	No	$Y = 1.84$
MILP	Safavi and Geranmehr [4]	No	$Y = 1.57$
SDE-GOBL	Liu et al. [47]	No	$Y = 1.53$
Loop-by-loop cutting algorithm and GA-DDDP	Haghighi and Bakhshipour [28]	Yes	$Y = 1.59$
Loop-by-loop cutting algorithm and TS	Haghighi and Bakhshipour [20]	Yes	$Y = 1.43$
Reliability - DDDP	Haghighi and Bakhshipour [48]	Yes	$Y = 2.41$
ACOA-TGA-NLP	Moeini and Afshar [31]	Yes	$Y = 1.39$
MIP and DP	**(Present work)**	**Yes**	$Y = \mathbf{1.29}$

Our layout could be improved with more iterations to gain a better approximation of the costs in terms of the flow and pipe direction. A non-linear cost approximation for the LS problem could also improve the results of the layout and, therefore, the entire sewer network design algorithm.

4.2. Benchmark Moeini and Afshar

We obtained a design for the benchmark network studied in Moeini and Afshar [31] that attains a total cost of ¥ 524,687 after 30 iterations of the algorithm, using an elevation change of $\Delta Z = 0.1$ m. Additionally, we conducted an experiment in which the layout is fixed to that of Moeini and Afshar [31] and solved only the hydraulic design problem. This design costs ¥ 569,334, using the same elevation change of $\Delta Z = 0.1$ m. Both scenarios assume the same constraints, cost function and conditions described in Li and Matthew [34]. Total costs include the construction costs and maintenance costs (Table 3). Column 1 describes the statistics of interest; column 2 shows the results reported in Moeini and Afshar [31], column 3 shows the results when we only apply the HD component of our algorithm on the layout reported by Moeini and Afshar [31]; and finally, column 4 shows the results of our algorithm.

Table 3. Comparison of the total cost for the benchmark proposed by Moeini and Afshar [31].

	ACOA-TGA-NLP [31]	DP (Present work) with Fixed Layout [31]	MIP and DP (Present Work)
Costs			
Construction Cost	¥ 640,845 *	¥ 400,940	¥ 369,498
Maintenance Cost	¥ 269,155 *	¥ 168,395	¥ 155,189
Total Cost	**¥ 910,000 ***	**¥ 569,335**	**¥ 524,687**
Computational effort			
Iterations [–]	10	1	30
Time [min]	198	4	115
Physical characteristics			
Average diameter [m]	0.331 **	0.262	0.267
Average depth [m]	Not reported	2.07	1.95
Outfall diameter [m]	1.05 **	0.7	0.8
Outfall depth [m]	Not reported	6.6	5.4

* Estimated from the reported total cost. ** Based on the reported design solution [31].

4.3. Real Network: Chicó

Figure 7 shows the layout obtained after 25 iterations using an elevation change of $\Delta Z = 0.1$ m. The hydraulic design for this layout uses the same constraints, cost function and conditions described in Li and Matthew [34]. The construction and maintenance costs are ¥ 699,097 and ¥ 293,621, respectively. Therefore, the total cost is ¥ 992,718. The design has a maximum excavation depth of 15.9 m and a diameter in the outfall of 1.05 m. All iterations took a computational time of 113 min to solve both the LS and HD. The detailed hydraulic design can be found in the Supplementary Material as well as the distribution of the coefficient of determination R^2 of the linear approximation of the cost function for each pipe after 25 iterations. The average R^2 is 0.84, with a standard deviation of 0.23. The R^2 across all the individual linear regressions varied from a minimum of 0.004 and a maximum of 1. Pipes with $R^2 = 1$ represent those pipes that had the exact same design along the iterations.

4.4. Convergence Curves

We close this section with a convergence analysis of the algorithm. For each case study, Figure 8 shows the construction cost of the sewer network design obtained at each iteration of our algorithm. Our algorithm behaves somewhat similarly across the three case studies. The first design leads to a relatively high cost due to a poor initial layout that is obtained with random coefficients in Equation (1). However, all subsequent iterations benefit from the cost function approximation that is employed in the LS problem. Moreover, as the algorithm progresses, the cost function approximation is refined with new hydraulic design solutions that populate the linear regression. In short, these convergence curves highlight a core component of our framework that enables the HD problem to provide feedback to the LS problem.
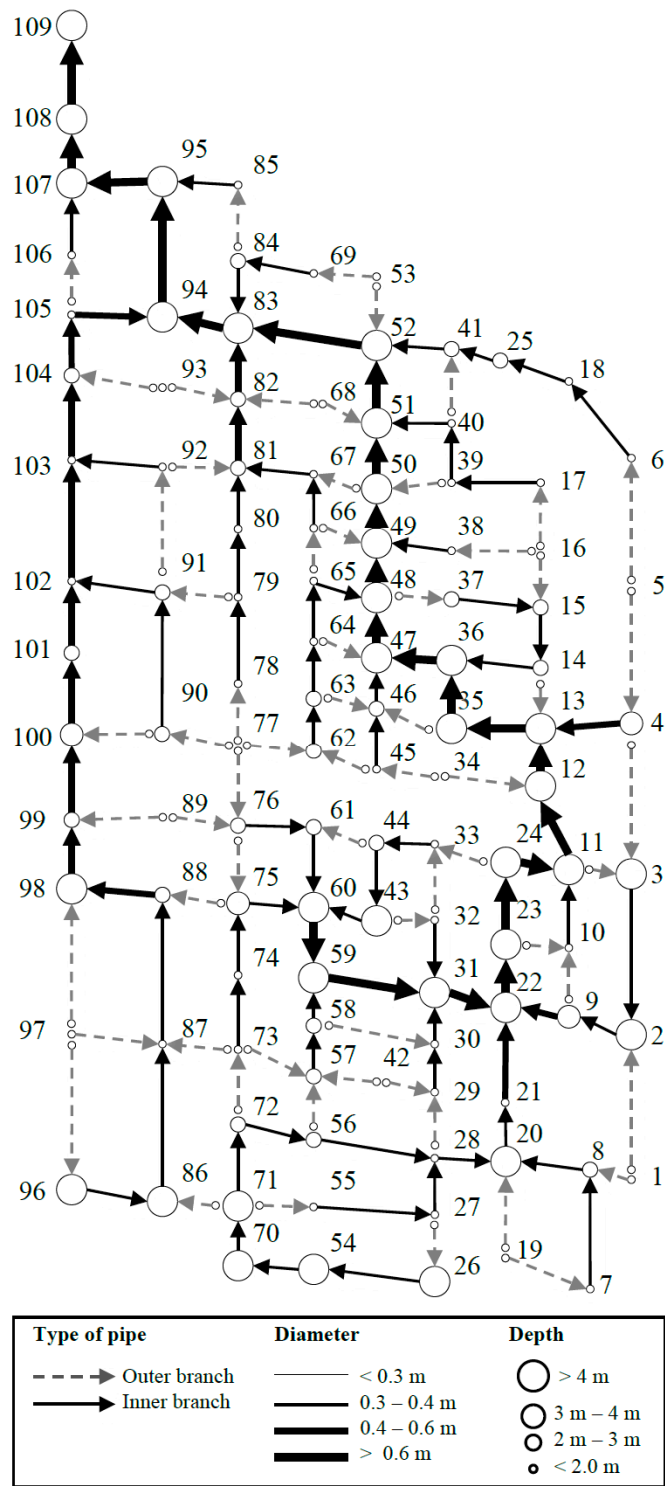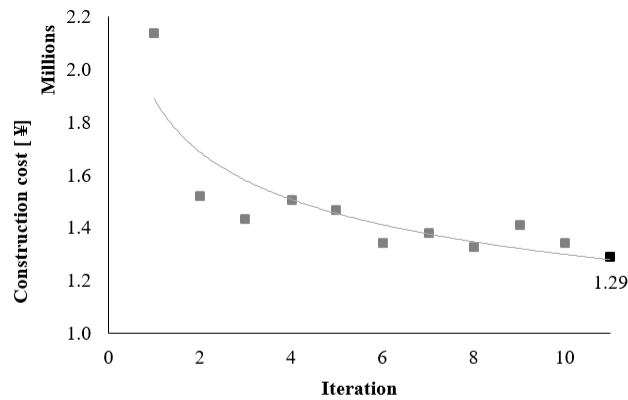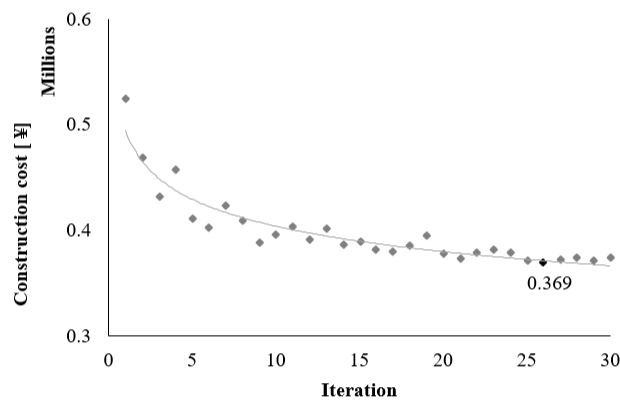
Figure 7. Layout and hydraulic design of Chicó sewer network.

Li and Matthew (1990)

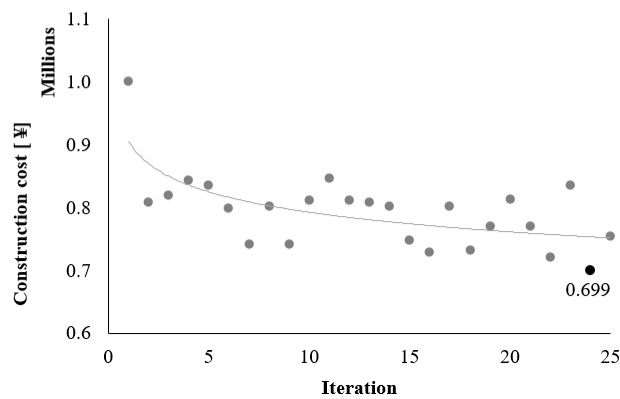**Moeini and Afshar (2019)**

Chicó

Figure 8. Convergence curves for each case study concerning construction costs.

5. Discussion

For all case studies, the proposed methodology found a better solution than those reported in the literature. These are near-optimal solutions with respect to construction cost that satisfy all hydraulic constraints and ensure feasible pipe connections in a sewer system. The LS is modelled as an MIP and is solved exactly (for a particular objective function approximation) with an off-the-shelf optimization solver. The cost function approximation for the LS problem is enhanced by the results of new hydraulic designs, as shown with the convergence curves in Figure 8. Hydraulic design solutions of previous iterations are used as additional data points in a linear regression model that estimates construction costs as a linear function of the flow rate and the flow direction. These solutions are not intended to be the global optimal solution for the network design, since the layout selection problem uses a

proxy for construction costs that might not be the best approximation of the cost. Nevertheless, once a near-optimal layout is selected, we obtained an optimal hydraulic design for the specific near-optimal layout. As expected, during the first iterations, there is a big improvement of the LS cost function and then the algorithms start to converge to a constant value. The minimum cost hydraulic design represents the best layout obtained for that network within the number of iterations or accepted tolerance for the design.

The comparison of all the results obtained by different authors (Table 2) over the well-known benchmark proposed by Li and Matthew [34] shows that our methodology obtains the best cost, even if the layout optimization is not optimal. Some authors used the exact same layout proposed by Li and Matthew [34] and only solved the hydraulic design problem with the best result cost ¥ 1.53×10^6 using the self-adaptive differential evolution algorithm [47]. Later researchers tackled the LS problem as well, aiming for better results for the final hydraulic design. Previously, the best solution for both the LS and HD problems cost ¥ 1.39×10^6 using ACOA-TGA-NLP [31]. Our methodology shows a big improvement by saving up to 16% of the costs compared to the lowest cost found in the literature, with a final cost of ¥ 1.29×10^6.

Regarding the results for the second benchmark proposed by Moeini and Afshar [31] presented in Table 3, we obtained up to 37% savings on the HD using Moeini and Afshar's layout, and up to 42% savings when solving both the LS and HD with our approach. As expected, we have a high computational effort given that our methodology makes an exhaustive search among all possible design combinations. Nonetheless, the computational time spent by our algorithm was lower than the result obtained using ACOA-TGA-NLP [31]. Therefore, we can say that our computational time is acceptable considering the improvement we obtain in the total construction cost. The last case study was proposed as a new benchmark using a network with real topographic data with mild slopes.

Our methodology can be improved by getting a better proxy for the cost function used to solve the LS problem. The current cost function is obtained through a linear regression of previous design options for the network. A non-linear cost approximation for the LS problem could also improve the results of the layout and, therefore, the entire sewer network design algorithm. This improvement might also have a positive impact on the computational time required to obtain a near-optimal sewer network design.

6. Conclusions

We proposed a mathematical optimization framework for the design of sewer networks, with the objective of minimizing construction costs while ensuring proper performance of the sewer network. A graph modelling framework was proposed to represent the different features of a sewer network design. The layout selection graph (Figure 1) represents the type and flow direction of each pipe that are the decision variables for the LS problem. A second graph represents the layout as an open network, with a tree structure that avoids loops. The last graph represents the hydraulic design problem, modelling the diameters and invert elevations of the pipes.

The LS is modelled as an MIP, which is solved for a particular objective function with an off-the-shelf optimization solver. The objective function approximation is enhanced by new hydraulic design solutions, which are used as additional data points in a linear regression model that estimates construction costs as a linear function of the flow rate and the flow direction. On the other hand, the extension of the hydraulic design methodology proposed by Duque et al. [9] ensures the global optimal hydraulic design for the given layout and design conditions. The iterative scheme significantly reduces construction costs during the first iterations and it encounters a near-optimal solution as the algorithm progresses. Better approximations of the real cost function for the layout selection model might translate into further improvement of the overall design. Moreover, the method can be extended to consider pumping stations for flat terrain by using additional arcs in the design graph.

Since the size of the problem increases exponentially with the number of pipes in the network, there is a computational capacity burden in that fewer iterations can be executed for large-scale

networks. Future research is needed to formulate better cost function approximations for the layout selection (e.g., non-linear functions) to produce a better representation of the construction cost in this stage of the design. Furthermore, our methodology could leverage parallel computing to speed up certain components of the framework such as the hydraulic design.

Nomenclature

Layout Selection

$\mathcal{G}$	is the graph that represents the layout selection problem.
$\mathcal{M}$	is the set of nodes representing manholes.
$\mathcal{E}$	is the set of undirected edges representing links between two nodes $m_i \in \mathcal{M}$ and $m_j \in \mathcal{M}$.
Q_i	is the inflow at manhole $m_i \in \mathcal{M}$.
Q_K	is the total flow in the system reaching the outfall manhole $m_k \in \mathcal{M}$.
Z_i	is the ground elevation at manhole $m_i \in \mathcal{M}$.
$\mathcal{T}$	is the set of possible types of pipes, containing outer-branch pipes (t_1) and inner-branch pipes (t_2).
$\mathcal{A}_L$	is the set of *directed* links between two manholes, m_i and m_j, so that $\left(m_i, m_j\right) \in \mathcal{E}$.
	x_{ijt} is the binary decision variable that represents the flow direction and connection type in the network layout, for all $\left(m_i, m_j\right) \in \mathcal{A}_L$ and $t \in \mathcal{T}$.
q_{ijt}	is the continuous decision variable that represents the flow through arc $\left(m_i, m_j\right)$ of type t, for all $\left(m_i, m_j\right) \in \mathcal{A}_L$ and $t \in \mathcal{T}$.
M	is a large positive number.

Tree-structured Layout

$\mathcal{G}_T$	is the graph that represents the selected layout as a tree-structured network.
$\mathcal{N}_T$	is the set of nodes in the tree-structured graph $\mathcal{G}_T$.
l_k	is a label that represents the index of the manhole associated to tree node $v_k \in \mathcal{N}_T$
$\mathcal{N}_T^i$	is a subset of $\mathcal{N}_T$ that contains tree nodes associated with manhole $m_i \in \mathcal{M}$.
$\mathcal{A}_T$	is the set of arcs in the tree-structured graph $\mathcal{G}_T$.

Hydraulic Design

$\mathcal{G}_D$	is the auxiliary graph used to represent the hydraulic design problem.
$\mathcal{N}_D$	is the set of nodes in the hydraulic design graph $\mathcal{G}_D$, which is divided in subsets of nodes $\mathcal{N}_D^k$ related with the tree node $v_k \in \mathcal{N}_T$.
$\mathcal{A}_D$	is the set of arcs of the hydraulic design graph $\mathcal{G}_D$.
$\mathcal{D}$	is the discrete set of commercially available pipe diameters.

$\delta\left(v_i^k\right)$	is the discrete decision variable that represents a possible diameter for an upstream pipe of tree node $v_k \in \mathcal{N}_T$ and manhole l_k.
$Z\left(v_i^k\right)$	is the continuous decision variable that represents the elevation above a reference level of a node of the graph $\mathcal{G}_D$.
s	is the slope for each pipe, fully determined by the invert elevations $Z\left(v_i^k\right)$ and $Z\left(v_j^k\right)$ at the extremes of the arc $\left(v_i^k, v_j^{k'}\right) \in \mathcal{A}_D$ which represents a pipe.
k_s	is the absolute roughness of the pipes.

References

1. Elimam, A.A.; Charalambous, C.; Ghobrial, F.H. Optimum design of large sewer networks. *J. Environ. Eng.* **1989**, *115*, 1171–1190. [CrossRef]
2. Swamee, P.K.; Sharma, A.K. Optimal design of a sewer line using linear programming. *Appl. Math. Model.* **2013**, *37*, 4430–4439. [CrossRef]
3. Gupta, M.; Rao, P.; Jayakumar, K. Optimization of integrated sewerage system by using simplex method. *VFSTR J. STEM* **2017**, *3*, 2455-2062.
4. Safavi, H.; Geranmehr, M.A. Optimization of sewer networks using the mixed-integer linear programming. *Urban Water J.* **2017**, *14*, 452–459. [CrossRef]
5. Holland, M.E. Computer Models of Waste-Water Collection Systems. Ph.D. Thesis, Harvard University, Cambridge, MA, USA, May 1966.
6. Mansouri, M.; Khanjani, M. Optimization of sewer networks using nonlinear programming. *J. Water Wastewater.* **1999**, *10*, 20–30.
7. Mays, L.W.; Yen, B.C. Optimal cost design of branched sewer systems. *Water Resour. Res.* **1975**, *11*, 37–47. [CrossRef]
8. Walters, G.A. The design of the optimal layout for a sewer network. *Eng. Optim.* **1985**, *9*, 37–50. [CrossRef]
9. Duque, N.; Duque, D.; Saldarriaga, J. A new methodology for the optimal design of series of pipes in sewer systems. *J. Hydroinf.* **2016**, *18*, 757–772. [CrossRef]
10. Kulkarni, V.S.; Khanna, P. Pumped wastewater collection systems optimization. *J. Environ. Eng.* **1985**, *111*, 589–601. [CrossRef]
11. Afshar, M.H. Rebirthing genetic algorithm for storm sewer network design. *Sci. Iran.* **2012**, *19*, 11–19. [CrossRef]
12. Haghighi, A.; Bakhshipour, A.E. Optimization of sewer networks using an adaptive genetic algorithm. *Water Resour. Manag.* **2012**, *26*, 3441–3456. [CrossRef]
13. Palumbo, A.; Cimorelli, L.; Covelli, C.; Cozzolino, L.; Mucherino, C.; Pianese, D. Optimal design of urban drainage networks. *Civ. Eng. Environ. Syst.* **2014**, *31*, 79–96. [CrossRef]
14. Cozzolino, L.; Cimorelli, L.; Covelli, C.; Mucherino, C.; Pianese, D. An innovative approach for drainage network sizing. *Water* **2015**, *7*, 546–567. [CrossRef]
15. Cisty, M. Hybrid genetic algorithm and linear programming method for least-cost design of water distribution systems. *Water Resour. Manag.* **2010**, *24*, 1–24. [CrossRef]
16. Hassan, W.H.; Jassem, M.H.; Mohammed, S.S. A GA-HP model for the optimal design of sewer networks. *Water Resour. Manag.* **2018**, *32*, 865–879. [CrossRef]
17. Afshar, M.H. A parameter free continuous ant colony optimization algorithm for the optimal design of storm sewer networks: Constrained and unconstrained approach. *Adv. Eng. Softw.* **2010**, *41*, 188–195. [CrossRef]
18. Moeini, R.; Afshar, M.H. Arc based ant colony optimization algorithm for optimal design of gravitational sewer networks. *Ain Shams Eng. J.* **2017**, *8*, 207–223. [CrossRef]
19. Yeh, S.F.; Chang, Y.J.; Lin, M.D. Optimal design of sewer network by tabu search and simulated annealing. In Proceedings of the 2013 IEEE International Conference on Industrial Engineering and Engineering Management, Bangkok, Thailand, 10–13 December 2013; Institute of Electrical and Electronics Engineers (IEEE): Piscataway, NJ, USA, 2013; pp. 1636–1640.
20. Haghighi, A.; Bakhshipour, A.E. Deterministic integrated optimization model for sewage collection networks using tabu search. *J. Water Resour. Plan. Manag.* **2015**, *141*. [CrossRef]
21. Ahmadi, A.; Zolfagharipoor, M.A.; Nafisi, M. Development of a hybrid algorithm for the optimal design of sewer networks. *J. Water Resour. Plan. Manag.* **2018**, *144*. [CrossRef]

22. Afshar, M.; Zaheri, M.; Kim, J. Improving the efficiency of cellular automata for sewer network design optimization problems using adaptive refinement. *Procedia Eng.* **2016**, *154*, 1439–1447. [CrossRef]

23. Rohani, M.; Afshar, M.H. Optimal design of Sewer network using Cellular Automata. *SCIREA J. Hydraul. Eng.* **2016**, *1*, 1–17.

24. Zaheri, M.M.; Ghanbari, R.; Afshar, M.H. A two-phase simulation–optimization cellular automata method for sewer network design optimization. *Eng. Optim.* **2019**, *52*, 620–636. [CrossRef]

25. Navin, P.K.; Mathur, Y.P.; Kumar, D. Layout optimization of sewer network using minimum cumulative flow in the sewer network. In *Advances in Waste Management*; Springer Science and Business Media LLC.: Berlin, Germany, 2018; pp. 333–343.

26. Hsie, M.; Wu, M.Y.; Huang, C.Y. Optimal urban sewer layout design using Steiner tree problems. *Eng. Optim.* **2019**, *51*, 1980–1996. [CrossRef]

27. Walters, G.A.; Smith, D.K. Evolutionary design algorithm for optimal layout of tree networks. *Eng. Optim.* **1995**, *24*, 261–281. [CrossRef]

28. Haghighi, A. Loop-by-loop cutting algorithm to generate layouts for urban drainage systems. *J. Water Resour. Plan. Manag.* **2013**, *139*, 693–703. [CrossRef]

29. Afshar, M.H.; Mariño, M.A. Application of an ant algorithm for layout optimization of tree networks. *Eng. Optim.* **2006**, *38*, 353–369. [CrossRef]

30. Moeini, R.; Afshar, M.H. Layout and size optimization of sanitary sewer network using intelligent ants. *Adv. Eng. Softw.* **2012**, *51*, 49–62. [CrossRef]

31. Moeini, R.; Afshar, M.H. Extension of the hybrid ant colony optimization algorithm for layout and size optimization of sewer networks. *J. Environ. Informatics* **2019**, *33*, 68–81. [CrossRef]

32. Navin, P.K.; Mathur, Y.P. Layout and component size optimization of sewer network using spanning tree and modified PSO algorithm. *Water Resour. Manag.* **2016**, *30*, 3627–3643. [CrossRef]

33. Mays, L.W.; Wenzel, H.G. Optimal design of multilevel branching sewer systems. *Water Resour. Res.* **1976**, *12*, 913–917. [CrossRef]

34. Li, G.; Matthew, R.G.S. New approach for optimization of urban drainage systems. *J. Environ. Eng.* **1990**, *116*, 927–944. [CrossRef]

35. Diogo, A.F.; Graveto, V.M. Optimal layout of sewer systems: A deterministic versus a stochastic model. *J. Hydraul. Eng.* **2006**, *132*, 927–943. [CrossRef]

36. Rohani, M.; Afshar, M.; Moeini, R. Layout and size optimization of sewer networks by hybridizing the GHCA model with heuristic algorithms. Scientia Iranica. *Trans. A Sch. J.* **2015**, *22*, 1742–1754.

37. Bakhshipour, A.E.; Makaremi, Y.; Dittmer, U. Multiobjective design of sewer networks. *J. Hydraul. Struct.* **2017**, *3*, 49–56. [CrossRef]

38. Nocedal, J.; Wright, S.J. *Numerical Optimization*, 2nd ed.; Springer: Berlin/Heidelberg, Germany; New York, NY, USA, 2006.

39. Smith, D.K.; Ahuja, R.K.; Magnanti, T.L.; Orlin, J.B. Network Flows: Theory, Algorithms, and Applications. *J. Oper. Res. Soc.* **1994**, *45*, 1340. [CrossRef]

40. Butler, D.; Davies, J.W. *Urban Drainage*, 3rd ed.; Spon Press: Milton Park, UK; Abingdon, UK, 2011.

41. Bertsimas, D.; Tsitsiklis, J.N. *Introduction to Linear Optimization*; Athena Scientific: Belmont, MA, USA, 1997; Volume 6, pp. 479–530.

42. Johnson, D.S.; Lenstra, J.K.; Kan, A.R. The complexity of the network design problem. *Networks* **1978**, *8*, 279–285. [CrossRef]

43. FICO. Fair Isaac Corporation FICO. In *Xpress-MP optimizer: Reference Manual*; Leamington Spa, Release 20.0; FICO: Warwickshire, UK, 2009; pp. 19–23. Available online: https://www.msi-jp.com/xpress/learning/square/optimizer.pdf (accessed on 8 June 2020).

44. *Gurobi Optimizer, 5.0.* Available online: http://www.gurobi.com (accessed on 8 June 2020).

45. Bellman, R. On a routing problem. *Q. Appl. Math.* **1958**, *16*, 87–90. [CrossRef]

46. Pan, T.C.; Kao, J.J. GA-QP Model to optimize sewer system design. *J. Environ. Eng.* **2009**, *135*, 17–24. [CrossRef]

47. Liu, C.; Han, H.G.; Wang, C.; Qiao, J. An adaptive differential evolution algorithm for sewer networks design. In Proceedings of the 11th World Congress on Intelligent Control and Automation, Shenyang, China, 29 June–4 July 2014; Institute of Electrical and Electronics Engineers (IEEE): Piscataway, NJ, USA, 2014; pp. 3577–3583.

48.   Haghighi, A.; Bakhshipour, A.E. Reliability-based layout design of sewage collection systems in flat areas. *Urban Water J.* **2016**, *13*, 790–802. [CrossRef]

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.