

Article

Cost–Benefit Prediction of Asset Management Actions on Water Distribution Networks

Amir Nafi ^{1,2,*}  and Jonathan Brans ¹

¹ Unité Mixte de Recherche Gestion Territoriale de l'Eau et de l'Environnement (GESTE) IRSTEA-ENGEES, 1 quai Koch, 67070 Strasbourg, France

² CSIP-ICube, Université de Strasbourg, 3-5, rue de l'Université, 67084 Strasbourg, France

* Correspondence: amir.nafi@engees.unistra.fr; Tel.: +33-388-248-293

Received: 19 June 2019; Accepted: 15 July 2019; Published: 25 July 2019



Abstract: The potential costs and benefits of a combination of asset management actions on the water distribution network are predicted. Two types of actions are considered: maintenance actions and renewal actions. Leak detection and repair of failures on connections and pipes define the set of potential maintenance actions to be carried out. Renewal actions concern connections, pipes, and meters. All these actions represent the model's decision variables in order to determine a trade-off between two objectives: (i) the maximization of the water efficiency rate and (ii) the minimization of the total cost of actions to be carried out on the water system. The assessment of objective functions is ensured by an artificial neural network (ANN) trained on a French mandatory database «SISPEA». A non-dominated sorting genetic algorithm (NSGA-II) is coupled to the ANN to reach the set of compromised solutions representing potential actions to achieve. Applied to a real water distribution system in the southeast of France, the proposed decision model indicates that the improvement of water efficiency rate (*WER*) in the short term requires increasing operation expenditures (*OPEX*), which represent 99% of the total cost. Results show the existence of a threshold effect that implies to use the budget in a certain way to improve performance. A potential solution can be chosen by the decision maker among the generated Pareto front with regard to the constraint on the budget and the targeted *WER*.

Keywords: actions; asset management; ANN; prediction; performance; water utility; water system; NSGA-II

1. Introduction

Water utility performance monitoring is widely addressed in the literature. IWA initiative carried out by Ref. [1] to build key performance indicators (KPIs) led to the emergence of national mandatory databases in several countries in order to improve the management of water utility and ensure transparency against stakeholders and users. However, KPIs are generally measured on an ex-post basis in order to assess the ability of conducted policy to achieve planned goals; otherwise, corrective actions can be planned in the case of a mismatch. This way of management could be expensive in terms of time and money.

One possible improvement to avoid this mismatch is the use of a decision-aiding model to predict KPIs based on potential decisions and a set of explanatory data. A possible shortfall concerns the absence of data collection at the scale of the water utility, which renders it difficult to train and fit a prediction model. The existence of an information system (IS) seems to be a prerequisite for the assessment and the prediction of KPIs. This shortfall tends to be solved. In fact, in the last 2 decades, we observe the development of sensors technologies and information and communications technology (ICT) that encourage water utility to install smart devices in order to monitor water systems in real

time and collect information about their operation. The relevance of adopting smart water systems and the potential benefits in terms of leak management, water quality monitoring, and energy savings are discussed in Ref. [2]. Smart systems generate an important quantity of data which are not always exploited in the decision-making process. Data gathering improves the water utility information system (IS) and constitutes a prerequisite for prospective analysis. The current research addresses the assessment of KPIs in an ex-ante way based on the exploitation of data due to the emergence of mandatory databases and the deployment of smart devices in the water systems. The current paper aims at answering the following question: How can the existing data collections or IS be exploited for prospecting asset management actions and assessing their costs and benefits in an ex-ante way?

For any planning of asset management actions, the assessment of expected costs and benefits is recommended because it allows decisions mitigation. The importance of cost–benefit quantification in the determination of optimal maintenance time is underlined in Ref. [3]. Models for asset management of water pipes seem to be driven by the estimation of the optimal date of renewal based on the deterioration of the asset, the assessment of whole life costing [4], the achievement of a critical threshold for the number of breaks [5] or the rendered service (pressure, flow, quality) under economic or technical constraints [6,7]. Pipes renewal planning considering multiple objectives can be achieved by genetic algorithms [8]. The problem of water pipe renewal planning based on a cost–benefit approach is addressed in Ref. [9]. Authors define five items of benefit. Items calculate the benefit of reduction of the repair cost, the benefit from avoiding potential damages of water suspension for domestics and non-domestics, and the benefit from avoiding the social cost in case of roads unavailability. The optimal time for pipe renewal is reached when expected benefits are greater than costs.

The use of genetic programming for pipe breaks prediction is discussed in Ref. [10]. Authors develop an economic-based model for pipes replacement. They assume that there exist two categories of models for pipe breaks prediction: The physically-based models that aim at identifying physical causes of breaks and statistical models that analyze historical data to identify explanatory variables.

The use of machine learning seems relevant to tackle prediction problems. Between 2006 and 2016, the use of Artificial Neural Networks (ANNs) has increased in the drinking water sector, particularly for modeling the infrastructure and water quality [11]. ANNs address water quality problems by modeling chlorine concentration [12]. To improve leakage management, hydraulic and water quality data collected from sensors are used to fit ANNs for detecting and locating leakage in Yorkshire Water's Keighley distribution system [13]. A principal component analysis (PCA) and ANN was carried out to predict the leakage ratio in the drinking water system using six effective parameters: pipe deterioration ratio, the volume of water supplied, pipe length, mean pipe diameter, the number of leaks, and an energy ratio [14]. Authors show the advantage of coupling ANN with PCA. To estimate the magnitude and the location of leaks, ANNs were trained on different sets of input data (pressure and flow rate) collected from sensors installed in the piping network [15].

It appears from the literature review that despite the output variable to predict, the training of ANNs in the drinking water sector is done at the local scale by using a series of monitoring data collected by sensors disseminated in the network. What can be done in case of the absence of monitoring data? A partial answer is given by Ref. [16], who investigated the training of ANNs not on monitoring data but on aggregated data or KPIs, representing high-level data gathered in mandatory databases. Authors establish cause–effect relationships between KPIs. They compare the use of ANN or multiple regression analysis (MRA) for calibrating a decision model that is able to predict the water efficiency ratio from a set of nine mandatory indicators considered as input variables.

In the context of absence or paucity of low-level monitoring data, the current work improves the model developed in Ref. [16] by prospecting asset management actions based on high-level data represented by ex-post KPIs measured at the scale of the water utility.

We assume that the proposed model can be adapted in the context of smart water systems where monitoring data are available at a low-level scale. The main added value of the proposed model is its ability to prospect asset management actions by measuring KPIs in an ex-ante way using an adaptation of ANNs and a multi-objective genetic algorithm. The prediction model can be fitted with a multiset of data from several water utilities or a national database of mandatory KPIs as SISPEA (French context) and the IS of the water utility. This can be very helpful in case of absence of enough monitoring data at the scale of the water utility.

The paper is organized into five sections. The current section proposes a literature review of asset management of water pipes and the use of ANN for KPI's prediction and genetic algorithm for problem optimization. Section 2 defines the objective functions and the mathematical formulation of the considered problem. The characteristics of the ANN and NSGA II are also detailed. Section 3 illustrates the use of the developed model on a real case study and shows how it is carried out. Section 4 discusses the results and the main added value of the model. Finally, the last section concludes the paper.

2. Materials and Methods

This paper focuses on the prediction of two KPIs considered as objective functions: (1) the water efficiency rate considered as a benefit and (2) the total cost obtained by the sum up of *OPEX* and capital expenditures (*CAPEX*). Considered costs are the result of the implementation of asset management actions: renewal of pipes, connections and meters on one hand; and leak detection, connections and pipes reparation on the other hand. The prediction of KPIs is ensured by an adaptation of ANNs coupled with a multi-objective genetic algorithm NSGA II [17].

2.1. The Water Efficiency Rate (WER)

In the French context, the *WER* is a mandatory KPI calculated for each water utility according to the decree of May 2007 [18]. It measures the ratio between the billed and distributed water. The prediction model uses the theoretical model developed in Ref. [16] to establish relationships between *WER* (output) and nine other mandatory KPIs (Input) considered as explanatory variables. Table 1 lists the explanatory variables with their corresponding code (taken from SISPEA) and their link with asset management actions.

Table 1. Explanatory variables for efficiency rate.

Asset Management Actions	SISPEA Code	Explanatory Variables—Indicators
Metering and metering error	VP.056	Number of users
	VP.228	Linear density of users
	VP.063	Billed metered domestic consumption
	VP.221	Volume of unmetered consumption
	VP.232	Billed metered consumption
Leakage and water losses	VP.234	Volume produced + Volume imported
	VP.225	Average network efficiency rate over last 3 years
	P106.3	Linear leakage index on distribution mains (LLI)
Pipes renewal	P107.2	Average renewal rate of water mains over the last 5 years

The assessment of *WER* requires the analysis of the yearly hydraulic balance of the whole network. Table 2 lists the required variables.

Table 2. List of variables required for hydraulic balance.

Symbol	Definition of the Variable	Unit
W_b	Annual volume of water-billed metered consumption	m^3
W_m	Annual volume of water loss due to metering error	m^3
W_p	Annual volume of water loss due to leaks on main pipes	m^3
W_c	Annual volume of water loss due to leaks on connections	m^3
W_i	Annual volume of water loss due to invisible leaks	m^3
ε_m	Metering error in percentage	%
\overline{Age}_m	Average age of meters	#
$MTTR_{vl}$	Mean time to repair visible leak	s
$MTTR_{inv}$	Mean time to repair hidden leak	s
d_p	Average flow rate for a leak on pipe	L/s
d_c	Average flow rate for a leak on connection	L/s
d	Average flow rate for a hidden leak	L/s
n_{inv}	Number of invisible breaks/leaks	#
n_p	Number of breaks/leaks on pipes per year	#
n_{lc}	Number of breaks/leaks on connections per year	#
n_c	Number of connections	#
r_b	Pipe breakage rate	#/km
r_{cb}	Connection breakage rate	#/km
r_d	leak detection efficiency rate	%
α	Invisible leakage rate on main pipes/connections	%
L_{net}	Network length	km

To be able to calculate WER , the listed explanatory variables in Table 1 should be calculated or estimated. WER can be indirectly estimated from the linear leakage, which encompasses four types of losses: losses due to metering errors W_m , losses due to leaks on main pipes W_p , losses due to leaks on connections W_c , and losses due to invisible leaks W_i . We assume that losses due to metering errors $W_m(t)$ can be calculated by Equation (1):

$$W_m(t) = W_b(t)(t) \times \varepsilon_m(t) \quad (1)$$

with:

$$\varepsilon_m(t) = \varepsilon_m(t-1) \times \frac{\overline{Age}_m(t)}{\overline{Age}_m(t-1)} \quad (2)$$

By considering the meter renewal rate, $\varepsilon_m(t)$ is calculated by Equation (3):

$$\varepsilon_m(t) = \varepsilon_m(t-1) \times (1 - r_m) \quad (3)$$

where r_m is the rate of annual meter renewal in percentage per year as listed in Table 3.

Table 3. List of required variables for cost calculations.

Symbols	Definition of the Variable
C_{rep}	Cost of a leak reparation in € per unit
C_{det}	Cost of leak detection in € per km
C_{meter}	Cost of a meter in € per unit
C_{con}	Cost of a connection renewal in € per unit
C_p	Cost of pipe renewal in € per km
n_{lc}	Number of leaks on connections per year
n_p	Number of leaks on pipes per year
n_d	Number of leaks detected by leak detection
n_m	Number of installed meters
l_{net}	Length of the network
l_{det}	Length of the network investigated by leak detection
r_c	Rate of annual connections renewal in percentage per year
r_p	Rate of annual pipe renewal in percentage of the length renewed per year
r_m	Rate of annual meter renewal in percentage per year

Losses due to leaks on pipes are computed by taking into account the estimated number of leaks on pipes from which the effect of the pipe renewal is subtracted:

$$W_p(t) = MTTR_{ol}(t) \times d_p(t) \times [n_p(t) - r_p(t) \times L_{net}(t) \times r_b(t)] \quad (4)$$

Analogously, losses due to leaks on connections at a given year $W_c(t)$ are computed by taking into account the estimated number of leaks on connection minus the effect of connections renewal:

$$W_c(t) = MTTR_{ol}(t) \times d_c(t) \times [n_{lc}(t) - r_c(t) \times n_{ct}(t) \times r_{cb}(t)] \quad (5)$$

The model also involves water losses $W_i(t)$ caused by invisible leaks. Equation (6) indicates how they are calculated:

$$W_i(t) = MTTR_{inv}(t) \times d(t) \times [n_{inv}(t) \times (1 - \alpha \times r_p(t) - (1 - \alpha) \times r_c(t)) - r_d(t) \times L_{net}(t)] \quad (6)$$

Asset management actions in terms of renewal (pipe, connections) and leak detection have an impact on leaks. Actions decrease the number of invisible leaks and the mean time to repair; this assumption is introduced by Equation (6). The total water loss for year t , $W_l(t)$, is obtained by the sum up of all types of water losses as shown in Equation (7):

$$W_l(t) = W_m(t) + W_p(t) + W_c(t) + W_i(t) \quad (7)$$

Based on previous equations, it is possible to compute the linear leakage index according to Equation (7).

$$LLI(t) = \frac{W_l(t)}{L_{net}(t)} \quad (8)$$

The average renewal rate of water mains over the 5 last years $\bar{r}_p(t)$ (code: P107.2) measures the mean value of the annual renewal rate of water pipes (without connections) over the last 5 years. This includes renewed, reinforced and rehabilitated pipes but does not take into account maintenance

actions as pipes repairation. The average renewal rate of water mains over the last 5 years is calculated by Equation (9):

$$\bar{r}_p(t) = \sum_{i=0}^3 \frac{r_p(t-i-1) + r_p(t)}{5} \tag{9}$$

with $r_p(t-i-1)$ for $i \in [0,3]$ being the annual renewal rate of pipes from the previous 4 years (known); and $r_p(t)$ is the annual renewal rate envisaged.

The remaining explanatory variables: number of users (VP.056), linear density of users (VP.228), billed metered domestic consumption (VP.063), volume of unmetered consumption (VP.221), billed metered consumption (VP.232), volume produced + volume imported (VP.234) are estimated based on water utility manager opinion, historical data and Monte Carlo analysis using a uniform distribution function as explained in Ref. [16].

In the context of a lack of low level data, we advise to use Equation (7) to estimate the mean and standard deviation of the following parameters: leakage flow rate, the number of hidden leaks and repair time for both pipes and connections over an observation period of at least 5 years. Obtained values represent a set of feasible solutions that satisfy the yearly hydraulic balance on the observation period.

The number of visible breaks and leaks on pipes and connections are supposed to be available as local data from the water utility. To involve the uncertainty of estimation, a Monte Carlo analysis is implemented using Equation (7), where a set of parameters and variables of the equation are randomly generated as shown in Figure 1. In the absence of data concerning the characteristics of leaks, normal distribution functions are used to randomly generate the flow rate, the number of leaks and time to repair. The achievement of this analysis provides a potential range of values for parameters of Equation (7) that make the estimation of water losses possible for prediction purposes. Figure 1 illustrates the required steps to estimate annual water losses.

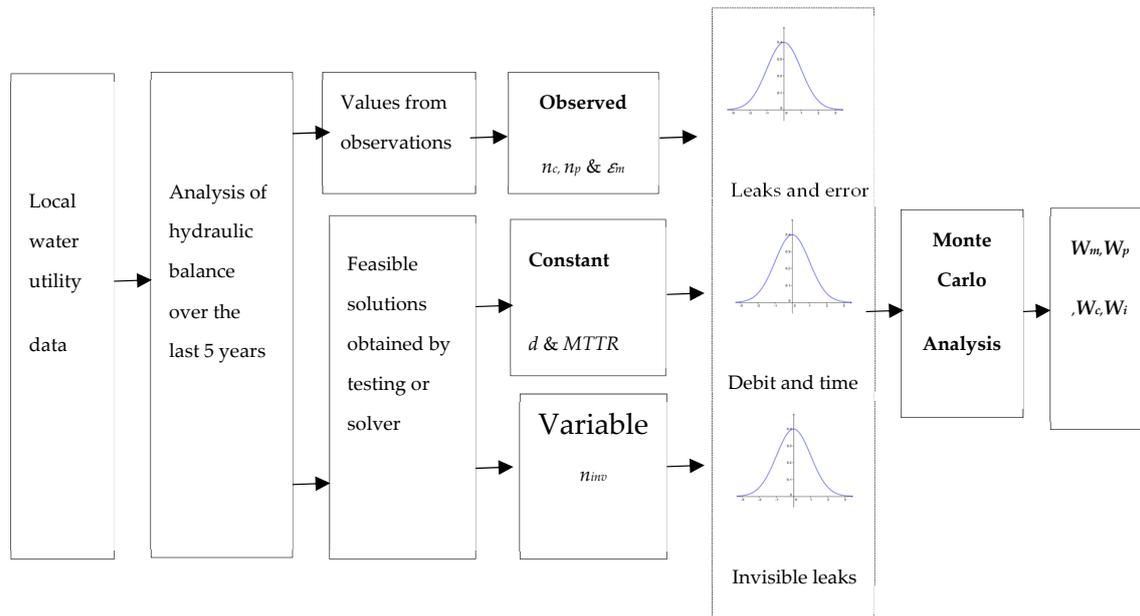


Figure 1. Steps for annual water losses estimation, adapted from Ref. [16].

2.2. The Total Annual Cost

The total annual cost (C_{Tot}) of decisions or a policy defined by asset management actions is calculated by Equation (10). Required variables for cost calculations are resumed in Table 3.

$$C_{Tot} = CAPEX + OPEX \tag{10}$$

OPEX are derived from curative maintenance actions of repairing pipes and connections, on the one hand, and preventive maintenance actions of leak detection, on the other hand; Equation (11) summarizes the annual maintenance costs as follows:

$$OPEX = C_{pipe_reparation} + C_{connection_reparation} + C_{leak_detection} \tag{11}$$

Each component of the maintenance cost is displayed in Equation (12) as follows:

$$OPEX = C_{rep} \times (n_p + n_d) + C_{rep} \times n_{lc} + C_{det} \times l_{det} \tag{12}$$

CAPEX measure the cost of asset management actions in terms of pipes, connections and meters renewal as indicated in Equation (13):

$$CAPEX = C_{pipe_renewal} + C_{connection_renewal} + C_{meter_renewal} \tag{13}$$

Equation (13) becomes as follows when each component of investment cost is displayed:

$$CAPEX = C_p \times r_p \times l_{net} + C_{con} \times r_c \times n_{lc} + C_{meter} \times r_m \times n_m \tag{14}$$

2.3. The Artificial Neural Network (ANN)

A neural network is composed of multiple perceptron and is called a deep neural network when the number of hidden layers is greater than or equal to 2 [19]. We use a multiple layers neural network in order to predict the WER based on nine KPIs considered as input [16]. Figure 2 illustrates a perceptron representing a layer in an ANN.

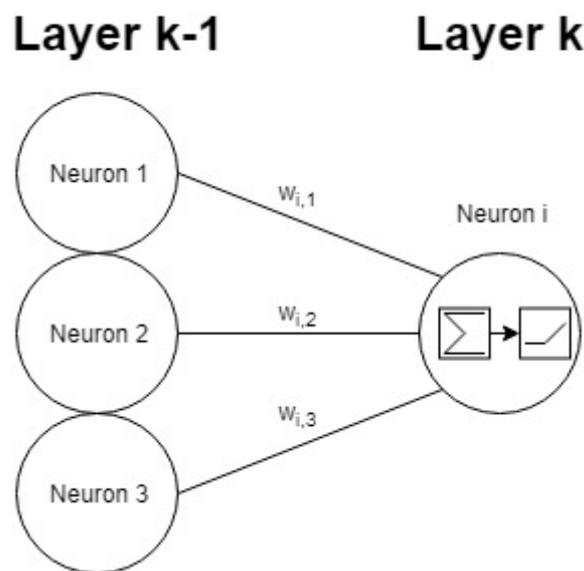


Figure 2. Example of a single perceptron.

The value assigned to neuron i in Figure 2 can be calculated by Equation (15) as follows:

$$neuron_i^k = relu(w_{i,1} \times neuron_1^{k-1} + w_{i,2} \times neuron_2^{k-1} + w_{i,3} \times neuron_3^{k-1} + b_i^{k-1}) \tag{15}$$

The rectified linear unit function $relu$ is given by Equation (16):

$$relu(x) = \{0 \text{ for } x < 0 ; x \text{ for } x \geq 0\} \tag{16}$$

The vector $neuron^k$ that groups all the values assigned to the neurons in layer k is calculated as follows:

$$neuron^k = \text{relu} \left[\begin{matrix} w_{0,0} & \dots & w_{0,n} \\ \dots & \ddots & \vdots \\ w_{i,0} & \dots & w_{i,n} \end{matrix} \right] \begin{bmatrix} neuron_0^{k-1} \\ \vdots \\ neuron_n^{k-1} \end{bmatrix} + \begin{bmatrix} b_0^{k-1} \\ \vdots \\ b_n^{k-1} \end{bmatrix} \quad (17)$$

Equation (17) becomes:

$$neuron^k = \text{relu} \left(M_w^{k-1} \times neuron^{k-1} + b^{k-1} \right) \quad (18)$$

where:

- M_w is the matrix of weights;
- $neuron$ is the vector of neuron values;
- k is the index of the layer;
- i is the number of neurons in the k th layer;
- n is the number of neurons in the $(k - 1)$ th layer;
- b is the bias vector.

The output value of the ANN can be computed by Equation (18). In our case, it is a single neuron which produced the water efficiency rate WER . The value of this neuron depends on the values of the previous neuron layers and the associated weights and biases.

Values of the previous layers also depend on weights and biases as well as input variables. The input variables are known, the objective is to determine the optimal values of weights and biases to give a good prediction.

To do this, during the learning phase, the prediction is compared to the real value. Weights and bias are adjusted until a satisfactory error is obtained. Error is commonly calculated with a Loss function noted L . For regression problems, the function L corresponds to the mean square error which computes the square difference between the observed and predicted value:

$$L(y_i, \hat{y}_i) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (19)$$

where n is the number of input values, y_i is the value of input i , and \hat{y}_i is the corresponding predicted value.

To minimize the loss function L , we use an optimization function *Adagrad* which modifies weights and bias in order to minimize the error. *Adagrad* was introduced by Ref. [20] and it is called so for adaptive gradient algorithm. During the learning process, the weights are updated considering Equation (20):

$$\Delta w_i(t) = - \frac{\eta}{\sqrt{G_i(t) + \epsilon}} \times \frac{\partial L}{\partial w_i}(t) \quad (20)$$

with:

$$\begin{cases} G_i(t) = G_i(t-1) + \left(\frac{\partial L}{\partial w_i}(t) \right)^2 \\ G_i(0) = 0 \end{cases} \quad (21)$$

The term $\frac{\eta}{\sqrt{G_i(t) + \epsilon}}$ is the effective learning rate, with η being the initial learning rate. The term $\frac{\partial L}{\partial w_i}(t)$ is the gradient (partial derivative of loss function with respect to weights). By this definition, G_i is a monotone increasing function. So, the effective learning rate is monotonously decreasing. Note that G_i and the effective learning rate are different for each weight.

Figure 3 illustrates the ANN built for our prediction model. It is designed for the nine input explanatory variables representing KPIs (with French mandatory codes), two hidden layers with the

same number of neurons as the input layer. The output layer considered as the output of the model is only composed of the neuron corresponding to the water efficiency rate, WER (code: P104.3).

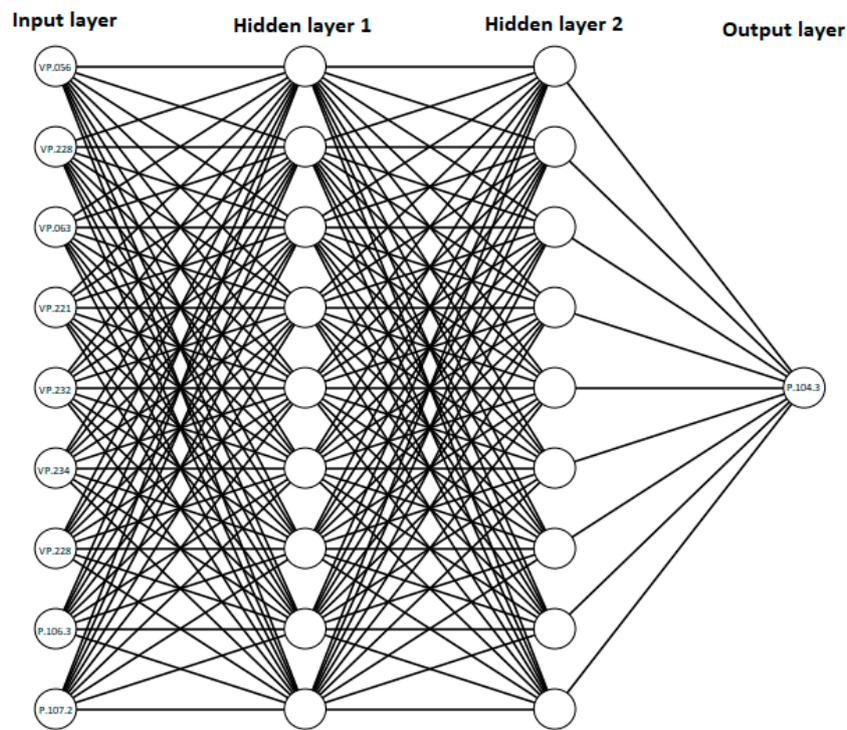


Figure 3. Neural network configuration with two hidden layers.

Since the number of input samples is more than 10,000, the model is trained with a batch size of 200. The batch size corresponds to the number of samples that will be propagated through the neural network. After propagation, weights and biases are updated in order to decrease the error. Once all training samples are passed once through the network, this counts as 1 epoch. The network training is done by performing multiple epochs.

2.4. NSGA II and the Problem Formulation

The problem to solve concerns the optimization of asset management actions in an ex-ante way in order to maximize the WER and minimize the annual total costs. The decision variables measure the level of actions in terms of pipes, connections, and meters maintenance and renewal. We consider that the following variables l_{det} , r_c , r_p , and r_m are the most relevant for the decision maker in terms of asset management. The problem can be formulated as the following:

$$\text{Maximize } f_1(l_{det}, r_c, r_p, r_m) = WER(t) \tag{22}$$

$$\text{Maximize } f_2(l_{det}, r_c, r_p, r_m) = \frac{1}{C_{Tot}(t)} \tag{23}$$

constrained by:

$$l_{det_min} \leq l_{det} \leq l_{det_max} \tag{24}$$

$$r_{p_min} \leq r_p \leq r_{p_max} \tag{25}$$

$$r_{c_min} \leq r_c \leq r_{c_max} \tag{26}$$

$$r_{m_min} \leq r_m \leq r_{m_max} \tag{27}$$

The value of upper and lower limits of decision variables are defined according to the water utility manager expectations. By considering the two fitness functions f_1 and f_2 , NSGA II will attempt to find the best 4-tuple $(l_{det}, r_c, r_p, \text{ and } r_m)$ from a population of potential solutions. The population size is set in advance and the values of the 4-tuple elements are generated randomly between the upper and lower boundaries to initialize the population as shown in Figure 4.

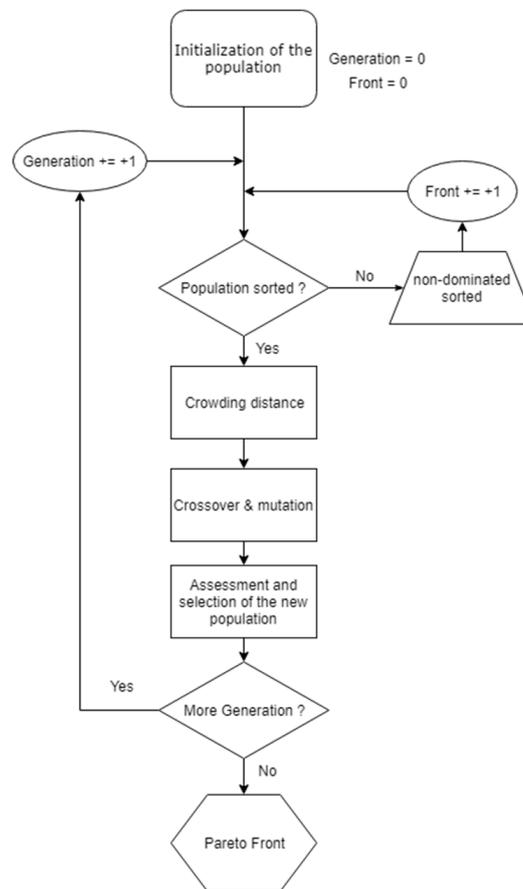


Figure 4. Flowchart of the fast, elitist, non-dominated sorting genetic algorithm (NSGA-II).

2.4.1. The Concept of Non-Dominance

The NSGA II implements the concept of dominance to reach potential solutions. The concept of dominance is well defined in Ref. [21]. Two definitions can be considered. The first one considers two solutions, that solution X_1 dominates solution X_2 if both conditions are true: (i) solution X_1 is not worse than X_2 for all the objectives, and (ii) solution X_1 is strictly better than X_2 for at least one objective. Conditions are resumed in Equations (28) and (29).

$$\forall i \in \{1,2\}: f_i (X_1) \geq f_i (X_2) \tag{28}$$

$$\exists j \in \{1,2\}: f_j (X_1) > f_j (X_2) \tag{29}$$

The second definition considers as non-dominated solutions those that are not dominated by any member of the considered population.

2.4.2. The Crowding Distance

To sort solutions, NSGA II uses a crowding distance [17,22,23]. It is used to estimate the density of solutions surrounding an individual in the population by considering the difference of the objective values of the nearest neighbor as shown in Figure 5. It is an estimate of the size of the largest cuboid

enclosing point k , without including any other point in the population. In the following sections, the term individual designates a potential solution.

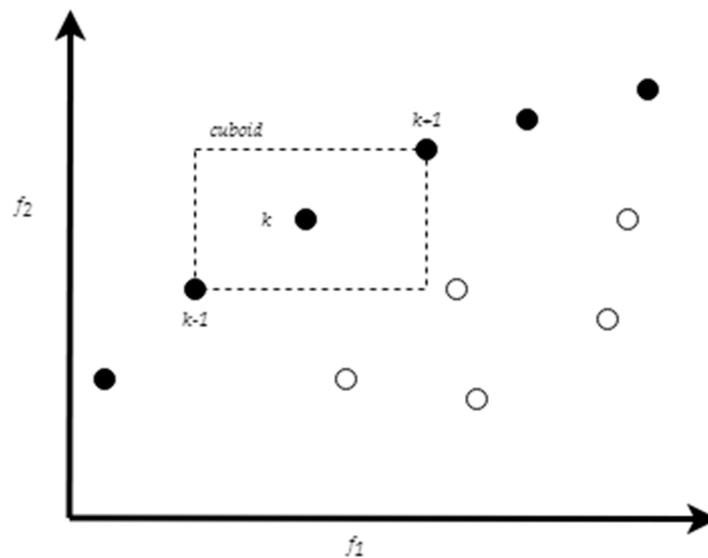


Figure 5. The crowding distance of individual in the front. Amended from Ref. [17].

Let's consider F the size of the front, for individuals, the crowding distance is calculated by the difference between the objective values of the two nearest neighbors:

$$d_i = \sum_{m=1}^M \frac{f_{i+1}^m - f_{i-1}^m}{f_{max}^m - f_{min}^m} \tag{30}$$

The edge, the first individual and the last individual in the rank, are assigned with a large distance to ensure that boundary points will always be selected as shown by Equation (31).

$$d_0 = d_{F-1} = \infty \tag{31}$$

where M is the number of objectives, f_i^m is i th fitness values in the m th objective, and f_{max}^m and f_{min}^m are the maximum and minimum objective values of the m th objective (in the non-dominated set).

This formulation maintains diversity in the population by eliminating redundant individuals but suffers from a loss of both vertical and horizontal diversity as explained by Ref. [22]. To improve the diversity in the final front, an improvement of the crowding distance has been proposed by Ref. [23] by defining a dynamic crowding distance:

$$D_{d_i} = \frac{d_i}{\log\left(\frac{1}{V_i}\right)} \tag{32}$$

with:

$$V_i = \sum_{m=1}^M (|f_{i+1}^m - f_{i-1}^m| - d_i)^2 \tag{33}$$

The dynamic crowding distance is computed for each individual in the non-dominated set. The individual which has the lowest dynamic crowding distance is removed. The dynamic crowding distance is updated after each removal. These operations are repeated until the size of the non-dominated set is equal to the population size.

2.4.3. The Selection Method

Once individuals have been assessed and sorted, k Elements of the population are taken as candidates for the mating pool, where k designates the tournament size [17]. Random selection is a particular case of tournament selection when $k = 1$. For $k > 1$, the selection method is called tournament selection. The k individuals are compared to each other based on their rank and crowding distance. The best individual is added to the mating pool. The operation is repeated a second time to obtain two individuals in the mating pool as shown in Figure 6. Selected solutions are subject to crossover and mutations to create offspring. Tournament selection is repeated until the number of created offsprings is sufficient.

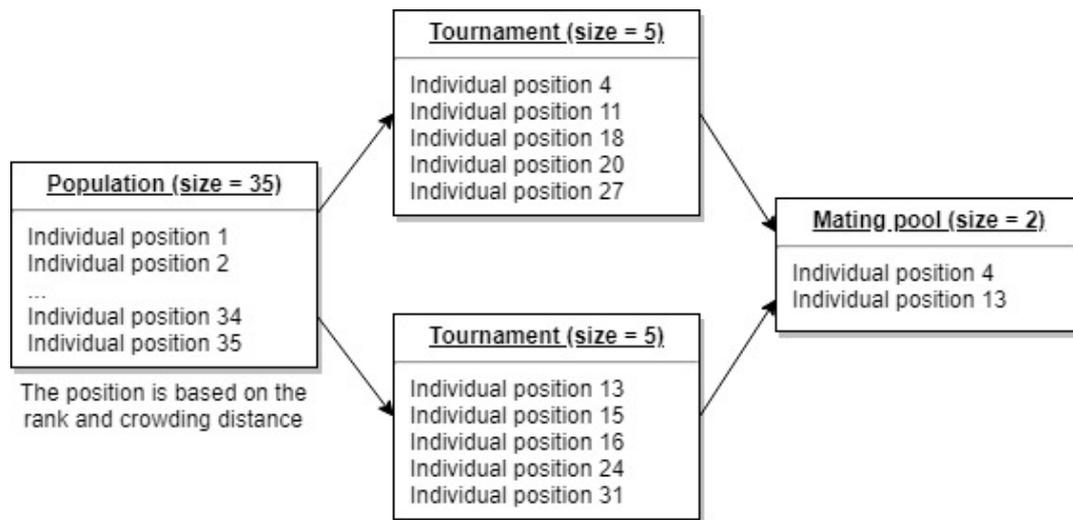


Figure 6. The tournament selection.

There exists other selection operators where individuals are chosen based on their proportional fitness value, as the roulette wheel selection (RWS). The individual is selected according to a probability of selection calculated by the ratio between its fitness value and the sum up of fitness values of individuals in the mating pool [24].

2.4.4. The Crossover

Realizing a crossover is a way of using the information of two parents in the population to obtain one child [25]. There are different possible recombinations and several authors have compared them to each other in different problems [26,27]. There is no consensus in the literature concerning the effectiveness of single point crossover or multi-point crossover. This depends on the particularities of the problem. The danger of algorithms comparison on a small sample according to their performance is underlined in Ref. [28]. The authors advise to integrate problem-specific knowledge into the functioning of the algorithm; this integration can also concern crossover operators. In our case, there are two objective functions and the only constraints in this problem are upper and lower bounds of the 4-tuple variable. Hence, we choose to use the flat crossover which is a widely used crossover method [29]. Considering two parents in the current population:

$$Parent_1 = (l_{det1}, r_{p1}, r_{c1}, r_{m1}) \tag{34}$$

$$Parent_2 = (l_{det2}, r_{p2}, r_{c2}, r_{m2}) \tag{35}$$

and a random vector:

$$r = (r_1, r_2, r_3, r_4) \tag{36}$$

with random values $r_i \in [0, 1]$. The i th child $(l_{det}^i, r_p^i, r_c^i, r_m^i)$ is a linear combination of the two parents:

$$l_{det}^i = r_1 \times l_{det1} + (1 - r_1) \times l_{det2} \tag{37}$$

$$r_p^i = r_3 \times r_{p1} + (1 - r_3) \times r_{p2} \tag{38}$$

$$r_c^i = r_2 \times r_{c1} + (1 - r_2) \times r_{c2} \tag{39}$$

$$r_m^i = r_4 \times r_{m1} + (1 - r_4) \times r_{m2} \tag{40}$$

Table 4 shows an example of the offspring that two parents can give by applying this crossover.

Table 4. Crossover and offspring generation.

Individual	l_{det}	r_p	r_c	r_m
Parent ₁	74.18	0.0042	0.0083	0.028
Parent ₂	82.07	0.0031	0.0068	0.035
Random vector, r	0.61	0.77	0.75	0.55
Offspring ₁	77.27	0.0040	0.0079	0.031
Offspring ₂	78.98	0.0033	0.0072	0.032

2.4.5. The Mutation

The mutation is an operator that modifies an individual to explore the entire search space [25] and to escape from local optima thanks to small changes in the values of the 4-tuple variables. It is used to maintain diversity in the population of potential solutions. We use the polynomial mutation introduced by Ref. [30]. For each variable, there is a mutation probability. The mutation probability is set at 1/4 since each solution is represented by a 4-tuple. There is one mutation per offspring on average.

2.4.6. The Selection of Offspring

Once the crossovers and mutations have been achieved, we end up with a population of P individuals and a population of P offspring. The total size of the selection is $2P$ and this must be reduced to P individuals. This selection is made by keeping the best individuals as requested by NSGA II [17]. In this way, the next generation will be better than the previous generation or equivalent if no individual from the descendants is better than the current population. This is called elitism selection. To select the best individual, we defined an operator \geq_n basis on individual domination rank $rank_p$ and dynamic crowding distance D_{dp} . The partial order \geq_n is defined as:

$$p \geq_n q \text{ if } (rank_p < rank_q) \text{ or } ((rank_p = rank_q) \text{ and } D_{dp} > D_{dq}) \tag{41}$$

The individual with the lower rank, according to the non-dominated sorting algorithm, is preferred. If two individuals have the same rank, the one which is located in the lower density of solutions is preferred.

The selection of $2P$ individual is first sorted in the ascending order with respect to their rank obtained by the non-dominated sorting algorithm. Then, individuals are sorted with respect to the dynamic crowding distance in descending order. The next generation is thus generated until there are P individuals in the new population.

2.4.7. Performance Metrics

The effectiveness of the model depends on its ability to ensure diversity, a good distribution and spread of solutions. To evaluate the distribution, we use the Spacing index (SP) introduced by Ref. [31]. To be able to assess the spread in a population of P individuals (potential solutions), we need

to calculate d_i which is the minimum of the sum of the absolute difference in objective fitness values between the i th solution and any other solution as shown in Equation (42), and \bar{d} , the mean value of d_i calculated by Equation (43):

$$d_i = \min_{i, i \neq k} \left\{ \sum_{m=1}^M |f_i^m - f_k^m| \right\} \quad (42)$$

$$\bar{d} = \sum_{i=1}^{|P|} \frac{d_i}{|P|} \quad (43)$$

Therefore, SP is obtained by Equation (44):

$$SP = \sqrt{\frac{1}{|P|-1} \sum_{i=1}^{|P|} (d_i - \bar{d})^2} \quad (44)$$

SP is used to evaluate the spacing between the different solutions. If the distance between each solution is the same, then the SP value will be zero. Thus, a value of zero or near zero indicates a good distribution of solutions on the Pareto front. The spread index was proposed by Ref. [17]:

$$\Delta = \frac{d_f + d_l + \sum_{i=1}^{|P|} (d_i - \bar{d})}{d_f + d_l + (|P| - 1) \times \bar{d}} \quad (45)$$

d_f and d_l are the Euclidean distances between the extreme solutions and the obtained Pareto solutions. A Δ value close to 0 means that the solutions are well dispersed along the Pareto front.

3. Case Study

The model is implemented on a real water distribution network in the south of France. According to the data of the year 2016, the water system delivers 700,000 m³ of drinking water for about 6300 users with a network length of 82 km. We consider the actual asset management actions implemented by the water utility as a baseline solution. It can be resumed by a leak detection of the entire network once ($l_{det} = 82$ km) that allows detecting 14 leaks on average. Annual renewal rates are: $r_p = 0.71\%$, $r_c = 2.5\%$ and $r_m = 10\%$. Thanks to these actions, $WER = 76.90\%$ with a total cost equals to 551,493 €, shared between 70% in $CAPEX$ and 30% in $OPEX$. We aim at improving WER by conducting alternative asset management actions at lower costs than commonly used strategies. Before searching compromise solutions, ANN is fitted thanks to the SISPEA database. SISPEA is a mandatory French database that gathers 26 KPIs from more than 12,000 water utilities between 2006 and 2016. Data were split into two samples, 70% of the data is split to fit the Ann model, and the remaining 30% is used for validation.

3.1. Artificial Neural Network Fitting

The calibration of the ANN requires the definition of a set of parameters that improve its accuracy. As discussed in Ref. [16], many simulations are carried out in order to determine the most appropriate values for the number of hidden layers, the number of neurons per layer, and the type of activation number. The selected ANN is built by three hidden layers with 144, 36, and 9 neurons at each layer, respectively. The chosen activation function is the function *relu* for all neurons. The estimation of required variables for water losses estimation for the year ($N + 1$) at the local scale (see Equation (7)) is generated based on expert opinion and Monte Carlo analysis. Table 5 compares the observed and predicted values of WER for the period between 2010 and 2016.

Table 5. Comparison between observed and predicted WER between 2010 and 2016.

Year	Observed	Predicted-ANN (%)	Estimation Error (%)
2010	65.9	65.7	−0.30%
2011	63.2	62.4	−1.27%
2012	61.4	60.7	−1.14%
2013	75	75.1	0.13%
2014	75.1	75.1	0.00%
2015	76.9	76.8	−0.13%
2016	76.6	76.6	0.00%

According to Table 5, ANN seems to predict the WER with a high accuracy; the estimation error oscillates between −1.27% and +0.13%.

3.2. Parameters of the NSGA II

The implementation of NSGA II requires the definition of the type of tournament to consider and to set the population size of potential solutions. According to performance indicators resumed in Table 5, we compare between the tournament selection method with tournament size $k = 2$, random selection, and roulette wheel selection. The mean values and standard deviation are calculated on 10 tests for each selection method; results are resumed in Table 6. The population size is fixed at 200, and the number of generations is set at 25.

Table 6. Performance comparison of three selection methods in terms of distribution and spread.

Selection Method	Performance Metrics	Space Index	Spread Index
Random	Mean	0.127	0.313
	Std	0.025	0.091
Roulette Wheel	Mean	0.113	0.259
	Std	0.023	0.074
Tournament	Mean	0.115	0.286
	Std	0.035	0.130

Tournament selection and roulette wheel selection seem more efficient than random selection in terms of both distribution and spread.

The method with the best results is roulette wheel selection. Both the average and the standard deviation of the two performance metrics are the lowest. Note that the standard deviation for tournament selection is greater than for random selection.

3.3. Population Evolution

The objective of the model is to get closer to the true and unknown Pareto front. Over the generations, the population should move closer to the Pareto front. The starting population is generated randomly between the limits set for the 4-tuple values of the decision variables as presented in Table 7.

Table 7. Definition of decision variable constraints.

Decision Variables	Lower Bound	Upper Bound
l_{det}	0	$12 \times L_{net}$
r_p	0	5%
r_c	0	5%
r_m	0	10%

Figure 7 shows the evolution of potential solutions composing the population (size $P = 200$) using the roulette wheel selection after 25 generations. The optimal front is quickly reached; the population improves significantly in the first generations, and then very slowly over the last five generations, the front stabilizes for the last generations. The obtained front confirms the relevance of using the roulette wheel selection.

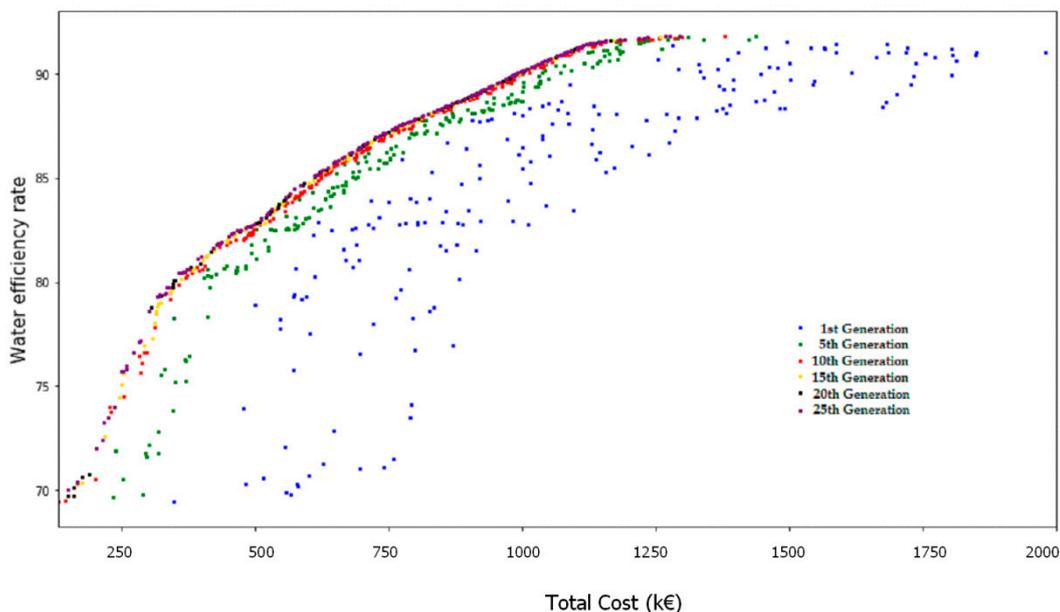


Figure 7. Evolution of Pareto front after 25 generations.

3.4. Problem Resolution

Performed tests guide the choice of the type of selection, crossover and mutation operators. To solve the considered problem (see Equations (22) and (23)), the NSGA II is implemented with an initial population generated randomly with 1000 individuals. The crossover probability is set to 90% to generate the offspring using the flat crossover. The polynomial mutation is used with an index polynomial mutation of 1 and a mutation probability of 25%. The new population is selected from roulette wheel selection. The number of performed generations is 25.

Figure 8 illustrates the Pareto front obtained for the considered objectives. The blue dots forming the front in the middle correspond to the average value of the water efficiency rate predicted by the ANN model. The red dots forming the upper and lower fronts define the limits of the 95% confidence interval. Each point of the front represents the 4-tuple of the constrained decision variables of the problem: the rate of pipes renewal, the rate of connections renewal, meters renewal, and length proven by leak detection. Table 8 details some of the solutions composing the Pareto front.

Table 8. Trade-off solutions with regard to considered constraints and objective.

r_p	r_c	r_m	L_{net} (km)	Total Cost (€)	CAPEX (€)	OPEX (€)	WER (%)
4×10^{-5}	8.1×10^{-5}	4.1×10^{-3}	86.5	223,302	2837	220,465	74.5
4.6×10^{-5}	8.2×10^{-4}	7.3×10^{-3}	162.6	309,838	3935	305,903	79.0
5.4×10^{-5}	9.5×10^{-4}	8.4×10^{-3}	260.2	413,868	4536	409,332	81.6

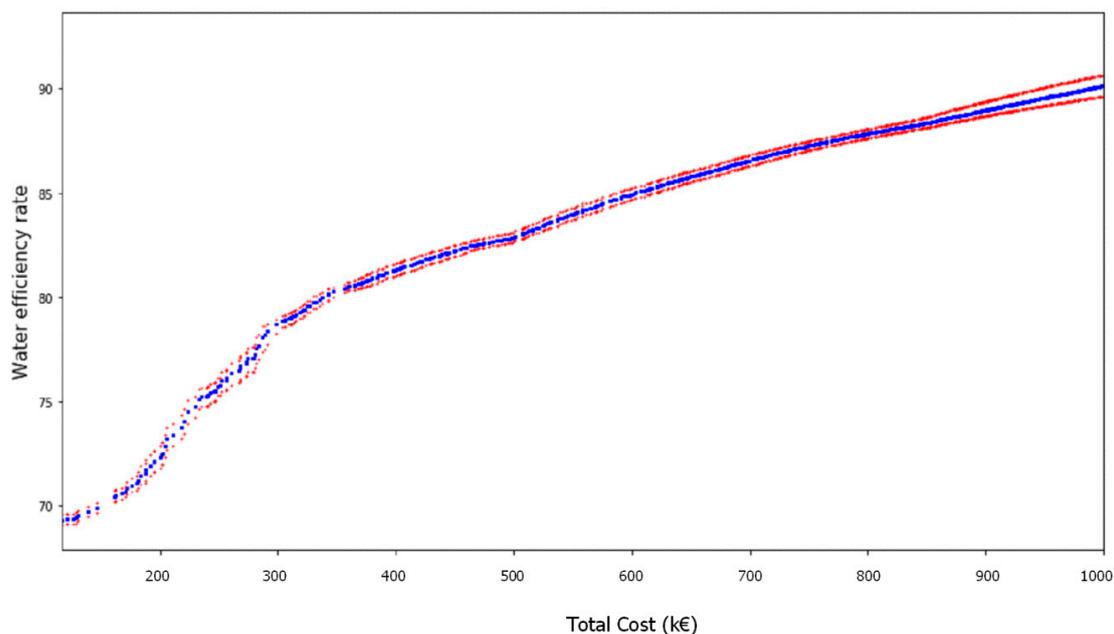


Figure 8. The Pareto front obtained from 1000 individuals and 25 generations.

The comparison of the baseline solution (actual practice) to the proposed solutions indicates that actual practice does not offer a compromise between cost and performance. Its costs more than all solutions listed in Table 8 with a value of $WER = 76.9\%$. Another interesting analysis concerns the repartition of expenditures between *CAPEX* and *OPEX*. Water utility privileges investment by increasing *CAPEX* (70 %) where our model advises to increase *OPEX* to 99% of total expenditures (according to Table 8).

Results show a significant influence of the leak detection and reparation actions on the WER . This is an intuitive result but the main advantage of the proposed model is its capacity to predict the effects of actions on the WER . The length of the water system under consideration is about 86 km. The values of length proven by leak detection per year contained in Table 8 correspond to approximately one, two and three times the total length of the network. The total cost is shared into two parts, *CAPEX* and *OPEX*. *OPEX* are largely due to the leak detection and asset reparations while *CAPEX* value is low due to low investments. In the short term period, the model shows that the main way to improve WER when the value of the efficiency is already high is the investigation for leaks. Indeed, leak detection allows improving more efficiently the water efficiency rate for an acceptable cost, compared to renewal actions. Renewal actions start to have an impact when performance values and asset condition are low.

4. Discussion

Predictions of WER (outputs) obtained for management actions (inputs) seem to be coherent with practice. In fact, Table 8 shows a positive correlation between WER and total cost, which confirms that it is required to spend more money to enhance performance. There seems to exist a threshold effect between expenditures and performance, even if we double the budget (from 2.23 k€ to 4.13 k€) performance increases only by 7%. This result is important because it indicates that even if the budget is available, it has to be spent in a certain manner and shared adequately between investment and maintenance actions. Another point of interest concerns the share of *OPEX* in the total expenditures. *OPEX* represents 99% of the total costs, which implies that if we aim at improving WER in the short term, it is recommended to spend more money for maintenance actions than investment. The advantage of the proposed approach is to drive the decision by indicating the type of maintenance actions to implement. For the studied case, Table 8 indicates that the leak detection and reparation of leaks seem to be efficient. The values of upper limits for decision variables were defined as the following: The limit

of leak detection rate corresponds to a total inspection of the entire network each month (12 per year), renewal rate of pipes and connections is limited to 5% (5 times the actual rate) per year considering an average lifespan for asset of 50 years (ambitious), and the renewal rate of meters is limited to 10% (two times the actual rate) which corresponds to a lifespan for meters of 10 years on average.

The definition of NSGA II parameters requires expertise and should be driven by tests. Results show the relevance of comparing different operators of selection and crossover. For the current study, roulette wheel seems more pertinent than other methods. The size of the population and the number of generations are also an important parameter to fit. The followed procedure aims at driving the implementation of the approach by: (i) defining the first suitable operators for a fixed-size population ($P = 200$); (ii) test the range of values for the number of generations (5 to 25); and (iii) increase the size of population for a given number of generations from 200 to 1000. Even if we cannot generalize the obtained results, it seems that this procedure leads to improve the shape of the Pareto front and to make it less discontinuous and more uniform. It can be interpreted as an improvement of the consistency of the front as shown in Figure 8.

The variety of solutions offered by the Pareto front constitutes a set of potential actions to implement depending on the context, constraints, and objectives to reach. This constitutes a valuable mitigation tool for decision makers and stakeholders.

Another advantage provided by the prospecting model is its capacity to be coupled with NSGA II in order to guide the search for the most relevant solutions. Even if results are really encouraging, some aspects have to be investigated. The dynamics of the model are not actually addressed: how is it possible to improve the planning of actions from year to year by updating input data? Another important aspect concerns the effect of asset management actions in the long term; it appears that maintenance actions significantly improve the value of *WER* with a low total cost. This can be considered as relevant in the short term, but it is not supposed to encourage the non-investment actions. A risk can be faced by the water utility due to an under-investment, which is the deterioration of the asset and the delivered service. One possible improvement is to constrain the rate of asset renewal when the solutions are searched for in order to avoid an important asset aging due to disinvestment.

5. Conclusions

The actual research is considered an encouraging improvement of our model based on ANN for predicting KPI's. The proposed improvement confirms that it is possible to predict and optimize KPIs for water utilities by coupling ANN and NSGA II in the context of the lack of local data. Many aspects should be checked in relation to the characteristics and parameters of ANN like the number of hidden layers and the number of neurons and activation functions. For NSGA II, the set of population and type of selection, crossover, and mutation have to be fixed before implementing the prediction model. All these aspects can render the model difficult to implement by the water utility because it requires specific skills. The actual model should be improved to gain simplicity for easy implementation by water utilities.

The absence of local data is encountered by the use of a national mandatory database and Monte Carlo simulations; this can be useful in the short term. We demonstrate to the water utility managers the usefulness of using data for prediction; this should encourage the water utility manager to improve their IS and converge to a smart water system in order to catch real-time data for supporting the decision making. The interpretation of results should take into account the context of the water utility. The preference of implementing maintenance actions versus renewal actions can be relevant when the value of *WER* is high. That means that the condition of the asset is good and does not require renewal. This can be acceptable in certain conditions but not adequate when the asset has deteriorated. For example, if assets are in a good condition and the water system is young, it is not necessary to check the network by leak detection. The context and condition of the network have to be considered when the boundaries of decision variables are set. Their range of variation may consider as low boundaries thresholds different from 0 to avoid the aging of assets in the long term.

The variation of *WER* and cost shown by the Pareto front seem realistic and offers a variety of potential solutions to the decision maker which is valuable.

Further research will explore the reproducibility of the developed approach for other KPIs by defining the set of input variables and how the ANN model and NSGA II can help to predict them. For example, the SISPEA database contains 25 additional KPIs that merit to be predicted in the same way as *WER*. We intend to explore the possibility to adapt the current model and make a general methodology for predicting water utility KPIs.

Author Contributions: A.N. conceived the methodology, analyzed and interpreted results, contributed to the paper writing, reviewing and editing. J.B. implemented the methodology, performed data processing, models fitting and participated to the paper writing.

Funding: The work presented is part of the French project “SPHEREAU”, grant number AAP FUI n° 22. It was funded by “bpifrance”, the basin water agency “Agence de l’Eau Rhin Meuse”, French regional authorities “Région Centre Val de Loire” and “la région Grand Est”.

Acknowledgments: We are grateful to the manager of the utility who helped us in our research and provided us with data, information and advice.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Alegre, H.; Hirnir, W.; Baptisia, J.M.; Parena, R. *Performance Indicators for Water Supply Services. Manual of Best Practice Series*; IWA Publishing: London, UK, 2000; p. 146.
2. Mutchek, M.; Williams, E. Moving Towards Sustainable and Resilient Smart Water Grids. *Challenges* **2014**, *5*, 123–137. [[CrossRef](#)]
3. Walski, T.M.; Pelliccia, A. Economic analysis of water main breaks. *J. Am. Water Work. Assoc.* **1982**, *74*, 3140–3147. [[CrossRef](#)]
4. Skipworth, P.J. *Whole Life Costing for Water Distribution Network Management*; Thomas Telford: London, UK, 2002.
5. Park, S.W.; Loganathan, G.V. Methodology for economically optimal replacement of pipes in water distribution systems: 1. Theory. *KSCE J. Civ. Eng.* **2002**, *6*, 539–543. [[CrossRef](#)]
6. Shamir, U.; Howard, C.D.D. Analytical approach to scheduling pipe replacement. *J. AWWA* **1979**, *71*, 248–258. [[CrossRef](#)]
7. Kleiner, Y.; Adams, B.J.; Rogers, J.S. Selection and scheduling of rehabilitation alternatives for water distribution systems. *Water Resour. Res.* **1998**, *34*, 2053–2061. [[CrossRef](#)]
8. Halhal, D.; Walters, G.A.; Ouazar, D.; Savic, D.A. Water Network Rehabilitation with Structured Messy Genetic Algorithm. *J. Water Resour. Plan. Manag.* **1997**, *123*, 137–146. [[CrossRef](#)]
9. Kim, K.; Seo, J.; Hyung, J.; Kim, T.; Kim, J.; Koo, J. Economic-based approach for predicting optimal water pipe renewal period based on risk and failure rate. *Environ. Eng. Res.* **2018**, *24*, 63–73. [[CrossRef](#)]
10. Xu, Q.; Chen, Q.; Ma, J.; Blanckaert, K. Optimal pipe replacement strategy based on break rate prediction through genetic programming for water distribution network. *J. Hydro Environ. Res.* **2013**, *7*, 134–140. [[CrossRef](#)]
11. O’Reilly, G.; Bezuidenhout, C.; Bezuidenhout, J.J. Artificial neural networks: Applications in the drinking water sector. *Water Sci. Technol. Water Supply* **2018**, *18*, 1869–1887. [[CrossRef](#)]
12. Cuesta Cordoba, G.A.; Tuhovcak, L.; Taus, M. Using Artificial Neural Network Models to Assess Water Quality in Water Distribution Networks. *Procedia Eng.* **2014**, *70*, 399–408. [[CrossRef](#)]
13. Mounce, S.R.; Machell, J. Burst detection using hydraulic data from water distribution systems with artificial neural networks. *Urban Water J.* **2006**, *3*, 21–31. [[CrossRef](#)]
14. Dongwoo, J.H.P.; Gyewoon, C. Estimation of Leakage Ratio Using Principal Component Analysis and Artificial Neural Network in Water Distribution Systems. *Sustainability* **2018**, *10*, 750. [[CrossRef](#)]
15. Caputo, A.C.; Pelagagge, P.M. An inverse approach for piping networks monitoring. *J. Loss Prev. Process. Ind.* **2002**, *15*, 497–505. [[CrossRef](#)]
16. Nafi, A.; Brans, J. Prediction of Water Utility Performance: The Case of the Water Efficiency Rate. *Water* **2018**, *10*, 1443. [[CrossRef](#)]

17. Deb, K.; Pratap, A.; Agarwal, S.; Meyarivan, T. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans. Evolut. Comput.* **2002**, *6*, 182–197. [[CrossRef](#)]
18. Legifrance. Available online: <https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=ORFTEXT000000274838URL> (accessed on 12 June 2019).
19. Kamiński, K.; Kamiński, W.; Mizerski, T. Application of Artificial Neural Networks to the Technical Condition Assessment of Water Supply Systems. *Ecol. Chem. Eng. S* **2017**, *24*, 31–40. [[CrossRef](#)]
20. Duchi, J.; Hazan, E.; Singer, Y. Adaptive Sub-gradient Methods for Online Learning and Stochastic Optimization. *J. Mach. Learn. Res.* **2011**, *12*, 2121–2159.
21. Ding, L.; Zeng, S.; Kang, L. A fast algorithm on finding the non-dominated set in multi-objective optimization. In Proceedings of the Congress on Evolutionary Computation (CEC'03), Canberra, ACT, Australia, 8–12 December 2003.
22. Yang, L.; Guan, Y.; Sheng, W. A novel dynamic crowding distance based diversity maintenance strategy for MOEAs. In Proceedings of the International Conference on Machine Learning and Cybernetics (ICMLC), Ningbo, China, 9–12 July 2017; IEEE: Piscataway, NJ, USA, 2017; Volume 1, pp. 211–216.
23. Luo, B.; Zheng, J.; Xie, J.; Wu, J. Dynamic crowding distance—A new diversity maintenance strategy for MOEAs. In Proceedings of the Fourth International Conference on Natural Computation, Jinan, China, 18–20 October 2008; IEEE: Piscataway, NJ, USA, 2008; Volume 1, pp. 580–585.
24. Deb, K. *Multi-Objective Optimization Using Evolutionary Algorithms*; John Wiley & Sons: Hoboken, NJ, USA, 2001.
25. Lim, S.M.; Sultan, A.B.; Sulaiman, N.; Mustapha, A.; Leong, K.Y. Crossover and Mutation Operators of Genetic Algorithms. *Int. J. Mach. Learn. Comput.* **2017**, *7*, 9–12. [[CrossRef](#)]
26. Mendes, J.A. Comparative study of crossover operators for genetic algorithms to solve the job shop scheduling problem. *WSEAS Trans. Comput.* **2013**, *12*, 164–173.
27. Liagkouras, K.; Metaxiotis, K. An Elitist Polynomial Mutation Operator for Improved Performance of MOEAs in Computer Networks. In Proceedings of the 22nd International Conference on Computer Communication and Networks (ICCCN), Nassau, Bahamas, 30 July–2 August 2013; pp. 1–5.
28. Wolpert, D.H.; Macready, W.G. No Free Lunch Theorems for Optimization. *IEEE Trans. Evolut. Comput.* **1997**, *1*, 67–82. [[CrossRef](#)]
29. Sharapov, R.R. Genetic Algorithms: Basic Ideas, Variants and Analysis. In *Vision Systems: Segmentation and Pattern Recognition*; Goro, O., Ashish, D., Eds.; Springer: Berlin, Germany, 2009; pp. 407–422.
30. Deb, K.; Deb, D. Analysing mutation schemes for real-parameter genetic algorithms. *Int. J. Artif. Intell. Soft Comput.* **2014**, *4*, 1–28. [[CrossRef](#)]
31. Schott, R.J. Fault Tolerant Design Using Single and Multicriteria Genetic Algorithm Optimization. Master's Thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, 1995.

