water

MDPI

Article

Lost in Optimization of Water Distribution Systems: Better Call Bayes

Antonio Candelieri [1,*], Andrea Ponti [2,3], Ilaria Giordani [2,4] and Francesco Archetti [3,4]

[1] Department of Economics, Management and Statistics, University of Milano-Bicocca, 20126 Milan, Italy
[2] Oaks s.r.l., 20125 Milan, Italy; a.ponti5@campus.unimib.it (A.P.); ilaria.giordani@unimib.it (I.G.)
[3] Consorzio Milano Ricerche, 20126 Milan, Italy; francesco.archetti@unimib.it
[4] Department of Computer Science, Systems and Communication, University of Milano-Bicocca, 20126 Milan, Italy
* Correspondence: antonio.candelieri@unimib.it

Abstract: The main goal of this paper is to show that Bayesian optimization can be regarded as a general framework for the data-driven modelling and solution of problems arising in water distribution systems. Scenario-based hydraulic simulation and Monte Carlo are key tools in modelling in water distribution systems. The related optimization problems fall into a simulation/optimization framework in which objectives and constraints are often black box. Bayesian optimization (BO) is characterized by a surrogate model, usually a Gaussian process but also a random forest, as well as neural networks and an acquisition function that drives the search for new evaluation points. These modelling options make BO nonparametric, robust, flexible, and sample efficient, making it particularly suitable for simulation/optimization problems. A defining characteristic of BO is its versatility and flexibility, given, for instance, by different probabilistic models, in particular different kernels, different acquisition functions. These characteristics of the Bayesian optimization approach are exemplified by two problems: cost/energy optimization in pump scheduling and optimal sensor placement for early detection of contaminant intrusion. Different surrogate models have been used both in explicit and implicit control schemes, showing that BO can drive the process of learning control rules directly from operational data. BO can also be extended to multi-objective optimization. Two algorithms are proposed for multi-objective detection problems using two different acquisition functions.

Keywords: pump scheduling optimization; Bayesian optimization; optimal sensor placement; Wasserstein distance; robustness

check for updates

Citation: Candelieri, A.; Ponti, A.; Giordani, I.; Archetti, F. Lost in Optimization of Water Distribution Systems: Better Call Bayes. *Water* **2022**, *14*, 800. https://doi.org/10.3390/w14050800

Academic Editor: Francesco De Paola

Received: 13 January 2022
Accepted: 25 February 2022
Published: 3 March 2022

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

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Optimization problems arising in environmental modelling are usually very challenging. One reason is the presence of several usually conflicting objectives: the financial cost and resilience in designing a water distribution network, and in general the issue of sustainability. Optimizing the detection time, cost, and probability of detection in designing a network to monitor water quality is another instance of multi-objective optimization. Multi-objective (MO) problems do not typically have a single best solution: The goal is to identify the set of Pareto optimal solutions such that any improvement in one objective means deteriorating another. Another reason is that we are dealing with simulation–optimization problems in a reference scenario where the objectives are expensive-to-evaluate black-box functions with no known analytical expression and no observable gradients. Another challenge is that the system performance has to be optimized in different conditions, which adds one more element of computational complexity.

The reference problem is:

$$\begin{gathered}\min H(x) = (F_1(x), \ldots, F_m(x)) \\ x \in \mathrm{X} \subset \mathbb{R}^d\end{gathered} \tag{1}$$

where:

$$F_i(x) = \int f_i(x, w)\ p(w) dw \text{ in the continuous case} \tag{2}$$

$$F_i(x) = \sum_{w=1}^{n} f_i(x, w) \text{in the discrete case} \tag{3}$$

Here, X is a simple compact (e.g., hyperrectangle, simplex, or a finite collection of points), w is a vector belonging to a set W, and p is finite, has a known analytical form, and is inexpensive to evaluate. Regarding $f_i(x, w)$ in Equations (2) and (3), we have no information on linearity or convexity; their evaluation is expensive and does not provide derivatives, making (2) and (3) black-box global optimization problems.

Contexts in which problems (Equation (2)) and (Equation (3)) naturally arise are:

- Optimizing average case performance of a system where $f_i(x, w)$ is the performance of design x under the environmental condition w and $p(w)$ represents the probability (or the fraction of time) that condition w occurs.

This is the case of optimal sensor placement in a monitoring network, where x is the placement of a number of sensors over the nodes, the environmental condition w is the injection of a contaminant at a node in a water distribution network, $f_i(x, w)$ is the detection time of the contamination event or of the fake news, and F_i is the detection time averaged over all contamination events.

- Optimizing the expected value of systems modelled by a discrete event simulation $f(x, w)$, where w is a random variable. In this case, $F_i(x, w) = \mathbb{E}(F_i(x, w^*|)$ and the objective becomes either (Equation (2)) if w is continuous or (3) if w is discrete. We can obtain noisy evaluations of $f_i(x, w)$ by simulating $f_i(x, w^*)$, where w^* is drawn from its conditional distribution given w.

This is the case, for instance, in the design of a water distribution and transmission system to maximize several performance metrics subject to stochastic patterns of water availability and demands.

The problem in Equation (3) arises in hyperparameter optimization in a machine learning algorithm. In this application, $f_i(x, w)$ is the i-metric (predictive accuracy, fairness, explainability) on fold w using hyperparameter x, and our goal is to minimize $\sum\ f_i(x, w)$. In this paper we do not consider this problem, but it has some water specific applications—for instance, water demand forecasting [1].

Even if the basic models (Equations (2) and (3)) share some properties, the computational strategies are different and, in this paper, we focus on the discrete case. The resulting combinatorial problems are typically NP-hard and cannot be solved in reasonable computing time. Approximation methods have been found to be very efficient in finding near-optimal solutions to a wide range of problem settings. In the following, the main types of methods are described.

Metaheuristics, in particular those that incorporate randomization and simulation and can be characterized as simheuristics.

Multi-objective evolutionary algorithms (MOEA), which offer the advantages of simplicity, applicability, versatility, and robustness. MOEAs can be naturally extended to multi-objective optimization using either a strategy of non-dominated sorting (NSGA) [2] that maintains an approximation of the Pareto set or a decomposition/strategy. Zhang et al. [3] proposed a parametrized aggregation of the objective functions whose optimizers for various parameter values span the Pareto set. MOEAs usually require a large number of function evaluations, which might make them unfeasible for very expensive functions. In order to improve the sample efficiency, many recent versions of EA incorporate a surrogate model: GP is often chosen as a surrogate since the seminal paper [4] and more recently in [5].

Bayesian optimization (BO) methods [6] are based on a surrogate model and offer unparalleled sample efficiency at the cost of a significant computational overhead, which makes them the methods of choice for very expensive function evaluations. BO methods are usually based on GPs whose kernel allows principled uncertainty evaluation, but random forests are also used, and more recently, neural networks. Multi-objective Bayesian optimization (MOBO) has recently been the subject of intense research. Rahat et al. [7] analyzed the two main strategies for generalizing the surrogate model to the multi-objective context. The first is termed "multi-surrogate", in which each objective function is modelled using a GP and the combined models induce a multivariate predictive distribution from which an acquisition function can be derived [8]. The other approach, "mono-surrogate", aggregates the objectives functions to generate a scalarized model. Both modelling options can be used to derive the acquisition functions to the multi-objective case [9], and more recently have been given to multi-objective EI [10] and LCB [11]. An analysis of MOBO is given in Section 2.3.

The main aim of this paper is to show that the flexibility of BO makes it an effective tool to solve a number of optimization problems in WDNs. This feature is exemplified by two basic problems in WSNs.

The first topic addressed is the energy optimization of pump scheduling. BO can be used to model explicit control, where the decision variables are the status of the pumps, and implicit control, where the optimization variables are the trigger thresholds in learning the pressure-based control rules. The Bayesian approach can also be used when only SCADA data are available. The problem of black-box constraints can also be handled in Bayesian optimization.

The second topic specifically addressed is the design of a sensor network for the detection of intruding contaminants in a WDN. It is shown that the optimal sensor placement in the presence of different contamination scenarios can be formulated as a sample average approximation (SAA) multi-objective optimization problem and is shown to be amenable to BO with specific acquisition functions based on hypervolume improvement. This is, to the authors' knowledge, the first case in which hypervolume improvement has been used in an optimal sensor problem. The focus of this paper is not on the computational performance per se, but rather on the analysis of a benchmark problem.

1.1. Related Works

The literature on energy optimization of pump scheduling is extremely broad.

A recent survey [12] under the evocative title of "Lost in optimisation of water distribution systems?" presented a wide-ranging literature review. MOEA algorithms have been widely used since the seminal paper by [13]. Castro-Gama et al. [14] gave an application of NSGA-II for a large scale WDN. Bayesian optimization was first considered in [15]. The BO approach was subsequently applied in [16] in order to learn implicit control strategies in water distribution networks. A further step was outlined in [17], where optimal controls for pump scheduling optimization are learned based only on SCADA data. A particularly relevant issue in the optimization of WDS is that the objective function may be undefined outside the feasible set: The EPANET simulator in the case of the violation of some constraints does not yield any results, making the application of penalty methods difficult. A specific method for this problem, also called "crash constraints", was proposed in [18] and was generalized in [16], which addresses the issue in general terms, modelling it as a classification/optimization problem.

The literature on optimal sensor placement is extremely broad.

This is also a key problem in water research literature that has been published in the last two decades following different approaches. Early contributions are [19], which used Gaussian processes, and [20], which defined "the battle of the water sensor networks (BWSN)". An early application of BO for sensor set selection is [21], in which the metric to be optimized was the predictive accuracy of the air temperature across the UK. The use of BO for contaminant source localization given information at each monitoring well was been proposed in [22]. A related management problem is ground water remediation, in which the pumping rates of "pump and treat" wells were fine-tuned using Bayesian optimization [23]. Recent contributions to optimal sensor placement used several methodological approaches. He et al. [24] proposed a multi-objective optimization method for sensor placement, allowing for contamination probability variations. Naserizade et al. [25] proposed a multi-objective model that includes risk directly in the objective functions. Zhang et al. [26] proposed an assessment of the quality of sensor placement strategies by a resilience index with respect to sensor failures. A new approach to optimal sensors was proposed in [27]. The novelty of this approach is that a sensor placement is represented as a histogram, which captures better than the sample average all the information gathered during the simulation. The Wasserstein (WST) distance between histograms enabled the design of a new multi-objective evolutionary algorithm (MOEA/WST), which has shown remarkable performance in solving the optimal sensor placement problem.

1.2. Our Contributions

- An analysis of the constraints in pump scheduling optimization, focusing on the black-box case and when the objective function cannot be computed at some points outside the feasible region.
- Embedding hydraulic simulation in a Bayesian optimization framework, exploiting its sample efficiency to deal with expensive function evaluation.
- BO can drive the learning process of the pressure control rules in implicit models of pump scheduling.
- The sample efficiency is carried over to a multi-objective case, where a hyper-volume-based acquisition function is shown to outperform the approach based on Chebyshev scalarization and expected improvement.

2. Bayesian Optimization

This section contains a basic description of the key components of Bayesian optimization (BO) for the single objective case (Sections 2.1 and 2.2) and the outline of its generalization to the multi-objective problem (Section 2.3).

Although BO can work for any kind of optimization problem, its substantial computational overhead makes it suitable when the evaluation of the objective functions is expensive (as in the case of the execution of a simulation model), for the analytical form of objectives, and constraints and derivatives are not available. These features of BO, along with its versatility, make it a de-facto standard in black-box situations (Figure 1). BO generates a sequence of points at which to evaluate the objectives, with the aim of determining a near-optimal approximation of the global minimizer after a small number of evaluations, which can also be noisy.

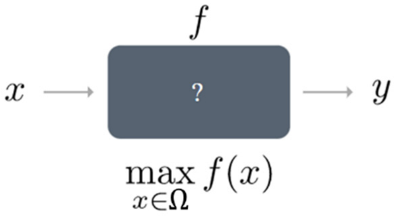


Figure 1. The general scheme of a problem underlying BO.

Bayesian optimization is a statistically principled approach to adaptively and sequentially select these query points (or batches of query points), which enables an effective trade-off between evaluating f in regions of good observed performance and regions of high uncertainty.

In order to be sample efficient—that is, optimize f within a small number of evaluations—we need a way of extrapolating our belief about what f looks like at points not yet evaluated. In Bayesian optimization (Figure 2), this is enabled by a surrogate model, which should also be able to quantify the uncertainty of its predictions in the form of a posterior distribution over function values $f(x)$ at points x. Posteriors represent the belief a model has about the function values about points not yet observed. Therefore, the posterior is the distribution over the outputs conditioned on the data observed so far. When using the Gaussian process model, the posterior is given explicitly as a multivariate distribution.

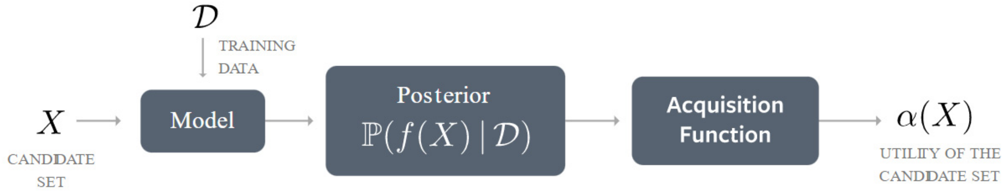


Figure 2. The general scheme of a BO framework.

2.1. Probabilistic Model

A Gaussian process (GP) is a probability distribution over functions denoted as $f(x) \sim GP(\mu(x),\ k(x,\ x'))$, where $\mu(x)$ is the mean function of the GP and $k(x,x')$ is the covariance function (a.k.a., kernel). Therefore, GP is a collection of random variables, any finite number of which have a joint Gaussian distribution, and $f(x)$ can be considered a sample from a multivariate normal distribution [28].

Let $\mathrm{X}_{1:n} = \left\{x^{(i)}\right\}_{i=1,\ldots,n}$ with a set of n points in $\Omega \subset \Re^d$ and $y_{1:n} = \left\{f\left(x^{(i)}\right) + \varepsilon\right\}_{i=1,\ldots,n}$ denote the associated function values, possibly noisy with ε being a zero-mean Gaussian noise $\varepsilon \sim \mathcal{N}(0,\lambda^2)$. Then the posterior predictive mean $\mu(x)$ and standard deviation $\sigma^2(x)$, conditioned on $\mathrm{X}_{1:n}$ and $y_{1:n}$, are given by the following equations:

$$\mu(x) = \mathrm{k}(x,\mathrm{X}_{1:n})\ \left[\mathrm{K} + \lambda^2 I\right]^{-1} y_{1:n} \tag{4}$$

$$\sigma^2(x) = k(x,x) - \mathrm{k}(x,\mathrm{X}_{1:n})\ \left[\mathrm{K} + \lambda^2 I\right]^{-1} k(\mathrm{X}_{1:n},x) \tag{5}$$

where $\mathrm{k}(x,\mathrm{X}_{1:n}) = \left\{k\left(x,x^{(i)}\right)\right\}_{i=1,\ldots,n}$ and $\mathrm{K} \in \Re^{n\times n}$ with entries $\mathrm{K}_{ij} = k\left(x^{(i)},x^{(j)}\right)$.

The choice of kernel should reflect prior beliefs over the structural properties of $f(x)$, specifically its smoothness. Almost every kernel has its own hyperparameters to tune—usually via maximum log-likelihood estimation (MLE) or maximum a posteriori probability (MAP)—for reducing the potential mismatches between prior smoothness assumptions and the observed data. Common kernels for GP regression are:

- Squared exponential: $k_{SE}(x,x') = e^{-\frac{||x-x'||^2}{2\ell^2}}$
- Exponential: $k_{EXP}(x,x') = e^{-\frac{||x-x'||}{\ell}}$
- Power exponential: $k_{PE}(x,x') = e^{-\frac{||x-x'||^p}{\ell^p}}$
- Matérn3/2: $k_{M3/2}(x,x') = \left(1+\frac{\sqrt{3}\,||x-x'||}{\ell}\right)e^{-\frac{\sqrt{3}\,||x-x'||}{\ell}}$
- Matérn5/2: $k_{M5/2}(x,x') = \left[1+\frac{\sqrt{5}\,||x-x'||}{\ell}+\frac{5}{3}\left(\frac{||x-x'||}{\ell}\right)^2\right]e^{-\frac{\sqrt{5}\,||x-x'||}{\ell}}$

2.2. Acquisition Functions

The acquisition function is the mechanism behind the sample efficiency of the Bayesian optimization method. It manages the balance between exploration and exploitation. It drives the search of the new evaluation points towards points with potential high values of objective function either because the value of $\mu(x)$ is high or the uncertainty represented by $\sigma^2(x)$ is high (or both). This is also called the exploitation/exploration dilemma: Exploiting means considering the area that provides a higher chance of improving over the best value observed so far, and exploring pushes the search towards less explored regions of the search space in order to acquire new knowledge moving. Acquisition functions have been widely studied in the Bayesian optimization and machine learning community, as has artificial intelligence, in general. They are the basis of the learning process itself, and many different approaches have been proposed [6,29]. The two considered in this paper are detailed in Appendix A.

2.3. Multi-Objective Bayesian Optimization

As far as the surrogate model is concerned, there are two main strategies for generalizing BO to the multi-objective context [7]. The first is termed "multi-surrogate", in which each objective function is modelled using a GP, and the combination of these models induces a multivariate predictive distribution from which an acquisition function can be derived. The second, "mono-surrogate", aggregates the objectives into a scalarized model that is used to compute the acquisition function. The weight of the m objectives are sampled from a uniform distribution to construct a single objective function whose optimizers can be shown to span, under wide conditions, the whole Pareto set [30]. As far as the acquisition functions are concerned, a survey of multi-objective versions is given in [9]. Specific versions have been more recently given to multi-objective expected improvement [10] and lower confidence bound [11].

The hypervolume is an indicator of the solution quality that measures the objective space between a non-dominated set and a predefined reference vector. Figure 3 displays the mechanism of hypervolume improvement in the case of max$F(x) = (f^{(1)}(x), f^{(2)}(x))$. Figure 3a displays four points l_i, where $i = 1, \ldots, 4$ non-dominated in the objective space (the dominated space is in red), a new point x_1 dominates the previous ls but l_1 brings about an improvement of the dominate hypervolume (in blue, Figure 3b), and x_2 dominates l_1,l_2,l_3 and is not dominated by l_4, bringing about a further improvement (in green, Figure 3c).

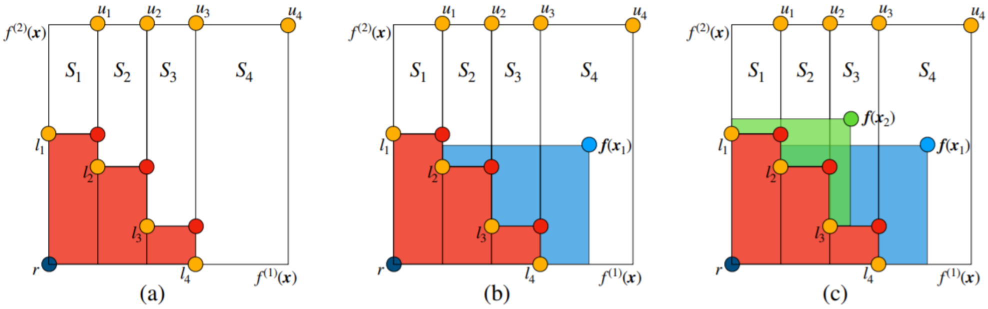


Figure 3. The initial Hypervolume (**a**). The improvement adding the point x_1 (**b**). The improvement adding the point x_2 (**c**).

Indeed, it can be shown that the maximization of the dominated hypervolume is equivalent to locating the optimal Pareto set and gives approximate Pareto fronts with excellent coverage [31].

As such, the expected hypervolume improvement (EHVI) is also a natural acquisition function. A drawback of the hypervolume is its computational cost, an even more critical factor if it is to be repeatedly optimized as in acquisition functions. A number of computational improvements focused on efficient methods for approximating EHVI have been suggested [32–34], but its computation is still very expensive. Very recently, the issue of MOBO was analyzed also in connection with multi-fidelity and multiple information sources [8,35].

The situation is further complicated when constraints are also black box or even undefined outside the feasible set. In this case, MOEA algorithms are not easily made suitable for this kind of problem. A general solution is proposed in [36].

2.4. BoTorch

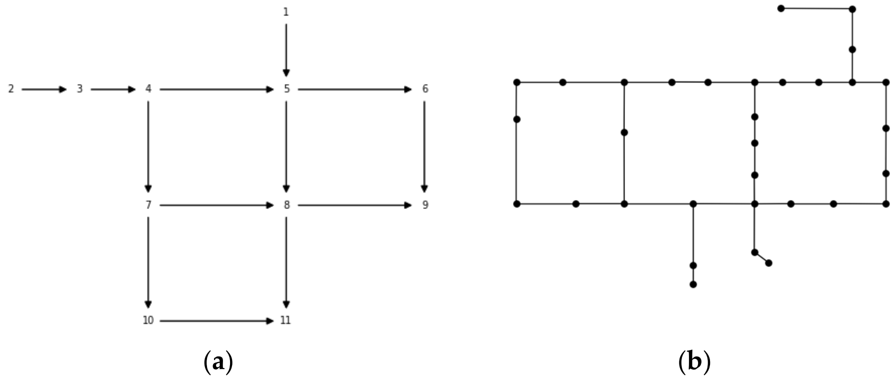
Two libraries were used in this paper: mlrMBO for pump scheduling optimization (Section 5) and Ax-BoTorch [33,37] for the optimal sensor placement (Section 6).

BoTorch is a library for Bayesian optimization built on top of PyTorch and is part of the PyTorch ecosystem. In this section, we present a short analysis of its main features. The reader is referred to [33] for a detailed exposition. BoTorch is particularly suitable to exploiting the defining characteristic of BO as its versatility and flexibility, given by different probabilistic models—in particular, different kernels and different acquisition functions. BoTorch supports a wider set of kernels than the basic set in Section 2.1 and therefore a wider set of surrogate models. Beyond the acquisition functions EI and L/U confidence bounds in Section 2.2, BoTorch supports the knowledge gradient and max volume entropy search. The knowledge gradient is also available in its multi-fidelity version. BoTorch allows novel approaches that do not admit analytic solutions, including batch acquisition functions and multi-output models with multiple correlated outcomes. In particular, BoTorch utilizes Monte-Carlo methods in the computation of acquisition functions, lifting the restrictive assumptions about the underlying model. It also supports the optimization of acquisition functions over any kind of posterior distribution, as long as it can be sampled from. BoTorch supports batch versions of EI and UCB. In batch BO, q design points are selected for parallel evaluation. BoTorch also supports black-box constraints using variants of the expected improvement, in which the improvement of the objective is weighted by the probability of feasibility [38]. BoTorch supports the computation of the Pareto front and of the expected hypervolume improvement (EHVI) as acquisition function, computing its gradients by auto-differentiation.

3. Data and Software Resources

3.1. Data Resources

Net1 is a toy network provided in WNTR. Hanoi [27] is a benchmark used in the literature whose associated graph consists of 32 nodes and 34 edges (Figure 4).



(**a**) (**b**)

Figure 4. The graph of Net1 (**a**) and Hanoi (**b**).

The case study used for learning the pressure control rules is a WDN in the greater Milan, Italy, area, which supplies water to three municipalities: Bresso (around 20,000 inhabitants), Cormano (around 26,000 inhabitants), and Cusano-Milanino (around 19,000 inhabitants). The network consists of 7418 pipes, 8493 junctions, 14 reservoirs, 1381 valves, and 9 pumping stations, with 14 pumps overall. The piezometric level of the WDN ranges from 136 to 174 m (average: 148 m). A schematic representation of the WDN is reported in Figure 5.

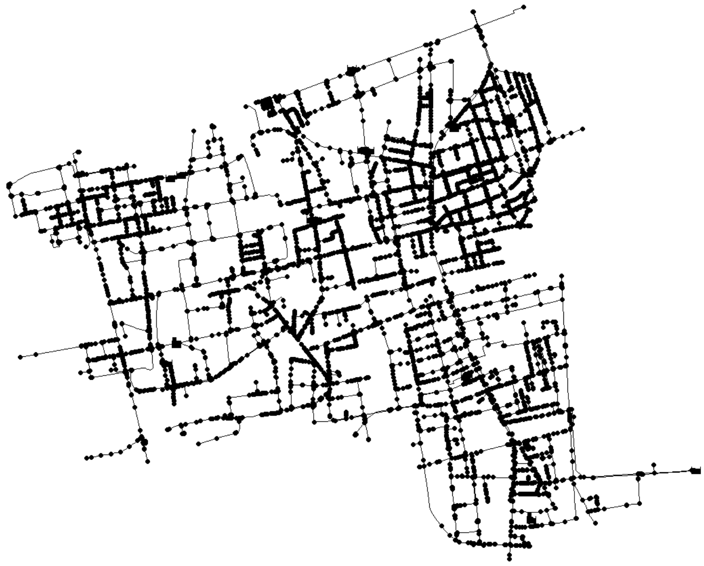


Figure 5. The graph of the BCM (Bresso–Cormano–Cusano-Milanino) network.

3.2. Software Resources

The pump scheduling optimization problem was solved using the R package named ″mlrMBO″. A toolbox for model-based optimization (MBO). mlrMBO can handle both single- and multi-objective optimization with mixed continuous, categorical, and conditional decision variables. It also offers different acquisition functions and methods for their optimization.

The optimal sensor placement problem was solved using the BoTorch library for Bayesian optimization as part of the PyTorch ecosystem. BoTorch provides an easy-to-use interface for defining, managing, and running sequential experiments and a modular interface for composing Bayesian optimization primitives as probabilistic models, acquisition functions, and optimizers. BoTorch organizes the computations in batches of size q.

EPANET 2.0 is the most well-known and largely adopted open-source software for simulating WDN. More recently, WNTR was proposed: It is coded in Python, based on EPANET, and provides a larger set of functionalities, especially for structural and resilience analysis of a WDN.

4. Pump Scheduling Optimization

The commonly used approach (pump scheduling optimization) is explicit control: The pump's status and speed are decided at prefixed times. Thus, the problem is to efficiently search among all the possible schedules (i.e., configurations of the decision variables) to optimize the objective function with a constraint of hydraulic feasibility. Many methods have been proposed from domains such as mathematical programming, meta-heuristics, and evolutionary and nature-inspired algorithms. Explicit control models require many decision variables for real-world water distribution networks, increasing with the number of pumps and the number of time intervals.

Another approach is the implicit control model: The pumps' status/speeds is controlled by pressure-related rules—a pump is activated if pressure (at specific locations) is lower than a minimum threshold, or it is deactivated if pressure exceeds a maximum threshold. Otherwise, the status/speed of the pump is not changed. These thresholds become the decision variables of the optimal implicit control problem. The search space for the implicit control has a lower dimensionality than in the explicit case. For instance, if one has to control n pumps, by deciding to simply switch them on/off on an hourly basis over H hours, the entire search space for finding an optimal explicit control (i.e., optimal pump schedule) consists of $2nH$ possible pump schedules within an -dimensional space. In contrast, the search space associated with implicit control, for instance, consisting of min–max ranges on the pumps' pressures, has just $2n$ dimensions. On the other hand, the structure of the associated search space is typically more complex due to the constraints/conditions among values of the decision variables.

The section demonstrates the versatility of the Bayesian optimization paradigm which, through different surrogate models and acquisition functions, can handle both explicit (Section 4.1) and implicit (Section 4.2) models.

4.1. Pump Scheduling Optimization through Bayesian Optimization

The reference paper is [15], which proposes a BO approach for the PSO problem by also comparing two different probabilistic surrogate models to approximate the black-box objective function, which are a Gaussian process (GP) regression and a random forest (RF).

The approach considers a simulation–optimization setting, where a simulation run of the software hydraulic model of the WDN, given a certain pump schedule, provides the associated value of the objective function (i.e., energy cost to be minimized) or its non-computability, meaning that some hydraulic constraints have been violated, leading to the impossibility of computing the objective function associated with a hydraulic feasible schedule. More precisely, hydraulic feasibility refers to "basic" computational constraints (e.g., the impossibility of having a negative pressure at some location, lack of convergence in the simulation, impossibility of satisfying water demand, etc.) as well as operational constraints that can be set by the user, such as specific min–max operating ranges for pressure at each node. Finally, feasibility also depends on the specific simulated scenario, such as the water demand patterns the user associates with each node according to historical data.

Although "penalizing" infeasible (aka non-computable) schedules proved to be a sufficiently workaround [15], the estimation of the unknown feasibility region was successively addressed in [18] in the case of PSO and in [17] in the case of a generic problem with a potentially non-computable black-box objective function.

4.2. Learning Optimal Control Rules as a Black-Box Optimization Problem

In implicit control, the pumps are controlled depending on pressure at some locations, usually their associated pressures. For the sake of simplicity, one can consider the easiest case, where the control rule is based on two thresholds: τ1 and τ2, the lowest threshold, implying to switch the pump on, and the highest threshold, implying to switch the pump off, respectively:

```
IF ($p$ < τ1 AND $S$ = OFF) THEN $S$ ← ON
ELSE
IF ($p$ > τ2 AND $S$ = ON) THEN $S$ ← OFF
```

where p is the current pressure value at the pump and S is the status of the pump (i.e., ON/OFF).

Figure 6 shows an example of this simple control rule for a single pump.

Thus, τ1 and τ2 are the decision variables to be optimized in the implicit control case, with the aim of minimizing the associated energy cost while constrained to the satisfaction of the water demand.

The possibility of generating infeasible (aka non-computable) control strategies can also be taken into account, leading to a simulation–optimization problem with a black-box objective function.

Bayesian optimization has been shown to also be able to handle unknown and crash constraints, which would make a penalty-based approach impossible, but can be handled by BO. Implicit control strategies learned through the BO approach proposed in [16] were not only optimal and efficient in terms of the number of simulation runs required to identify them, but also "hydraulically" robust with respect to perturbations in the water demand.

One further step was taken in the paper "Active learning of pressure control rules of water distribution networks" [17], in which it was shown that an accurate approximation of the outputs of the hydraulic simulation can be obtained by training a deep neural network on SCADA data. The main result was that the hydraulic simulation model can be "replaced" by a deep neural network (DNN): The knowledge hidden in the data can be leveraged into a low-cost source using SCADA data to train a DNN to predict the relevant outputs (i.e., energy and hydraulic feasibility), avoiding costs for the design, development, validation, and execution of a "virtual twin" provided by the hydraulic simulation of the water distribution network. Computational results in [17] showed that threshold-based rules can be learned through an active learning approach: The lower dimensionality of the search space, compared to explicit control, substantially reduces the computational cost.

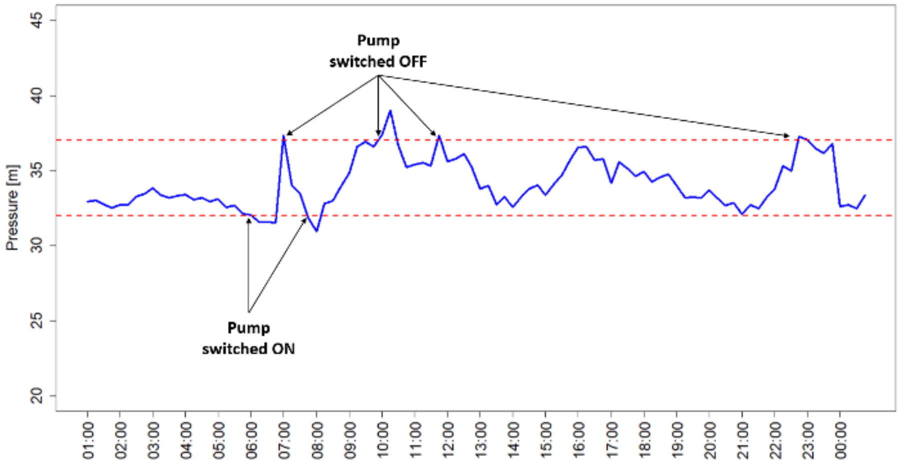


Figure 6. A schematic representation of an implicit pump control strategy based on min–max thresholds (red dotted lines) on the pressure value (in blue).

5. Optimal Sensor Placement

5.1. The Basic Model

Let $G = (V, E)$ be the graph underlying the water distribution network and $L \subseteq V$ the set of possible locations for deploying sensors. That is, a sensor placement (SP) is a subset of L, with the subset's size less or equal to p depending on the available budget, represented by a binary vector $s \in \{0,1\}^{|L|}$ whose components are $s_i = 1$ if a sensor is located at node i, $s_i = 0$. Otherwise, let $A \subseteq V$ be the set of contamination events $a \in A$ and d_{ai} the detection time of a sensor placed in node i of event a. A probability distribution is placed over possible contamination events $a \in A$. The "fitness" of a sensor placement is given by two metrics: the detection time and its standard deviation.

This brings about the following model of sensor placement:

$$P = \begin{cases} \min f_1(s) = \sum\limits_{a \in A} \alpha_a \sum\limits_{i=1,\ldots,|L|} d_{ai} x_{ai} \\ s.t. \\ \sum\limits_{i=1,\ldots,|L|} s_i \le p \\ s_i \in \{0,1\} \end{cases} \tag{6}$$

- α_a is the probability of event a.
- $x_{ai} = 1$ if $s_i = 1$, where i is the first sensor detecting event a, and 0 otherwise.

In our study, we assumed that all the events had the same chance of happening, that is, $\alpha_a = 1/|\mathrm{A}|$; therefore, $f_1(\mathrm{s})$ is:

$$f_1(s) = \frac{1}{|A|} \sum_{a \in A} \hat{t}_a \tag{7}$$

where $\hat{t}_a = \sum\limits_{i=1,\ldots,|L|} d_{ai} x_{ai}$ is the *MDT* of event a. General discrete distributions over $A \subseteq V$ can be also used. For each event a and sensor placement s the minimum detection time is defined as $MDT_a = \min_{i:\, s_i=1} d_{ai}$, with $\hat{t}_a$ being the minimum time step at which concentration exceeds a given threshold τ for event a. f_1 is the average over the events of the detection time.

As a measure of risk, we considered f_2 as the standard deviation of the sample average approximation of f_1.

$$f_2(s) = STD_{f_1}(s) = \sqrt{\frac{1}{|A|}\sum_{a\in A}\left(\hat{t}_a - f_1(s)\right)^2} \quad (8)$$

5.2. Multi-Objective Bayesian Optimization for Sensor Placement

Two Bayesian optimization algorithms, A1 and A2, were used in this paper for optimal sensor placement.

A1 was the mono-surrogate, based on Chebyshev scalarization of the objectives of $f^{(1)}$ and $f^{(2)}$. A GP of the aggregate function and the expected improvement (EI) were the acquisition function. A2 was the multi-surrogate based on GP models of the single objectives and the expected hypervolume improvement was the acquisition function.

The authors implemented them using software components from BoTorch. Both organized the computations in batches of size $q = 5$.

The "helper" function created the outcomes required for the scalarized objective and applied the scalarization and the constraints. The helper function initialized A1 and A2 and returned the batch $(x_1, x_2, \ldots, x_q)$ along with the observed values. In A1, each candidate was optimized in a sequential greedy fashion using a different random scalarization.

The BO loop for a batch of size q iterated the following steps:

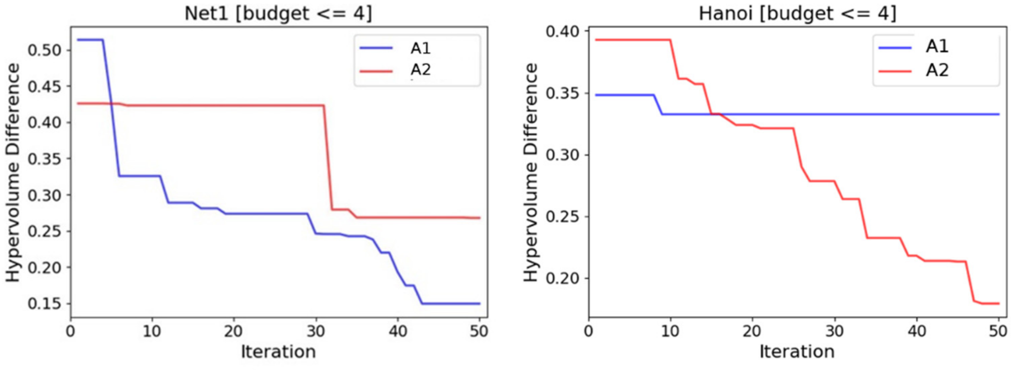
- Given a surrogate, choose a batch of q points.
- Observe $f(x)$ for each x.
- Update the surrogate model.

A1 could also be extended to the constrained case by weighting the EI by the probability of feasibility.

In A2, a list of acquisition functions was created, each with different scalarization weights. This list generated one candidate for each acquisition and conditioned the next candidate (and acquisition) on the previously selected candidates.

5.3. Computational Results

The plot below (Figure 7) shows, for Net 1 (left) and the Hanoi network (right), the difference in the hypervolume between the feasible true Pareto front and hypervolume of the observed Pareto front for algorithms A1 and A2.

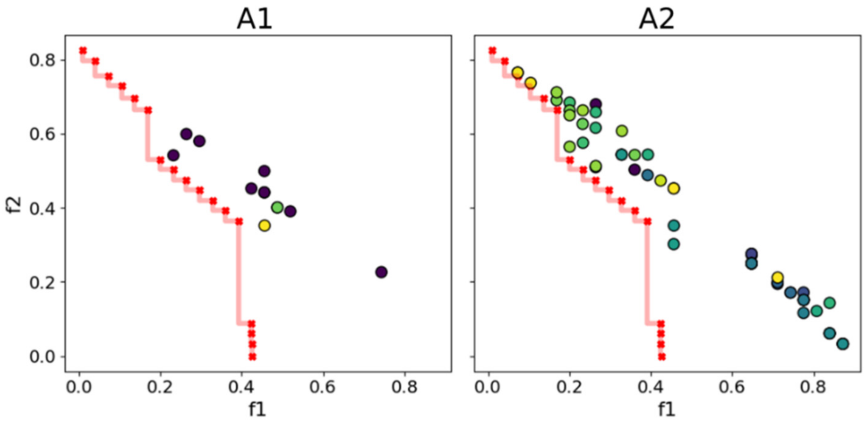


(**a**) (**b**)

Figure 7. Difference in hypervolume between the optimal Pareto front and the Pareto front approximated by A1 (blue) and A2 (red) over the iteration considering Net1 (**a**) and Hanoi (**b**).

It is apparent that A2's relative performance improved with the size of the problem.

An effective way to visualize the optimization process is to represent the observations by a color associated with the iteration count of A1 and A2. It is apparent from Figure 8 that A2 (right plot) was much quicker than A1 (left plot) at identifying the Pareto front.



(**a**) (**b**)

Figure 8. The objective space considering Hanoi network using algorithm A1 (**a**) and A2 (**b**). True Pareto front in red. Lighter colors denote later iterations.

An example of the sensor placement given by the algorithms is reported in Figures 9 and 10.

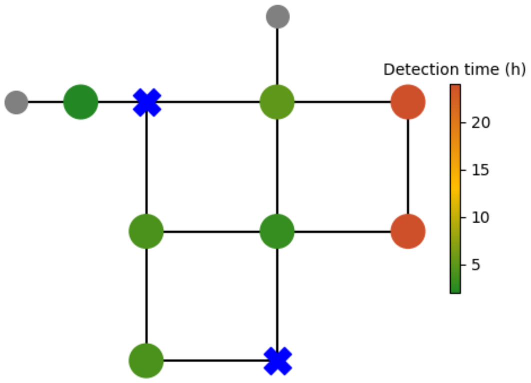


Figure 9. Optimal sensor placement on the Net_1 WDN. Blue crosses are sensor locations and the colors denote the detection time. Grey nodes are a reservoir (left) and a tank (top), which are not considered possible sensor locations.

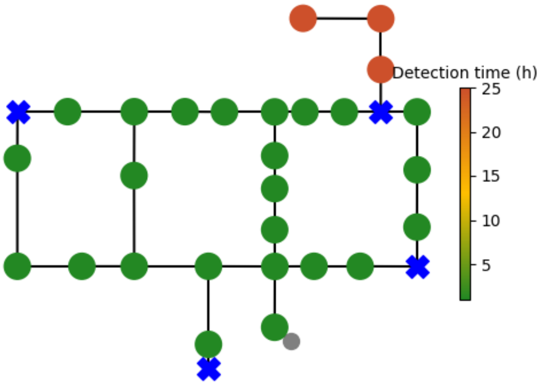


Figure 10. Optimal sensor placement on the Hanoi WDN. Blue crosses are sensor locations and the colors denote the detection time. The grey node is a reservoir, which is not considered a possible sensor location.

6. Conclusions

The main conclusion is that Bayesian optimization offers a versatile and comprehensive framework for the solution of a wide range of problems both for the design and operation of water distribution networks and other environmental domains. The key argument is that Bayesian optimization is sample efficient method, and this is particularly important in black-box problems when the optimization is driven by the output of a simulation model.

The evaluation of the objective functions is largely based on the simulation of the underlying system in different contexts and can result in computationally expensive black-box problems. The probabilistic surrogate model enables a quantification of predictive uncertainty, which is fed into an acquisition function that drives the selection of the next evaluation point. It was shown how this feature can be exploited in multi-objective problems by incorporating into the acquisition function the concept of hypervolume improvement. Evolutionary algorithms are often used for these problems, as well as in multi-objective cases, but they are not nearly as sample efficient as BO, despite their performance being improved by incorporating a surrogate model. Sample efficiency in BO is particularly important when the data and computational resources are severely constrained by the high computational cost of simulation experiments encompassing several different scenarios or Monte Carlo simulation.

BO libraries, as exemplified in this paper, allow for quick development of a first prototype requiring a negligible amount of coding and allow both the data scientist and the domain expert to choose those components (in particular, probabilistic models and acquisition functions) that best fit the target application and the available resources.

Specifically, the most relevant achievements were showing (a) that Bayesian optimization enables the optimal learning of pressure-based control rules for pump scheduling from the SCADA system, and (b) that the optimal placement of sensors for the early detection of contaminants can also be efficiently solved through Bayesian optimization.

The value of data hidden in the SCADA systems in order to approximate the hydraulic behavior and develop data-driven strategies for a number of operational problems is an interesting lesson that water utilities can draw from the results of this paper.

Author Contributions: All authors contributed equally to the paper. All authors have read and agreed to the published version of the manuscript.

Funding: This study was partially supported by the Italian project "PERFORM-WATER 2030"—Programma POR (Programma Operativo Regionale) FESR (Fondo Europeo di Sviluppo Regionale) 2014–2020 and the innovation call "Accordi per la Ricerca e l'Innovazione" ("Agreements for Research and Innovation") of Regione Lombardia (DGR N. 5245/2016—AZIONE I.1.B.1.3—ASSE I POR FESR 2014–2020)—CUP E46D17000120009. This study was also supported by the national project "ENERGIDRICA" MIUR ARS01_00625.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data of the water distribution networks are available from the authors on request.

Acknowledgments: We greatly acknowledge the DEMS Data Science Lab for supporting this work by providing computational resources (DEMS—Department of Economics, Management and Statistics).

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

Appendix A

Acquisition Functions

A widely used acquisition function is the expected improvement (EI), given by the expected improvement on $f(x)$ with respect to the predictive distribution of the surrogate model. The parameter ξ is related to the trade-off between exploitation and exploration. Ξ

should be adjusted dynamically to decrease as the optimization moves along, biasing the search toward exploitation in later stages of the optimization process:

$$EI(x) = \begin{cases} (\mu(x) - f(x^+) - \xi)\Phi(Z) + \sigma(x)\phi(Z)\ if\ \sigma(x) > 0 \\ 0\ if\ \sigma(x) = 0 \end{cases} \tag{A1}$$

where ϕ and Φ are the probability and the cumulative distribution functions, respectively, and

$$Z = \begin{cases} \frac{\mu(x) - f(x^+) - \xi}{\sigma(x)}\ if\ \sigma(x) > 0 \\ 0\ if\ \sigma(x) = 0 \end{cases} \tag{A2}$$

Another common acquisition is based on the confidence bound concept (lower confidence bound (LCB) and upper confidence bound (UCB) in the minimization and maximization case, respectively). The next point to evaluate is given by the minimizer of:

$$\mathrm{LCB}(x) = \mu(x) - \mathrm{k}\sigma(x) \tag{A3}$$

where $\mathrm{k} \geq 0$ is the parameter to manage the exploration/exploitation trade-off ($\mathrm{k} = 0$ means pure exploitation).

References

1. Candelieri, A.; Giordani, I.; Archetti, F. Automatic Configuration of Kernel-Based Clustering: An Optimization Approach. In Proceedings of the International Conference on Learning and Intelligent Optimization, Nizhny Novgorod, Russia, 19–21 June 2017; Springer: Berlin/Heidelberg, Germany, 2017; pp. 34–49.
2. Deb, K.; Pratap, A.; Agarwal, S.; Meyarivan, T. A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II. *IEEE Trans. Evol. Comput.* **2002**, *6*, 182–197. [CrossRef]
3. Zhang, Q.; Li, H. MOEA/D: A Multiobjective Evolutionary Algorithm Based on Decomposition. *IEEE Trans. Evol. Comput.* **2007**, *11*, 712–731. [CrossRef]
4. Zhang, Q.; Liu, W.; Tsang, E.; Virginas, B. Expensive Multiobjective Optimization by MOEA/D with Gaussian Process Model. *IEEE Trans. Evol. Comput.* **2009**, *14*, 456–474. [CrossRef]
5. Liu, F.; Zhang, Q.; Han, Z. MOEA/D with Gradient-Enhanced Kriging for Expensive Multiobjective Optimization. In Proceedings of the International Conference on Evolutionary Multi-Criterion Optimization, Shenzhen, China, 28–31 March 2021; Springer: Cham, Switzerland, 2021; pp. 543–554.
6. Archetti, F.; Candelieri, A. *Bayesian Optimization and Data Science*; Springer: Berlin/Heidelberg, Germany, 2019.
7. Rahat, A.A.-A.M.; Everson, R.M.; Fieldsend, J.E. Alternative Infill Strategies for Expensive Multi-Objective Optimisation. In Proceedings of the Genetic and Evolutionary Computation Conference, GECCO 2017, Berlin, Germany, 15–19 July 2017; Bosman, P.A.N., Ed.; ACM: New York, NY, USA, 2017; pp. 873–880.
8. Belakaria, S.; Deshwal, A.; Jayakodi, N.K.; Doppa, J.R. Uncertainty-Aware Search Framework for Multi-Objective Bayesian Optimization. In Proceedings of the Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, 7–12 February 2020; AAAI Press: Palo Alto, CA, USA, 2020; pp. 10044–10052.
9. Emmerich, M.T.M.; Giannakoglou, K.C.; Naujoks, B. Single- and Multiobjective Evolutionary Optimization Assisted by Gaussian Random Field Metamodels. *IEEE Trans. Evol. Comput.* **2006**, *10*, 421–439. [CrossRef]
10. Zhan, D.; Qian, J.; Liu, J.; Cheng, Y. Pseudo Expected Improvement Matrix Criteria for Parallel Expensive Multi-Objective Optimization. In Proceedings of the World Congress of Structural and Multidisciplinary Optimisation, Braunschweig, Germany, 5–9 June 2017; Springer: Berlin/Heidelberg, Germany, 2017; pp. 175–190.
11. Sun, G.; Li, L.; Fang, J.; Li, Q. On Lower Confidence Bound Improvement Matrix-Based Approaches for Multiobjective Bayesian Optimization and Its Applications to Thin-Walled Structures. *Thin-Walled Struct.* **2021**, *161*, 107248. [CrossRef]
12. Mala-Jetmarova, H.; Sultanova, N.; Savic, D. Lost in Optimisation of Water Distribution Systems? A Literature Review of System Design. *Water* **2018**, *10*, 307. [CrossRef]
13. Van Zyl, J.E.; Savic, D.A.; Walters, G.A. Operational Optimization of Water Distribution Systems Using a Hybrid Genetic Algorithm. *J. Water Resour. Plan. Manag.* **2004**, *130*, 160–170. [CrossRef]
14. Castro-Gama, M.; Pan, Q.; Lanfranchi, E.A.; Jonoski, A.; Solomatine, D.P. Pump Scheduling for a Large Water Distribution Network. Milan, Italy. *Procedia Eng.* **2017**, *186*, 436–443. [CrossRef]
15. Candelieri, A.; Perego, R.; Archetti, F. Bayesian Optimization of Pump Operations in Water Distribution Systems. *J. Glob. Optim.* **2018**, *71*, 213–235. [CrossRef]

16. Candelieri, A.; Ponti, A.; Archetti, F. Data Efficient Learning of Implicit Control Strategies in Water Distribution Networks. In Proceedings of the 2021 IEEE 17th International Conference on Automation Science and Engineering (CASE), Lyon, France, 23–27 August 2021; pp. 1812–1816.
17. Candelieri, A.; Perego, R.; Giordani, I.; Archetti, F. Active Learning of Optimal Controls for Pump Scheduling Optimization. In Proceedings of the EGU General Assembly Conference Abstracts, Online, 19–30 April 2021; p. EGU21-12610.
18. Tsai, Y.-A.; Pedrielli, G.; Mathesen, L.; Zabinsky, Z.B.; Huang, H.; Candelieri, A.; Perego, R. Stochastic Optimization for Feasibility Determination: An Application to Water Pump Operation in Water Distribution Network. In Proceedings of the 2018 Winter Simulation Conference (WSC), Gothenburg, Sweden, 9–12 December 2018; pp. 1945–1956.
19. Guestrin, C.; Krause, A.; Singh, A.P. Near-Optimal Sensor Placements in Gaussian Processes. In Proceedings of the 22nd International Conference on Machine Learning, Bonn, Germany, 7–11 August 2005; pp. 265–272.
20. Ostfeld, A.; Uber, J.G.; Salomons, E.; Berry, J.W.; Hart, W.E.; Phillips, C.A.; Watson, J.-P.; Dorini, G.; Jonkergouw, P.; Kapelan, Z. The Battle of the Water Sensor Networks (BWSN): A Design Challenge for Engineers and Algorithms. *J. Water Resour. Plan. Manag.* **2008**, *134*, 556–568. [CrossRef]
21. Garnett, R.; Osborne, M.A.; Roberts, S.J. Bayesian Optimization for Sensor Set Selection. In Proceedings of the 9th ACM/IEEE International Conference on Information Processing in Sensor Networks, Stockholm, Sweden, 12–16 April 2010; pp. 209–219.
22. Pirot, G.; Krityakierne, T.; Ginsbourger, D.; Renard, P. Contaminant Source Localization via Bayesian Global Optimization. *Hydrol. Earth Syst. Sci.* **2019**, *23*, 351–369. [CrossRef]
23. Pourmohamad, T.; Lee, H.K. *Bayesian Optimization with Application to Computer Experiments*; Springer: Berlin/Heidelberg, Germany, 2021.
24. He, G.; Zhang, T.; Zheng, F.; Zhang, Q. An Efficient Multi-Objective Optimization Method for Water Quality Sensor Placement within Water Distribution Systems Considering Contamination Probability Variations. *Water Res.* **2018**, *143*, 165–175. [CrossRef]
25. Naserizade, S.S.; Nikoo, M.R.; Montaseri, H. A Risk-Based Multi-Objective Model for Optimal Placement of Sensors in Water Distribution System. *J. Hydrol.* **2018**, *557*, 147–159. [CrossRef]
26. Zhang, Q.; Zheng, F.; Kapelan, Z.; Savic, D.; He, G.; Ma, Y. Assessing the Global Resilience of Water Quality Sensor Placement Strategies within Water Distribution Systems. *Water Res.* **2020**, *172*, 115527. [CrossRef]
27. Ponti, A.; Candelieri, A.; Archetti, F. A Wasserstein Distance Based Multiobjective Evolutionary Algorithm for the Risk Aware Optimization of Sensor Placement. *Intell. Syst. Appl.* **2021**, *10*, 200047. [CrossRef]
28. Rasmussen, C.E.; Williams, C.K.I. *Gaussian Processes for Machine Learning*; Adaptive Computation and Machine Learning; MIT Press: Cambridge, UK, 2006; ISBN 978-0-262-18253-9.
29. Frazier, P.I. Bayesian Optimization. In *Recent Advances in Optimization and Modeling of Contemporary Problems*; INFORMS: Catonsville, MD, USA, 2018; pp. 255–278.
30. Paria, B.; Kandasamy, K.; Póczos, B. A Flexible Framework for Multi-Objective Bayesian Optimization Using Random Scalarizations. In Proceedings of the Thirty-Fifth Conference on Uncertainty in Artificial Intelligence, UAI 2019, Tel Aviv, Israel, 22–25 July 2019; Globerson, A., Silva, R., Eds.; AUAI Press: Arlington, VA, USA, 2018; Volume 115, pp. 766–776.
31. Couckuyt, I.; Deschrijver, D.; Dhaene, T. Fast Calculation of Multiobjective Probability of Improvement and Expected Improvement Criteria for Pareto Optimization. *J. Glob. Optim.* **2014**, *60*, 575–594. [CrossRef]
32. Yang, K.; Emmerich, M.; Deutz, A.; Bäck, T. Multi-Objective Bayesian Global Optimization Using Expected Hypervolume Improvement Gradient. *Swarm Evol. Comput.* **2019**, *44*, 945–956. [CrossRef]
33. Balandat, M.; Karrer, B.; Jiang, D.; Daulton, S.; Letham, B.; Wilson, A.G.; Bakshy, E. BoTorch: A Framework for Efficient Monte-Carlo Bayesian Optimization. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 21524–21538.
34. Zhang, R.; Golovin, D. Random Hypervolume Scalarizations for Provable Multi-Objective Black Box Optimization. In Proceedings of the 37th International Conference on Machine Learning, ICML 2020, Virtual Event, 13–18 July 2020; Volume 119, pp. 11096–11105.
35. Khatamsaz, D.; Peddareddygari, L.; Friedman, S.; Allaire, D. Bayesian Optimization of Multiobjective Functions Using Multiple Information Sources. *AIAA J.* **2021**, *59*, 1964–1974. [CrossRef]
36. Candelieri, A. Sequential Model Based Optimization of Partially Defined Functions under Unknown Constraints. *J. Glob. Optim.* **2019**, *59*, 1964–1974.
37. Bakshy, E.; Dworkin, L.; Karrer, B.; Kashin, K.; Letham, B.; Murthy, A.; Singh, S. AE: A Domain-Agnostic Platform for Adaptive Experimentation. In Proceedings of the Conference on Neural Information Processing Systems, Montréal, QC, Canada, 3–8 December 2018; pp. 1–8.
38. Letham, B.; Karrer, B.; Ottoni, G.; Bakshy, E. Constrained Bayesian Optimization with Noisy Experiments. *Bayesian Anal.* **2019**, *14*, 495–519. [CrossRef]