Lost in optimization of water distribution systems: better call Bayes

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

[1] University of Milano-Bicocca, Department of Economics, Management and Statistics, Italy

[2] University of Milano-Bicocca, Department of Computer Science, Systems and Communication, Italy

Abstract. The main goal of this paper is to show that Bayesian optimization could be regarded as a general framework for the data driven modelling and solution of problems arising in water distribution systems. Hydraulic simulation, both scenario based, and Monte Carlo is a key tool in modelling in water distribution systems. The related optimization problems fall in 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 and increasingly neural networks and an acquisition function which drives the search for new evaluation points. These modelling options make BO nonparametric, robust, flexible and sample efficient 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 the two problems: cost/energy optimization in pump scheduling and optimal sensor placement for early detection on 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 have been proposed for multi-objective detection problem using two different acquisition functions.

Keyword: Pump scheduling optimization; Bayesian optimization; Optimal sensor placement; Wasserstein distance; Robustness.

1 Introduction

Optimization problems arising in environmental modelling are usually very challenging. One reason is the presence of several objectives usually conflicting: instances are 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 have typically 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:

$$\min H(x) = \left(F_1(x), \dots, F_m(x)\right) \tag{1}$$

$x \in X \subset \mathbb{R}^d$ 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, p is finite, has a known analytical form and is inexpensive to evaluate. About $f_i(x, w)$ in Equation (2) and (3) we have no information as 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 (2) and (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.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 January 2022 doi:10.20944/preprints202201.0047.v1

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 and $f_i(x,w)$ can be the detection time of the contamination event or of the fake new 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 (2), if w is discrete, or (3) if w is continuous and 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 systems to maximize several performance metrics subject to stochastic patterns of water availability and demands.

Problem 2 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 shall not consider this problem, but it has some water specific applications for instance water demand forecasting (Candelieri, 2017; Shabani et al., 2018)

Even if the basic models (2) and (3) share some properties the computational strategies are different and, in this paper, we shall focus on the discrete case. The resulting combinatorial problems typically are 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 problems settings. In the following the main types of methods are described.

Metaheuristics in particular those which incorporate randomization and simulation and can be characterised as simheuristics (Calvet et al., 2016).

Multiobjective evolutionary algorithms (MOEA) which offer the advantages of simplicity, applicability, versatility, and robustness MOEAs can be naturally extended to multiobjective optimization either using a strategy of non-dominated sorting (NSGA) (Deb et al., 2000) which maintains an approximation of the Pareto set or a decomposition / strategy. (Zhang et al., 2009) which is 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 (Zhang et al., 2009) and more recently in (Liu et al., 2021).

Bayesian optimization (BO) methods (Archetti, F., & Candelieri, A. (2019). 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 morel recently neural networks. Multiobjective Bayesian Optimization (MOBO) has been recently the subject of intense research. (Rahat et al., 2017) analyses the 2 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 (Belakaria et al., 2020). 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 (Emmerich et al., 2016) and more recently have been given to multi-objective EI (Zhan et al., 2017) and LCB (Sun et al., 2021). An analysis of MOBO will be given in 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. Also the problem of black-box constraints can be handled in Bayesian optimization.

The second topic specifically addressed is the design of a sensor network for the detection of intrusion of contaminants in a WDN: it is shown that the optimal sensor placement in presence of different contamination scenarios can be formulated as a sample average approximation (SAA) multi-objective optimization problem and shown to be amenable to BO with a specific acquisition functions. based on hypervolume improvement. This is at the authors' knowledge the first case in which Hypervolume improvement has been used in the 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.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 January 2022 doi:10.20944/preprints202201.0047.v1

1.1 Related works

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

A recent survey (Mala-Jetmarova et al., 2017) under the evocative title of "Lost in optimisation of water distribution systems?" presents a wide ranging literature review. MOEAs algorithms have been widely used since the seminal paper of (Van Zyl et al., 2004). Castro-Gama et al. (2015) gives an application of NSGA-II for a large scale WDN. Bayesian optimization has been first considered in (Candelieri et al., 2018). The BO approach has been subsequently applied in (Candelieri et al., 2021), towards data efficient learning of implicit control strategies in Water Distribution Networks. A further step has been outlined in (Candelieri et al., 2021-April) where active learning of optimal controls for pump scheduling optimization is performed based only on SCADA data. The issue of crash constraints is particularly relevant in the optimization of WDS: the EPANET simulator in case of violation of some constraints does not yield any results. A specific method for this problem is proposed in (Tsai et al., 2018) where stochastic optimization for feasibility determination is applied to water pump operation in water distribution network. This approach has been generalized in (Candelieri, 2021) which addresses the issue in general terms modelling it as a classification/optimization problem.

The literature on Optimal sensor placement is extremely wide.

This is also a key problem in water research literature which has been studied in the last 2 decades following different approaches. Early contributions are (Guestrin et al., 2005) which models the optimal sensor placement using gaussian processes and (Ostfeld et al., 2008) which defined "The battle of the water sensor networks (BWSN)". An early application of BO for sensor set selection is (Garnett et al., 2010) in which the metric to be optimized is predictive accuracy of the air temperature across the UK. The use of BO for contaminant source localization, given information at each monitoring wells has been proposed in (Pirot et al., 2019). A related management problem is ground water remediation in which the pumping rates of "pump and treat" wells are fine-tuned using Bayesian optimization (Pourmohamad and Lee, 2021). Recent contributions to the optimal sensor placement are using several methodological approaches. He et al. (2018) proposes a multi-objective optimization method for water quality sensor placement allowing contamination prob-ability variations. Naserizade et al. (2018) proposes a risk-based multi-objective model for optimal placement of sensors. Zhang et al. (2020) proposes to assess the quality of a sensor placement strategies by a resilience index with respect to sensor failures. A new approach to optimal sensor has been proposed in (Ponti et al., 2021). The novelty of this approach is that a sensor placement is represented as a histogram which captures better than the sample average of all the information gathered during the simulation. The Wasserstein (WST) distance between histograms enables the design of a new algorithm called MOEA/WST which has shown a remarkable performance in solving the optimal sensor placement problem.

1.2 Our Contribution

- An analysis of the constraints in pump scheduling optimization focusing on the black-box case also when the objective function is undefined outside the feasible region.
- Hydraulic simulation can be embedded in a BO 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 the multi-objective case where an hyper-volume based acquisition function is shown to outperform the approach based on Chebyshev scalarization and expected improvement.

1.3 Organization of the paper

Sect. 2 outlines the basic notion of BO explaining the probabilistic model, the acquisition function and the basis of multi-objective optimization.

Sect. 3 outlines the features of BoTorch a Bayesian optimization library which will be used in the solution of the optimal sensor placement.

Sect. 4 contains the data and software resources used in this paper.

Sect. 5 contains two formulation of the problem of energy/cost optimization of pump scheduling 3 and the computational results on two networks.

Sect. 6 gives the multi-objective formulation of the optimal sensor placement problem. The two BO algorithms developed and the computational results.

Sect. 7 is devoted to conclusions and perspectives.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 January 2022 doi:10.20944/preprints202201.0047.v1

2 Bayesian Optimization

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

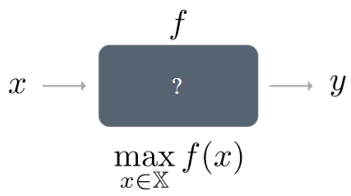

Figure 1. *The general scheme of a problem underlying BO.*

Although BO can work for any kind of optimization problems, its substantial computational overhead has been making it a de-facto standard in black-box situations (Figure 1) where the objective functions are expensive to evaluate, the analytical form of objectives and constraints and derivatives is available, and we can evaluate f at a sequence of points with the aim of determining a near optimal approximation after a small number of evaluations. which can be also noisy.

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

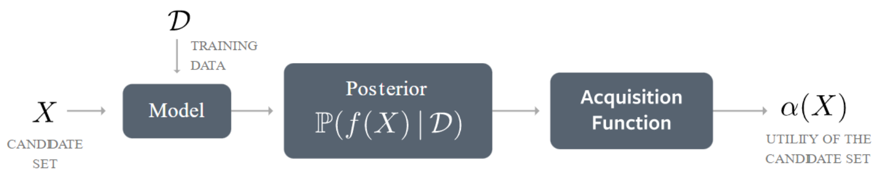

Figure 2. *The general scheme of a BO framework.*

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 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.

2.1 Probabilistic model

A GP is a random distribution over functions $f: \Omega \subset \Re^d \to \Re$ denoted with $f(x) \sim GP(\mu(x), k(x, x'))$ where $\mu(x) = \mathbb{E}(f(x)): \Omega \to \Re$ is the mean function of the GP and $k(x, x'): \Omega \times \Omega \to \Re$ is the kernel or covariance function. One way to interpret a GP is as a collection of correlated random variables, any finite number of which have a joint Gaussian distribution, so $f(x)$ can be considered as a sample drawn from a multivariate normal distribution (Williams and Rasmussen 2006). GPs provide probabilistic predictions by conditioning $\mu(x)$ and $\sigma^2(x)$ on a set of available data/observations.

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

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

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

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

The choice of the kernel establishes prior assumptions over the structural properties of the underlying (aka latent) function $f(x)$, specifically its smoothness. Moreover, 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 – considered in this paper – 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 to implement the trade-off between exploration and exploitation in the BO. In particular, any acquisition function aims to guide the search of the optimum towards points with potential high values of objective function either because the prediction of $f(x)$, based on $\mu(x)$ is high or the uncertainty is high (or both). While *exploiting* means to consider the area providing more chance to improve the current solution (with respect to the current surrogate model), *exploring* means acquiring new knowledge moving towards less explored regions of the search space. The literature on acquisition functions is very large: the reader is referred to (Archetti and Candelieri, 2019; Frazier, 2018).

The expected Improvement (EI) measures the expectation of the improvement on $f(x)$ with respect to the predictive distribution of the surrogate model. The parameter ξ is used to handle the trade-off between exploitation and exploration. When exploring, points associated to high uncertainty of the surrogate model are chosen, while when exploiting, points associated to high value of the mean of the surrogate model are selected. *ξ should be adjusted dynamically to decrease as the optimization moves along.*

$$EI(x) = \begin{cases}(\mu(x) - f(x^+) - \xi)\Phi(Z) + \sigma(x)\phi(Z) \; if \; \sigma(x) > 0 \\ 0 \; if \; \sigma(x) = 0\end{cases} \tag{6}$$

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

$$Z = \begin{cases}\frac{\mu(x) - f(x^+) - \xi}{\sigma(x)} \; if \; \sigma(x) > 0 \\ 0 \; if \; \sigma(x) = 0\end{cases} \tag{7}$$

The other acquisition function selects the next point to evaluate according to the Confidence Bound concept. Lower Confidence Bound – LCB – and Upper Confidence Bound – UCB – in the minimization and maximization case, respectively. For the case of minimization, LCB is given by:

$$\mathrm{LCB}(x) = \mu(x) - \mathrm{k}\sigma(x) \tag{8}$$

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

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. (Rahat et al., 2017) 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 which 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 (Paria et al., 2020). As far as the acquisition functions are concerned a survey of multi-objective versions is given in (Emmerich et al., 2016). Specific versions have been more recently given to multi-objective expected improvement (Zhan et al., 2017) and lower confidence bound (Sun et al., 2021).

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 4 points l_i with $i = 1, \ldots, 4$ non dominated in the objective space (the dominated space is in red), a new point x_1 dominates the previous ls, but l_1 bringing about an improvement of the dominate hypervolume (in blue, Figure 3b), x2 dominates l_1,l_2,l_3 and is not dominated by l_4 and brings about a further improvement (in green, Figure 3c).

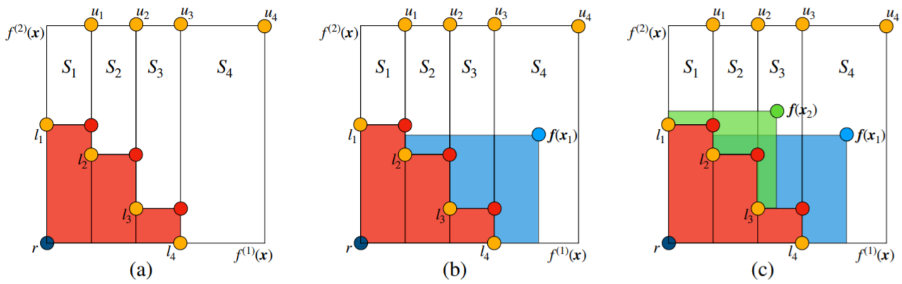

Figure 3. *Hypervolume improvement.*

Indeed, it can be shown that the maximization of the dominated hypervolume is equivalent to locating the optimal Pareto set and gives anyway approximate Pareto fronts with excellent coverage (Couckuyt et al., 2014).

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 (Yang et al., 2019; Balandat et al., 2020; Golovin and Zhang 2020) but its computation is still very expensive. Very recently the issue of MOBO has been analyzed also in connection with multi-fidelity and multiple information sources (Belakaria et al., 2020; Khatamsaz et al., 2021).

The situation is further complicated when constraints are also black-box or even undefined outside the feasible set. In this case MOEA algorithm are not easily made suitable for this kind of problem. A general solution is proposed in (Antonio, 2021).

2.4 BoTorch

Two libraries are used in this paper: mlrMBO for Pump Scheduling Optimization (Sect. 5) and Ax-BoTorch (Bakshy, 2018; Balandat, 2020) for the optimal sensor placement (Sect. 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 (Balandat et al., 2020) for a detailed exposition. BoTorch is particularly suitable to exploit 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 Sect. 2.1 and therefore a wider set of surrogate models. Beyond the acquisition functions EI and L/U confidence bounds in Sect. 2.2, BoTorch supports 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 supports also black-box constraints using variants of the expected improvement in which the improvement of the objective is weighted by the probability of feasibility (Letham et al., 2019). 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

Hanoi is a benchmark used in the literature whose associated graph consists in 32 nodes and 34 edges.

Figure 4. *The graph of Abbiategrasso network.*

Abbiategrasso (Figure 4) refers to a pressure management zone in Milan (namely, Abbiategrasso) with an associated graph consisting of 1213 nodes and 1391 edges, analyzed in the European project Icewater (Candelieri et al., 2015).

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

The case study used for learning the pressure control rules is a WDN in the larger- Milano, Italy, which supplies water to three municipalities: Bresso (around 20000 inhabitants), Cormano (around 26000 inhabitants) and Cusano-Milanino (around 19000 inhabitants). The network consists of 7418 pipes, 8493 junctions, 14 reservoirs, 1381 valves, 9 pumping stations with 14 pumps overall. Piezometric level of the WDN ranges in 136 to 174 meters (average: 148 meter). A schematic representation of the WDN is reported in Figure 5.

3.2 Software resources

In order to implement the BO framework proposed for solving the pump scheduling optimization problem, the R package named “mlrMBO” has been adopted: it is a flexible and comprehensive toolbox for model-based optimization (MBO). This toolbox has been designed for both single- and multi-objective optimization with mixed continuous, categorical and conditional decision variables, and infill criteria and infill optimizers are easily exchangeable.

BoTorch is a library for Bayesian Optimization built on top of PyTorch and is part of the PyTorch ecosystem. BoTorch is supported by an extensive documentation. BoTorch is best used in tandem with Ax, Facebook's open-source adaptive experimentation platform, which provides an easy-to-use interface for defining, managing, and running sequential experiments. BoTorch provides a modular and easily extensible interface for composing Bayesian optimization primitives, including probabilistic models, acquisition functions, and optimizers. BoTorch organizes the computations in batches of size q. BoTorch provides first-class support for state-of-the art probabilistic models in GPyTorch, a library for efficient and scalable GPs implemented in PyTorch

As hydraulic simulation software we have used WINTR, a recent Python-based open-source based on EPANET 2.0, the most widely adopted hydraulic simulation software for pressurized WDNs.

4 Pump scheduling optimization

Two approaches are possible in optimal pump management: explicit and implicit control. The first assumes that decision variables are pump statuses/speeds to be set up at prefixed time, and it is usually also known as Pump Scheduling Optimization (PSO). Thus, the problem is to efficiently search among all the possible schedules (i.e., configurations of the decision variables) to optimize the objective function – typically minimization of the energy-related costs – while satisfying hydraulic feasibility. A plethora of methods have been proposed such as mathematical programming, meta-heuristics, evolutionary and nature-inspired algorithms. However, explicit control typically implies many decision variables for real-world water distribution networks, increasing with the number of pumps and the number of time intervals for actuating the control. The resulting high dimensionality of the search space consequently im-plies to evaluate a huge number of possible solutions (i.e., schedules), each one requiring performing a hydraulic simulation run.

On the contrary, implicit control aims at controlling pump status/speeds depending on some control rules related, for instance, to pressure into the network: 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, status/speed of the pump is not modified. These thresholds become the decision variables of the optimal implicit control problem. Compared to explicit control, implicit control approaches allow to significantly reduce the number of decision variables, at the cost of making more complex the search space, due to the introduction of further constraints and conditions among decision variables (e.g., the pressure value implying to switching a pump off cannot be lower than the pressure value implying to switching the same pump on).

4.1 Pump Scheduling Optimization through Bayesian Optimization

The reference paper is (Candelieri et al., 2018) which proposes a BO approach for the PSO problem by also comparing two different probabilistic surrogate models to approximate the black-box objective function, that 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 not-computability, meaning that some hydraulic constraints has been violated leading to the impossibility to compute the objective function associated to a hydraulic feasible schedule. More precisely, hydraulic feasibility refers to "basic" computational constraints (e.g., the impossibility to have a negative pressure at some location, lack of convergence in the simulation, impossibility to satisfy water demand, etc.) as well as operational constraints which 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 to each node according to historical data.

Although "penalizing" infeasible (aka not-computable) schedule proved to be a sufficiently workaround (Candelieri et al., 2018), successively the estimation of the unknown feasibility region has been successively addressed in (Tsai et al., 2018) in the case of PSO, and in (Candelieri, 2021) in the case of a generic problem with a potentially not-computable black-box objective function.

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

4.2 Learning optimal control rules as a black-box optimization problem

Implicit control has been less explored and when combined with BO it can successfully unlock data efficiency in learning optimal control strategies, as empirically proved in (Candelieri et al., 2021).

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 a H-dimensional space. In contrast, the search space associated to implicit control, for instance consisting of min-max ranges on 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.

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 associated to a pump is based on two

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 January 2022 doi:10.20944/preprints202201.0047.v1

different thresholds, τ1 and τ2, all day long, respectively the lowest threshold implying to switch the pump on and the highest threshold implying to switch the pump off:

IF ($p<\tau_1$ AND S=*OFF*) THEN $S \leftarrow$ *ON*

ELSE

IF ($p>\tau_2$ AND S=*ON*) THEN $S \leftarrow$ *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.

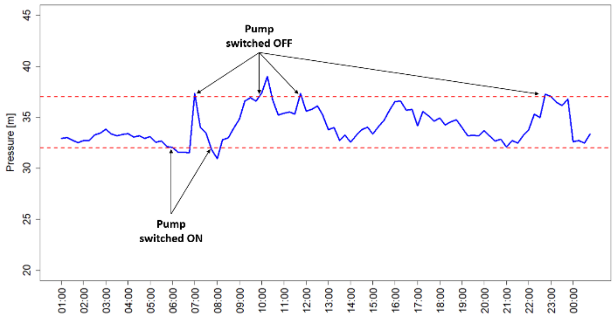

Figure 6. *A schematic representation of and implicit pump control strategy based on min-max thresholds (red dotted lines) on the pressure value (in blue).*

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

Also in this case, the possibility to generate infeasible (aka not-computable) control strategies was taken into account, leading to a simulation-optimization problem having a black-box objective function and subject to unknown constraints. Computational results on a real-life WDN also showed that implicit control strategies learned through the proposed BO approach were not only optimal and efficient in terms on number of simulation runs required to identify them, but also “hydraulically” robust with respect to perturbations in the water demand.

One further step has been taken in the paper “Active learning of pressure control rules of water distribution networks” EGU’21 in which it is shown that an accurate approximation of the outputs of the hydraulic simulation, like EPANET, can be obtained training a deep neural network on SCADA data. whose effectiveness and feasibility have been tested on a real-life mid-scale WDN.

The main result is 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” of the real-world water distribution network.

Another result is that thresholds-based rules for implicit control can be learned through an active learning approaches well-suited for the implicit control setting: the lower dimensionality of the search space, compared to explicit control, substantially improves computational efficiency.

5 Optimal sensor placement

5.1 The basic model

We consider a graph $G = (V, E)$ We assume a set of possible locations for placing p sensors, that is $L \subseteq V$. Thus, a sensor placement (SP) is a subset of sensor locations, with the subset’s size less or equal to p depending on the available budget. An SP is 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. Thus, an SP is given by the nonzero components of s.

For a WDN the vertices in V represent junctions, tanks, reservoirs or consumption points, and edges in E represent pipes, pumps, and valves.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 January 2022 doi:10.20944/preprints202201.0047.v1

Let $A \subseteq V$ denote the set of contamination events $a \in A$ which must be detected by a sensor placement s, and d_{ai} is the detection time of a sensor placed in node I for a contamination event in node a.

A probability distribution is placed over possible contamination events associated to the nodes. In the computations we assume – as usual in the literature – a uniform distribution, but in general discrete distributions are also possible. In this paper we consider as objective functions the detection time and its standard deviation.

We consider a general model of sensor placement,

$$P = \begin{cases} \min f_1(s) = \sum_{a \in A} \alpha_a \sum_{i=1,..,|L|} d_{ai} x_{ai} \\ s.t. \\ \sum_{i=1,..,|L|} s_i \leq p \\ s_i \in \{0,1\} \end{cases} \quad (9)$$

- α_a is the probability for the contaminant to enter the network at node a.
- $x_{ai} = 1$ if $s_i = 1$, where i is the first sensor detecting the contaminant injected at node a; 0 otherwise.

In our study we assume that all the events have the same chance of happening, that is $\alpha_a = 1/|A|$, therefore $f_1(s)$ is:

$$f_1(s) = \frac{1}{|A|} \sum_{a \in A} \hat{t}_a \quad (10)$$

where $\hat{t}_a = \sum_{i=1,..,|L|} d_{ai} x_{ai}$ is the MDT of event a.

f_1 is the average over the contamination events of the detection time for each event. 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 $\hat{t}_a$ the minimum time step at which concentration reaches or exceeds a given threshold τ for the event a.

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

$$f_2(s) = STD_{f_1}(s) = \sqrt{\frac{1}{|A|} \sum_{a \in A} \left(\hat{t}_a - f_1(s)\right)^2} \quad (11)$$

5.2 Multi-Objective Bayesian Optimization for sensor placement

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

A1 is 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) as acquisition function.

A2 is multi-surrogate based on GP models of the single objectives and the Expected Hypervolume Improvement as acquisition function).

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

The "helper" function creates the outcomes required for the scalarized objective and applies the scalarization and the constraints.

The helper function initializes A1 and A2 and returns the batch $(x_1, x_2, \ldots, x_q)$ along with the observed values. In A1 each candidate is optimized in a sequential greedy fashion using a different random scalarization.

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

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

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

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 January 2022 doi:10.20944/preprints202201.0047.v1

In A2 a list of acquisition functions are created each with different scalarization weights. This list generates one candidate for each acquisition and conditions 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 of the hypervolume between the feasible true Pareto front and hypervolume of the observed Pareto front for algorithms A1 and A2.

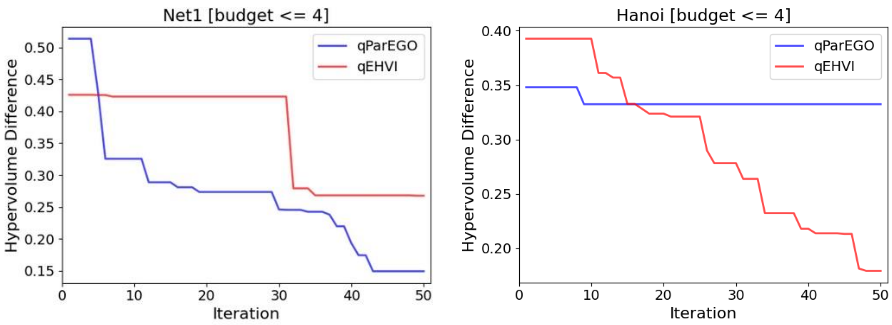

Figure 7. *Difference in hypervolume between the optimal Pareto front and the Pareto front approximated by qParEgo (blue) and qEHVI (red) over the iteration.*

It is apparent that A2 relative performance improves with the size of the problem.

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

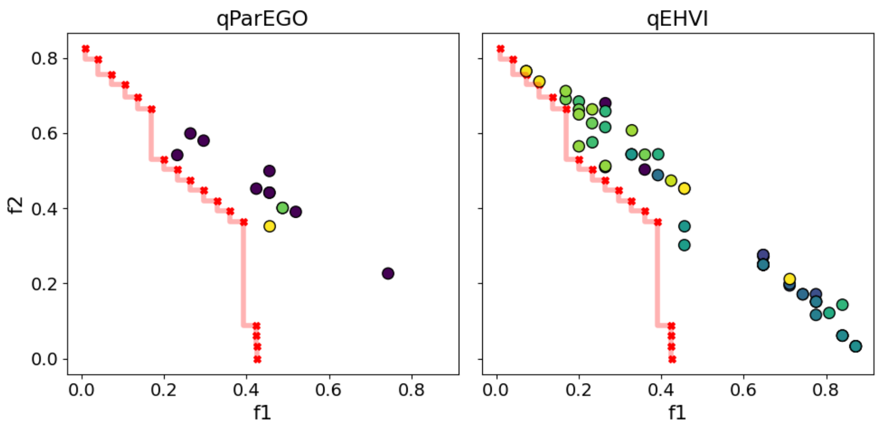

Figure 8. *Hanoi network. True Pareto front in red. Lighter colours denote later iterations.*

6 Conclusions

The main conclusion is that Bayesian Optimization offers a 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 evaluation of the objective functions are 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 which drives the selection of the next evaluation point. It is shown how this feature can be exploited in multi-objective problem incorporating in the acquisition function the concept of hypervolume improvement. Evolutionary algos are often used for these problems, also in the multi-objective case, but they are not nearly as sample efficient as BO, albeit their performance can be improved incorporating a surrogate model. The sample efficiency is 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 the paper, allow a quick development of a first prototype requiring a negligible amount of coding and allowing both the data scientist and the domain expert to choose those components (in particular probabilistic models and acquisition functions) more befitting the target application and the available resources.

References

1. Antonio, C. (2021). Sequential model based optimization of partially defined functions under unknown constraints. Journal of Global Optimization, 79(2), 281-303.
2. Archetti, F., & Betro, B. (1979). A probabilistic algorithm for global optimization. Calcolo, 16(3), 335-343.
3. Balandat, M., Karrer, B., Jiang, D. R., Daulton, S., Letham, B., Wilson, A. G., & Bakshy, E. (2020, January). BoTorch: A Framework for Efficient Monte-Carlo Bayesian Optimization. In NeurIPS.
4. Belakaria, S., Deshwal, A., Jayakodi, N. K., & Doppa, J. R. (2020, April). Uncertainty-aware search framework for multi-objective Bayesian optimization. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 34, No. 06, pp. 10044-10052).
5. Candelieri 2021. Sequential model based optimization of partially defined functions under unknown constraints. Journal of Global Optimization, 79(2), 281-303.
6. Candelieri, A., Giordani, I., & Archetti, F. (2017, June). Automatic configuration of Kernel-based clustering: an optimization approach. In International Conference on Learning and Intelligent Optimization (pp. 34-49). Springer, Cham.
7. Candelieri, A., Ponti, A., & Archetti, F. (2021, August). Data efficient learning of implicit control strategies in Water Distribution Networks. In 2021 IEEE 17th International Conference on Automation Science and Engineering (CASE) (pp. 1812-1816). IEEE.
8. Candelieri, A.; Soldi, D.; Archetti, F. Cost-Effective Sensors Placement and Leak Localization–the Neptun Pilot of the ICeWater Project. J. Water Supply Res. Technol. AQUA 2015, 64, 567–582.
9. Castro Gama, M. E., Pan, Q., Salman, S., & Jonoski, A. (2015). Multivariate optimization to decrease total energy consumption in the water supply system of Abbiategrasso (Milan, Italy). Environmental Engineering & Management Journal (EEMJ), 14(9).
10. Couckuyt, I., Deschrijver, D., & Dhaene, T. (2014). Fast calculation of multiobjective probability of improvement and expected improvement criteria for Pareto optimization. Journal of Global Optimization, 60(3), 575-594.
11. Golovin, D., & Zhang, Q. (2020). Random hypervolume scalarizations for provable multi-objective black box optimization. arXiv preprint arXiv:2006.04655.
12. Guestrin, C., Krause, A., & Singh, A. P. (2005, August). Near-optimal sensor placements in gaussian processes. In Proceedings of the 22nd international conference on Machine learning (pp. 265-272).
13. He, G., Zhang, T., Zheng, F., & Zhang, Q. (2018). An efficient multi-objective optimization method for water quality sensor placement within water distribution systems considering contamination probability variations. Water research, 143, 165-175.
14. Liu, F., Zhang, Q., & Han, Z. (2021, March). MOEA/D with Gradient-Enhanced Kriging for Expensive Multiobjective Optimization. In International Conference on Evolutionary Multi-Criterion Optimization (pp. 543-554). Springer, Cham.
15. Mala-Jetmarova, H., Sultanova, N., & Savic, D. (2017). Lost in optimisation of water distribution systems? A literature review of system opera-tion. Environmental modelling & software, 93, 209-254.
16. Naserizade, S. S., Nikoo, M. R., & Montaseri, H. (2018). A risk-based multi-objective model for optimal placement of sensors in water distribution system. Journal of hydrology, 557, 147-159.
17. Ostfeld, A., Uber, J. G., Salomons, E., Berry, J. W., Hart, W. E., Phillips, C. A.,... & Walski, T. (2008). The battle of the water sensor networks (BWSN): A design challenge for engineers and algorithms. Journal of Water Resources Planning and Management, 134(6), 556-568.
18. Paria, B., Kandasamy, K., & Póczos, B. (2020, August). A flexible framework for multi-objective bayesian optimization using random scalarizations. In Uncertainty in Artificial Intelligence (pp. 766-776). PMLR.
19. Ponti, A., Candelieri, A., & Archetti, F. (2021). A Wasserstein distance based multiobjective evolutionary algorithm for the risk aware optimization of sensor placement. Intelligent Systems with Applications, 10, 200047.
20. Shahriari, B., Swersky, K., Wang, Z., Adams, R. P., & de Freitas, N. (2016). Taking the Human Out of the Loop: A Review of Bayesian Optimization. Proceedings of the IEEE, 104(1).

 doi:10.20944/preprints202201.0047.v1

21. Tsai, Y. A., Pedrielli, G., Mathesen, L., Zabinsky, Z. B., Huang, H., Candelieri, A., & Perego, R. (2018, December). Stochastic optimization for feasibility determination: an application to water pump operation in water distribution network. In 2018 Winter Simulation Conference (WSC) (pp. 1945-1956). IEEE.
22. Van Zyl, J. E., Savic, D. A., & Walters, G. A. (2004). Operational optimization of water distribution systems using a hybrid genetic algorithm. Journal of water resources planning and management, 130(2), 160-170.
23. Yang, K., Emmerich, M., Deutz, A., & Bäck, T. (2019). Multi-objective Bayesian global optimization using expected hypervolume improvement gradient. Swarm and evolutionary computation, 44, 945-956.
24. Zhang, Q., Liu, W., Tsang, E., & Virginas, B. (2009). Expensive multiobjective optimization by MOEA/D with Gaussian process model. IEEE Transactions on Evolutionary Computation, 14(3), 456-474.
25. Zhang, Q., Zheng, F., Kapelan, Z., Savic, D., He, G., & Ma, Y. (2020). Assessing the global resilience of water quality sensor placement strategies within water distribution systems. Water research, 172, 115527.