







Learning Optimal Control of Water Distribution Networks Through Sequential Model-Based Optimization

Antonio Candelieri¹(✉) , Bruno Galuzzi² , Ilaria Giordani² ,
and Francesco Archetti² 

¹ Department of Economics, Management and Statistics,
University of Milano-Bicocca, 20126 Milan, Italy
antonio.candelieri@unimib.it

² Department of Computer Science Systems and Communication,
University of Milano-Bicocca, 20126 Milan, Italy

Abstract. Sequential Model-based Bayesian Optimization has been successfully applied to several application domains, characterized by complex search spaces, such as Automated Machine Learning and Neural Architecture Search. This paper focuses on optimal control problems, proposing a Sequential Model-based Bayesian Optimization framework to learn optimal control strategies. The strategies are synthesized by pressure-based rules, whose parameters are the design variables of the optimization problem whose black-box objective is the energy cost. A Bayesian optimization framework is presented which handles a quite general formalization of the control problem including multiple constraints, also black box. Relevant results on a real-life Water Distribution Network are reported, comparing different possible choices for the proposed framework.

Keywords: Sequential Model-based Bayesian optimization · Optimal control · Water distribution networks

1 Introduction

Sequential Model-based Bayesian Optimization (SMBO) is a sample-efficient strategy for global optimization (GO) of black-box, expensive and multi-extremal functions [1–4], where the solution of the problem is traditionally constrained to over a box-bounded search space X :

$$\min_{x \in X \subset \mathbb{R}^d} f(x) \quad (1)$$

SMBO has been successfully applied in several domains, ranging from design problems (new materials, drugs, software, structural design) to robotics, control and finance (a brief overview about application domains is provided in Chap. 7 of [5]).

In the Machine Learning (ML) community, it recently became the standard strategy for Automated Machine Learning (AutoML) [6] and Neural Architecture Search

(NAS) [7], which are usually characterized by a search space being more complex than a box-bounded domain. More precisely, x can consist of mixed (continuous, integer, categorical) components as well as “conditional”, where conditional means that the value of a component x_i depends on the value of another component x_j , with $i \neq j$. An example of complex search space in AutoML is related to the optimization of ML pipelines, such as the one presented in [8].

Starting from the SMBO advances in the ML domain, we consider in this paper an optimal control problem sharing many characteristics with AutoML. More precisely, we addressed the optimal definition of control rules regulating the ON/OFF switching of pumps in a Water Distribution Network (WDN). The objective is the minimization of the energy-related costs while guaranteeing the supply of the water demand.

We remark that the problem is black-box because the evaluation of both the objective function and the constraints is based on a hydraulic simulator: moreover the constraints on decision variables make the search-space complex (i.e., analogously to AutoML, decision variables are both mixed – numeric and discrete – and possibly conditional).

The rest of the paper is organized as follows: in Sect. 2, the methodological background about SMBO and optimization of operations in WDNs is presented. Section 3 provides the mathematical formulation of the optimal control problems, along with the proposed solution. Section 4 defines the experimental setting and Sect. 5 summarizes the results obtained. Finally, conclusions and discussion on advantages and limitations of the proposed approach are provided.

2 Background

2.1 Sequential Model-Based Bayesian Optimization

To solve problem (1), SMBO uses two key components: a *probabilistic surrogate model* of $f(x)$, sequentially updated with respect to new function evaluations, and an *acquisition function* (aka *infill criterion* or *utility function*), driving the choice of the next promising point x where to evaluate $f(x)$ while dealing with the exploitation-exploration dilemma. A typical choice for the probabilistic surrogate model is a Gaussian Process (GP) [9] (in this case, SMBO is also known as GP-based optimization or Bayesian Optimization [5, 10]). An alternative probabilistic surrogate model is a Random Forest (RF) [11], an ensemble learning approach which, by construction, can deal with mixed and conditional components of x , making RFs more well-suited than GPs to solve problems with these characteristics.

The probabilistic surrogate model – whichever it is – should provide an estimate of $f(x)$ along with a measure of uncertainty about such an estimate, with $x \in X$. These two elements are usually the mean and standard deviation of the prediction provided by the probabilistic surrogate model, denoted by $\mu(x)$ and $\sigma(x)$, respectively.

With respect to the acquisition function, several alternatives have been proposed, implementing different mechanisms to balance exploitation and exploration (i.e., $\mu(x)$ and $\sigma(x)$, respectively) [5, 10]. In this paper we focused on a subset of acquisition functions, reported in the experimental setting section.

Due to the sequential nature of SMBO, at a generic iteration n we can denote the set of function evaluations performed so far by $D_{1:n} = \{(x^{(i)}, y^{(i)})\}_{i=1, \dots, n}$, where $y^{(i)} = f(x^{(i)}) + \varepsilon$, and $\varepsilon \sim \mathcal{N}(\mu_\varepsilon, \sigma_\varepsilon)$ in the case of a noisy objective function.

The probabilistic surrogate model is learned at every iteration, providing the updated $\mu^{(n)}(x)$ and $\sigma^{(n)}(x)$. Then, $x^{(n+1)}$, is chosen by solving the auxiliary problem:

$$x^{(n+1)} = \operatorname{argmax}_{x \in X \subset \mathbb{R}^d} \alpha^{(n)}(x) \quad (2)$$

where $\alpha^{(n)}(x)$ is the acquisition function, typically $\alpha^{(n)}(x, \mu^{(n)}(x), \sigma^{(n)}(x))$. This auxiliary problem is usually less expensive than the original one, and can be solved by gradient-based methods (e.g., L-BFGS) – in the case that the analytical form of $\mu^{(n)}(x)$ and $\sigma^{(n)}(x)$ is given (i.e., when a GP is used as probabilistic surrogate model) – or GO approaches (e.g., DIRECT, Random Search and its recent variants, evolutionary meta-heuristics, etc.) – in the case that $\mu^{(n)}(x)$ and $\sigma^{(n)}(x)$ are also black-box (i.e., when a RF is used as a probabilistic surrogate model).

Then, the objective function is evaluated at $x^{(n+1)}$, leading to the observation of $y^{(n+1)}$ and the update $D_{1:n+1} = D_{1:n} \cup \{(x^{(n+1)}, y^{(n+1)})\}$. The process is iterated until some termination criterion is achieved, such as a maximum number of function evaluations has been performed.

2.2 Constrained SMBO

Real life optimization problems have most often constraints making the search space more complex than simply box-bounded [10] and the “vanilla” SMBO not well suited for solving them. In constrained SMBO, the problem (1) can be rewritten as:

$$\begin{aligned} \min_{x \in X \subset \mathbb{R}^d} f(x) \\ g_i(x) \leq 0 \quad i = 1, \dots, n_g \end{aligned} \quad (3)$$

Solving approaches can be categorized depending on the nature of the constraints: they can be known a-priori and given in analytical form or, on the contrary, they are unknown (aka hidden, black-box). With respect to the first case, several approaches have been proposed in the GO community [12–14], while the second case is more related to simulation-optimization and AutoML [15–19].

A further consideration, with respect to unknown constraints, is that the objective function could be not computable in association with the violation of one or more constraint, leading to the global optimization of partially defined functions [20–22]. Recently, a two-stage approach has been proposed in [23], using Support Vector Machine (SVM) to estimate the portion of the box-bounded search space where the objective function is defined (aka computable), depending on a set of unknown constraints. In the second stage a constrained Bayesian Optimization task is performed on the estimated feasible region. This paper makes use of this approach.

3 Problem Definition and Solution Approach

3.1 Optimization of Operations in Water Distribution Networks

Optimization of WDNs' operations has been an active research field in the last decades. Optimal pump operation, aimed to minimize energy related costs due to pumping water, has been one of the most relevant topics. A systematic review on WDNs' operations optimization has been recently provided in [24], where approaches for optimal pumps management are categorized into: (i) *explicit control* of pumps by times to operate and (ii) *implicit control* by pumps' pressures, flows or speeds, or tanks levels. Although *explicit control* solutions were the most frequently adopted, the optimization problem (also known as Pump Scheduling Optimization, PSO) could be characterized by a huge number of decision variables in the case that the WDN has many pumps and/or times to operate (e.g., decisions about pump activation every hour on a daily horizon).

Most of the *explicit control* solutions proposed use meta-heuristics, mainly evolutionary strategies, such as in [25–27]. However, contrary to SMBO, these strategies are not sample efficient, requiring a huge number of hydraulic simulation runs to identify an optimal pump schedule. More recently, an SMBO approach to PSO has been initially proposed in [28] and then extended in [18] to include unknown constraints on the hydraulic feasibility of the pump schedules proposed by SMBO.

On the other hand, *implicit control* strategies allow reduction of the number of decision variables, but make more complex the search space, due to the introduction of further constraints on and conditions among decision variables. Another important advantage offered by implicit control solutions is that they do not require specification of times to operate; they usually work by applying simple (control) rules depending on the values of collected measurements. Thus, time to operate is given by the data acquisition rate instead of prefixed timestamps as in *explicit control* solutions.

3.2 Learning Optimal Control Rules as a Black-Box Optimization Problem

We consider the case of an *implicit control* solution, where pumps are controlled depending on the associated pressure values. In the simplest case, control for a given pump is defined by two different thresholds, x_1 and x_2 , and the following rule:

IF (pump's pressure $< x_1$ AND pump is OFF) THEN pump is switched ON
 ELSE
 IF (pump's pressure $> x_2$ AND pump is ON) THEN pump is switched OFF

This means that the pump is activated if its pressure is lower than a minimum threshold, x_1 , it is deactivated if its pressure exceeds a maximum threshold, x_2 , and remains in the current status (ON/OFF) otherwise. Clearly, x_1 and x_2 are the decision variables to optimize with respect to the minimization of energy cost, constrained to water demand satisfaction. A graphical representation of this kind of simple control for a single pump is reported in Fig. 1.

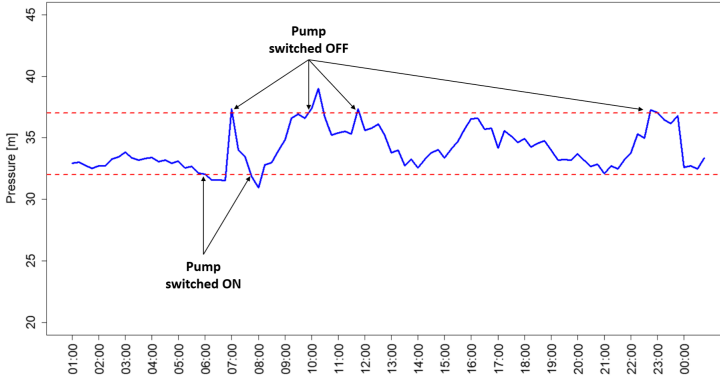


Fig. 1. A schematic representation of implicit pump control based on thresholds (red dotted lines) on pressure (in blue). If pressure goes below/over the lower/upper threshold the pump is switched ON/OFF, respectively. Pressure value could not change immediately after the pump switch because it also depends on the status of the other pumps in the WDN (Color figure online)

It is important to highlight that both energy costs and water demand satisfaction (as well as any other relevant constraints related to the hydraulic behavior of the WDN, such as min/max tanks levels) are black-box, because they can be only evaluated after having fixed the values of the decision variables. Moreover, an analytical constraint must be added, modelling that the minimum threshold x_1 cannot be greater or equal than the maximum one x_2 . As follows, we define the optimization problem in the more general case consisting of more than a pair of thresholds. This situation is quite common in real-life WDNs, having more than one pump and/or requiring different control thresholds over the day (e.g., during morning and evening) for a given pump:

$$\begin{aligned}
 & \min_{x \in X \subset \mathbb{R}^d} f(x) \\
 & x_i \in S_i \quad i = 1, \dots, 2\tau \quad (c_1) \\
 & x_j - x_{j+\tau} \leq 0 \quad j = 1, \dots, \tau \quad (c_2) \\
 & g(x) = 0 \quad (c_3)
 \end{aligned} \tag{4}$$

where $f(x)$ is the energy cost associated to the control rule defined by the x_i value, S_i is the set of possible values for the thresholds, $S_i = \{s_{i1}, \dots, s_{iN_j}\}$, τ is the number of thresholds pairs to be set up (leading to $d = 2\tau$) and $g(x)$ is related to the hydraulic feasibility: it is unknown/black-box and makes $f(x)$ partially defined. Thus, both $f(x)$ and $g(x)$ are black-box and are computed via hydraulic software simulation, typically over a simulation horizon of a day. The open-source EPANET 2.0 is the most widely adopted tool for simulating the hydraulic behavior of a pressurized WDN, so that the search for the optimal values of the control thresholds is sequentially performed on the software model of the WDN. A single simulation run, referred to a specific set up of the thresholds, involve computational costs; SMBO is a sample-efficient strategy to identify an optimal set up within few simulation runs (i.e., function evaluations).

Thresholds are modelled as discrete variables to consider the resolution of the monitoring sensors (i.e., in the case study analyzed in this paper, measurements are acquired with a resolution of 0.5[m]). In the case of continuous variables, constraint c_1 turns into $x_i \in [\min S_i, \max S_i]$. According to (4), the optimal definition of an implicit control strategy, based on pressure values, shares common characteristics with AutoML: decision variables are discrete (c_1) and conditional (c_2) – such as many Machine and Deep Learning algorithms’ hyperparameters – and $g(x)$ is black-box – such as a constraint on resources (i.e., memory usage) for a trained Machine/Deep Learning algorithm.

4 Experimental Setting

4.1 Case Study Description

The case study considered in this paper refers to a WDN in Milan, Italy, supplying water to three different municipalities: Bresso (around 20’000 inhabitants), Cormano (around 26’000 inhabitants) and Cusano-Milanino (around 19’000 inhabitants). The overall WDN 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 m (average: 148 m). Moreover, this WDN is also interconnected with the WDNs of other three municipalities (namely, Paterno Dugnano, Sesto San Giovanni and Cinisello Balsamo). The hydraulic software models of these further municipalities were not available, the hydraulic behavior at the interconnections was modelled through three reservoirs with levels varying over time according to historical data about the flow from the WDN to the other three municipalities and vice versa (Fig. 2).

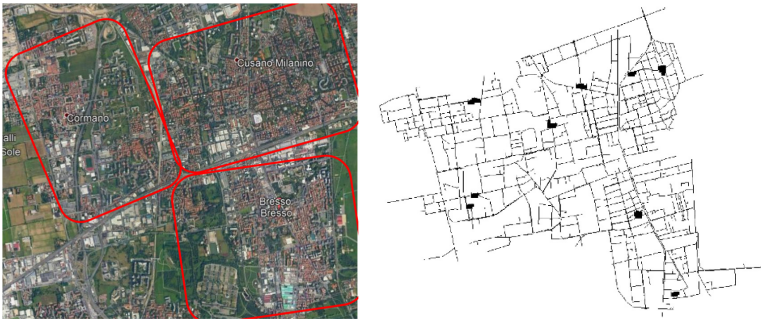


Fig. 2. The three municipalities considered in the study (on the left) and the hydraulic software model, developed in EPANET, of the associated WDN (on the right)

4.2 SMBO Setting

In this section we provide all the details about the setting of our experiments, organized in two different sub-sections. The first one provides all the details about the SMBO

process applied to problem (4). In the second, we decided to relax the constraint related to the discreteness of the decision variables. The aim is to evaluate which could be the difference between the optimal solution identified for the problem (4) and a more optimistic one, in the hypothetical case that the numerical precision of the actual control could be finer. In both cases, RF have been used as probabilistic surrogate model due to the presence of conditional decision variables.

Since the actual global optimizer x^* is unknown, we cannot use performance measures such as regret [29] or Gap metrics [30], but just looking at the best value observed over SMBO iterations, the so called “best seen”:

$$y^{+(n)} = \min_{i=1,\dots,n} \left\{ f(x^{(1)}), \dots, f(x^{(n)}) \right\}$$

Finally, it is important to highlight that, in this study, we have evaluated all the control strategies identified through SMBO by simulating them over the same “test day”. This means that we have considered an unnoisy setting, so $y^n = f(x^{(n)})$, for every $n = 1, \dots, N$ and with N the maximum number of function evaluations.

4.2.1 RF-Based SMBO

As mentioned in Sect. 4.1, the WDN has 14 pumps, overall. However, 4 are only used to support supply during peak-hours. They are controlled by time and will not be part of the optimization. With respect to the other 10 pumps, 8 of them requires the identification of optimal control thresholds which can be different during the day (i.e., 06:00–23:00) and the night (i.e., 23:00–06:00). This means that we have optimized 2 thresholds for 2 pumps and 4 thresholds for 8 pumps, leading to 36 decision variables overall (i.e., thresholds x_i) for the problem (4) – that is $\tau = 18$. It is important to highlight that the number of decision variables is significantly higher in the case of *explicit control*: the optimization of hourly-based schedules on the same case study would require 240 decision variables (that is 10 pumps time 24 h).

The possible discrete values for all the *lower* thresholds, that are the sets $S_{i=1,\dots,\tau}$, range from 21[m] to 32[m], with a step of 0.5[m] (i.e., 23 possible values). The possible discrete values for all the *upper* thresholds, that are the sets $S_{i=\tau+1,\dots,2\tau}$, range from 26[m] to 44[m], with a step of 0.5[m] (i.e., 23 possible values). These two sets instantiate the constraint (c_1) of the problem (4). Due to the nature of the problem, involving discrete and conditional decision variables, the most suitable probabilistic surrogate model is a RF. Initialization of the probabilistic surrogate model was performed by randomly sampling 10 initial vectors of control thresholds (“initial design”). More precisely, a Latin Hypercube Sampling (LHS) procedure has been applied. Remaining budget (i.e., function evaluations) has been set to 200.

We decided to compare three different acquisition functions, namely Lower Confidence Bound (LCB), Expected Improvement (EI) [5, 10] and Augmented Expected Improvement (AEI) [31] – the last usually replaces EI in the noisy setting. Although in this paper we solve the case study deterministically (i.e., in the noise-free setting), we have decided to include AEI just to evaluate how much the assumption of working in a noisy setting – while the problem is noise-free – could affect the final solution. We plan

to use AEI to extend our approach to the noisy setting by considering the water demand as a random variable, instead of known (or predicted) a priori.

$$\begin{aligned}
 LCB(x) &= \mu(x)^{(n)} - \beta^{(n)} \sigma(x)^{(n)} \\
 EI(x) &= \begin{cases} \left(y^+ - \mu(x)^{(n)}\right) \Phi(Z) + \sigma(x)^{(n)} \phi(Z) & \text{if } \sigma(x)^{(n)} > 0 \\ 0 & \text{otherwise} \end{cases} \\
 AEI(x) &= \begin{cases} \left(y^+ - \mu(x)^{(n)}\right) \Phi(Z) + \sigma(x)^{(n)} \phi(Z) \left(1 - \frac{\sigma_\varepsilon}{\sqrt{\sigma_\varepsilon^2 + (\sigma(x)^{(n)})^2}}\right) & \text{if } \sigma(x)^{(n)} > 0 \\ 0 & \text{otherwise} \end{cases}
 \end{aligned}$$

where $\beta^{(n)}$ manages the exploitation-exploration trade-off, $Z = \frac{y^+ - \mu(x)^{(n)}}{\sigma(x)^{(n)}}$ and y^+ is the best seen up to n . Then, $x^{(n+1)}$ is selected by minimizing LCB and maximizing EI and AEI.

Since we are using a RF as probabilistic surrogate model, the acquisition functions are also black-box. A global-local method has been used to solve the auxiliary problem (2) and identify the next promising $x^{(n+1)}$. More precisely, the global-local method used is known as “focus-search” [32]: it can handle with numeric, discrete and mixed search spaces, also involving conditional variables. Other approaches, also recent evolutionary methods such as Covariance Matrix Adaptation Evolution Strategy (CMA-ES) [33] are suitable for dense spaces, but the conditionality of the search spaces makes their usage still problematic. Focus-search is an adaptive Random Search strategy: it starts from a large set of random points where the acquisition function is evaluated. Then, it shrinks the search space around the current best point and perform a new random sampling of points within the “focused space”. The shrinkage operation is iteratively performed until a maximum number of iterations and the entire procedure can be restarted multiple times to mitigate the risk to converge to a local optimum. Finally, the best point over all restarts and iterations is returned as the solution of the auxiliary problem (2). The R package mlrMBO [32] provides an implementation of focus-search.

Although different acquisition functions have been used, all the associated SMBO processes started from the same initial design. Furthermore, to mitigate the effect of randomness, we have performed 20 experiments with 20 different initial designs.

4.2.2 RF Based SMBO with Relaxation of the Discreteness Constraint

In this experiment we have relaxed the problem (4) by removing the constraint about the discreteness of the decision variable (c_1). This makes the initial box-bounded search space continuous, even if it remains complex due both to the presence of conditional decision variables (c_2) and the black-box hydraulic feasibility constraint (c_3). The rest of the experimental setup is identical to what reported in the previous sub-section.

5 Results

This section summarizes the most relevant results. Figure 3 shows how the “best seen” changes over function evaluations: solid lines are the averages over 20 different runs, while the shaded areas represent the standard deviations (almost 0). The first value, at iteration 0, is the best seen observed within the initial design.

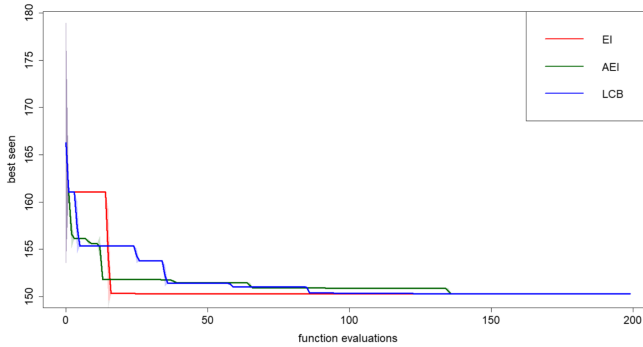


Fig. 3. Best seen over function evaluations of RF-based SMBO using three acquisition functions: EI (red), AEI (green) and LCB (blue). Solid lines and shaded areas represent, respectively, mean and standard deviation (that is almost 0) of the best seen over 20 different runs (Color figure online)

With respect to the second experiment – related to the relaxation of the discreteness constraint (c_1) – Fig. 4 shows how the “best seen” changes over function evaluations. Visualization is limited to the first 20 evaluations – out of the overall 200 – because, already after two function evaluations, no further improvements have been obtained.

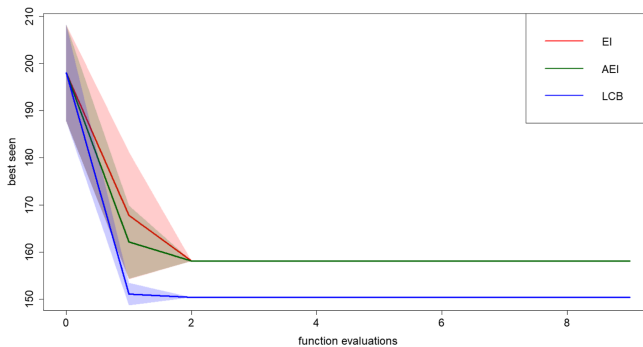


Fig. 4. Best seen over function evaluations of RF-based SMBO with relaxation of the discreteness constraint. Comparison between three acquisition functions: EI (red), AEI (green) and LCB (blue). Solid lines and shaded areas represent, respectively, mean and standard deviation (that is almost 0) of the best seen over 20 different runs (Color figure online)

Finally, we have evaluated the improvement, in terms of energy costs reduction, provided by SMBO with respect to the energy cost implied by the current pressure-based control operated by the water utility, that is 332,30€/day. In Table 1, the best cost over 20 runs has been selected for every acquisition function and separately for the two types of experiments described in Sect. 4.2.1 and 4.2.2.

Table 1. Optimal energy costs obtained via SMBO and associated costs reduction with respect to the current cost implied by the current pressure-based control operated by the water utility

	Original Problem (4)			Relaxation of (c_1)		
	EI	AEI	LCB	EI	AEI	LCB
Energy cost [€]	<u>150.29</u>	150.32	150.31	158.12	154.91	150.42
Cost reduction w.r.t. the control strategy currently operated [€]	<u>182.01</u>	181.98	181.99	174.18	177.39	181.88

The relaxation of discreteness constraint does not provide any improvement. This could be due to the use of RF, which can result less effective on continuous variables than discrete ones, also depending on the smoothness of the objective function. Probably, in the second experiment, the approach was not able to escape from some plateau, within the maximum number of function evaluations allowed.

As an example, we report in Fig. 5 the activation pattern of two pumps, A and B, according to the control currently operated by the WDN (“curr” suffix) versus the new activation implied by the new control optimized through SMBO (“opt” suffix).

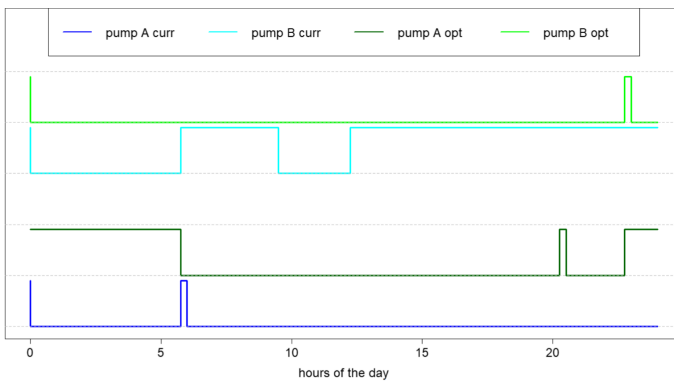


Fig. 5. Activation of two pumps according to current and optimized implicit control

6 Conclusions and Discussion

We have presented a SMBO approach for solving optimal control problems characterized by black-box objective functions and complex, partially unknown, search spaces. A general formalization of the problem was provided along with an instantiation on a specific real-life application, that is the optimal control of pumps in water distribution networks. The use of a hydraulic simulation software, EPANET, makes both objective function and constraints – related to hydraulic feasibility of the identified control rules – black-box. Using SMBO to search for an optimal *implicit* control allowed us to work with a dimensionality which is significantly lower than the one required by the (more widely adopted) *explicit* controls. A more realistic experimentation should consider different “simulation days”, characterized by random water demands whose empirical distribution is generated from historical data. This requires evaluating the robustness of the implicit control rules proposed by SMBO and to move towards a “distributionally robust” SMBO.

Acknowledgements. This study has been partially supported by the Italian project “PerFORM WATER 2030” – programme POR (Programma Operativo Regionale) FESR (Fondo Europeo di Sviluppo Regionale) 2014–2020, 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.

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

References

1. Kushner, H.J.: A new method of locating the maximum point of an arbitrary multipeak curve in the presence of noise. *J. Basic Eng.* **86**(1), 97–106 (1964)
2. Močkus, J.: On Bayesian methods for seeking the extremum. In: Marchuk, G.I. (ed.) *Optimization Techniques 1974*. LNCS, vol. 27, pp. 400–404. Springer, Heidelberg (1975). https://doi.org/10.1007/3-540-07165-2_55
3. Zhigljavsky, A., Zilinskas, A.: *Stochastic Global Optimization*, vol. 9. Springer, Heidelberg (2007). <https://doi.org/10.1007/978-0-387-74740-8>
4. Hutter, F., Hoos, H.H., Leyton-Brown, K.: Sequential model-based optimization for general algorithm configuration. In: Coello, C.A.C. (ed.) *LION 2011*. LNCS, vol. 6683, pp. 507–523. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-25566-3_40
5. Archetti, F., Candelieri, A. (eds.): *Bayesian Optimization and Data Science*. SO. Springer, Cham (2019). <https://doi.org/10.1007/978-3-030-24494-1>
6. Hutter, F., Kotthoff, L., Vanschoren, J.: *Automated Machine Learning-Methods, Systems, Challenges*. Springer, Heidelberg (2019). <https://doi.org/10.1007/978-3-030-05318-5>
7. Elsken, T., Metzger, J.H., Hutter, F.: Neural architecture search: a survey. *J. Mach. Learn. Res.* **20**(55), 1–21 (2019)
8. Candelieri, A., Archetti, F.: Global optimization in machine learning: the design of a predictive analytics application. *Soft. Comput.* **23**(9), 2969–2977 (2018). <https://doi.org/10.1007/s00500-018-3597-8>
9. Williams, C.K., Rasmussen, C.E.: *Gaussian Processes for Machine Learning*. MIT Press, Cambridge (2006). 2(3)

10. Frazier, P.I.: Bayesian optimization. In: *INFORMS TutORials in Operations Research*, pp. 255–278 (2018)
11. Ho, T.K.: Random decision forests. In: *Proceedings of the 3rd International Conference on Document Analysis and Recognition*, pp. 278–282 (1995)
12. Sergeyev, Y.D., Pugliese, P., Famularo, D.: Index information algorithm with local tuning for solving multidimensional global optimization problems with multiextremal constraints. *Math. Program.* **96**(3), 489–512 (2003)
13. Paulavičius, R., Žilinskas, J.: Advantages of simplicial partitioning for Lipschitz optimization problems with linear constraints. *Optim. Lett.* **10**(2), 237–246 (2014). <https://doi.org/10.1007/s11590-014-0772-4>
14. Strongin, R.G., Sergeyev, Y.D.: *Global Optimization with Non-convex Constraints: Sequential and Parallel Algorithms*, vol. 45. Springer, Berlin (2013)
15. Digabel, S.L., Wild, S.M.: A taxonomy of constraints in simulation-based optimization. arXiv preprint [arXiv:1505.07881](https://arxiv.org/abs/1505.07881) (2015)
16. Gramacy, R.B., Lee, H.K.M., Holmes, C., Osborne, M.: Optimization under unknown constraints. *Bayesian Stat.* **9**, 229 (2012)
17. Bernardo, J., et al.: Optimization under unknown constraints. *Bayesian Stat.* **9**(9), 229 (2011)
18. Candelieri, A., Archetti, F.: Sequential model based optimization with black-box constraints: feasibility determination via machine learning. In: *AIP Conference Proceedings*, vol. 2070, no. 1, p. 020010. AIP Publishing LLC (2019)
19. Hernández-Lobato, J.M., Gelbart, M.A., Hoffman, M.W., Adams, R.P., Ghahramani, Z.: Predictive entropy search for Bayesian Optimization with unknown constraints. In: *Proceedings of the 32nd International Conference on Machine Learning*, p. 37 (2015)
20. Strongin, R.G., Sergeyev, Y.D.: *Global Optimization with Non-convex Constraints: Sequential and Parallel Algorithms*. Kluwer Academic Publishers, Dordrecht (2000)
21. Bachoc, F., Helbert, C., Picheny, V.: Gaussian process optimization with failures: classification and convergence proof. HAL id: hal-02100819, version 1 (2019)
22. Sacher, M., et al.: A classification approach to efficient global optimization in presence of non-computable domains. *Struct. Multidisc. Optim.* **58**(4), 1537–1557 (2018). <https://doi.org/10.1007/s00158-018-1981-8>
23. Antonio, C.: Sequential model based optimization of partially defined functions under unknown constraints. *J. Glob. Optim.* **57**(3), 1–23 (2019). <https://doi.org/10.1007/s10898-019-00860-4>
24. Mala-Jetmarova, H., Sultanova, N., Savic, D.: Lost in optimisation of water distribution systems? A literature review of system operation. *Environ. Model Softw.* **93**, 209–254 (2017)
25. Bene, J.G., Selek, I., Hos, C.: Comparison of deterministic and heuristic optimization solvers for water network scheduling problems. *Water Sci. Technol. Water Supply* (2013)
26. Castro Gama, M.E., Pan, Q., Salman, S., Jonoski, A.: Multivariate optimization to decrease total energy consumption in the water supply system of Abbiategrosso (Milan, Italy). *Environ. Eng. Manag. J. (EEMJ)* **14**(9) (2015)
27. 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.* **186**, 436–443 (2017)
28. Candelieri, A., Perego, R., Archetti, F.: Bayesian optimization of pump operations in water distribution systems. *J. Glob. Optim.* **71**(1), 213–235 (2018). <https://doi.org/10.1007/s10898-018-0641-2>
29. Srinivas, N., Krause, A., Kakade, S.M., Seeger, M.W.: Information-theoretic regret bounds for Gaussian process optimization in the bandit setting. *IEEE Trans. Inf. Theory* **58**(5), 3250–3265 (2012)

30. Huang, D., Allen, T.T., Notz, W.I., Zheng, N.: Global optimization of stochastic black-box systems via sequential Kriging meta-models. *J. Glob. Optim.* **3**(34), 441–466 (2006)
31. Picheny, V., Wagner, T., Ginsbourger, D.: A benchmark of kriging-based infill criteria for noisy optimization. *Struct. Multidiscip. Optim.* **48**(3), 607–626 (2013)
32. Bischl, B., Richter, J., Bossek, J., Horn, D., Thomas, J., Lang, M.: mlrMBO: A modular framework for model-based optimization of expensive black-box functions. arXiv preprint [arXiv:1703.03373](https://arxiv.org/abs/1703.03373) (2017)
33. Hansen, N., Müller, S.D., Koumoutsakos, P.: Reducing the time complexity of the derandomized evolution strategy with covariance matrix adaptation (CMA-ES). *Evol. Comput.* **11**(1), 1–18 (2003)