



Article

Bayesian Optimization Using Simulation-Based Multiple Information Sources over Combinatorial Structures

Antonio Sabbatella ¹, Andrea Ponti ² , Antonio Candelieri ^{2,*} and Francesco Archetti ¹

¹ Department of Computer Science, Systems and Communications, University of Milan-Bicocca, 20126 Milan, Italy; a.sabbatella@campus.unimib.it (A.S.); francesco.archetti@unimib.it (F.A.)

² Department of Economics, Management, and Statistics, University of Milan-Bicocca, 20126 Milan, Italy; a.ponti5@campus.unimib.it

* Correspondence: antonio.candelieri@unimib.it

Abstract: Bayesian optimization due to its flexibility and sample efficiency has become a standard approach for simulation optimization. To reduce this problem, one can resort to cheaper surrogates of the objective function. Examples are ubiquitous, from protein engineering or material science to tuning machine learning algorithms, where one could use a subset of the full training set or even a smaller related dataset. Cheap information sources in the optimization scheme have been studied in the literature as the multi-fidelity optimization problem. Of course, cheaper sources may hold some promise toward tractability, but cheaper models offer an incomplete model inducing unknown bias and epistemic uncertainty. In this manuscript, we are concerned with the discrete case, where $f(\mathbf{x}, \mathbf{w}_i)$ is the value of the performance measure associated with the environmental condition \mathbf{w}_i and $p(\mathbf{w}_i)$ represents the relevance of the condition \mathbf{w}_i (i.e., the probability of occurrence or the fraction of time this condition occurs). The main contribution of this paper is the proposal of a Gaussian-based framework, called augmented Gaussian process (AGP), based on sparsification, originally proposed for continuous functions and its generalization in this paper to stochastic optimization using different risk profiles for combinatorial optimization. The AGP leverages sample and cost-efficient Bayesian optimization (BO) of multiple information sources and supports a new acquisition function to select the new source–location pair considering the cost of the source and the (location-dependent) model discrepancy. An extensive set of computational results supports risk-aware optimization based on CVaR (conditional value-at-risk). Computational experiments confirm the actual performance of the MISO-AGP method and the hyperparameter optimization on benchmark functions and real-world problems.

Keywords: Bayesian optimization; simulation; combinatorial optimization; value-at-risk; network design; multi-fidelity; information sources



Citation: Sabbatella, A.; Ponti, A.; Candelieri, A.; Archetti, F. Bayesian Optimization Using Simulation-Based Multiple Information Sources over Combinatorial Structures. *Mach. Learn. Knowl. Extr.* **2024**, *6*, 2232–2247.

<https://doi.org/10.3390/make6040110>

Academic Editor: Javier Del Ser Lorente

Received: 24 June 2024

Revised: 21 September 2024

Accepted: 26 September 2024

Published: 5 October 2024



Copyright: © 2024 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

Simulation-based optimization problems are usually black-box and computationally expensive and have been receiving increasing attention for their relevance in ubiquitous applications [1]. Bayesian optimization (BO), due to its flexibility and sample efficiency, has become a standard approach for simulation optimization. The computational cost, notwithstanding its sample efficiency, can still represent an obstacle to a wider diffusion. To mitigate this problem, in many situations, one can resort to cheaper surrogates of the objective function such as the output of a computer simulation. Examples are ubiquitous, including experimental design in protein engineering or material science, where the “ground truth” is given by a physical prototype as extremely expensive synthesis and characterization of a new material in a laboratory. In other cases, sources of different fidelities are given by the output of a partial differential equation solver using different discretization parameters. Sources of different fidelities can be also exploited tuning machine learning

algorithms. Rather than using the full dataset, one could use a smaller related dataset [2] or terminate the training procedure early as in [3]. Cheap information sources in the optimization scheme have been studied in the literature as the multi-fidelity optimization problem. Specific methods have been developed to leverage cheaper sources in more efficient methods [4]. Of course, cheaper sources may hold some promise toward tractability, but cheaper models offer an incomplete model inducing unknown bias and epistemic uncertainty. Multi-fidelity optimization methods require that sources are hierarchical organized. This means that once a source has been queried at location \mathbf{x} , no further knowledge can be obtained querying any other source of lower fidelity at any location [5]. Moreover, hierarchical source organization relies on the assumption that information sources are unbiased, admitting only aleatoric uncertainty that must be independent across sources.

To overcome these limitations, the multi-fidelity setting was generalized under different headings as multi-task BO [2], non-hierarchical multifidly [6], or multiple information source BO [7,8]. In [9], it is shown how more cost-effective sources of information can be integrated with more accurate, as in computational chemistry, in material discovery. Another application of BO for material optimization is [10].

The above difficulties were first addressed in [11], which considered a single model integrating the different information sources into a single model with relative discrepancies between each source and the function to optimize depending on the location proposed and change across the search space. Moreover, [8] introduces a general notion of location-dependent model discrepancy to quantify the difference between each source and the objective function. Under these general assumptions, sources are no longer necessarily unbiased and allow for epistemic error.

Another feature of simulation-based optimization is that the evaluations of the objective functions are noisy (aleatoric or observational errors) and can be affected by uncertain (epistemic) errors and model uncertainty. The usual solution considers as the objective function the sample average approximation (SAA), as is done in the cross-validation procedure in machine learning. The reference problem is:

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

Real-world optimization problems tend to have stochastic elements in the objective function, the constraints, or the context of the problem. This is the case when querying the objective functions requires the execution of a stochastic simulation model accounting for different scenarios, but also the choice of a stochastic optimizer of the loss function or the randomness in the initialization of the optimization algorithm.

A more general formulation considers the different sources of randomness synthesized by a random variable \mathbf{w} . Consequently, the objective function $F(\mathbf{x})$ in (1) becomes a random function $F(\mathbf{x}) = f(\mathbf{x}, \mathbf{w})$, and the problem (1) becomes:

$$\min_{\mathbf{x} \in \mathcal{X} \subset \mathbb{R}^d} \sum_i f(\mathbf{x}, \mathbf{w}_i) p(\mathbf{w}_i) \quad (2)$$

If $F(\mathbf{x})$ is a performance metric of a system, this defines the optimization of the average performance. In this manuscript, we are concerned with the discrete case where $f(\mathbf{x}, \mathbf{w}_i)$ is the value of the performance measure associated with the environmental condition \mathbf{w}_i . $p(\mathbf{w}_i)$ represents the relevance of condition \mathbf{w}_i (i.e., the probability of occurrence or the fraction of time this condition occurs).

This is for instance the case of optimal sensor placement in a network where the integer variable \mathbf{x} corresponds to the placement of a number of sensors over the nodes of a network, while the environmental condition \mathbf{w}_i is the injection of a contaminant at a node and $F(\mathbf{x})$ is a performance score of the placement \mathbf{x} to monitor the propagation process and detect the detection/intrusion as early and effectively as possible. $F(\mathbf{x})$ is the sample average approximation of the detection time corresponding to \mathbf{x} .

A relevant limitation of SAA is that it is a risk-neutral measure, while infrastructure networks like water, energy, or transport, among others, must weigh specifically the downside risk. The networks considered in this paper use a different risk profile given by value-at-risk (Var) and conditional VaR (CVaR), borrowed from financial analysis.

Among simulation optimization problems, combinatorial domains present challenges due to the generalization of the Gaussian process to combinatorial structures and the combinatorial optimization of the acquisition function. In this paper, the authors consider a “naïve” solution given by a continuous embedding of a solution. A continuous relaxation allows for an efficient optimization of the acquisition function, but it does not account for the discretization needed before the next function evaluation.

The general objective of this paper is to propose a Gaussian-based framework, called augmented Gaussian process (AGP), based on sparsification, originally proposed in [7], for continuous functions, and to show that it can be generalized to stochastic optimization using different risk profiles for combinatorial optimization. Some approaches to deal with integer and categorical variables are analyzed in [12,13].

The AGP, used in [14] for fine-tuning the hyperparameters of a machine learning model to optimize simultaneously accuracy and fairness while also reducing energy consumption, is shown in this paper to provide a solution that can be generalized to simulation-based combinatorial and network problems. The AGP leverages into sample and cost-efficient BO over multiple information sources and supports a new acquisition function to select the new source–location pair which combines the AGP confidence bound, the cost of the source, and the (location-dependent) model discrepancy between the source-specific GP and the AGP model.

An extensive set of computational results supports risk-aware optimization based on CVaR. The multiple information source acquisition function avoids variance starvation, premature convergence to local optima, as well as ill-conditioning in the GP training.

Computational experiments confirm the performances of the MISO-AGP (multiple information source optimization through AGP) methods on both benchmark functions and real-world problems.

1.1. Related Works

Multi-fidelity and multiple information source BO have been a thriving research domain. Many approaches have been proposed and leveraged into effective algorithms among which only a few are here commented. The case of unreliable information sources is considered in [15], where a methodology is proposed which makes multi-fidelity BO robust meaning that a theoretical guarantee is given to the effect that the addition of an auxiliary information source will not lead to worse performance than “vanilla” BO. Also, [16] proposes multi-fidelity BO with the acquisition function max-value entropy search and analysis of a parallel version. A general framework for multi-fidelity BO based on mutual information and the greedy strategy (namely, MF-MI-greedy) is proposed in [17], specifically requiring that strict relations between the quality and the cost of a lower fidelity function are likely to lead to sub-optimal experiment design and to limit their practicality. Moreover, it is proposed that a simple notion of regret which incorporates the cost of different fidelities and proves that (MF-MI-Greedy) achieves a low regret. Another strategy for adaptive sampling of multi-fidelity GP is proposed in [18] to reduce predictive uncertainty as well as the cost of the execution of the simulation runs.

The key approach proposed in this paper is the AGP [7], which proposes sparsifying over multiple information sources. The strategy is to “augment” the observations of the high-fidelity source with only the “reliable” ones coming from the cheaper sources, and to extend the acquisition function to the selection among sources which can be considered reliable.

Furthermore, transfer learning as a tool for multi-fidelity optimization is addressed in [19], which proposes an acquisition function based on across-task transferable max-value entropy that balances the need to acquire information about the current task with the goal of acquiring information transferable to future tasks. Also, [20] considers the effects of

heterogeneous errors on multi-fidelity BO and proposes a method to learn a noise model for each data source and leverage highly biased low-fidelity sources which are only locally correlated with the high-fidelity source.

A seminal paper for BO on combinatorial structure is [21], which proposes an approximate optimizer of the acquisition function to overcome the difficulty of many acquisition functions to large combinatorial domains. Another approach is [22], which provides a wide analysis of BO over combinatorial spaces and samples discrete variables upon continuous relaxation. The surrogate model is a Bayesian neural network with Thompson sampling and variational optimization of the acquisition function. An entirely different approach is based on autoencoders and deep learning to generate high-dimensional discrete objects. Ref. [23] uses the epistemic uncertainty of the decoder to guide the exploration of new points. The algorithm proposed in [24] integrates deep metric learning and a variational autoencoder and provides vanishing regret guarantees. Another approach for solving combinatorial problems was proposed in [25], which introduces a new learning-to-search approach that employs a combinatorial search over a combinatorial space where each discrete structure is represented by discrete variables. Heuristics are used to select good starting spaces while machine learning is adopted to improve global knowledge. A different approach was proposed in [26] based on recent advances in submodular relaxation for solving binary quadratic programming. The approach is inspired by parametrized submodular relaxation which makes it possible to optimize efficiently the acquisition function via minimum graph cut algorithms.

In [27], a new approach based on Mercer features for combinatorial Bayesian optimization is proposed, based on diffusion kernels and using Thompson sampling as an acquisition function. Finally, the method proposed in [28] maps the structural information of the combinatorial space into a corresponding latent space, where the optimization takes place. The next candidate latent solution is decoded into a discrete one to evaluate it. The superiority of the method, especially in small-data setting, is empirically proven.

BO has been applied to a wide set of problems. In this manuscript, we focused on problems characterized by main features such as combinatorial search spaces of discrete variables, simulation-based optimization with stochastic elements, and multiple information sources. Several application domains fit into this framework. Optimal sensor placement in networks, which will be considered in our experiments, is one application. Other problems considered in the experiments are the combinatorial optimization binary quadratic problems, and standard multi-fidelity benchmarks.

Epidemic scenarios also fit into the simulation optimization setting. Given a network of interacting people, the problem is to choose a small set of people whose surveillance enables the early detection of any disease outbreak when very few people are already infected. In the domain of the web, bloggers publish posts and use hyperlinks to other content on the web. We want to select a set of links to most of the stories that propagate in the blogosphere.

1.2. Our Contributions

The key contribution of this paper is a new decision-theoretic approach based on the AGP for generating a single model on different information sources which can be used also for combinatorial and network design problems. The proposed acquisition function, to select the new source–location pair combines the AGP confidence bound, the cost of the source, and the (location-dependent) model discrepancy between the source-specific GP and the AGP. A genetic operator was also proposed for the optimization of the acquisition function over combinatorial structures.

The focus of the proposed method was on simulation optimization models which typically generate black-box expensive optimization problems. The risk profile of the problem has been accounted for, in the case of network design, using the risk measures VaR and CVaR. The multiple information source acquisition function avoids variance starvation, premature convergence to local optima as well as ill-conditioning in the GP training. Computational experiments confirm the actual performance of the MISO-AGP method for hyperparameter optimization on benchmark functions and real-world problems.

1.3. Organization of the Paper

The rest of the paper is organized as follows. Section 2 provides the background on the GP-based BO. Section 3 summarizes the MISO-AGP framework initially proposed in [7] for continuous optimization problems. Section 4 presents the structure of BoTorch (<https://botorch.org/>, accessed on 4 October 2024). The standard reference library for BO in which MISO-AGP was recently included <https://github.com/pytorch/botorch/pull/2152>, accessed on 4 October 2024. Section 5 provides the computational results of MISO-AGP applied to a binary quadratic programming problem from literature. Then, Section 6 regards the adoption of MISO-AGP for solving a real-world application, specifically the optimal sensor placement in a water distribution network. Finally, Section 7 summarizes concluding remarks, perspectives, and limitations.

2. Background

2.1. Gaussian Processes and Bayesian Optimization in Brief

BO is a sequential and sample-efficient algorithm for solving the global optimization problem (1). Briefly, the basic BO algorithm consists of two components, a probabilistic surrogate model approximating $F(\mathbf{x})$, depending on the candidate solutions evaluated so far, and an acquisition function whose optimization drives the identification of the next candidate solution by dealing with the exploitation–exploration dilemma [29–31].

We refer to the most common setting, which uses a Gaussian process as probabilistic surrogate model and the GP's lower confidence bound (GP-LCB) as acquisition function.

Denote with $\mathbf{X}_{1:n} = \{\mathbf{x}^{(i)}\}_{i=1,\dots,n}$ the candidate solutions evaluated so far, and with $\mathbf{y}_{1:n} = \{F(\mathbf{x}^{(i)})\}_{i=1,\dots,n}$ the corresponding observed values of the objective function. The GP's equations related to the prediction of $F(\mathbf{x})$, namely $\mu(\mathbf{x})$, and the square of the associated predictive uncertainty, with respect to the dataset $\mathbf{D} = (\mathbf{X}_{1:n}, \mathbf{y}_{1:n})$, are:

$$\mu(\mathbf{x}) = \mathbf{k}(\mathbf{x}, \mathbf{X}_{1:n}) \mathbf{K}^{-1} \mathbf{y}_{1:n} \quad (3)$$

$$\sigma^2(\mathbf{x}) = k(\mathbf{x}, \mathbf{x}) - \mathbf{k}(\mathbf{x}, \mathbf{X}_{1:n}) \mathbf{K}^{-1} \mathbf{k}(\mathbf{X}_{1:n}, \mathbf{x}) \quad (4)$$

where $k(\mathbf{x}, \mathbf{x}')$ is a kernel function, $\mathbf{k}(\mathbf{x}, \mathbf{X}_{1:n}) = \{k(\mathbf{x}, \mathbf{x}^{(i)})\}_{i=1:n}$, and \mathbf{K} is a $n \times n$ matrix with entries $\mathbf{K}_{ij} = k(\mathbf{x}^{(i)}, \mathbf{x}^{(j)})$. The choice of kernel function implies structural assumptions on the shape of $\mu(\mathbf{x}) \simeq F(\mathbf{x})$, which is adjusted by tuning the kernel hyperparameter(s). In this paper, we consider the kernel Matérn-5/2. Observations can be considered noisy, that is $y^{(i)} = F(\mathbf{x}^{(i)}) + \varepsilon^{(i)}$, with $\varepsilon^{(i)} \sim \mathcal{N}(0, \lambda^2)$, leading to replace \mathbf{K}^{-1} with $[\mathbf{K} + \lambda^2 \mathbf{I}]^{-1}$ in (3) and (4).

Finally, we also report the equation of the GP-LCB acquisition function, which combines $\mu(\mathbf{x})$ and $\sigma(\mathbf{x}) = \sqrt{\sigma^2(\mathbf{x})}$ to identify the next candidate solution:

$$\text{GP-LCB}(\mathbf{x}) = \mu(\mathbf{x}) - \beta \sigma(\mathbf{x}) \quad (5)$$

The parameter β regulates the so-called uncertainty bonus and, consequently, the trade-off between exploitation and exploration. In our experiments, we used $\beta = 3$.

2.2. Dealing with Integer Variables

When the problem contains some categorical or integer-valued variables the objective function output of the GP cannot be evaluated at all potential input locations. A commonly chosen solution, for instance followed in the BO in Spearmint (<https://github.com/HIPS/Spearmint>, accessed on 4 October 2024) and suggested in BoTorch, is to optimize the acquisition function assuming that all variables can take continuous values and subsequently replace the continuous minimizer of the acquisition function by the closest integer. In this paper, we propose a different approach in which the minimization of the acquisition is performed directly over the integer variables.

We set up the discrete solutions as real-valued in the GP, while the optimization of the acquisition function is performed in the combinatorial space, by considering the possible (discrete) solutions only. This means that there is no need to modify the probabilistic model (i.e., the AGP) or the acquisition function. The only required modification regards the method used to optimize the acquisition function. Genetic algorithms (GA) were used. The Pymoo implementation of GA was used to optimize the acquisition function. As GA’s mutation operator, a standard bitflip mutation was used, while as a crossover operator, the problem-specific operator previously and successfully proposed in [32], was used.

3. Multi-Information Source Optimization and Augmented Gaussian Processes

Multiple Information Source (Bayesian) Optimization

Assuming that we worked with S information sources, each identified by an index $s \in 1, \dots, S$ and having a different cost for evaluating a candidate solution \mathbf{x} . For simplicity, and without loss of generality, we assume that the sources are sorted by descending cost, so that $s = 1$ denotes the ground truth. Then, denote with $\mathbf{D}_s = \left\{ \left(\mathbf{x}^{(i)}, \mathbf{y}^{(i)} \right) \right\}_{i=1, \dots, n_s}$ the current set of n_s solutions and associated values collected on the information source s .

In the MISO framework, the goal is to solve (1) by using the cheaper information sources to reduce the cost accumulated over the entire optimization process. The common idea of the MISO approach is to model every source independently—through an independent GP—and then combine them into a unique model. In this paper, we used one of the most recent, promising, and efficient MISO approaches, namely MISO-AGP, which uses a particular combination technique called augmented Gaussian process (AGP).

We summarize here the basics of the MISO-AGP method. Starting from all the observations collected on the ground truth, this dataset is augmented by including observations from other information sources that can be considered reliable. Reliability is computed in terms of the discrepancy between the GP modeling the ground truth and the GP modeling the cheap source. The set of augmented locations, denoted with \mathbf{D}_{AUG} , is computed as $\mathbf{D}_{AUG} \leftarrow \mathbf{D}_1 \cup \bar{\mathbf{D}}$, where \mathbf{D}_1 is the dataset of the observations on the ground truth and $\bar{\mathbf{D}}$ is:

$$\bar{\mathbf{D}} = \{ (\bar{\mathbf{x}}, \bar{y}) : \exists \bar{s} : (\bar{\mathbf{x}}, \bar{y}) \in \mathbf{D}_{\bar{s}} \wedge |\mu_1(\bar{\mathbf{x}}) - \bar{y}| < m\sigma_1(\bar{\mathbf{x}}) \} \tag{6}$$

In simpler terms, an observation $(\bar{\mathbf{x}}, \bar{y})$ from a cheap source $\bar{s} \neq 1$ is considered reliable and hence included into the augmented set \mathbf{D}_{AUG} if the difference between the prediction on the ground truth—i.e., $\mu_1(\bar{\mathbf{x}})$ —and the value observed on the cheap source \bar{s} —is lower than m times the predictive uncertainty on the ground truth.

While the main differences between the AGP and the *fusing*-GPs method, initially proposed for multiple information source optimization, were deeply addressed in [7], it is useful here to also consider the differences with cokriging methods. Cokriging uses covariance between two or more regionalized variables that are related, and it is a suitable method to deal with GP-based modeling when the main variable of interest is sparse and, on the contrary, related secondary information is abundant. The mutual spatial behavior of the regionalized variables is also named co-regionalization. Cokriging requires the same assumptions as kriging but entails a higher computational cost. Furthermore, cokriging requires the computation of the spatial covariance model of the primary attribute (i.e., ground truth source), the spatial covariance model of the secondary attribute (i.e., the cheap source), and the spatial cross-covariance model of primary and secondary attributes.

Figure 1 shows what the AGP looks like on the well-known Forrester problem considering two sources.

To select the next source-solution pair to evaluate, denoted with (s', \mathbf{x}') , the MISO-AGP approach uses a revised formulation of the GP-LCB, able to deal with multiple information sources. Formally:

$$\text{AGP-LCB}_s(\mathbf{x}) = \frac{\hat{y}^+ - (\hat{\mu}(\mathbf{x}) - \beta \hat{\sigma}(\mathbf{x}))}{c_s(1 + |\hat{\mu}(\mathbf{x}) - \mu_s(\mathbf{x})|)} \tag{7}$$

with \hat{y}^+ the best function value observed among those in the current \mathbf{D}_{AUG} , and $\hat{\mu}(\mathbf{x})$ and $\hat{\sigma}(\mathbf{x})$ the prediction and associated uncertainty provided by the AGP.

Thus, AGP-LCB considers, for every information source s and every candidate solution \mathbf{x} , the most optimistic improvement with respect to \hat{y}^+ , penalized by both the source-specific cost c_s and the discrepancy between the predictions provided by the AGP and the GP modeling the source s , that is $|\hat{\mu}(\mathbf{x}) - \mu_s(\mathbf{x})|$.

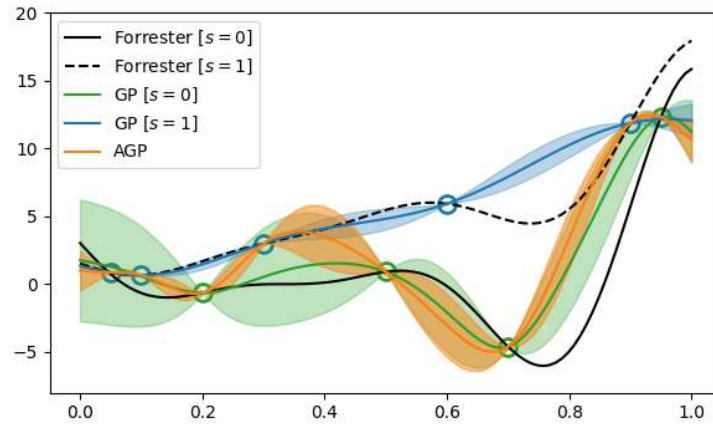


Figure 1. Forrester function with ground truth (solid black line) and one cheap source (dashed black line), along with two GPs (green and blue solid lines for predictive means and shaded areas for predictive uncertainty) individually modeling the two sources depending on source-specific observations. Finally, the resulting AGP model is depicted (orange solid line). Three out of four observations from the cheap source are considered reliable and used to augment the set of observations on the ground truth, leading to the AGP.

In Figure 2, we summarize the MISO-AGP approach into a flow chart. In the case of a combinatorial problem, only the optimization of the acquisition function (7) is different. Specifically, we used an evolutionary algorithm whose cross-over operator was designed to guarantee feasibility of the solutions, as better detailed in the next section.

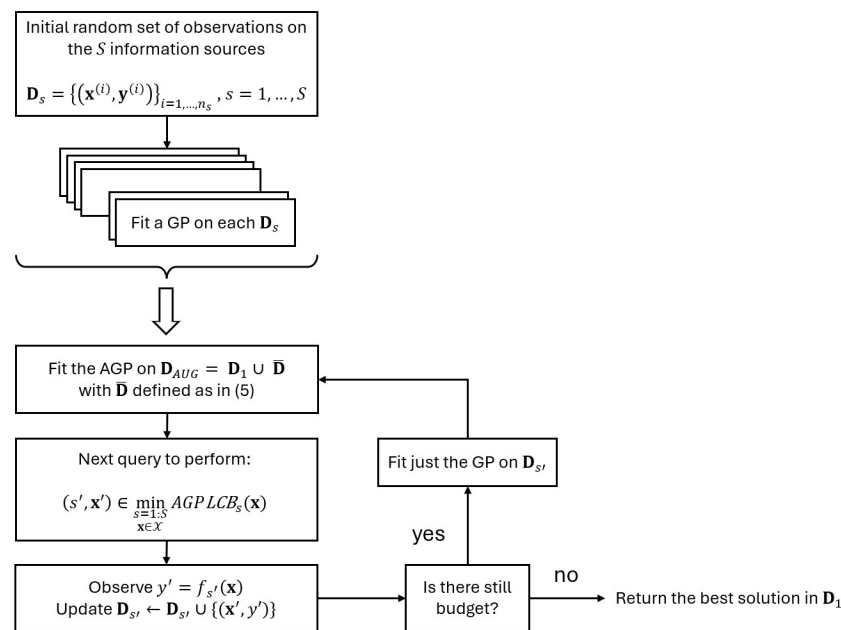


Figure 2. A flow chart summarizing the MISO-AGP approach. In the case of a combinatorial problem just the optimization in “next query to perform” box is different. Specifically, we use an evolutionary algorithm whose cross-over operator has been designed to guarantee feasibility of the solutions.

4. The Software Environment

4.1. BoTorch and Competing Algorithms

This section introduces the implementation of MISO-AGP, and its acquisition function based on GP-LCB for multi-information source problems. The key idea of the AGP is to fit a GP model for each information source and augment the observations on the high-fidelity source with those from cheaper sources which can be considered reliable.

The AGP implementation in BoTorch is based on the SingleTaskGP class. Each source is implemented as an independent SingleTaskGP and all the reliable observations are used to fit the SingleTaskGP representing the AGP, namely SingleTaskAugmentedGP. The AGP-LCB implementation is based on the UpperConfidenceBound—by default BoTorch solves maximization problem, so it uses GP_UCB. It is sufficient to solve maximization of $-F(\mathbf{x})$ to solve the minimization of $F(\mathbf{x})$ —but it is penalized by the cost of the source and the discrepancy between the source-specific GP and the AGP.

The competing algorithms considered in this paper are max-value entropy search (MES) [33,34] and general-purpose information-based Bayesian optimization (GIBBON) [35]. Both use an information-theoretic perspective to select the next solution to evaluate and the next information source. The key idea behind MES is to maximize the information gain of the optimal function value. The multi-fidelity implementation, namely MF-MES, considers the information gain over the optimal function value on the highest fidelity. The multi-fidelity version MF-GIBBON is a lightweight version of the multi-fidelity MES and uses a determinantal point process-based formulation to allow it to have a fully analytical expression.

4.2. Code Availability

To ensure reproducibility of experiments and results, we made our code and data freely available at the following GitHub repository, where all the other technical details and settings can be retrieved: <https://github.com/andreaponti5/miso-bocs> (accessed on 4 October 2024).

5. Test Problem: Binary Quadratic Programming

The MISO-AGP approach was compared against two state-of-the-art information-based multi-fidelity approaches, whose implementations are available in the BoTorch platform. As far as the MISO-AGP is concerned, all the GPs, including the AGP, use a Matern 5/2 kernel. Moreover, to prevent over-reliance on the cheap information source, a minimum number of evaluations on the ground truth was established.

Throughout the optimization process, if the threshold was violated, then the algorithm was forced to evaluate the ground truth.

To mitigate the effect of randomness in the initialization of the three algorithms, five independent runs were performed. For each run, the three algorithms shared the same set of initial random solutions.

The objective in the binary quadratic programming problem was a quadratic function with regularization.

$$F(\mathbf{x}) - \lambda * P(\mathbf{x}) = \mathbf{x}^T \mathbf{Q} \mathbf{x} - \lambda \|\mathbf{x}\|_1$$

where $\mathbf{Q} \in \mathbb{R}^{d \times d}$ is a random matrix with zero-mean Gaussian entries, multiplied element-wise by a matrix $\mathbf{K} \in \mathbb{R}^{d \times d}$ with entries $\mathbf{K}_{i,j} = e^{-\left(\frac{i-j}{L_c}\right)^2}$ which decays smoothly away from the diagonal at a rate determined by the correlation length.

According to the literature, we set $d = 10$, and sampled 50 independent realizations for \mathbf{Q} . Every algorithm was run 10 times on each instance for each realization of \mathbf{Q} . The tests were performed for the two cases: $L_c = 10$, $\lambda = 0$ and $L_c = 100$ and $\lambda = 1$. For the cheaper source cost, we considered 50% and 10% of the high fidelity.

Experiment with $\lambda = 0$, $L_c = 10$, and cheap source' cost: 50% of the ground truth's cost.

As depicted in Figure 3, MISO-AGP achieves, on average, a lower best-seen value (Figure 2 on the left) and a smaller accumulated query cost (Figure 2 on the right). Although

the final best-seen of MISO-AGP was lower than the other two approaches, there was not a statistically significant difference, as evaluated via a Wilcoxon test (p -value > 0.05). Further, MISO-AGP is significantly more efficient than MF-MES and MF-GIBBON.

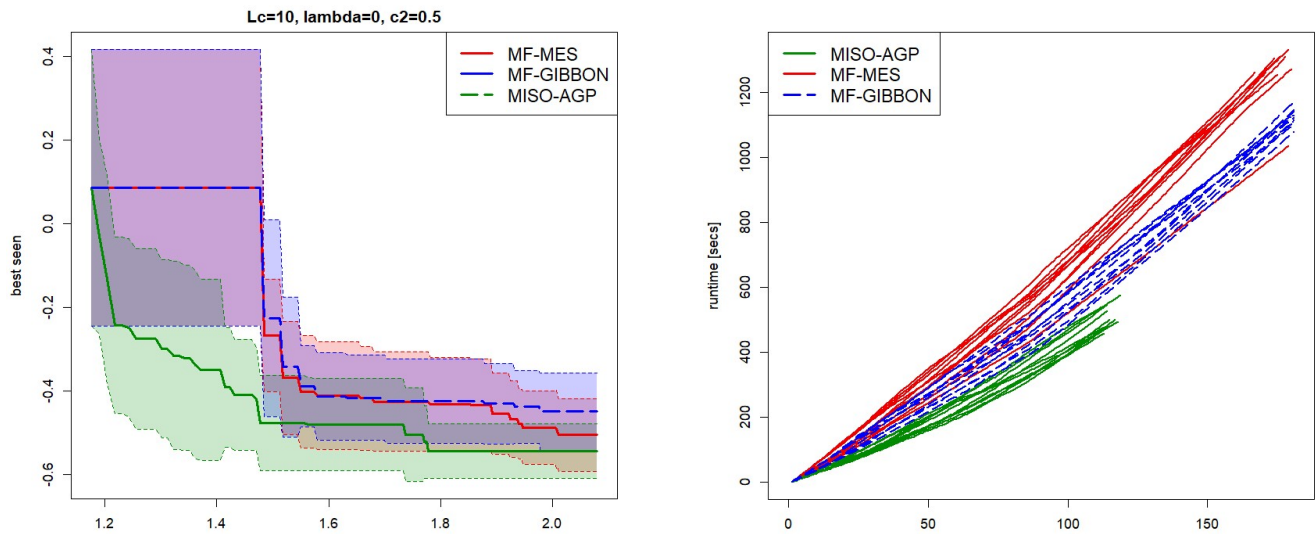


Figure 3. Best-seen with respect to the number of queries (left) and cumulated runtime (on the right). Curves represent average performance on different runs, shaded areas are for standard deviation.

Finally, MISO-AGP uses the ground truth in 79% of the total queries, against 25% for MF-MES, and 20% for MF-GIBBON. This behavior is motivated by a relevant discrepancy between the ground truth and the cheap source, leading the AGP to rely on the expensive source instead of the cheap one. This is crucial because, contrary to other standard methods for combining GPs (e.g., fusing GPs), the AGP discards cheap observations if the two sources are—even locally—uncorrelated. This crucial property of the AGP model—at the core of its design—was specifically and carefully addressed in [7].

Experiment with $\lambda = 0$, $L_c = 10$, and cheap source' cost: 10% of the ground truth's cost.

As depicted in Figure 4, MISO-AGP achieves, on average, a lower best-seen (on the left) and a smaller accumulated query cost (on the right).

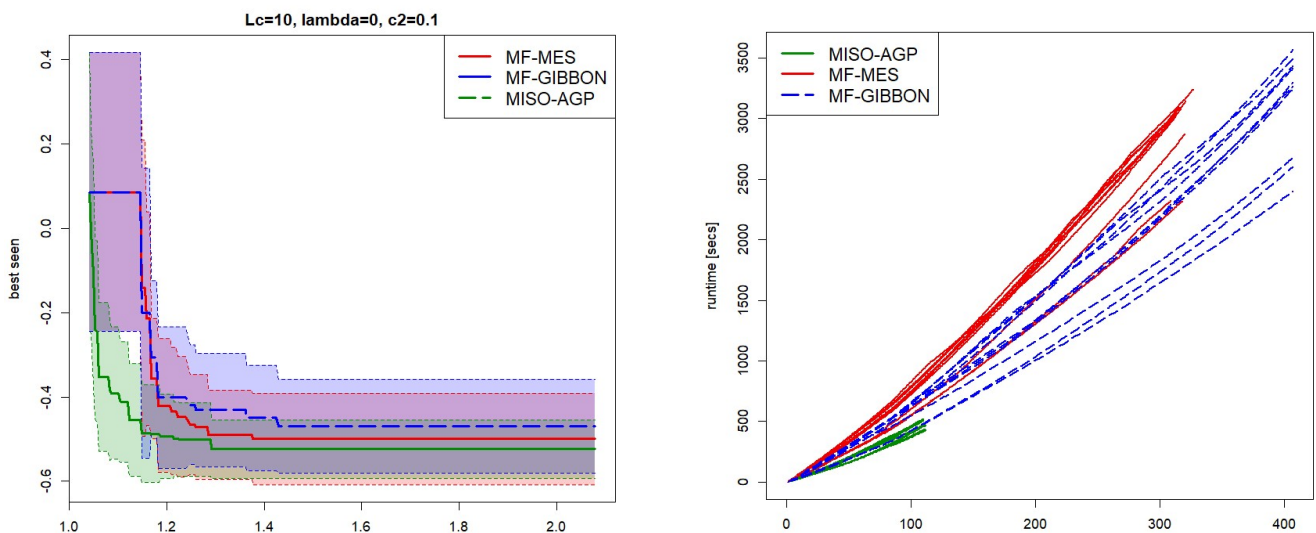


Figure 4. Best-seen with respect to the number of queries (left) and cumulated runtime (on the right). Curves represent average performance on different runs, shaded areas are for standard deviation.

In this case, MISO-AGP uses the ground truth in 91% of the total queries, against 29% for MF-MES and 20% for MF-GIBBON. It is important to remark that both MISO-AGP and MF-MES have increased the number of queries on the ground truth, even if the query cost of the cheap source decreased from 50% to 10% of the ground truth's query cost. Both the algorithms increased the number of queries on the cheap source in the first iterations—due to its small cost—leading them to understand that it is poorly correlated to the ground truth and, consequently rely only on the expensive source for most of the remaining queries.

Experiment with $\lambda = 1$, $L_c = 100$, and cheap source' cost: 50% of the ground truth's cost.

For this specific experiment, MISO-AGP shows worse results than the other two approaches, with a significantly larger value of the final best-seen (Wilcoxon test: against MF-MES, p -value = 0.0143; against MF-GIBBON, p -value = 0.0141). However, the cumulated runtime of MISO-AGP was still significantly lower than those of the other two methods, as depicted in Figure 5.

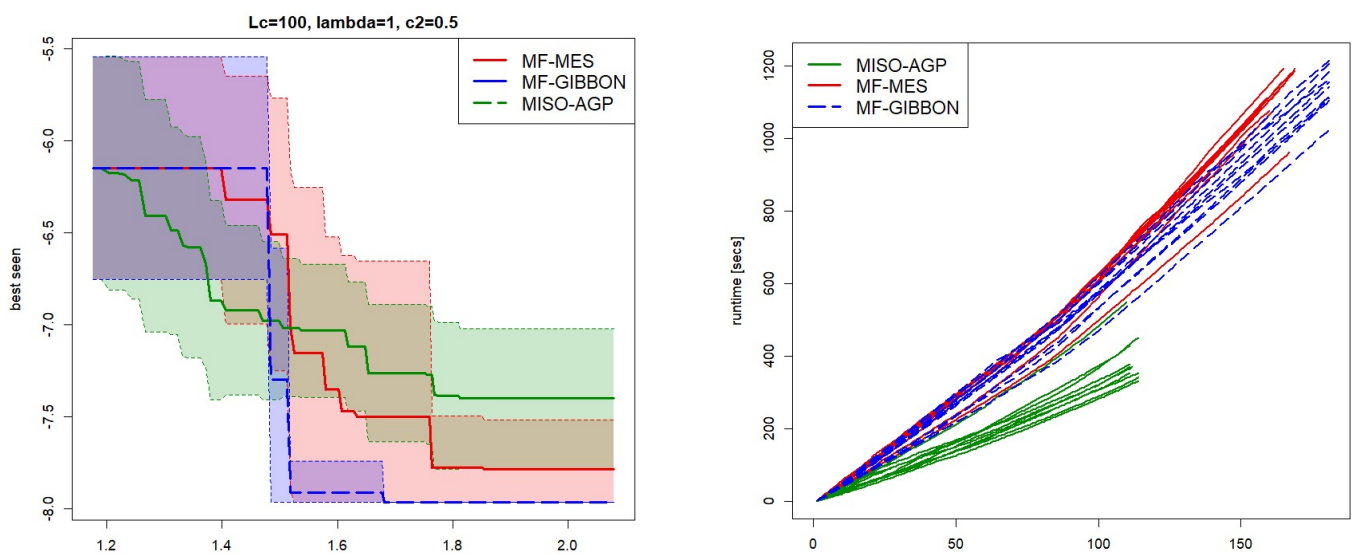


Figure 5. Best-seen with respect to the number of queries (left) and cumulated runtime (on the right). Curves represent average performance on different runs, shaded areas are for standard deviation.

Anyway, MISO-AGP queried the ground truth in 83% of the iterations, against 31% for MF-MES and 20% for MF-GIBBON. Again, MISO-AGP is more capable of understanding that the two sources are—locally—poorly correlated.

Experiment with $\lambda = 1$, $L_c = 100$, and cheap source' cost: 10% of the ground truth's cost

As depicted in Figure 6, MISO-AGP, again, achieves on average a lower best-seen value and at a lower accumulated query cost. Moreover, the final value of the best-seen is significantly smaller than those provided by the other two approaches (Wilcoxon test).

Finally, MISO-AGP uses the ground truth in 87% of the total queries (slightly increase with respect to the previous experiment), against 33% for MF-MES and 20% for MF-GI. The underlying motivation is the one already provided for the previous experiments.

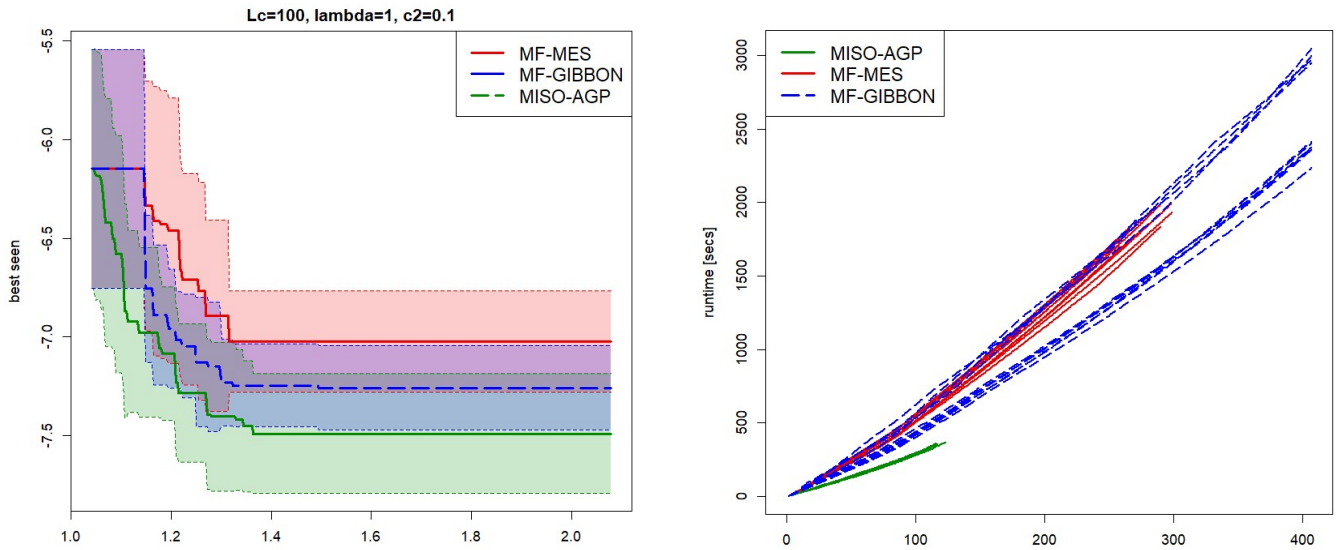


Figure 6. Best-seen with respect to the number of queries (left) and cumulated runtime (on the right). Curves represent average performance on different runs, shaded areas are for standard deviation.

6. A Real-Life Application: Risk-Averse Optimal Sensor Placement in Water Distribution Network

6.1. Conditional Value-at-Risk (CVaR)

CVaR is based on the value-at-risk (VaR), which is the maximum potential value of a metric of interest, at a certain confidence level α . Formally:

$$P(\mathcal{D}_x \leq \text{VaR}) \geq \alpha \tag{8}$$

where \mathcal{D}_x is the distribution of the metric of interest with respect to a given solution x .

When the distribution \mathcal{D}_x is discrete, VaR is easily computed as the q -quantile of the distribution, with $q = 1 - \alpha$. A general framework for Bayesian quantile and expectile optimization is established in [36]. A BO approach for CVaR is given [37], which received a BoTorch implementation. An application of the CVaR metric to water distribution networks was given in Naseridaze [38] using genetic algorithms.

Then, CVaR is the expected value of the metric of interest, given that it is beyond the VaR. For discrete distributions, CVaR is computed as:

$$\text{CVaR} = \frac{1}{j_{\text{VaR}}} \sum_{j=1}^{j_{\text{VaR}}} d_j^* \tag{9}$$

where j_{VaR} is the CVaR index, which is the position indicating the values exceeding the VaR threshold, within the sorted samples $d_1^*, \dots, d_{j_{\text{VaR}}}^*$.

6.2. The Optimal Sensor Placement Problem

The optimal sensor problem (OSP) problem aims at selecting a subset of locations where a fixed number of sensors are to be deployed to minimize an impact measure. There is not a unique impact metric because the final choice strictly depends on the specific case. Some examples of frequently used impact measures are (i) the time required to detect a contamination (aka detection time), (ii) the amount of contaminated water consumed up to the detection as well as the affected number of inhabitants, or (iii) the probability of detecting a contamination.

In this paper, we consider the detection time, and more precisely the CVaR of the detection times under a set of scenarios. We briefly introduce some required notation and then present the formalization of the OSP problem.

A water distribution network was modeled as a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where the node set \mathcal{V} contains junctions and consumption points, while the edge set \mathcal{E} consists of all the pipes connecting pairs of nodes.

A sensor placement is defined as a binary vector $\mathbf{x} \in \{0, 1\}^{|\mathcal{L}|}$, where $\mathcal{L} \subseteq \mathcal{V}$ is the subset of nodes where sensors can be possibly deployed. Specifically, each component of the vector \mathbf{x} refers to a location in the set \mathcal{L} , thus $x_i = 1$ if a sensor is deployed at the correspondent i th location, $x_i = 0$ otherwise. The number of sensors to deploy is fixed in advance as b .

Now, we introduce the stochastic component of the problem, which is the definition of simulation scenarios referred to different contamination events. Specifically, a set of contamination events is denoted with $\mathcal{A} \subseteq \mathcal{V}$, which is a subset of nodes where a contaminant is, in turn, injected. Each contamination event requires a simulation run and, therefore, is uniquely associated with a scenario. Thus, we referred to scenarios or contamination events indifferently.

6.3. Combinatorial Multi-Information Source Optimization (MISO) for Risk-Averse Optimal Sensor Placement

As far as the MISO setting was concerned, we used two sets of scenarios, that were, respectively, $\mathcal{A}_1 = \mathcal{V}$ and $\mathcal{A}_2 \subset \mathcal{A}_1$ with $|\mathcal{A}_2| = |\mathcal{A}_1|/2 = |\mathcal{V}|/2$. Consequently, computing CVaR by using \mathcal{A}_1 (the higher fidelity source i.e., the ground truth) led to a sampling cost twice as large as that required for computing CVaR on \mathcal{A}_2 (i.e., the cheaper source).

The optimal sensor placement \mathbf{x}^* is the one that optimizes the CVaR over all the contamination events \mathcal{A}_1 , so we wanted to solve the following problem:

$$\begin{aligned} \mathbf{x}^* &= \operatorname{argmin}_{\mathbf{x} \in \{0,1\}^{|\mathcal{L}|}} \operatorname{CVaR}(\mathbf{x}|\mathcal{A}_1) \\ &\text{s.t.} \\ &\sum_{i=1, \dots, |\mathcal{L}|} x_i \leq b \end{aligned} \tag{10}$$

where $\operatorname{CVaR}(\mathbf{x}|\mathcal{A}_1)$ denotes the conditional value-at-risk of the detection times observed on the \mathcal{A}_1 scenarios under the deployment \mathbf{x} .

Specifically, the detection time for one event is the lowest time needed to detect the contamination through any of the sensors in the placement \mathbf{x} . This leads to as many detection times as the number of scenarios, and their distribution is used to compute $\operatorname{CVaR}(\mathbf{x}|\mathcal{A}_1)$.

Since we are considering a MISO setting, we wanted to solve (9) by generating a sequence of solutions that also involves evaluations on the cheap source (i.e., uses the scenario set \mathcal{A}_2), with the aim to converge to the optimum with a low cumulative cost. Indeed, denote the sequence of generated solutions with $\left\{ \left(s^{(1)}, \mathbf{x}^{(1)} \right), \dots, \left(s^{(n)}, \mathbf{x}^{(n)} \right) \right\}$, then the generic $s^{(j)}$ can be $s^{(j)} = 1$ if CVaR must be computed by using \mathcal{A}_1 (i.e., the ground truth, entailing a nominal cost 1) or $s^{(j)} = 2$ if CVaR must be computed by using \mathcal{A}_2 (i.e., the cheap source, entailing a nominal cost 0.5).

Our Search Space is $\{0, 1\}^{|\mathcal{L}|} \times \{1, 2\}$, where the first $|\mathcal{L}|$ dimensions refer to the sensor placement \mathbf{x} and the last dimension refers to the information source to use for computing the objective function.

At a generic iteration of the MISO-AGP algorithm, the minimization of the acquisition function (6) is performed under the following two constraints:

$$(x, s) \in \{0, 1\}^{|\mathcal{L}|} \times \{1, 2\} \text{ and } \sum_{i=1}^{|\mathcal{L}|} x_i \leq b.$$

To solve this constrained combinatorial optimization problem, a Pymoo implementation of a genetic algorithm was used. As a mutation operator, a standard bitflip mutation

was used with a probability of $1/|\mathcal{L}|$. As crossover operator, the problem specific operator previously proposed in [32] was used.

It is briefly summarized here. Consider the example in Figure 7: Each offspring, O_1 and O_2 , takes in turn a random sensor from each parent, P_1 and P_2 , until no more sensors are available. This strategy guarantees to produce feasible offspring when using feasible parents, i.e., the offspring will have the same number of sensors as the parents.

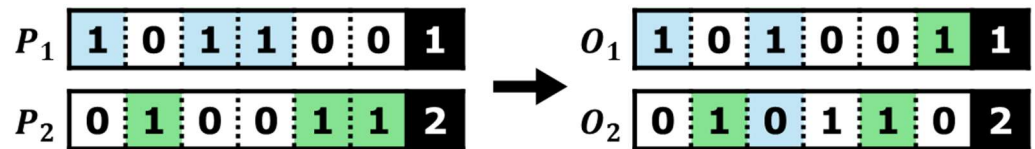


Figure 7. Example of the sensor placement crossover used. The parents P_1 and P_2 produces the offspring O_1 and O_2 . Colors are used to show from which parent each gene comes.

6.4. Numerical Results

A contamination event on each node was simulated using WNTR v1.1.0 (a Python wrapper of EPANET, a water distribution network simulator). The simulations lasted 24 (simulated) hours and the contamination concentration in each node was registered hourly. Sensors could be placed only on a subset of nodes \mathcal{L} identified by sampling nodes uniformly on their coordinates, to attain good coverage of the entire water distribution network.

The network considered in the study was named Apulian, whose number of nodes is 1364. The number of allowed sensor locations was 63, and the number of sensors allowed was set to $b = 15$.

In Figure 8, we report the best-seen value, which is the lowest CVaR value observed so far, with respect to (top of the figure) the cumulative evaluation cost and (bottom of the figure) the overall wall-clock time.

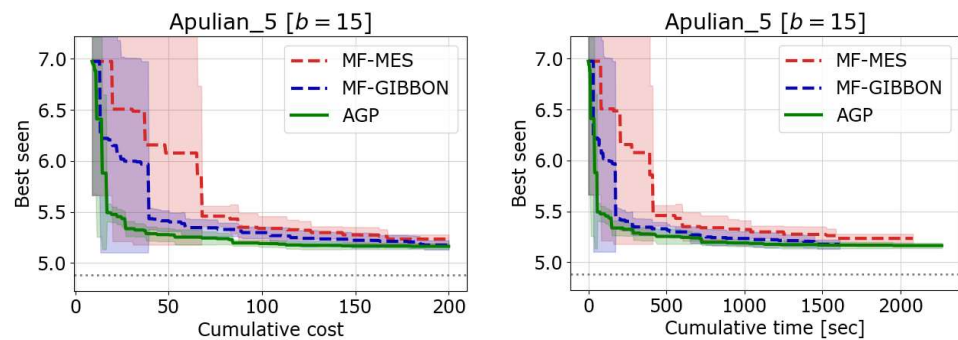


Figure 8. Performance graph for MISO-AGP, MF-MES, MF-GIBBON ($b = 15$). Curves represent average performance on different runs, shaded areas are for standard deviation.

The proposed MISO-AGP and MF-GIBBON were aligned in terms of performance, while MF-MES results were slightly worse than the other two methods.

The main advantage of the proposed approach is the significantly lower standard deviation over the different runs, making MISO-AGP a more robust framework than MF-GIBBON and MF-MES. A drawback of MISO-AGP is the slightly higher wall-clock time.

7. Conclusions, Limitations, and Perspectives

We presented an extension of the basic BO algorithm to a distributionally aware, constrained, and combinatorial multiple information source optimization setting.

The method proposes a new mechanism for generating a single model on the information sources based on GP sparsification and a decision-theoretic approach based on the MISO-AGP framework, initially and successfully tested on many test and real-world

continuous optimization problems [7,14]. The extension to the combinatorial case was quite straightforward, basically requiring modifying the way in which the MISO-AGP's acquisition function is optimized.

Specifically, the real world addressed in this paper is the optimal sensor placement in water distribution networks, required to optimize the MISO-AGP's acquisition function via a genetic algorithm, whose cross-over operator was designed to address the feasibility of the generated solutions, with respect to the combinatorial nature of the problem.

It is important to remark that, to account for risk measures that are non-neutral, CVaR was considered as the objective function of the optimal sensor placement problem. Computational experiments—also on a test problem from the literature—confirm the previous results obtained on continuous optimization problems.

Examples of other combinatorial optimization problems which could benefit from the approach proposed in this paper are epidemic source detection in contact tracing networks [39] and fake news detection using a graph-based approach [40].

Although high-dimensionality is out-of-scope in our paper, authors are aware that the scalability of MISO-AGP for high-dimensional problems is crucial. Fortunately, there are many available GP-based methods for high-dimensional Bayesian optimization (HDBO), such as TuRBO [41] and its recent extensions BxUS [42] and BOUNCE [43], which are able to perform scalable BO in high-dimensional spaces, directly working within the general GP-based BO framework. Thus, equipping one of these algorithms with the AGP, with the aim to target a MISO problem, would lead to a scalable MISO-AGP implementation. Moreover, it is important to remark that the AGP is based on a GP sparsification technique (i.e., by insertion of relevant observations), so the resulting AGP is fitted on a subset of all the observations collected over all the information sources, leading to a lower computational cost for training it, contrary to cokriging and fusing GPs methods.

Author Contributions: Conceptualization, F.A., A.C and A.P.; methodology, F.A. and A.C.; software, A.P. and A.S.; validation, all.; data curation, A.P. and A.S.; writing—original draft preparation, all; writing—review and editing, all; visualization, A.P. and A.S.; supervision, F.A. and A.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: To ensure reproducibility of experiments and results, we made our code and data freely available at the following GitHub repository, where all the other technical details and settings can be retrieved: <https://github.com/andraponti5/miso-bocs>.

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

References

1. Hunter, S.R.; Applegate, E.A.; Arora, V.; Chong, B.; Cooper, K.; Rincón-Guevara, O.; Vivas-Valencia, C. An introduction to multi-objective simulation optimization. *ACM Trans. Model. Comput. Simul. (TOMACS)* **2019**, *29*, 7. [CrossRef]
2. Swersky, K.; Snoek, J.; Adams, R.P. Multi-Task Bayesian Optimization. *Advances in Neural Information Processing Systems*. 2013, p. 26. Available online: <https://www.cs.princeton.edu/~rpa/pubs/swersky2013multi.pdf> (accessed on 4 October 2024).
3. Li, L.; Jamieson, K.; DeSalvo, G.; Rostamizadeh, A.; Talwalkar, A. Hyperband: A novel bandit-based approach to hyperparameter optimization. *J. Mach. Learn. Res.* **2018**, *18*, 1–52.
4. Choudhury, R.; Swamy, G.; Hadfield-Menell, D.; Dragan, A.D. On the utility of model learning in hri. In *Proceedings of the 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, Daegu, Republic of Korea, 11–14 March 2019; IEEE: New York, NY, USA, 2019.
5. Willcox, K.; Marzouk, Y. *Large-Scale Optimization for Bayesian Inference in Complex Systems*; No. Final report; Massachusetts Inst. of Technology (MIT): Cambridge, MA, USA, 2013.
6. Lam, R.; Willcox, K.; Wolpert, D.H. Bayesian Optimization with a Finite Budget: An Approximate Dynamic Programming Approach. *Advances in Neural Information Processing Systems*. 2016, p. 29. Available online: https://www.researchgate.net/publication/323884383_Bayesian_optimization_with_a_finite_budget_An_approximate_dynamic_programming_approach (accessed on 4 October 2024).
7. Candelieri, A.; Archetti, F. Sparsifying to optimize over multiple information sources: An augmented Gaussian process-based algorithm. *Struct. Multidiscip. Optim.* **2021**, *64*, 239–255. [CrossRef]

8. Poloczek, M.; Wang, J.; Frazier, P. Multi-Information Source Optimization. *Advances in Neural Information Processing Systems*. 2017, p. 30. Available online: https://proceedings.neurips.cc/paper_files/paper/2017/file/df1fd20ee86704251795841e6a9405a-Paper.pdf (accessed on 4 October 2024).
9. HeHerbol, H.C.; Poloczek, M.; Clancy, P. Cost-effective materials discovery: Bayesian optimization across multiple information sources. *Mater. Horiz.* **2020**, *7*, 2113–2123. [[CrossRef](#)]
10. Valladares, H.; Li, T.; Zhu, L.; El-Mounayri, H.; Hashem, A.M.; Abdel-Ghany, A.E.; Tovar, A. Gaussian process-based prognostics of lithium-ion batteries and design optimization of cathode active materials. *J. Power Sources* **2022**, *528*, 231026. [[CrossRef](#)]
11. Lam, R.; Allaire, D.L.; Willcox, K.E. Willcox. Multifidelity optimization using statistical surrogate modeling for non-hierarchical information sources. In *Proceedings of the 56th AIAA/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, Kissimmee, FL, USA, 5–9 January 2015.
12. Daulton, S.; Wan, X.; Eriksson, D.; Balandat, M.; Osborne, M.A.; Bakshy, E. Bayesian optimization over discrete and mixed spaces via probabilistic reparameterization. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 12760–12774.
13. Garrido-Merchán, E.C.; Hernández-Lobato, D. Dealing with categorical and integer-valued variables in bayesian optimization with gaussian processes. *Neurocomputing* **2020**, *380*, 20–35. [[CrossRef](#)]
14. Candelieri, A.; Ponti, A.; Archetti, F. Fair and green hyperparameter optimization via multi-objective and multiple information source Bayesian optimization. *Mach. Learn.* **2024**, *113*, 2701–2731. [[CrossRef](#)]
15. Mikkola, P.; Martinelli, J.; Filstroff, L.; Kaski, S. Multi-fidelity bayesian optimization with unreliable information sources. In *Proceedings of the International Conference on Artificial Intelligence and Statistics*, Valencia, Spain, 25–27 April 2023; pp. 7425–7454.
16. Takeno, S.; Fukuoka, H.; Tsukada, Y.; Koyama, T.; Shiga, M.; Takeuchi, I.; Karasuyama, M. Multi-fidelity Bayesian optimization with max-value entropy search and its parallelization. In *Proceedings of the International Conference on Machine Learning, Virtual Event*, 13–18 July 2020.
17. Song, J.; Chen, Y.; Yue, Y. A general framework for multi-fidelity bayesian optimization with gaussian processes. In *Proceedings of the The 22nd International Conference on Artificial Intelligence and Statistics*, Okinawa, Japan, 16–18 April 2019; pp. 3158–3167.
18. Ghosh, S.; Kristensen, J.; Zhang, Y.; Subber, W.; Wang, L. A strategy for adaptive sampling of multi-fidelity gaussian processes to reduce predictive uncertainty. In *Proceedings of the International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, Anaheim, CA, USA, 18–21 August 2019; Volume 59193, p. V02BT03A024, American Society of Mechanical Engineers.
19. Zhang, Y.; Park, S.; Simeone, O. Multi-Fidelity Bayesian Optimization with Across-Task Transferable Max-Value Entropy Search. *arXiv* **2024**, arXiv:2403.09570.
20. Foumani, Z.Z.; Yousefpour, A.; Shishehbor, M.; Bostanabad, R. On the Effects of Heterogeneous Errors on Multi-fidelity Bayesian Optimization. *arXiv* **2023**, arXiv:2309.02771.
21. Baptista, R.; Poloczek, M. Bayesian optimization of combinatorial structures. In *Proceedings of the International Conference on Machine Learning*, Stockholm, Sweden, 10–15 July 2018.
22. Wu, T.C.; Flam-Shepherd, D.; Aspuru-Guzik, A. Bayesian Variational Optimization for Combinatorial Spaces. *arXiv* **2020**, arXiv:2011.02004.
23. Notin, P.; Hernández-Lobato, J.M.; Gal, Y. Improving black-box optimization in VAE latent space using decoder uncertainty. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 802–814.
24. Grosnit, A.; Tutunov, R.; Maraval, A.M.; Griffiths, R.R.; Cowen-Rivers, A.I.; Yang, L.; Zhu, L.; Lyu, W.; Chen, Z.; Wang, J.; et al. High-dimensional Bayesian optimisation with variational autoencoders and deep metric learning. *arXiv* **2021**, arXiv:2106.03609.
25. Deshwal, A.; Belakaria, S.; Doppa, J.R.; Fern, A. Optimizing discrete spaces via expensive evaluations: A learning to search framework. *Proc. AAAI Conf. Artif. Intell.* **2020**, *34*, 3773–3780. [[CrossRef](#)]
26. Deshwal, A.; Belakaria, S.; Doppa, J.R. Scalable combinatorial Bayesian optimization with tractable statistical models. *arXiv* **2020**, arXiv:2008.08177.
27. Deshwal, A.; Belakaria, S.; Doppa, J.R. Mercer features for efficient combinatorial Bayesian optimization. *Proc. AAAI Conf. Artif. Intell.* **2021**, *35*, 7210–7218. [[CrossRef](#)]
28. Deshwal, A.; Doppa, J. Combining Latent Space and Structured Kernels for Bayesian Optimization over Combinatorial Spaces. *Advances in Neural Information Processing Systems*. 2021, p. 34. Available online: <https://ask.qcloudimg.com/draft/8436237/7jou7wsmpp.pdf> (accessed on 4 October 2024).
29. Frazier, P.I. Bayesian Optimization. *Recent Advances in Optimization and Modeling of Contemporary Problems*. *Inform.* 2018, pp. 255–278. Available online: <https://pubsonline.informs.org/doi/abs/10.1287/educ.2018.0188> (accessed on 4 October 2024).
30. Archetti, F.; Candelieri, A. *Bayesian Optimization and Data Science*; Springer: Cham, Switzerland, 2019.
31. Candelieri, A. A gentle introduction to bayesian optimization. In *Proceedings of the 2021 Winter Simulation Conference (WSC)*, Phoenix, AZ, USA, 12–15 December 2021; IEEE: New York, NY, USA, 2021.
32. 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](#)]
33. Wang, Z.; Jegelka, S. Max-value entropy search for efficient Bayesian optimization. In *Proceedings of the International Conference on Machine Learning*, Sydney, Australia, 6–11 August 2017.
34. Takeno, S.; Nomura, M.; Karasuyama, M. Towards practical preferential Bayesian optimization with skew Gaussian processes. In *Proceedings of the International Conference on Machine Learning*, Honolulu, HI, USA, 23–29 July 2023.

35. Moss, H.B.; Leslie, D.S.; Gonzalez, J.; Rayson, P. Gibbon: General-purpose information-based bayesian optimisation. *J. Mach. Learn. Res.* **2021**, *22*, 1–49.
36. Picheny, V.; Moss, H.; Torossian, L.; Durrande, N. Bayesian quantile and expectile optimisation. In Proceedings of the Uncertainty in Artificial Intelligence, Eindhoven, The Netherlands, 1–5 August 2022.
37. Cakmak, S.; Astudillo Marban, R.; Frazier, P.; Zhou, E. Bayesian optimization of risk measures. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 20130–20141.
38. NaseriNaserizade, 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](#)]
39. Yu, P.D.; Tan, C.W.; Fu, H.L. Epidemic source detection in contact tracing networks: Epidemic centrality in graphs and message-passing algorithms. *IEEE J. Sel. Top. Signal Process.* **2022**, *16*, 234–249. [[CrossRef](#)]
40. Gangireddy, S.C.R.; P, D.; Long, C.; Chakraborty, T. Unsupervised fake news detection: A graph-based approach. In Proceedings of the 31st ACM Conference on Hypertext and Social Media, Virtual Event, 13–15 July 2020.
41. Eriksson, D.; Pearce, M.; Gardner, J.; Turner, R.D.; Poloczek, M. Scalable Global Optimization via Local Bayesian Optimization. *Advances in Neural Information Processing Systems*. 2019, p. 32. Available online: <https://proceedings.neurips.cc/paper/2019/file/6c990b7aca7bc7058f5e98ea909e924b-Paper.pdf> (accessed on 4 October 2024).
42. Papenmeier, L.; Nardi, L.; Poloczek, M. Increasing the scope as you learn: Adaptive Bayesian optimization in nested subspaces. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 11586–11601.
43. Papenmeier, L.; Nardi, L.; Poloczek, M. Bounce: Reliable high-dimensional Bayesian optimization for combinatorial and mixed spaces. *Adv. Neural Inf. Process. Syst.* **2023**, *36*, 1764–1793.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.