

Monitoring Water Resources through a Bayesian Optimization-based Approach using Multiple Surface Vehicles: The Ypacarai Lake Case Study

1st Federico Peralta

*Department of Electronic Engineering
Technical School of Engineering of Seville
Seville, Spain
fperalta@us.es*

2nd Samuel Yanes Luis

*Department of Electronic Engineering
Technical School of Engineering of Seville
Seville, Spain
syanes@us.es*

3rd Daniel Gutiérrez Reina

*Department of Electronic Engineering
Technical School of Engineering of Seville
Seville, Spain
dgutierrezreina@us.es*

4th Sergio Toral

*Department of Electronic Engineering
Technical School of Engineering of Seville
Seville, Spain
storal@us.es*

Abstract—Bayesian optimization is a popular sequential decision strategy that can be used for environmental monitoring. In this work, we propose an efficient multi-Autonomous Surface Vehicle system capable of monitoring the Ypacarai Lake (San Bernardino, Paraguay) (60 km²) using the Bayesian optimization approach with a Voronoi Partition system. The system manages to quickly approximate the real unknown distribution map of a water quality parameter using Gaussian Processes as surrogate models. Furthermore, to select new water quality measurement locations, an acquisition function adapted to vehicle energy constraints is used. Moreover, a Voronoi Partition system helps to distributing the workload with all the available vehicles, so that robustness and scalability is assured. For evaluation purposes, we use both the mean squared error and computational efficiency. The results showed that our method manages to efficiently monitor the Ypacarai Lake, and also provides confident approximate models of water quality parameters. It has been observed that, for every vehicle, the resulting surrogate model improves by 38%.

Index Terms—Bayesian Optimization, Gaussian Processes, Data Acquisition, Environmental Monitoring, Multi-robot Informative Path Planning, Autonomous Vehicles, Voronoi Diagrams

I. INTRODUCTION

Water resources, such as lakes and lagoons, are a vital component in life. They serve as wildlife bed and act as water reservoirs. A failure in maintaining water resources

This work has been partially funded by the Spanish “Ministerio de Ciencia, innovación y Universidades, Programa Estatal de I+D+i Orientada a los Retos de la Sociedad” under the Project “Despliegue Adaptativo de Vehículos no Tripulados para Gestión Ambiental en Escenarios Dinámicos RTI2018-098964-B-I00”, and by the regional government Junta de Andalucía under the Projects “Despliegue Inteligente de una red de Vehículos Acuáticos no Tripulados para la monitorización de Recursos Hídricos US-1257508” and “Despliegue y Control de una Red Inteligente de Vehículos Autónomos Acuáticos para la Monitorización de Recursos Hídricos Andaluces PY18-RE0009”.

free from contamination triggers unhealthy situations for the environment. This is the current state of Ypacarai Lake [1] (San Bernardino, Paraguay). Its water has periodical blooms of cyanobacteria caused by the rapid enrichment of nutrients, known as eutrophication [1]. Eutrophication as well as cyanobacteria are harmful for the fauna and flora; consequently, the current scenario is not desirable and actions to significantly improve the situation should be taken. The first step towards eradicating this toxic alga is to continuously monitor the state of water quality parameters of the Ypacarai Lake, e.g., levels of pH, turbidity, CO₂, Dissolved Oxygen, etc. Collecting data about such variables can enhance the decision-making towards cleaning and recovering the Lake.

Traditional monitoring systems of water resources include manual water sampling and continuous supervision using fixed stations. Recently, the use of Autonomous Surface Vehicle (ASVs) has been proposed [2], [3] for continuous monitoring and water quality measurements. In addition, using ASVs for monitoring tasks presents various advantages when compared with monitoring stations and manual water sampling. First, an ASV can autonomously move to any location within of a Lake, while fixed stations can only measure in one place. Second, in hazardous environments (locations with high levels of toxicity), an ASV may be in direct contact with the water, while humans cannot. However, a major inconvenient arises when using ASVs for monitoring: the performance of the system depends directly on the autonomy of the vehicle [4], [5]. A practical way to overcome this issue is to monitor the lake using a multi-robot system.

Using a fleet of ASVs to accomplish the objective of water quality monitoring greatly improves the results when compared to single ASV monitoring systems. Nevertheless, several important aspects should be considered: i) it is required

an appropriate intelligent system that distributes the work among the available ASVs, ii) redundancy on data collection should be avoided to speed up the procedure. Therefore, data acquisition should be cooperative. In this work, we propose an intelligent multi-ASV system, where the ASVs are in charge of efficiently performing measurements of a certain water quality parameter using a Bayesian Optimization-based approach. Moreover, in order to assure cooperation, a Voronoi Partition system is included. The Voronoi Partition or Voronoi Diagram (VD) is a mathematical expression, which defines regions according to a set of generators. If the positions of the ASVs serve as generators, each region has an ASV assigned to it and the Bayesian Optimization (BO) technique can be utilized seamlessly so that each ASV picks new measurement locations within each one of the defined regions.

In the context of this work, BO can be defined as a sequential decision strategy that manages to optimize a black-box or unknown real function, in our case, water quality profile of the Ypacarai Lake (e.g., spatial distribution maps of pH, temperature, turbidity, etc.). BO infers a supposed model of the real function and uses it to define a point where the function is optimized according to an exploration/exploitation balance. Given these two main characteristics, BO for environmental monitoring can be used not only as an intelligent informative path planning system but also as a water quality mapping system. In general, BO provides results with the use of i) a surrogate model and ii) an acquisition function. The surrogate model corresponds to an expression that approximates or mimics the real distribution (water quality parameter distribution) using functions that are less expensive to evaluate than the original [6]. Without any prior information, a Gaussian Process (GP) as surrogate model can offer a supposed response of the black-box function, or in this case, of a certain water quality model. However, if any information (i.e., measurements of a water quality parameter of the Ypacarai Lake) is known, GPs can be updated in order to infer a more realistic surrogate model. Finally, due to the fact that GP Regressions (GPR) are based on normal distributions [6], GPs as surrogate models have the additional value of providing uncertainty levels or confidence measures of the produced regressions.

In order to provide GPs with updated information, new measurements locations within the Ypacarai Lake should be chosen. The process of selecting new informative measurement locations is a task executed by Acquisition Functions (AFs). AFs are commonly used in BO approaches because they can weight the future rewards of selecting points for evaluation. For ASVs monitoring systems, classical AFs (Expected Improvement, Probability of Improvement, etc. [6]) can provide unreachable locations (due to energy constraints), so adaptations of the AFs are needed in this system. A study of possible adaptations is shown in a previous work [7], and, here, the selection is based on its conclusions. Moreover, this work is based on the mentioned previous work [7], which proposes a BO approach for environmental monitoring using a single ASV.

The ASV team can use the mentioned system as is and

start monitoring the Lake, but eventually, they will start to perform measurements on the same locations (or very near locations). To solve this issue, the VD is used to avoid redundant information gathering and also to allow robustness and scalability. Our implementation carries out this task through a Centralized Coordinator (CC), designed so that the multi-robot system cooperates and performs asynchronous measurements of the water quality of the studied Lake.

The proposed system includes the CC in order to maintain an organized sequential decision-making system: i) VD partitioning, ii) GP regression iii) AF evaluation and iv) performing new measurements. But only the last step is performed asynchronously, i.e., each ASV can perform measurements independently of the state of the other vehicles.

The main contributions of this work are:

- A Voronoi-based coordination system for multiple ASVs goal definition.
- A multi-ASV BO-based system for environmental monitoring and parameter modelling.

The present paper continues as follows: First, the related works are included in Section II. Next, Section III describes the problem as well as some assumptions and the roadmap to real life applications. Section IV defines our proposed approach, the mathematical and the algorithmic design. In Section V, the simulation setup and the performed tests are presented and discussed. Finally, Section VI contains the conclusion and future works.

II. RELATED WORK

Using autonomous vehicles for environmental monitoring is an ongoing research topic. Usually, the vehicles are used to execute Multi-robot Informative Path Planning (MIPP) [5], [8], which describes paths that maximize information gathering. For example, in [8], a MIPP under continuous connectivity constraints is designed. In [5], the MIPP is used for explicit environmental monitoring applying adaptive routing. Another environmental monitoring MIPP design is presented in [4], where continuous region partitioning is used. Region partitioning helps the multi-robot system to achieve better workload balance [4]. A recent study [9] shows that the design of a Multi-Autonomous Underwater Vehicle (AUV) coordinator can be based on dynamic VD as well as GPs. The mentioned work uses Root Mean Squared Error and Negative Log-Likelihood to test their proposed method. In [10], accurate inference and prediction of plants phenotype is made using a multi-Autonomous Ground Vehicle (AGV) system based on GPs and MIPPs. Multiple-ASV are used for large-scale environmental monitoring in [11], where a Travelling Salesman Problem (TSP) based algorithm is proposed. Table I shows a brief summary of Multi-Robot Environmental Monitoring systems using autonomous vehicles. The works present different types of MIPPs. As shown in Table I, while they all manage to acquire information, the specific objective requires different monitoring approaches.

Some of the environmental monitoring approaches using a single vehicle found in the literature can also be adapted to

TABLE I: Brief summary of MIPP using autonomous vehicles.

Ref.	Specific Objective	Monitoring Algorithm	Vehicle	Year
[8]	Exploration under Connectivity Constraints	Bipartite Graph Formulation	Generic Robot	2019
[5]	Ocean temperature environmental monitoring	Dynamic graph-based routing	Underwater Vehicle	2018
[9]	Algae Bloom Monitoring	Decentralized adaptive informative sampling	Underwater Vehicle	2017
[10]	Plant phenotyping	Entropy-based MIPP	Ground Vehicle	2018
[11]	Complete coverage of known environments	Traveling salesman problem k-TSP-formulation	ASV	2018

multi-ASV systems. Such is the case of the work found in [12], where an adaptive Informative Path Planning (IPP) system is proposed using a single ASV and a roadmap to multi-ASV implementation. In [13], a single Autonomous Aerial Vehicle (AAV) is used to monitor terrain environments using a covariance matrix adaptation evolution strategy as main algorithm and GPs as the underlying surrogate model.

Using a single ASV for monitoring the Ypacarai Lake are proposed by several works. In [14], the problem is solved by applying a TSP modelling with evolutionary algorithms. The mentioned work is expanded in [15], using Eulerian circuits so that efficient continuous monitoring is performed. In [3], a Deep Reinforcement Learning method is proposed so that an optimal patrolling policy provides efficient monitoring. These works proposes that multi-robot systems can benefit from their proposed methods due to the fact that large-scale water bodies such as Ypacarai Lake cannot be exhaustively monitored using a single ASV. In [7], a single ASV uses BO approach to efficiently perform measurements on the Ypacarai Lake. The present work is based upon [7].

The related works included in this section provide theoretical bases to design and implementation of our proposed method. Although most of the MIPPs shown in Table I proposed GP as surrogate models, none of them implement BO procedures to obtain information. This work also takes into account energy constraints, so that the performance evaluations can provide insights for real life experiments.

III. STATEMENT OF THE PROBLEM

The water quality parameters of the Ypacarai Lake must be sampled efficient and intelligently, so that confident water quality models can be obtained using a system of multiple ASVs. The region to be monitored corresponds to a subspace of the \mathbb{R}^2 space. Therefore, the vehicles can be located anywhere within the boundaries defined by the model of the Ypacarai Lake. The proposed system provides (see (1)) a confident approximation $\hat{f}(\mathbf{x})$ of a real water quality behavior $f(\mathbf{x})$, where \mathbf{x} is the full set of locations (x, y) within the map model.

$$f(\mathbf{x}) \approx \hat{f}(\mathbf{x}) \quad (1)$$

In order to acquire data, there is a total of k available ASVs for the mission. They are equipped with positioning systems that provide their location $p \in \mathbf{x}$ within the mentioned map model. Each ASV is also equipped with a sensor that measures and stores the value s of a water quality parameter, as well as the position where the measurement has been performed. For performance evaluations, the measurement is stored in a set $D = \{(p_i, s_i) \mid i = 1, 2, \dots, N\}$, which contains a total of N collected data. Additionally, the data is used to represent discrete values of the real function, as shown below

$$s_i = f(p_i) \quad (2)$$

If the surrogate model regression is appropriately fitted with data D , the complete water quality map can be approximated using (1) and (2). The fitness level can be expressed using the Mean Squared Error (MSE) function, as shown in (3). This expression returns the inverse value of how well the surrogate model behaves, accentuating on the differences between good and bad approximations.

$$\text{MSE}(f(\mathbf{x}), \hat{f}(\mathbf{x})) = \frac{1}{n} \sum_{i=1}^n (f(\mathbf{x}_i) - \hat{f}(\mathbf{x}_i))^2 \quad (3)$$

A. Assumptions, Considerations and Design

Some assumptions are considered so that the scope of the work is appropriately defined. Fig. 1. shows the system design, where two main components are defined: i) the centralized coordinator and ii) the ASV system. Next, the system design as well as the assumptions and considerations are explained in detail.

- **Ypacarai Lake Model:** A matrix \mathcal{M} is used to model the monitoring space. The matrix \mathcal{M} assumes a grid-like configuration model so that each element $\mathcal{M}_{i,j}$ of the matrix corresponds to a square with side d . Each square has an associated real-world location which corresponds to the center of the square and also has a occupancy state value. It is set with a value of 1 if the real square is occupied by an obstacle or it is terrain and a value of 0 if the square corresponds to a navigable water surface area. The map model of the Ypacarai Lake is shown in the next section (Fig. 2.) where it is in use by system.
- **Multi-robot coordinator system:** Fig. 1 shows the components of the CC. The CC is responsible for data storage, surrogate model or Gaussian process fitting, and evaluation of acquisition functions. As a result, new measurement locations are chosen. This centralized coordination ensures that all vehicles share the same information, and as a consequence, the vehicles will not make redundant measurements. It can be observed that the proposed multi-robot system is decentralized only for performing measurements (each vehicle should perform measures independent of the state of the other vehicles) so that the system has an increased robustness.
- **Guidance, Navigation and Control (GNC) system:** Each vehicle has a dedicated GNC system (as shown in

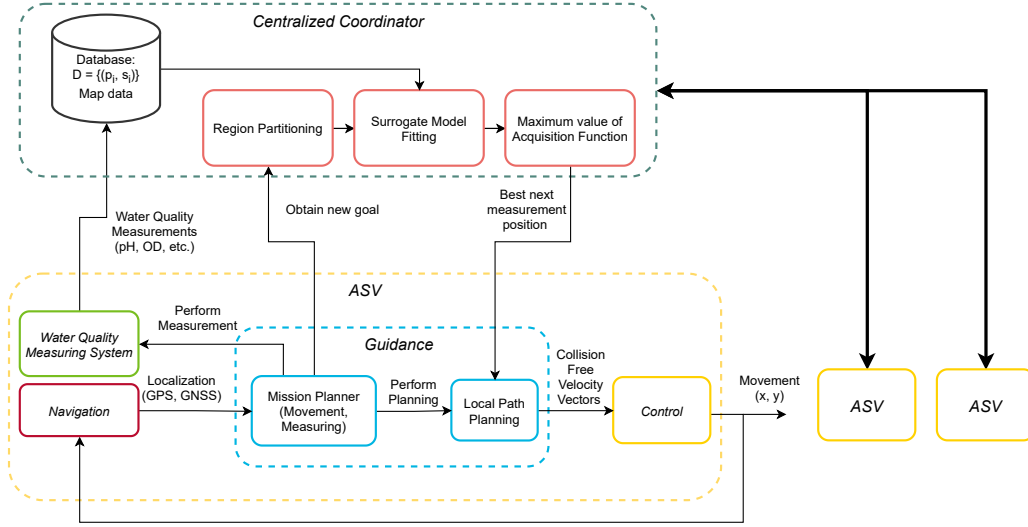


Fig. 1: Proposed multi-ASV system. The centralized coordinator is in charge of obtaining new measurement positions. The coordinator receives and processes the information but the guidance, navigation and control is processed on each of the vehicles.

Fig. 1.) that ensures that the vehicle can travel between locations on the surface of the Lake. The Navigation subsystem allows for positioning, but not perfectly (it is assumed to have a resolution $\approx d$), as real life positioning systems are susceptible to errors. For simulation purposes, the positioning error is simulated by shifting the final position of vehicles before performing a measurement by a few meters (< 25).

- **Water quality measuring system:** The ASV has a subsystem that can measure several water quality parameters (power of Hydrogen, Dissolved Oxygen, Turbidity, etc.) and store their values in a database. However, in this work only one value is used by the BO system. Moreover, the s_i value is the result of preprocessing the measurement performed so that all the possible values s_i , $\forall i \in N$ conform a set with standard attributes ($\mu(\mathbf{x}) = 0$, $\sigma(\mathbf{x}) = 1$). The preprocessing algorithm should follow the Central Limit Theorem (CLT) so that this assumption is true for real cases.
- **ASV autonomy:** The vehicle is equipped with a battery that provides enough autonomy for 2.1 hours, which translates to approximately 15000 linear meters of autonomy, according to [7]. An ASV that travels a total distance of 15000 meters cannot perform any movement nor measurements.

IV. PROPOSED APPROACH

The general procedure of our centralized monitoring system is shown in Algorithm 1. Whenever an ASV needs a new goal (line 4), a Voronoi region is defined using the positions of the available ASVs (line 5). Next, with the available data D the surrogate model is obtained (line 6) and used to find the optimal measurement location within the Voronoi region of each ASV (line 7). Additionally, the CC stores new data

$D_{i+1} \leftarrow (s_{i+1}, p_{i+1}) = (s_k, p_k)$ if it is available and proceeds to fit the model (lines 9-11).

Algorithm 1: Centralized Coordinator procedure.

```

1 initialize:  $D \leftarrow \text{Database}$ ,  $\mathcal{M} \leftarrow \text{Map Data}$ ;
2 while  $ASVs\_available()$  do
3   for  $asv$  in  $ASVs$  do
4     if  $asv.needs\_new\_goal()$  then
5        $k \leftarrow asv.id$ ;
6        $R_k \leftarrow VD(ASVs, asv, \mathcal{M})$ ;
7        $\hat{f}(\mathbf{x}) \leftarrow GPR(\mathbf{x})$ ;
8        $p_k \leftarrow \text{find\_max}(AF(\hat{f}(\mathbf{x})), R_k)$ ;
9     end
10    if  $asv.performed\_new\_measurement()$  then
11       $D_{i+1} \leftarrow asv.measurement$ ;
12       $GP.fit(D)$ ;
13    end
14  end
15 end

```

Finally, each vehicle executes asynchronously Algorithm 2. Whenever a new goal location is obtained (line 4), an ASV manages to perform a movement in order to reach the destination goal (line 10). If the ASV reaches the optimal position, it performs a measurement and sends the data to update the general water quality model of the lake (lines 6-8).

A. Initial Vehicle Positioning

Due to the fact that VDs will be used to define regions of coverage, the ASVs should be positioned so that the initial regions of all the vehicles are similar in size and shape. A mathematical approach to solve this problem is to distribute the vehicles in the perimeter of a circle of radius r centered at the

Algorithm 2: Mission planning component of the Guidance subsystem of the ASV.

```

1 initialize:  $k \leftarrow \text{asv.id}$ ;
2 while  $\text{asv.is\_online}()$  do
3   if  $\text{asv.has\_no\_goal}()$  then
4      $\text{asv.goal} \leftarrow p_k$ ;
5   else
6     if  $\text{asv.reached\_goal}()$  then
7        $s_k \leftarrow \text{obtain\_new\_measurement}()$ ;
8        $\text{asv.measurement} \leftarrow (s_k, p_k)$ ;
9     else
10       $\text{asv.perform\_movement}(\text{goal})$ ;
11    end
12  end
13 end

```

center of the Lake, evenly separated angle-wise, in an attempt to have each vehicle aimed to cover different regions. Due to the fact that the ASVs will be evenly spaced, the VD will provide similar initial regions for all the vehicles. However, for real scenarios, it is not desirable that the ASVs start the mission at the center of the Lake, therefore this work extends the proposed idea and relocates the initial positions to the shore of the Lake taking into account the spacing between vehicles, having all vehicles are evenly separated angle-wise but with different distances in respect to the center of the lake according to the shoreline. In Fig. 2., four vehicles are located in the shoreline of the Ypacarai Lake, they are evenly distributed angle-wise. The regions, to be discussed in the next sub-section, are similar in size and shape.

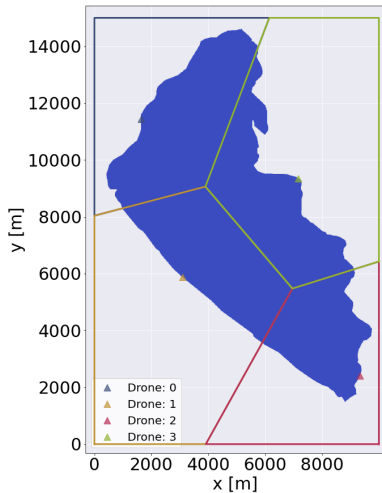


Fig. 2: Example of initial positioning using four ASVs. The vehicles are evenly distributed and, when used as generators, they generate similar Voronoi Regions. The sum of the surface of the regions is equal to the surface of the model of the Ypacarai Lake.

B. Voronoi Diagram

Having appropriately defining the initial positioning, the next procedure is to define the regions of the vehicles so that each one can obtain an optimal measurement location. In this work, we propose that this partition is done using the Voronoi partitioning or Voronoi Diagram (VD).

VD is a partition of a n -dimensional space into regions of the same dimension. In this work, the 2D space of the map \mathcal{M} is partitioned into m regions. Each region has a generator V_k , which corresponds to the position p_k of the k ASV. The VD expression has the form of

$$R_k = \{\mathbf{x} \in \mathbb{R}^2 \mid d(\mathbf{x}, V_k) \leq d(\mathbf{x}, V_j) \forall j \neq k\} \quad (4)$$

To adjust the generated regions so that all of them contain only points that are within the sub-space \mathcal{M} , more formally $\mathbf{x} \in \mathcal{M}$, additional virtual generators are introduced in such form that the initial regions contain all the points that define \mathcal{M} and only them. For example, for a set of 4 vehicles, the resulting initial VD is shown in Fig. 2. As mentioned, the regions are bounded to the map. However, these regions are updated when needed.

C. Bayesian Optimization

Once a vehicle V_k obtained a region R_k to monitor, the next step is to obtain an informative position of measurement. This position $\mathbf{x}_i = (x_i, y_i)$ corresponds to the maximum value of a function of our supposed model of the Lake. The surrogate water quality function of the Lake is modelled as a GP, as it can easily fit real data and produce an output regression with an associated uncertainty [6].

1) *Gaussian Process Regressions:* Gaussian Processes are stochastic or random processes, where each variable (or feature) has a multivariate random distribution. In this work, we define the variables as the components of 2-D plane that represent the locations on the real world. The main components of GP are the mean function $\mu(\mathbf{x})$ and the kernel or covariance function $k(\mathbf{x}, \mathbf{x}')$. The former represents the expected output of the black-box function, while the latter defines the correlations of the features or inputs [6].

Kernel Functions define similarities/covariances between different input values considered as random variables (\mathbf{x}). The kernel is usually presented as a semipositive definite matrix, where each element corresponds to the covariance value of two inputs $k_{\mathbf{x}, \mathbf{x}'}$. The hyperparameters of the kernel plays an important role and they are set up by using marginal likelihood maximization [16]. Our previous work [7] shows that for smooth functions, the best kernel for this particular matter is the Radial Basis Function (RBF), which has the following expression

$$k_{\mathbf{x}, \mathbf{x}'} = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|}{2\ell^2}\right) \quad (5)$$

RBF has a single hyperparameter ℓ , which defines the smoothness of the function. This characteristic length-scale provides a measure of how far apart two different locations

can be so that they influence each other. With higher values of ℓ , the posterior model regression is smoother and has a slower rate of change.

Having a defined kernel and a set of measured data, the GPR is developed by first fitting the input data [6], [16]. Finally, the unknown response $\mu(\mathbf{x}_*), \sigma(\mathbf{x}_*)$ of our surrogate model $\hat{f}(\mathbf{x}_*)$ given data D can be inferred using the next expressions:

$$\mu_{\hat{f}(\mathbf{x}_*)|D} = K_*^T K^{-1} f(\mathbf{x}) \quad (6)$$

$$\sigma_{\hat{f}(\mathbf{x}_*)|D} = K_{**} - K_*^T K^{-1} K_* \quad (7)$$

Where K , K_{**} and K_* are taken from the fitted (hyper-parameters updated with marginal likelihood maximization) kernel that includes covariances between known data $k(\mathbf{x}, \mathbf{x})$, unknown data $k(\mathbf{x}_*, \mathbf{x}_*)$ and between the known and unknown data $k(\mathbf{x}, \mathbf{x}_*)$.

$$K = \begin{bmatrix} K & K_* \\ K_*^T & K_{**} \end{bmatrix} = \begin{bmatrix} k(\mathbf{x}, \mathbf{x}) & k(\mathbf{x}, \mathbf{x}_*) \\ k(\mathbf{x}_*, \mathbf{x}) & k(\mathbf{x}_*, \mathbf{x}_*) \end{bmatrix} \quad (8)$$

2) *Acquisition Function*: Once the surrogate posterior model of the water quality parameter is obtained, the next step is to select a new measurement location according to a function that evaluates the information gain from the possible measurement locations within the lake. In the general case, the GP regression is used as argument of the Acquisition Function (AF) as GPs provide both an inference and uncertainty of the result. There are several AFs found in the literature [6], but our previous experiments found in [7] show that one of the best acquisition functions for water quality monitoring of the Ypacarai Lake is the Expected Improvement (EI). EI evaluates the improvement optimization by weighting current values of the surrogate model according to the Cumulative Density Function $\Phi(\mathbf{x})$ and the Probabilistic Density Function $\phi(\mathbf{x})$. In order to use these functions, the next expression is used to normalize the input arguments

$$Z = \frac{f(\mathbf{x}^+) - \mu(\mathbf{x}) - \xi}{\sigma(\mathbf{x})} \quad (9)$$

Where $f(\mathbf{x}^+)$ corresponds to the best value so far and ξ is an exploration/exploitation explicit parameter. Z appropriately normalizes the input and, with the use of (10), the $EI(\mathbf{x})$ of performing a water quality measurement on location \mathbf{x} can be obtained.

$$EI(\mathbf{x}) = \begin{cases} 0 & \text{if } \sigma(\mathbf{x}) = 0 \\ (f(\mathbf{x}^+) - \mu(\mathbf{x}) - \xi)\Phi(Z) & \\ +\phi(Z)\sigma(\mathbf{x}) & \text{else} \end{cases} \quad (10)$$

The predictive mean (found in the first term of (10)) manages the exploitation weights, while the predictive uncertainty $\sigma(\mathbf{x})$ provides more exploration weight as it increases.

In order to find the optimal location, usually an optimizer like the limited-memory Broyden-Fletcher-Goldfarb-Shanno algorithm (L-BFGS) is used, but in our case, the use of

efficient optimizers is discarded. The main reason is due to the fact that optimizers rely on continuous and unbounded functions. But our map \mathcal{M} is not only undefined in terrain zones but also presents discontinuity on the shores of the Lake. In that sense, the practical method of finding an optimal next measurement position is to value all navigable locations in $\mathcal{M} \cap R_k$ and select the maximum value as the optimal next location.

Furthermore, EI function as other classical AFs presents issues when used for monitoring tasks due to the fact that it does not account for distance travelled by the ASVs. If the locations are very far from each other, the potential total amount of measurements decreases. To overcome this situation, our previous work [7] studied different adaptations so that the ASV can perform multiple measurements. The truncated adaptation performed very well in the performance tests and it is the selected adaptation in this work. It consists on stopping the movement of the ASV after a distance l when travelling in the direction of the best position according to the AF. Whenever the ASV travels the defined distance, the ASV performs a measurement on its current position and updates the data. Generally, the new data modifies the generated surrogate regression, which provides a new optimal measurement location.

V. PERFORMANCE EVALUATION

The multi-ASV proposed system is evaluated in this section. The evaluation consists mainly on observing how well the GP fits the data and produces the surrogate posterior model using the MSE for comparison. First, the setup for the simulations is defined, next the results are shown. The code for this section was developed in Python 3.8.4 and is available online ¹. The simulations have been performed in a laptop computer with 32GB RAM, Intel i7 3.2 GHz processor.

A. Simulation Setup

The Ypacarai Lake map is modelled as \mathcal{M} with dimensions 1000×1500 , where each element $\mathcal{M}_{i,j}$ corresponds to a square with side $d \sim 10[m]$, centered at position \mathbf{x} . The associated real function of water quality maps, such as temperature, pH, turbidity levels, will be smooth due to the real-life behavior of fluid dynamics and wind conditions [1], [7]. In that sense, a simulated ground truth should be based on a smooth benchmark function such as the Bohachevsky or Shekel functions. The latter is very useful for testing generalized solutions, as the Shekel Function (SF) is a multimodal, continuous, multidimensional function. Additionally, the maximum locations are defined with two input arguments.

The SF has the expression shown in (11), where c_i and a_{ij} correspond to the elements of two given set of matrices c and A . This input matrices define the number and positions of the maximum locations (matrix A with size $M \times N$, M as the number of maximum locations, which are N -dimensional points) and their respective inverse importance value c_i grouped in vector c with size $M \times 1$.

¹https://github.com/FedePeralta/BO_ASVs

$$f_{\text{Shekel}}(\mathbf{x}) = \sum_{i=1}^M \frac{1}{c_i + \sum_{j=1}^N (\mathbf{x}_j - a_{ij})^2} \quad (11)$$

For generalization, we tested 10 different ground truth scenarios using the SF with $M \in [2, 6]$ random number of maximum locations. As for the c_i values, we chose a constant value of 5. However, the simulated values receive the same treatment as the assumption for real values explained in Section III-A. In Fig. 3, three out of the ten different ground truth scenarios are shown.

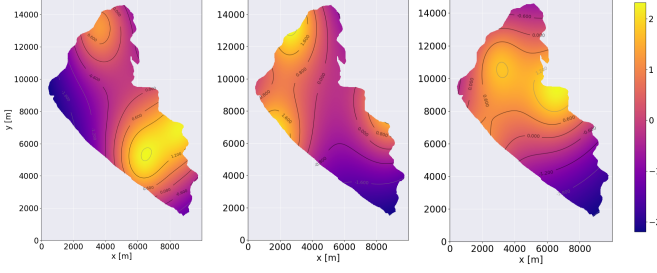


Fig. 3: Three examples of the random Shekel functions taken as ground truth for performance evaluations

The simulations were performed 50 times for each ground truth, using 1, 2, 3, 4 and 5 ASVs. The initial positions were randomized with the same seed. Therefore, for every scenario the same set of 50 initial positions were used. The initial positions also correspond to the first locations where the ASVs perform measurement so that the more ASVs are available at the start, the more initial data the system has. Each simulation stops whenever the average distance travelled by all the considered ASVs is greater than 15000m, but while this condition is not met, if an ASV travels a distance greater than 15000m, it will not be allowed to move any longer.

B. Simulations results

1) *Environmental Monitoring and obtaining the real water quality parameter function:* The resulting MSE per number of measurements performed is shown in Fig. 4. Provided with the VD, our method efficiently distributes the monitoring using all the available ASVs. For the initial measurements, there exists some advantages while using a low number of ASVs, but as the number of measurements increases, with more ASVs available, a better monitoring is achieved.

It has been observed that the system decreases the MSE by an average of 38% with every new ASV added to the system. Fig. 5 shows an example of the proposed method using different number of ASVs. It can be observed in Fig. 5 that the GPs manage to correctly fit the data and approximate the real function with ease. Furthermore, the VD enhances robustness and provides scalability. Without the VD, the vehicles will perform measurements on the same locations, and consequently, provide redundant information. With the proposed VD system, the fleet cooperates while monitoring.

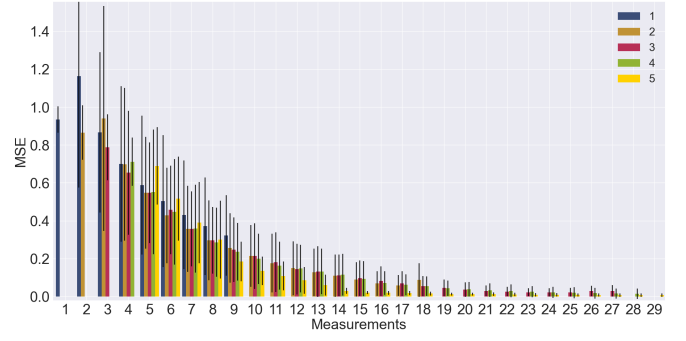


Fig. 4: MSE using different number of ASVs.

2) *Computational Efficiency:* For time efficiency comparisons, the computational time from measurement to measurement have also been evaluated. The obtained results are shown in Fig. 6. This time takes into account the full process of the proposed method (VD, GPR, AF). Therefore, the required time increases with the number of measurements performed, as GPs are non-parametric models with $\mathcal{O}(n^3)$ complexity.

VI. CONCLUSIONS AND FUTURE WORKS

To the best off the authors knowledge, the present work is the first one proposing a BO-based multi-ASV system for water quality monitoring. We improved our previous work [7] by including a centralized coordinator capable of defining Voronoi regions for each one of the available ASVs. The system has proven to be robust against redundancy since the search regions are continuously adapting to the locations of the available ASVs. Moreover, it has been shown that the proposed approach is scalable with the number of vehicles. For future works, the proposed method can be enhanced by taking into account noisy data from the sensors. Additional tests with ASV availability and communication constrains can also be performed.

REFERENCES

- [1] G. A. López Moreira, L. Hinegk, A. Salvatore, G. Zolezzi, F. Hölker, R. A. Monte Domecq S, M. Bocci, S. Carrer, L. De Nat, J. Escribá, et al. Eutrophication, research and management history of the shallow ypacaraí lake (paraguay). *Sustainability*, 10(7):2426, 2018.
- [2] M. Arzamendia, D.s Gregor, D. Gutierrez Reina, and S. Toral. An evolutionary approach to constrained path planning of an autonomous surface vehicle for maximizing the covered area of ypacaraí lake. *Soft Computing*, 23(5):1723–1734, 2019.
- [3] S. Yanes Luis, D. Gutiérrez Reina, and S. Toral Marín. A deep reinforcement learning approach for the patrolling problem of water resources through autonomous surface vehicles: The ypacaraí lake case. *IEEE Access*, 8:204076–204093, 2020.
- [4] A. Dutta, A. Bhattacharya, O. P. Kreidl, A. Ghosh, and P. Dasgupta. Multi-robot informative path planning in unknown environments through continuous region partitioning. *International Journal of Advanced Robotic Systems*, 17(6):1729881420970461, 2020.
- [5] K. Ma, Z. Ma, L. Liu, and G. S. Sukhatme. Multi-robot informative and adaptive planning for persistent environmental monitoring. In *Distributed Autonomous Robotic Systems*, pages 285–298. Springer, 2018.
- [6] F. Archetti and A. Candelieri. *Bayesian Optimization and Data Science*. Springer, 2019.

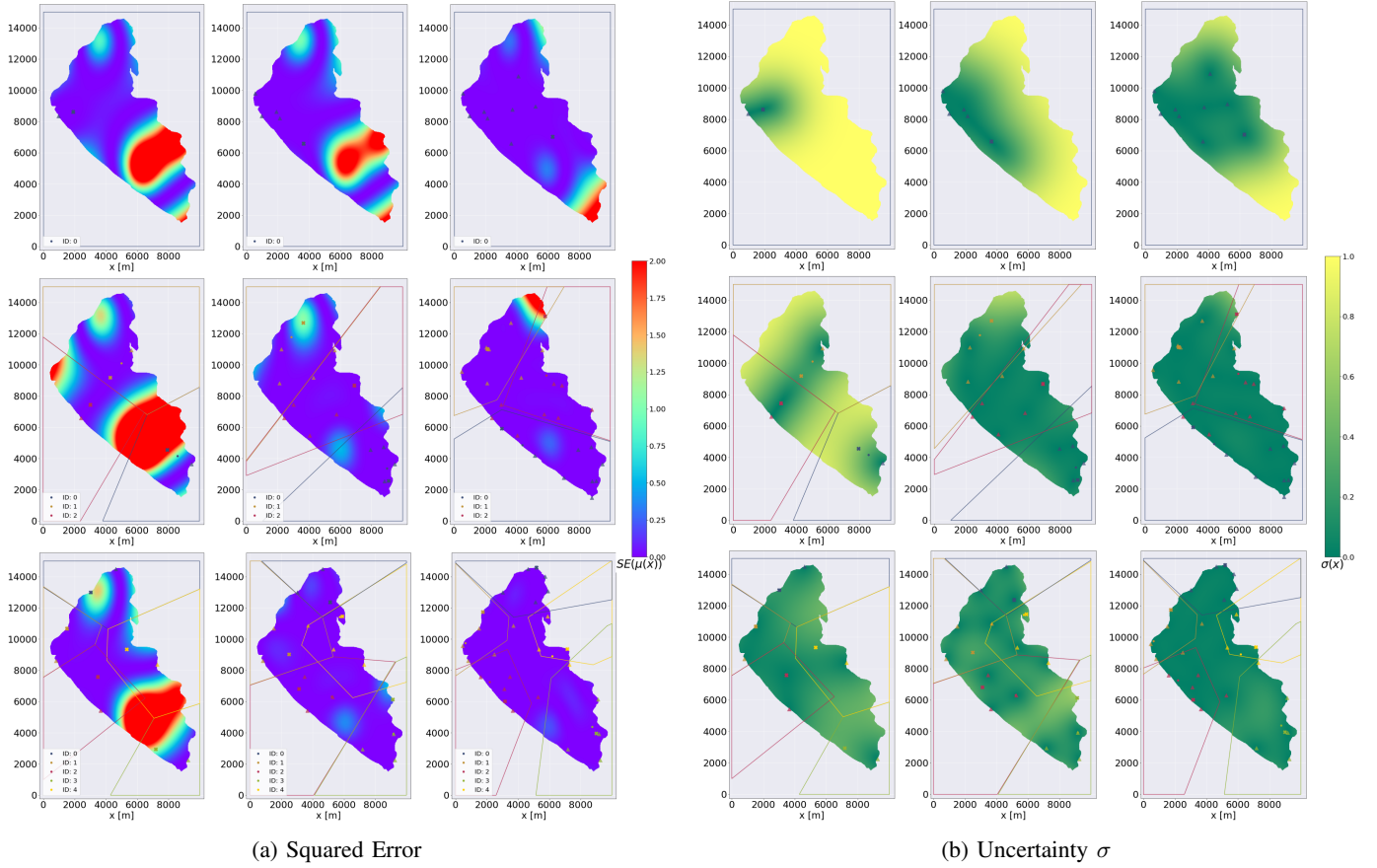


Fig. 5: Evolution of a) Squared Error and b) Uncertainty σ using different number of ASVs (one, three and five ASVs in the first, second and third row, respectively). The results after the first measurement is shown (1st col.) as well as in the middle of the mission (2nd col.) and at the end of the mission (3rd col.) The Voronoi regions are generated on demand, so the regions shown do not necessarily fit the map.

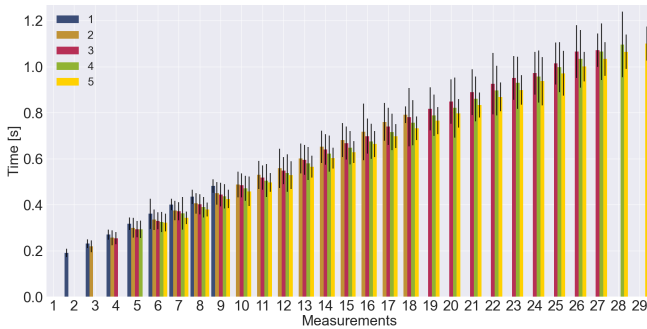


Fig. 6: Average computational time between measurements using different number of ASVs.

- [7] F. Peralta, D. G. Reina, S. Toral, M. Arzamendia, and D. Gregor. A bayesian optimization approach for water resources monitoring through an autonomous surface vehicle: The ypacarai lake case study. *IEEE Access*, 9:1–17, 2021.
- [8] A. Dutta, A. Ghosh, and O. P. Kreidl. Multi-robot informative path planning with continuous connectivity constraints. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 3245–3251. IEEE, 2019.
- [9] S. Kemna, J. G. Rogers, C. Nieto-Granda, S. Young, and G. S. Sukhatme.

- Multi-robot coordination through dynamic voronoi partitioning for informative adaptive sampling in communication-constrained environments. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2124–2130. IEEE, 2017.
- [10] T. Gao, H. Emadi, H. Saha, J. Zhang, A. Lofquist, A. Singh, B. Ganapathysubramanian, S. Sarkar, A. K. Singh, and S. Bhattacharya. A novel multirobot system for plant phenotyping. *Robotics*, 7(4):61, 2018.
- [11] N. Karapetyan, J. Moulton, J. S. Lewis, A. Q. Li, J. M. O’Kane, and I. Rekleitis. Multi-robot dubins coverage with autonomous surface vehicles. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2373–2379. IEEE, 2018.
- [12] G. Hitz, E. Galceran, M. Garneau, F. Pomerleau, and R. Siegwart. Adaptive continuous-space informative path planning for online environmental monitoring. *Journal of Field Robotics*, 34(8):1427–1449, 2017.
- [13] M. Popović, T. Vidal-Calleja, G. Hitz, J. Chung, I. Sa, R. Siegwart, and J. Nieto. An informative path planning framework for uav-based terrain monitoring. *Autonomous Robots*, pages 1–23, 2020.
- [14] M. Arzamendia, D. Gregor, D. Gutierrez Reina, and S. Toral. A path planning approach of an autonomous surface vehicle for water quality monitoring using evolutionary computation. In *Technology for Smart Futures*, pages 55–73. Springer, 2018.
- [15] M. Arzamendia, I. Espartza, D. Gutierrez Reina, S. Toral, and D. Gregor. Comparison of eulerian and hamiltonian circuits for evolutionary-based path planning of an autonomous surface vehicle for monitoring ypacarai lake. *Journal of Ambient Intelligence and Humanized Computing*, 10(4):1495–1507, 2019.
- [16] C. Williams and C. Rasmussen. *Gaussian processes for machine learning*, volume 2. MIT press Cambridge, MA, 2006.