Mathematical formulation for designing a monitoring strategy: Application to the design of a river quality monitoring system

L.J. Alvarez-Vázquez\textsuperscript{a}, G. Casal\textsuperscript{b}, A. Martínez\textsuperscript{a}, M.E. Vázquez-Méndez\textsuperscript{b}

\textsuperscript{a}Departamento de Matemática Aplicada II, Universidade de Vigo. E.E. Telecomunicación, 36310 Vigo. Spain
\textsuperscript{b}Departamento de Matemática Aplicada, Universidade de Santiago de Compostela. Escola Politécnica Superior, 27002 Lugo. Spain

Abstract

In this work we propose a technique to automatically optimize the monitoring of any distributed indicator (blood pressure of a patient over time, concentration of a substance along a river, etc.) for which a reliable estimate is previously available. As an illustrative case, we show the steps to follow in order to implement that strategy when designing a system for monitoring water quality in a river. From a mathematical point of view, the problem is based on obtaining a reliable estimate of the chosen indicator by numerical simulation, and then solving a multi-objective optimization problem (with mixed real and integer variables) whose solution must provide a satisfactory monitoring strategy. In final sections we present and analyze the results when applying the proposed technique to study a real case in the Neuse River (North Carolina, USA).

Keywords: Monitoring strategy, River quality monitoring system, Mathematical modeling, Multi-objective optimization, Sampling point.

Email addresses: lino@dma.uvigo.es (L.J. Alvarez-Vázquez), xerardo.casal@usc.es (G. Casal), aurea@dma.uvigo.es (A. Martínez), miguelernesto.vazquez@usc.es (M.E. Vázquez-Méndez)
1. Introduction

We understand as monitoring the systematic and planned measurement of quality indicators for a process under study. An important factor in the design of monitoring techniques is the location, spatial or temporal, of the points on which these indicators are measured, the so-called sampling points [13]. For instance, if the indicator being measured is a function $\rho$ defined over an interval $[a, b]$, either temporal (blood pressure or concentration of a drug in the blood of a patient) or spatial (concentration of a pollutant in a river), the sampling points should provide representative measurements of what happens over the entire interval [22].

Sometimes there is no a priori information about the function $\rho$, and the choice of those sampling points should be made using techniques that combine random aspects with various strategies designed by the experimenter. These techniques (known as sampling techniques) seek to achieve the necessary information to infer the behaviour of the indicator over the entire range (see, for example, [20]).

In other situations, the function $\rho$ can be estimated using experimental and/or numerical models sufficiently tested (many times obtained by numerical simulation or statistical inference from the information available on it) [10, 16]. In these cases, when a reliable estimate of the indicator $\rho$ is available, the measurements performed at the sampling points will serve as measures of quality control or diagnosis and, in that sense, the values obtained must be measures that provide an overview of the behaviour of $\rho$ in the whole interval.

By assuming the latter premise (the indicator $\rho$ has been previously estimated and is therefore known) the first aim of this paper is to design a monitoring strategy that allows us, from measures of $\rho$ at specific points, to know its behaviour over the entire range. We deal with this problem in Section 2, where, taking also into account economic aspects, the problem is formulated as a multi-objective optimization problem, with mixed real and integer variables.

The second objective of this paper is to exemplify our technique applying it to the design of a river quality monitoring system, improving the previous works of the authors in this field [7, 17]. For this purpose, we introduce this real-world example (Section 3), give a detailed mathematical formulation of the problem (Section 4) and, assuming certain preferences in the decision making, propose a complete algorithm to obtain a solution providing a satis-
factory monitoring strategy (Sections 5). Then, we apply this algorithm to a realistic problem introduced in [17], and numerical results are presented and discussed in Section 6. Finally, in Section 7 we present some conclusions and future work.

2. Mathematical formulation for designing a monitoring strategy

We consider a real function \( \rho \) defined in an interval \([a, b] \subseteq \mathbb{R}\) (the indicator \( \rho : x \in [a, b] \rightarrow \rho(x) \in \mathbb{R} \) may represent the concentration of a substance - for instance, a medicine or a virus - in the blood along the time, the concentration of a pollutant in a river, etc.) We are interested in designing a monitoring strategy that allows us, from the measures of \( \rho \) at several specific points, to know their behaviour throughout the whole interval \([a, b]\). In order to do this, we propose to divide the interval \([a, b]\) in a number \((N \in \mathbb{N} \text{, to be determined})\) of subintervals \([a_{i-1}, a_i], i = 1, \ldots, N, \text{ also to be determined}\) and measure for each subinterval the function \( \rho \) at one point \((p_i \in [a_i, a_{i+1}], \text{ to be determined too})\), with the final aim that the measure at that point can give us a global idea of the values of \( \rho \) in the entire subinterval. Bearing these objectives in mind, we must take into account that:

- For each \( i = 1, \ldots, N, \) the point \( p_i \) should be such that the value of \( \rho \) at this point is a good approximation of the average of \( \rho \) on the corresponding subinterval, that is, it must be verified that

\[
d_i = (\rho(p_i) - \bar{\rho}_i)^2 \longrightarrow 0 \tag{1}
\]

where \( \bar{\rho}_i = \frac{\int_{a_{i-1}}^{a_i} \rho(x)dx}{a_i - a_{i-1}} \) represents the mean value of the indicator \( \rho \) in the subinterval \([a_{i-1}, a_i]\).

- The average of \( \rho \) in each subinterval must be representative, i.e., deviations from the mean should be as small as possible:

\[
\sigma_i = \sqrt{\int_{a_{i-1}}^{a_i} (\rho(x) - \bar{\rho}_i)^2 dx} \longrightarrow 0 \tag{2}
\]

- The main goal is to design the best monitoring system at the lowest possible cost and, in that sense, we suppose known an increasing function \( f \) giving us the economical cost depending on the number of measurements \( N \).
Thus, if we define $a_0 = a$, $a_N = b$, and denote by $\delta > 0$, the minimum length of any sample subinterval, the problem is to choose the optimal values of $N \in \mathbb{N}$, $a \equiv a(N) = (a_1, \ldots, a_{N-1}) \in \mathbb{R}^{N-1}$, and $p \equiv p(N,a) = (p_1, \ldots, p_N) \in \mathbb{R}^N$, such that they verify:

\begin{align*}
  a_i - a_{i-1} &\geq \delta, \quad \forall i = 1, \ldots, N, \quad (3) \\
  p_i &\in [a_{i-1}, a_i], \quad \forall i = 1, \ldots, N, \quad (4)
\end{align*}

and they satisfy the following objectives:

1. The optimal choice $(N,a,p)$ minimizes the economical cost $f$, given by

$$ f : N \in \mathbb{N} \rightarrow f(N) \in [0, +\infty) $$

2. Suitable representative averages are obtained at every interval. As noted above, this amounts to minimize the functional

$$ J_\sigma(a) = q\|\sigma(a)\|_\infty + (1 - q)\|\sigma(a)\|_2, $$

where $\sigma(a) = (\sigma_1, \ldots, \sigma_N)$ is defined by (2), $q \in [0,1]$ is a weight parameter, and $\|\cdot\|_\infty$ and $\|\cdot\|_2$ denote, respectively, the maximum and the standard Euclidean norms in $\mathbb{R}^N$.

3. In all subintervals there exists one point (the sampling point) in which the average is well captured. As before, this fact corresponds to minimizing the functional

$$ J_d(a,p) = r\|d(a,p)\|_\infty + (1 - r)\|d(a,p)\|_2, $$

where $d(a,p) = (d_1, \ldots, d_N)$ is defined by (1), and $r \in [0,1]$ is another weight parameter.

From a formal viewpoint, for each $N \in \mathbb{N}$ we define

$$ U^N_{ad} = \{(a,p) \in \mathbb{R}^{N-1} \times \mathbb{R}^N \text{ verifying (3) and (4)}\}, $$

and we consider $U_{ad} = \bigcup_{N \in \mathbb{N}} \{N\} \times U^N_{ad}$. Then, defining

$$ J : (N,a,p) \in U_{ad} \rightarrow J(N,a,p) = (f(N), J_\sigma(a), J_d(a,p)) \in \mathbb{R}^3, $$

the problem is formulated as the following Mixed-Integer Multi-Objective Problem (MIMOP):

$$ \min_{(N,a,p) \in U_{ad}} J(N,a,p) $$

4
Remark 1. If the function $\rho$ is defined on a $n$-rectangle $R \subset \mathbb{R}^n$, $n > 1$, (consider, for example, the case where we want to control the pollution of a river over a time interval taking particular measures at several specific moments), the technique described above to formulate the problem can be extended without difficulty, and the only thing that will significantly increase is the dimension of the control variable.

If the indicator $\rho$ takes values in $\mathbb{R}^M$, $M > 1$, (consider, for instance, that in the aforementioned case of a river we want to control pollution by measuring the concentration of different substances) the technique proposed for the problem remains valid too, and in this case what will significantly increase is the number of cost functionals defining $J$.

Conversely, if the function $\rho$ takes values in a space of infinite dimension (consider, for example, that we want to control river pollution measuring the concentration of a pollutant - or several ones - over a whole time period) the technique is also valid, but the objective functional should be redefined to fit the situation (in particular, the terms $d_i$ and $\sigma_i$, given respectively by (1) and (2) for the discrete case, need to be redefined for the distributed one). In section 4.2 we will detail how to do this in a concrete problem.

Obviously, in order to evaluate the functionals $J_\sigma$ and $J_d$ is necessary to estimate the values of the indicator $\rho$ on $[a,b]$, and in this step the mathematical modeling and the numerical simulation can be extremely useful (for instance, in sections 4.1 and 5.1 we show how to use them to estimate a typical river water quality indicator). Moreover, it should be taken into account that, in general, the utopic solution of the problem (9) - the ideal solution simultaneously minimizing all the cost functionals - will not exist, and the way to treat the multi-objective character of (9) will depend on the nature of the particular problem. In what follows, we will see how to apply our technique to the design of a monitoring system for a river (particularly, in section 5.2 we propose an algorithm to obtain a satisfactory solution of problem (9) under these premises).

3. Design of a river quality monitoring system

As it is well known, surface water quality directly impacts communities that depend on these sources for potable use, agricultural supplies, recreation, and commercial fishing. Water available for these purposes can be
drastically impacted by contamination from municipal and industrial discharges. One method for the effective monitoring and management of surface water quality is the establishment of near-real time, in situ monitoring systems. Such systems provide the basis for future adaptive management schemes using data about the transport and fate of contaminants within and across environmental regions. Distributed monitoring systems are predicated on the development of new sensors capable of monitoring the contaminants of interest. A critical component of implementing such a network is the identification of the optimal locations to deploy environmental sensors or establish sampling sites [18, 19, 9].

In previous works developed by the authors during the last decade, it has been established [7], and applied to a real situation [17], a methodology for determining the optimal location of sampling stations in a river. However, in both works it was assumed that the river is already divided into a number of previously chosen segments, and is in each of those segments where the optimal location of each sampling station is sought. It is clear that a correct division of the river into segments is essential for the information provided by the sampling station can be extrapolated to the entire segment that contains it. That task, which should be made by an experienced engineer, depends on the number and the type of discharges suffering the river, and can present great difficulties and limit largely the application of the methodology described in [7]. In order to overcome these limitations, in the present work we propose to apply the ideas in the section 2, and develop a method that not only determines the optimal location of the sampling stations, but it also indicates how the river should be divided so that the information provided by these stations could be accurately extrapolated to the entire basin under study.

4. Mathematical modeling

4.1. Estimation of a river quality indicator

For our study we consider a river of $L$ meters in length, which includes in its course a number $N_E$ of tributaries and/or wastewater discharges, located at points $b_k \in [0, L]$, for $k = 1, \ldots, N_E$ (see Figure 1). We assume that the pollutant discharges suffered by the river are known in terms of its concentration of fecal coliform (FC), but that it can vary depending on the situation (for instance, seasonal) in which we find ourselves. That is, we suppose the possibility of $N_S$ different contamination scenarios (from mild to
extremely high contamination), and for \( k = 1, \ldots, N_E \), and \( j = 1, \ldots, N_S \), we assume known the functions \( m^j_k(t) \) giving the mass coliform flow rate of discharge/tributary \( k \) corresponding to the \( j \)-th simulated contamination event. If the pollution situation \( j \), has a time duration \( T^j \), the CF concentration in the river over time, \( \rho^j(x, t) \), can be estimated by solving the following problem of partial differential equations with boundary and initial conditions (see [7]):

\[
\begin{aligned}
\frac{\partial \rho^j}{\partial t} + u^j \frac{\partial \rho^j}{\partial x} + \kappa \rho^j &= \frac{1}{A^j} \sum_{k=1}^{N_E} m^j_k \delta(x - b_k) \quad \text{in } (0, L) \times (0, T^j), \\
\rho^j(0, t) &= \rho^j_0(t) \quad \text{in } [0, T^j], \\
\rho^j(x, 0) &= \rho^j(x) \quad \text{in } [0, L],
\end{aligned}
\]

(10)

where

- \( \delta(x - b) \) denotes de Dirac measure at point \( b \),
- for \( k = 1, \ldots, N_E \), \( b_k \in (0, L) \) represents a point where a tributary (or wastewater discharge) is located, and \( m^j_k(t) \) is the mass coliform flow rate at this point corresponding to the \( j \)-th simulated contamination event,
- \( A^j(x, t) \) and \( u^j(x, t) \) denote, for the \( j \)-th simulated contamination event, the area of the section occupied by water (usually known as wet area) and the averaged velocity in that section, respectively. These fields can be experimentally known, or also obtained by numerical simulation, solving the following hyperbolic system:
\[
\begin{aligned}
\frac{\partial A^j}{\partial t} + \frac{\partial (A^j u^j)}{\partial x} &= \sum_{k=1}^{N_E} q^j_k \delta(x - b_k) \quad \text{in } (0, L) \times (0, T^j), \\
\frac{\partial (A^j v^j)}{\partial t} + \frac{\partial (A^j (u^j)^2)}{\partial x} + g A^j \frac{\partial \eta^j}{\partial x} &= \\
&= \sum_{k=1}^{N_E} q^j_k U^j_k \cos(\alpha_k) \delta(x - b_k) + S_f \quad \text{in } (0, L) \times (0, T^j), \\
A^j(L, t) &= A^j_L(t) > 0 \quad \text{in } [0, T^j], \\
A^j(x, 0) &= A^j_0(x) \quad \text{in } [0, L], \\
u^j(0, t) &= u^j_0(t) \quad \text{in } [0, T^j], \\
u^j(x, 0) &= u^j_0(x) \quad \text{in } [0, L],
\end{aligned}
\] (11)

where, in addition to the initial conditions \( A^j_0(x), u^j_0(x) \), and the boundary conditions \( A^j_L(t), u^j_0(t) \) (that must be known), \( g \) represents the gravity acceleration, \( S_f \) denotes the bottom friction stress, \( \eta^j(x, t) \) is the height of the water surface with respect to a fixed reference level (depending on the geometry of the river, it can be also expressed in terms of the unknown \( A^j(x,t) \)) and, finally, for the \( j \)-th simulated contamination event, \( q^j_k(t) \) is the flow rate at point \( b_k \), \( U^j_k(t) \) its velocity, and \( \alpha_k \) the angle between the main river and the tributary (or the wastewater outfall) located at point \( b_k \) (see [8] for further details),

- \( \kappa \) is the loss rate for fecal coliform bacteria (see [7] for its detailed characterization in terms of mortality, settling and light effects),
- \( \rho^j_0(t) \) is the CF concentration in the source of the river (which it is assumed known) along the time interval,
- \( \rho^j_0(x) \) is the initial CF concentration in the river (also experimentally known).

4.2. Design of the monitoring strategy

To measure the water quality of the river, we will assume that will be installed, at specific points, sampling stations that measure the CF concentration in those points over time. Designing the monitoring system then passes through choosing the number \( N \in \mathbb{N} \) of segments in which the river
will be divided (in other terms, the number of sampling stations to be installed), determine those segments and, in each one of them, select the point at which to locate the corresponding station.

By following the methodology described in section 2, we define $a_0 = 0$, $a_N = L$, and denote by $\delta > 0$ the minimum length of any segment. Given that the stations measure the CF concentration over the whole time interval, that we have $N_S$ different contamination scenarios, and that we want the measurements obtained at the stations to be representative of the segments in all situations, we must adapt the values of $d_i$ and $\sigma_i$, initially given by (1) and (2), to our circumstances. So, if we define

$$\bar{\rho}_i^j(t) = \frac{\int_{a_{i-1}}^{a_i} \rho^j(x,t) dx}{a_i - a_{i-1}}, \quad i = 1, \ldots, N, \quad j = 1, \ldots, N_S,$$

we have different options for redefining $d_i$. For instance, we could take

$$d_i = \max_{j=1,\ldots,N_S} \left\{ \frac{1}{T^j} \int_0^{T^j} \left( \rho^j(p_i,t) - \bar{\rho}_i^j(t) \right)^2 dt \right\}, \quad (12a)$$

or

$$d_i = \frac{1}{N_S} \sum_{j=1}^{N_S} \frac{1}{T^j} \int_0^{T^j} \left( \rho^j(p_i,t) - \bar{\rho}_i^j(t) \right)^2 dt. \quad (12b)$$

In a similar way, $\sigma_i$ could be taken as

$$\sigma_i = \max_{j=1,\ldots,N_S} \left\{ \frac{1}{T^j} \int_0^{T^j} \sqrt{\int_{a_{i-1}}^{a_i} \left( \rho^j(x,t) - \bar{\rho}_i^j(t) \right)^2 dx} dt \right\}, \quad (13a)$$

or

$$\sigma_i = \frac{1}{N_S} \sum_{j=1}^{N_S} \frac{1}{T^j} \int_0^{T^j} \sqrt{\int_{a_{i-1}}^{a_i} \left( \rho^j(x,t) - \bar{\rho}_i^j(t) \right)^2 dx} dt. \quad (13b)$$

Moreover, the economic cost of the monitoring system is due essentially to the price of the sampling stations and, therefore, for the sake of simplicity,

$$f(N) = c N, \quad \forall N \in \mathbb{N}, \quad (14)$$

where $c$ represents the price of one station.

With all this in mind, the problem of optimally designing the monitoring system for the river can be reduced to solve the MIMOP problem (9), considering that $f$ is now given by (14), and that $d_i$ and $\sigma_i$ are defined by one of the alternative expressions in (12) and (13), respectively.
5. Numerical resolution

5.1. Numerical simulation

The first step to resolving the problem (9) is to estimate the concentration of coliforms expected in each one of the $N_S$ scenarios of contamination that will be studied. As already indicated, this can be done by numerical simulation, solving the systems (10) and (11) with a suitable numerical method. So, for the sake of completeness, we summarize the steps to be followed in the numerical simulation of the problem:

**Algorithm 1.**

- **Step 1.** Characterizing the geometry of the river: In order to solve the system (10)-(11) is previously necessary to know the geometry of the river and determine $\eta^j(x,t)$ (the height of the water surface with respect to a fixed reference level) in terms of the unknown $A^j(x,t)$ (see [8] for further details).

- **Step 2.** Determining initial and boundary conditions: For $j = 1, \ldots, N_S$, one needs to know $A_j^0(x)$, $w^0_j(x)$ and $\rho^0_j(x)$ in $[0, L]$, and $u_0^j(t)$, $\rho_0^j(t)$ and $A_L^j(t)$ in $[0, T^j]$.

- **Step 3.** Characterizing the tributaries and all other discharges to the river: For $k = 1, \ldots, N_E$ one must know the location $b_k$, the angle of entry $\alpha_k$, and also, for each $j = 1, \ldots, N_S$, the functions $q_k^j(t)$, $U_k^j(t)$ and $m_k^j(t)$ defined on $[0, T^j]$ differentiating tributaries and/or discharges on each scenario of contamination.

- **Step 4.** Solving the system (11): Once known the geometry of the river (Step 1), the initial/boundary conditions (Step 2) and the characteristics of each one of the discharges (Step 3), the system (11) can be solved using varied numerical methods. In this paper we propose the use of an algorithm that combines characteristics with Lagrange $P_1$ finite element, whose detailed analysis may be seen in [8] and [6].

- **Step 5.** Solving the system (10): From the initial and boundary conditions (Step 2), the coliform flows in each discharge (Step 3), and the wet sections $A^j(x,t)$ and the averaged velocities $u^j(x,t)$ (Step 4), the system (10) can now be solved by different numerical methods. In this paper we have used an implicit upwind finite difference scheme, which does not need to solve any linear equations system (see [7]).
5.2. Numerical optimization

Once known the functions \( \rho^j(x,t) \), for \( j = 1, \ldots, N_S \), the cost functional \( J(N,a,p) = (f(N), J_\sigma(a), J_d(a,p)) \) is very simple to estimate, and the resolution of problem (9) can be addressed by very diverse methods leading to a satisfactory optimal solution. For a problem like this, in which there are only three objective functions, the use of \textit{a posteriori} methods (weight method, method of \( \epsilon \)-constraints...) can allow a determination of the Pareto front surface, which can be very useful to assess the decision maker in the choice of the solution that will satisfy their interests in the most satisfactory way (in [12, 1, 2, 3, 21, 11, 4] the authors used this methodology to solve different problems of environmental control). Similarly, the combination of \textit{interactive} methods can improve the support given to the decision maker in order to choose the most satisfactory solution (recently, in [5], the authors combined the methods VIA and STEM for solving a multi-objective problem related to air pollution).

In this paper, we will assume that the maximum number \( N_{\text{max}} \) of sampling stations that can be placed in the river is low and that, moreover, the decision maker has already indicated among its preferences that he wants a monitoring system with the smallest possible number of stations, but ensuring that in all segments, the measured values are sufficiently representative. In response to this orientation we propose to proceed as follows:

\textbf{Algorithm 2.}

- \textit{Step 1.} For each \( N = 1, 2, \ldots, N_{\text{max}} \), solving the problem

\[
\begin{aligned}
\min J_\sigma(a) \\
\text{subject to} \quad a_i - a_{i-1} \geq \delta, \quad i = 1, \ldots, N,
\end{aligned}
\]

and showing the decision maker the minimum values of \( J_\sigma \) that are obtained in each case.

- \textit{Step 2.} Asking the decision maker that, by virtue of the information obtained in the previous step, choose the number \( N \) of segments that will divide the river, and setting a maximum threshold for the value of \( J_\sigma \), denoted by \( \sigma_{\text{max}} \).
Step 3. Solving the problem

\[
\begin{align*}
\min J_d(a,p) \\
\text{subject to} \quad & a_{i-1} \leq p_i \leq a_i, \quad i = 1, \ldots, N, \\
& a_i - a_{i-1} \geq \delta, \quad i = 1, \ldots, N, \\
& J_\sigma(a) \leq \sigma_{\max},
\end{align*}
\]

and showing the obtained solution to the decision maker.

Step 4. Asking the decision maker if the solution obtained is satisfactory. If it is, STOP. If not, returning to Step 2 and increasing the number of stations that are to be placed \((N)\) and/or decreasing the maximum threshold for \(J_\sigma (\sigma_{\max})\).

The application of this algorithm depends on knowing how to solve, for different values of \(N\), the problems \((15)\) and \((16)\). The dimension of the problem \((15)\) is \(N - 1\) and, admitting that this value is not too high, in this work we solve that problem with the classical method of Nelder-Mead [15], using a penalty function to deal with the linear constraints. However, the dimension of the problem \((16)\) increases to \(2N - 1\) and, in addition, the variables \(a\) and \(p\) play very different roles, which discourages the direct use of that method, even for values of \(N\) not very large. Although any other method of direct or evolutionary search may be useful to solve the problem \((16)\), the authors - taking into account the experience gained in a previous work [7] - propose the following algorithm to solve the problem \((16)\).

5.2.1. Ad hoc method for solving the problem \((16)\)

The problem \((16)\) depends on two variables \(a \in \mathbb{R}^{N-1}\) and \(p \in \mathbb{R}^N\), that not only have a distinct physical interpretation, but they also can be treated in a different manner when solving the problem. Indeed, it is easy to see that for a given \(a\) (fixed) satisfying constraint \((3)\) and \(J_\sigma(a) \leq \sigma_{\max}\), the value of \(p\) that minimizes the value of \(J_d(a,p)\) - satisfying constraint \((4)\) - can be obtained as the solution of \(N\) one-dimensional optimization problems. Specifically, for \(i = 1, \ldots, N\), it is enough to take \(p_i \in \mathbb{R}\) as the solution of the problem:

\[
\begin{align*}
\min d_i(q) \\
\text{subject to} \quad q \in [a_{i-1}, a_i]
\end{align*}
\]

where, for instance, \(d_i(q) = \max_{j=1,\ldots,N_S} \left\{ \frac{1}{T_j} \int_0^{T_j} \left( \rho_j^i(q,t) - \tilde{\rho}_j^i(t) \right)^2 dt \right\} \).
Thus, if, for \( i = 1, \ldots, N \), we denote by \( p_i(a) \in [a_{i-1}, a_i] \) the solution of the problem (17) and define \( J(a) = J_d(a, p(a)) \), the problem (16) reduces to

\[
\begin{align*}
\min J(a) \\
\text{subject to } a_i - a_{i-1} &\geq \delta, \quad i = 1, \ldots, N, \\
J_\sigma(a) &\leq \sigma_{max},
\end{align*}
\]  

so that, if \( a^* \) is the solution of the problem (18), the solution of the original problem (16) will be given by \((a^*, p(a^*))\).

In this paper, the problem (18) is solved, like (15), with the Nelder-Mead algorithm, using again a penalty function to treat both constraints. Each time one needs to evaluate the function \( J(a) \), the one-dimensional problems (17) are solved using the golden section search [14], which has already been successfully used by the authors in a very similar problem (cf. [7]).

6. Numerical results

6.1. Case study and experimental data

In this work, we select as a case study the final section of the Neuse River (North Carolina, USA), which has already been studied in a previous work (see [17]). In that article we admitted that in each case (that is, for each value of \( N \)), the segments in which the stretch of river was divided were known, and we analyzed only how varied the optimal location of the sampling stations, depending on \( N \) and on the different pollution scenarios.
To complete the study, in this work we will analyze how the stretch of river must be divided, and where the sampling stations must be located, so that the data that these provide are representative of the contamination in each segment, and, moreover, this remains valid for all the pollution scenarios considered.

The section of river under study (see Figure 2) has a length $L = 53534 \, m$, and the height of the river bottom and the approximation of the river bed that we have used (Step 1 of Algorithm 1) can be seen in Figure 3. In this example we have considered seven wastewater discharges ($N_E = 7$), and three different pollution situations ($N_S = 3$, corresponding to mild ($j = 1$), moderate ($j = 2$), and high ($j = 3$) contamination), which were simulated over a period of 36 hours ($T^1 = T^2 = T^3 = 1.296 \times 10^5 \, s$).

The initial and boundary conditions in each of the scenarios of contamination (Step 2 of Algorithm 1) is shown in Table 1 (the initial wet area $A^{j,0}(x)$ is computed, taking into account the geometry of the river (see Figure 3), from the given initial water height $\eta^{j,0}(x)$).

Finally, the discharges are characterized (Step 3 of Algorithm 1) in Tables 2–5: the locations $b_k$ and the angles of entry $\alpha_k$ of discharges are shown in Table 2, while the values of the functions $q^j_k(t)$, $U^j_k(t)$ and $m^j_k(t)$, characterizing the different scenarios, are shown in the Tables 3, 4 and 5, respectively.

With this information we perform the Steps 4 and 5 of Algorithm 1,
<table>
<thead>
<tr>
<th>Initial/Boundary condition</th>
<th>Mild cont.</th>
<th>Moderate cont.</th>
<th>High cont.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\eta^0(x)$ (m)</td>
<td>0.79</td>
<td>1.48</td>
<td>2.55</td>
</tr>
<tr>
<td>$u^0(x)$ (ms$^{-1}$)</td>
<td>0.19</td>
<td>0.23</td>
<td>0.27</td>
</tr>
<tr>
<td>$\rho^0(x)$ (um$^{-3}$)</td>
<td>$150 \times 10^6$</td>
<td>$150 \times 10^6$</td>
<td>$150 \times 10^6$</td>
</tr>
<tr>
<td>$u^0(t)$ (ms$^{-1}$)</td>
<td>0.19</td>
<td>0.23</td>
<td>0.27</td>
</tr>
<tr>
<td>$\rho^0(t)$ (um$^{-3}$)</td>
<td>$150 \times 10^6$</td>
<td>$150 \times 10^6$</td>
<td>$150 \times 10^6$</td>
</tr>
<tr>
<td>$A_L^0(t)$ (m$^2$)</td>
<td>118.45</td>
<td>223.12</td>
<td>387.65</td>
</tr>
</tbody>
</table>

Table 1: Step 2 of Algorithm 1: Initial and boundary conditions for the three different scenarios

<table>
<thead>
<tr>
<th>Discharge ($k$)</th>
<th>Location down river ($b_k$)</th>
<th>Angle of entry ($\alpha_k$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4993</td>
<td>$\pi/2$</td>
</tr>
<tr>
<td>2</td>
<td>5454</td>
<td>$\pi/2$</td>
</tr>
<tr>
<td>3</td>
<td>14493</td>
<td>$\pi/2$</td>
</tr>
<tr>
<td>4</td>
<td>17784</td>
<td>$\pi/2$</td>
</tr>
<tr>
<td>5</td>
<td>18293</td>
<td>$\pi/2$</td>
</tr>
<tr>
<td>6</td>
<td>32684</td>
<td>$\pi/2$</td>
</tr>
<tr>
<td>7</td>
<td>35063</td>
<td>$\pi/2$</td>
</tr>
</tbody>
</table>

Table 2: Step 3 of Algorithm 1: Location and angle for each discharge

<table>
<thead>
<tr>
<th>Discharge ($k$)</th>
<th>Mild ($j = 1$)</th>
<th>Moderate ($j = 2$)</th>
<th>High ($j = 3$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.03</td>
<td>0.06</td>
<td>0.12</td>
</tr>
<tr>
<td>2</td>
<td>0.02</td>
<td>0.03</td>
<td>0.07</td>
</tr>
<tr>
<td>3</td>
<td>0.11</td>
<td>0.22</td>
<td>0.44</td>
</tr>
<tr>
<td>4</td>
<td>0.05</td>
<td>0.11</td>
<td>0.22</td>
</tr>
<tr>
<td>5</td>
<td>0.02</td>
<td>0.04</td>
<td>0.07</td>
</tr>
<tr>
<td>6</td>
<td>0.04</td>
<td>0.09</td>
<td>0.18</td>
</tr>
<tr>
<td>7</td>
<td>0.01</td>
<td>0.02</td>
<td>0.04</td>
</tr>
</tbody>
</table>

Table 3: Step 3 of Algorithm 1: Water flow rates ($q^j_k(t)$) in m$^3$s$^{-1}$ for the three different scenarios
<table>
<thead>
<tr>
<th>Discharge (k)</th>
<th>Mild (j = 1)</th>
<th>Moderate (j = 2)</th>
<th>High (j = 3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.04</td>
<td>0.07</td>
<td>0.15</td>
</tr>
<tr>
<td>2</td>
<td>0.02</td>
<td>0.04</td>
<td>0.07</td>
</tr>
<tr>
<td>3</td>
<td>0.11</td>
<td>0.22</td>
<td>0.44</td>
</tr>
<tr>
<td>4</td>
<td>0.05</td>
<td>0.09</td>
<td>0.19</td>
</tr>
<tr>
<td>5</td>
<td>0.03</td>
<td>0.06</td>
<td>0.12</td>
</tr>
<tr>
<td>6</td>
<td>0.06</td>
<td>0.11</td>
<td>0.22</td>
</tr>
<tr>
<td>7</td>
<td>0.02</td>
<td>0.05</td>
<td>0.09</td>
</tr>
</tbody>
</table>

Table 4: Step 3 of Algorithm 1: Water velocities ($U^j_k$) in $ms^{-1}$ for the three different scenarios

<table>
<thead>
<tr>
<th>Discharge (k)</th>
<th>Mild (j = 1)</th>
<th>Moderate (j = 2)</th>
<th>High (j = 3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$2900 \times 10^6$</td>
<td>$5900 \times 10^6$</td>
<td>$11800 \times 10^6$</td>
</tr>
<tr>
<td>2</td>
<td>$1600 \times 10^6$</td>
<td>$3300 \times 10^6$</td>
<td>$6600 \times 10^6$</td>
</tr>
<tr>
<td>3</td>
<td>$10900 \times 10^6$</td>
<td>$21800 \times 10^6$</td>
<td>$43600 \times 10^6$</td>
</tr>
<tr>
<td>4</td>
<td>$5400 \times 10^6$</td>
<td>$10900 \times 10^6$</td>
<td>$21800 \times 10^6$</td>
</tr>
<tr>
<td>5</td>
<td>$1700 \times 10^6$</td>
<td>$3500 \times 10^6$</td>
<td>$7000 \times 10^6$</td>
</tr>
<tr>
<td>6</td>
<td>$4400 \times 10^6$</td>
<td>$8900 \times 10^6$</td>
<td>$17900 \times 10^6$</td>
</tr>
<tr>
<td>7</td>
<td>$900 \times 10^6$</td>
<td>$1800 \times 10^6$</td>
<td>$3700 \times 10^6$</td>
</tr>
</tbody>
</table>

Table 5: Step 3 of Algorithm 1: Mass coliform flow rates ($m^j_k$) in $um^{-3}$ for the three different scenarios
\[
\mathbf{a} = (a_0, a_1, \ldots, a_N) \quad \text{optimal} \\
J_\sigma(a) \quad \text{(minimal value)}
\]

<table>
<thead>
<tr>
<th>N</th>
<th>(a = (a_0, a_1, \ldots, a_N)) optimal</th>
<th>(J_\sigma(a)) (minimal value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>((0, L))</td>
<td>(2.1126 \times 10^6)</td>
</tr>
<tr>
<td>2</td>
<td>((0, 15872, L))</td>
<td>(1.6024 \times 10^6)</td>
</tr>
<tr>
<td>3</td>
<td>((0, 14521, 19958, L))</td>
<td>(8.6290 \times 10^5)</td>
</tr>
<tr>
<td>4</td>
<td>((0, 14440, 15637, 20899, L))</td>
<td>(7.6731 \times 10^5)</td>
</tr>
<tr>
<td>5</td>
<td>((0, 14439, 15603, 20547, 32613, L))</td>
<td>(7.6498 \times 10^5)</td>
</tr>
<tr>
<td>6</td>
<td>((0, 5205, 14464, 15709, 21368, 38977, L))</td>
<td>(6.7602 \times 10^5)</td>
</tr>
</tbody>
</table>

Table 6: Step 1 of Algorithm 2: Optimal solutions of problem (15) for \(N = 1, \ldots, 6\).

<table>
<thead>
<tr>
<th>(p_i \in [a_{i-1}, a_i], \ i = 1, \ldots, 4) optimal</th>
<th>(J_d(a, d)) (minimal value)</th>
<th>(J_\sigma(a))</th>
</tr>
</thead>
<tbody>
<tr>
<td>9134 (\in [0, 13737])</td>
<td>(2.025 \times 10^6)</td>
<td>(8.3935 \times 10^5)</td>
</tr>
<tr>
<td>14449 (\in [13737, 14737])</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19095 (\in [14737, 20507])</td>
<td></td>
<td></td>
</tr>
<tr>
<td>38315 (\in [20507, L])</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 7: Step 3 of Algorithm 2: Optimal solutions of problem (16) for \(N = 4\) and \(\sigma_{\text{max}} = 8.5 \times 10^5\).

and obtain the FC concentrations expected for each of the three scenarios considered (functions \(\rho^1(x, t)\), \(\rho^2(x, t)\) and \(\rho^3(x, t)\)). As above mentioned, knowledge of these functions allows us to evaluate the multi-objective cost functional \(J(N, a, p) = (f(N), J_\sigma(a), J_d(a, p))\) in a straightforward way and, in this way, we can address the problem (9) with the Algorithm 2. In next subsection we present the results achieved for this example.

### 6.2. Results and discussion

First (Step 1 of Algorithm 2) we solve the problem (15) for different values of \(N\) (in our case, from budget constraints, we have assumed \(N_{\text{max}} = 6\)). In Table 6 the reader can find the optimal solutions obtained with the method of Nelder-Mead (for 20 different random initial simplex), taking the minimum length of subintervals \(\delta = 1000\), and the weight parameters \(q = r = 0.8\).

In view of the results shown in Table 6, the decision maker decided (Step 2 of Algorithm 2) to place \(N = 4\) sampling stations, and take \(\sigma_{\text{max}} = 8.5 \times 10^5\) as maximal threshold for the value of \(J_\sigma\). The problem (16) is then solved (Step 3 of Algorithm 2), and we obtain the results shown in Table 7. For this solution, the division of the river into segments and the location of the
Figure 4: Satisfactory solution achieved (divisions of river section and optimal location of sampling stations) for $N = 4$ and $\sigma_{\text{max}} = 8.5 \times 10^5$, along with the location of discharge points.

Sampling stations are shown in Figure 4, where also the points of discharge are painted. The figure shows how the method used tends to separate discharge points in different segments, so that the most important discharge is encased into a single segment of minimum length.

In Figure 5 we compare, for each segment, the averaged concentrations and the concentrations in the corresponding sample points for the three different pollution scenarios considered. Sampling points seem to capture properly the mean of each segment in all scenarios of pollution and, consequently, the obtained solution is deemed satisfactory for the decision maker (Step 4 of Algorithm 2).

As already commented, in [17] the optimal location of the sampling stations on this same stretch of Neuse River was studied, but starting from a fixed given division of subintervals. In particular, for $N = 4$, the river section was divided into the following segments: $[0, 0.8994]$, $[0.8994, 1.7987]$, $[1.7987, 2.9980]$, and $[2.9980, L]$. The optimal sampling points were obtained for each pollution scenario, but in this case ($N = 4$), the locations were very similar in all three scenarios. For example, for a moderate contamination ($j = 2$), the sampling points obtained in [17] were $p_1 = 8700$, $p_2 = 17700$, $p_3 = 21800$, $p_4 = 38500$. Now, if we consider this solution, $\hat{a} = (8994, 17987, 29980)$ and $\hat{p} = (8700, 17700, 21800, 38500)$, and evaluate the objective functionals $J_\sigma$ and $J_d$ at these vectors, we obtain the values $J_\sigma(\hat{a}) = 1.5215 \times 10^6$ and $J_d(\hat{a}, \hat{p}) = 1.1965 \times 10^{12}$. Comparing these values with those of the optimal solution obtained in the present work (see Table 7) we note that the new proposal of river monitoring (given in Table 7, and illustrated in Figure 3) is significantly better in terms of the representativeness of the average ($J_\sigma$...
is much lower), and also with regard to the capture of these averages ($J_d$ is also much lower).

7. Conclusions

This paper proposes a new technique to automate the design of the monitoring of any indicator, under the availability of a reliable estimate on it. The proposed strategy is valid for any indicator - or indicators - defined on any given domain $\Omega \subseteq \mathbb{R}^n$. To illustrate this fact, the article shows in detail how to apply this technique to designing a system for monitoring water quality in a river. As a case study we present the final section of the River Neuse (North Carolina, USA), where can be noticed that the new results are significantly better than those achieved in previous work [17]. This represents not only a major in the field of water quality monitoring progress, but it also shows the proposed technique to be useful in many different fields of application. Indeed, the implementation of this strategy into monitoring issues in other fields is one of the ways in which progress could be made in the future.

Acknowledgements

This work was supported by funding from the project MTM2012-30842 of Ministerio de Economía y Competitividad (Spain) and FEDER.

References


Figure 5: Comparison between the averaged concentrations of CF and the concentrations at sampling points for all four segments considered