

Optimal spatio-temporal design for water quality monitoring networks in maximizing economic value of information

François Destandau, and Youssef Zaiter¹

Laboratoire GESTE, UMR Engees-Irstea MA-8101, Strasbourg, France

Abstract The Water Framework Directive 2000/60/EC draw attention to Water Quality Monitoring Networks (WQMN) that allows the acquisition of information regarding water streams. Information could be acquired by a spatial and/or temporal approach. However, there is a cost for monitoring water quality. Hence, to determine the spatio-temporal design of the network, the Economic Value of Information must be known to undertake a cost-benefit analysis. In this study, we show how the calculation of the EVOI can help the network manager to answer questions such as: is the cost of monitoring justified? How to allocate a budget between adding a monitoring station or increasing the frequency of measurement of existing stations?

1 Introduction

The WQMN has been the research topic for several studies. Part of these studies focused on the physical optimization of the WQMN, while the other part focused on the Economic Value of Information (EVOI) provided by the WQMN.

The first type focused on optimizing the network by taking hydrological considerations into account, with the aim of minimizing imprecisions regarding water quality monitoring. The optimization can be spatial and/or temporal. On the one hand, spatial issues comprise the optimization of the location and the number of monitoring stations, whereas temporal issues take into account the optimization of the sampling frequency. [1] determined the location of the monitoring stations that decreased the average deviation for the water pollution index. [2] focused on determining the location of the monitoring network that minimized the detection time for accidental pollution. In [3], the objective was to determine the pollution trend: upward, downward or no trend. In [4], the objective was to determine the seasonal variation of water quality. More recently, [6] tried to find the optimal spatio-temporal design of a WQMN in the Karkkeh reservoir. They presented a multi-criteria methodology based on maximizing a statistical value of information, minimizing the number of monitoring stations and the sampling frequency.

Other studies have focused on determining the EVOI provided by a WQMN. The EVOI analysis is based on the Bayesian method. Several authors have used this concept to determine the economic value of information provided by a WQMN. [7] tried to find the best policy to manage eutrophication with a satellite observation. [8] tried to choose between a spatialized or uniform policy to protect the Great Barrier Reef. [9] identified the parameters that have an impact on the EVOI: prior probabilities on states of nature, costs linked to a bad decision and the accuracy of additional information.

In this article, we combine the literature on physical network optimization and EVOI, in order to determine the spatio-temporal design of the WQMN that maximizes the EVOI.

Section 2 deals with the hypotheses and the methodology that we followed. In Section 3, we determine the EVOI according to spatial and temporal monitoring intensity. Section 4 includes a monitoring cost to estimate the net benefit of the network. In Section 5, we study the optimal spatio-temporal design of networks with a fixed budget. Section 6 concludes the paper.

2 Hypotheses

We assume a river represented by a segment $[0,1]$, where location $l = 0$ is the most upstream point (source) and location $l = 1$ is the most downstream point of the river (outlet).

We assume that accidental pollution could be emitted at any point in the stream with probability P . This pollution generates an environmental damage that is proportional to the distance between the location of the emission l_x and the outlet of the river. The environmental damage generated is computed as follows:

$$D_{l_x, out} = \delta(1 - l_x) \quad (1)$$

¹ Corresponding author : youssef.zaiter@engees.unistra.fr

where $D_{l_x,out}$ represents the environmental damage generated by the accidental pollution and δ is the marginal environmental damage.

However, if a monitoring station is located in l_y , downstream from the location of the emission, the station could detect the accidental pollution. In that case, an action a at a cost C will be implemented to stop environmental damage. Environmental damage then becomes $D_{l_x,l_y} = \delta(l_y - l_x)$. The effect of the action is therefore immediate but it cannot act retrospectively on the damage generated between the emission and the pollution detection. Thus, if pollution is detected at the outlet of the river, it will be too late to implement an action and stop environmental damage.

Nevertheless, the monitoring at each station is obviously not continuous. A probability α exists such that the monitoring stations may not detect pollution. This allows the introduction of the temporal issue in the design of the monitoring network. Obviously, this probability decreases when the sampling frequency increases. Moreover, we assume that the probability α is the same for all stations of the network and that there is independence between the events. Specifically, the probability for detection/non-detection at a station does not depend on the number of stations that do not detect this pollution upstream.

3 The EVOI and the location of the monitoring stations

3.1 The EVOI and the Bayesian Method

We determine the EVOI by implementing the Bayesian decision theory. If the DM is uncertain about the states of nature, he will try to obtain additional information on the likelihood of the potential states of nature and update his prior beliefs to make more appropriate decisions. The EVOI is calculated as the difference between utilities of the decisions after acquiring new information and decisions that could have been taken without additional information.

By considering K states of nature and K possible actions, the EVOI can be written as follows:

$$EVOI = \sum_{k=1}^K P(i_k) \cdot \left[\sum_{k'=1}^K P(s_{k'}/i_k) \cdot U_{a_k/s_{k'}} \right] \tag{2}$$

Where:

$P(i_k)$: Probability that the network indicates the state of nature s_k ;

$P(s_{k'}/i_k)$: Probability of being in the state of nature $s_{k'}$ when the network indicates i_k

$U_{a_k/s_{k'}}$: Utility of the action a_k (more appropriate for the state of nature k) when the state of nature is $s_{k'}$.

In our model, we assume that the WQMN comprises n monitoring stations. Hence, with n monitoring stations, we have $(n + 2)$ states of nature depending on the location of the emission of the accidental pollution. If $l_{y/n}$ is the location of the y^{th} monitoring station for n monitoring stations, then $s_{[l_{(y-1)/n}, l_{y/n}]}$ is the state of nature where accidental pollution is emitted between $l_{(y-1)/n}$ and $l_{y/n}$, $s_{[l_{n/n}, out]}$ is the state of nature where accidental pollution is emitted between the last station and the outlet of the stream, and s_\emptyset is the state of nature where no accidental pollution is emitted. The states of nature are then:

$$\left\{ s_{[0, l_{1/n}]}, s_{[l_{1/n}, l_{2/n}]}, \dots, s_{[l_{(y-1)/n}, l_{y/n}]}, \dots, s_{[l_{n/n}, out]}, s_\emptyset \right\}$$

The network can deliver $(n + 2)$ messages as well: pollution is detected by the y^{th} monitoring station: $i_{l_{y/n}}$, accidental pollution is detected at the outlet of the stream: i_{out} , and no pollution is detected: i_\emptyset . The messages are then:

$$\left\{ i_{l_{1/n}}, i_{l_{2/n}}, \dots, i_{l_{y/n}}, \dots, i_{out}, i_\emptyset \right\}$$

The probability of detecting accidental pollution depends on the state of nature. On the basis of our hypotheses, this can be written as follows:

$$P(i_{l_{y/n}}/s_{[l_{(y-1)/n}, l_{y/n}]}) = (1 - \alpha) ; (i_{(y+1)/n}/s_{[l_{(y-1)/n}, l_{y/n}]} = \alpha \cdot (1 - \alpha) ; P(i_{(y+k)/n}/s_{[l_{(y-1)/n}, l_{y/n}]} = \alpha^k \cdot (1 - \alpha) \tag{3}$$

Obviously, if any pollution is emitted downstream of the station, or if no pollution is emitted, the probability of detection is null. By applying Bayes' theorem, the EVOI (2) becomes:

$$EVOI = \sum_{k=1}^K P(i_k) \cdot \left[\sum_{k=1}^K \frac{P(s_k) \cdot P(i_k/s_k)}{P(i_k)} \cdot U_{a_k/k} \right] \tag{4}$$

Utility is evaluated based on the status-quo (no action is taken). We compute the utility as the difference between environmental damage saved thanks to the action $D_{l_{y/n},out}$, and the cost C of implementing the action in the event of detection. Hence, the EVOI from Equation (5) becomes:

$$EVOI = \sum_{k=1}^K \sum_{k=1}^K P(s_k) \cdot P(i_k/s_k) \cdot [D_{l_{y/n},out} - C] \tag{5}$$

On the basis of our hypotheses of uniformity of accidental pollution distribution (the same probability of accidental pollution for areas as well as for lengths) and Equation (3), the EVOI (5) for n stations and a probability of non-detection α becomes:

$$EVOI_{n,\alpha} = P \cdot (1 - \alpha) \cdot \sum_{y=1}^n [D_{l_{y/n},out} - C] \cdot \sum_{y'=1}^y [l_{y'/n} - l_{(y'-1)/n}] \cdot \alpha^{y-y'} \tag{6}$$

3.2 Location of the monitoring stations

Close to the outlet of the river, the cost of the action could be higher than the damage avoided. Consequently, no station is useful in this part of the river. We can write this condition as follows:

$$\delta(1 - l_{y/n}) > C \quad \forall n \in [1; +\infty[; \forall y \in [1, n] \tag{7}$$

$$\Leftrightarrow l_{y/n} < 1 - \frac{C}{\delta} \quad \forall n \in [1; +\infty[; \forall y \in [1, n] \tag{8}$$

It can be deduced from hypothesis (1) concerning environmental damage uniformity that the locations of the monitoring stations have to be at equal distances. Then:

$$l_{y/n} = \frac{y}{n+1} \left(1 - \frac{C}{\delta}\right) \quad \forall n \in [1; +\infty[; \forall y \in [1, n] \tag{9}$$

The environmental damage avoided when pollution is detected in $l_{y/n}$ is:

$$D_{l_{y/n},out} = \delta \cdot \left[1 - \frac{y}{n+1} \cdot \left(1 - \frac{C}{\delta}\right)\right] \tag{10}$$

By implementing the above conditions in Equation (6), we rewrite EVOI as follows:

$$EVOI_{n,\alpha} = P \cdot (1 - \alpha) \cdot \sum_{y=1}^n \left[\delta \cdot \left[1 - \frac{y}{n+1} \cdot \left(1 - \frac{C}{\delta}\right)\right] - C \right] \cdot \sum_{y'=1}^y [l_{y'/n} - l_{(y'-1)/n}] \cdot \alpha^{y-y'}$$

$$\Leftrightarrow EVOI_{n,\alpha} = \frac{P(\delta - C)^2}{\delta(n+1)^2} \sum_{y=1}^n (n+1-y)(1 - \alpha^y) \tag{11}$$

With $X = \frac{P(\delta - C)^2}{\delta}$, we note that:

$$\frac{\partial X}{\partial C} = -\frac{2P(\delta - C)}{\delta} < 0; \frac{\partial X}{\partial P} = \frac{(\delta - C)^2}{\delta} > 0; \frac{\partial X}{\partial \delta} = \frac{P(\delta - C)(\delta + C)}{\delta^2} > 0$$

On the basis of the first-order conditions above, we can deduce that if the parameter P or δ increases, the EVOI increases. Conversely, if the parameter C increases, the EVOI decreases.

4 Cost for monitoring water quality and the net benefit of the network

4.1 Cost for monitoring water quality

We assume a monitoring cost Θ_α for each station, given by the following function, $\Theta_\alpha = \frac{\lambda}{\alpha}$, with λ expresses the slope of the function. An increase in the value of this parameter means that the reduction of α is more expensive. The monitoring cost of a WQMN with n stations and a probability of non-detection α is:

$$\Theta_{n,\alpha} = n \left(\frac{\lambda}{\alpha} \right) \tag{12}$$

The function is increasing in $n, (1 - \alpha), \lambda$.

4.2 Net benefit of monitoring

We compute the net benefit of monitoring $\pi_{n,\alpha}$ as the difference between the Economic Value of Information and the monitoring cost: $\pi_{n,\alpha} = EVOI_{n,\alpha} - \Theta_{n,\alpha}$; with (11) and (12), the net benefit can be written:

$$\pi_{n,\alpha} = \frac{P(\delta - C)^2}{\delta(n + 1)^2} \sum_{y=1}^n (n + 1 - y)(1 - \alpha^y) - n \left(\frac{\lambda}{\alpha} \right) \tag{13}$$

Figure 1 is obtained by assuming that $\delta = 100,000$, $C = 1,000$, $P = 10\%$, and $\lambda = 100$.

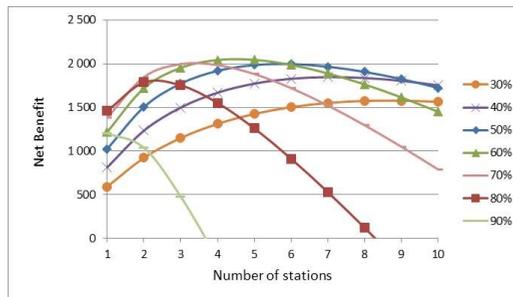


Figure 1. Net benefit according to the number of stations and probability of detection.

It can be observed in Fig.1 that the more (less) the probability of detection increases, the more the maximum of the net benefit curve is obtained with a lower (higher) number of monitoring stations. We can also note the trade-off between acting on the space parameter by increasing the number of stations and acting on time parameter by increasing the sampling frequency in order to increase the probability of detection by the stations.

5 Optimal design at a fixed budget

In this section, we consider a fixed budget, β , to establish the WQMN. In this case, a combination of the optimal number of monitoring stations (with optimized locations) and the optimal probability of detection has to be chosen. The optimal spatio-temporal design must be defined, i.e., the one that maximizes the EVOI under the constraint of fixed budget β . The methodology consists of finding the optimum by combining the iso-cost curves and the indifference curves.

5.2 Iso-cost curve

The iso-cost curve (14) is deduced from Equation (12):

$$\Theta_{n,\alpha} = n \left(\frac{\lambda}{\alpha} \right) = \beta \Leftrightarrow 1 - \alpha = 1 - \left(\frac{\lambda}{\beta} \right) n \tag{14}$$

On the basis of our hypotheses, the iso-cost curve is a line with a decreasing slope in β and an increasing slope in λ .

5.3 Indifference curves

An indifference curve is the combination of the number of monitoring stations and the probability of detecting accidental pollution that generates the same EVOI. Hence, the most efficient combination in an indifference curve is the one that reaches the lowest iso-cost curve. According to Equation (11), each indifference curve has to comply with:

$$EVOI_{n,\alpha} = \frac{P(\delta - C)^2}{\delta(n + 1)^2} \sum_{y=1}^n (n + 1 - y)(1 - \alpha^y) = \overline{EVOI} \quad (15)$$

where \overline{EVOI} is any value of EVOI.

The indifference curve has the form of the following function $f(\cdot)$: $(1 - \alpha) = f(n, P, \delta, C, \overline{EVOI})$

5.4 Optimal design

We determine the optimal WQMN design by superimposing the indifference curves and the iso-cost curves. In the bellow illustration (Fig. 2), we can observe that for a budget of 300, the optimal design consists of one monitoring station (located at 0.495 on the river/segment [0,1] according to Equation (9)) with a probability of detection of 70%. The EVOI generated by the network is 1,650. With the same budget, increasing the number of monitoring stations from one to two would require decreasing the probability of detection (from 70% to a little more than 30%) and the EVOI would decrease (the new EVOI would be between 1,000 and 1,650). In Fig. 2, we can also observe that for a budget of 500, the optimal design consists of two monitoring stations (located at 0.33 and 0.66 on the river/segment [0,1]) with a probability of detection of 60%. The EVOI generated by the network is 2,230. With the same budget, increasing the number of monitoring stations from two to three would require decreasing the probability of detection from 60% to 40%, and increasing the number of monitoring stations from three to four would require decreasing the probability of detection from 40% to 20%. It is interesting to note that an increasing budget from 300 to 500 leads to a lower probability of detection (from 70% to 60%). With the parameter values of our simulation, it would be preferable to act on a spatial dimension with an increasing budget.

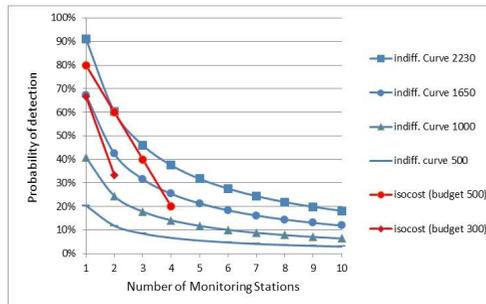


Figure 2. Optimal spatio-temporal design with a fixed budget.

Figure 2 makes it possible to understand that the slopes of iso-cost lines and indifference curves determine the location of the optimum combination. The more the optimum is at the top left (high slope of iso-cost line or low slope of indifference curve), the more the temporal dimension of the network must be intensified (higher sampling frequency). The more the optimum is at the bottom right (low slope of iso-cost line or high slope of indifference curve) of the figure, the more the spatial dimension of the network must be intensified (more monitoring stations).

6 Conclusion

In this article, we study the optimal spatio-temporal design of a WQMN. The objective of the network is to detect potential accidental pollution. The spatial approach consists in determining the number and the location of the monitoring stations. In our design, the temporal approach focuses on finding the probability of detection of

the accidental pollution for monitoring stations. This probability increases with the sampling frequency. The cost of monitoring obviously increases with the number of monitoring stations and the probability of detection.

In order to define the optimal design of the network the Economic Value of Information generated by the network must be known. Hence, the network manager, constrained by a fixed budget, has to choose between intensifying the monitoring either spatially or temporally.

Our results show that the network manager's choices must depend on the parameters of the model: the probability of the existence of accidental pollution, marginal environmental damage, the cost of action to stop environmental damage, the budget for the monitoring and the cost function form.

References

1. L. J. Alvarez-Vázquez, A. Martínez, M. E. Vázquez-Méndez, and M. A. Vilar, *Math Comput Simul*, **71** (2006)
2. I. T. Telci, K. Nam, J. Guan, and M. M. Aral, *J Environ Manag*, **90** (2009).
3. V. Naddeo, D. Scannapieco, T. Zarra, and V. Belgiorno, *Land use Policy*, **30** (2013).
4. G.-B. Kim, K.-K. Lee, J.-Y. Lee, and M.-J. Yi, *J Hydrol*, **342** (2007).
5. S. Pourshahabi, N. Talebbeydokhti, G. Rakhshandehroo, and M. R. Nikoo, *Water Resour Manag*, **32** (2018).
6. J. A. Bouma, H. J. van der Woerd, and O. J. Kuik, *J Environ Manag*, **90** (2009).
7. J. A. Bouma, O. Kuik, and A. G. Dekker, *Sci Total Environ*, **409** (2011).
8. F. Destandau and A. P. Diop, *JWARP*, **08** (2016).