

OPTIMAL SENSOR PLACEMENT FOR SIGNAL EXTRACTION

Fateme Ghayem, Bertrand Rivet, Christian Jutten

Univ. Grenoble Alpes, CNRS, Grenoble INP*,
GIPSA-lab, 38000 Grenoble, France

*Institute of Engineering Univ. Grenoble Alpes

Rodrigo Cabral Farias

Univ. Côte d’Azur, CNRS,
I3S, 06900 Sophia-Antipolis, France

ABSTRACT

This paper focuses on the optimal sensor placement problem with the purpose of signal extraction in an underdetermined noisy setting. Assuming prior information on the spatial gain of the measured signal and on the spatial noise correlation, we propose a sensor placement criterion based on the maximization of the average signal to noise ratio of the estimated signal. Our approach differs from classical Kriging based optimal sensor placement approaches, since the latter focus on best reconstruction of the spatial measured field and not on the estimation of an underlying signal. Performance analysis of the proposed criterion is presented along with synthetic results. It is observed that sensor placement with our criterion significantly outperforms other methods based on Kriging.

Index Terms— Optimal sensor placement, Kriging, Signal extraction.

1. INTRODUCTION

Optimal sensor placement plays an important role in a variety of domains such as industry, medicine, wireless communications, aerospace, biomedical and civil engineering, environmental studies, and robotics [1–7]. In such applications, one is usually dealing with data acquisition for the purpose of monitoring a spatial phenomenon. Due to resource constraints, the number of sensors is usually limited and an important task is to find the best positions to install them, such that the maximum information can be collected to provide the best estimation of the phenomenon.

In this paper, we target the problem of optimal sensor placement in the specific application of signal extraction from recordings. In other words, the aim of this study is to choose the sensor locations so that the estimation of a source signal of interest from the recordings leads to the best estimation in the signal-to-noise ratio (SNR) sense. We assume that the signal of interest is measured at each sensor with a time-invariant spatially-variant gain plus an additive Gaussian spatially-correlated noise. By using a stochastic spatial model on the gain and on the noise, we propose a robust sensor placement criterion based on the maximization of the average SNR. This criterion integrates not only the average gain and noise correlation but also the uncertainty on the gain.

This work can be seen as a different twist on optimal sensor placement using Kriging (spatial Gaussian processes), here we focus on reducing the uncertainty on the signal source to be extracted. This differs from the classical Kriging-based methods [8–10], since these methods focus on reducing the uncertainty directly on the spatial phenomenon, which would correspond to reconstruct the gain and not the signal itself. In the classical Kriging approach, sensor locations are selected according to criteria such as entropy [8, 9] or mutual information [10] on the gain, while in our case the criterion is the average SNR of the signal itself.

Our work can also be seen as a problem of optimal source extraction, thus being related to the domain of source separation [11]. We are here interested in extracting only a single source signal from a linear mixture with many other sources assuming the following: 1) prior information on the spatial gain of the target source is available; 2) the effect of the other sources on the sensors can be modelled by an additive spatially correlated Gaussian noise. It is important to stress that we do not focus on proposing a new algorithm for source extraction itself, but on choosing the sensor positions such that we obtain the best extraction performance with a given source estimator (linear estimator). It is worth noting that in the over-determined case (*i.e.* if the number of sensors is larger than or equal to the number of sources to be retrieved), there exist equivariant source separation methods [12], *i.e.* the quality of source retrieval is independent of the spatial gains and thus independent of the sensor positions. This property justifies why previous studies focused on providing the best algorithms to recover the sources but were not interested in choosing the best sensor locations. In the problem studied here, since the noise spatial correlation matrix is always full rank, the number of underlying sources composing the noise is always greater than the number of sensors. Therefore, such nice equivariance property is lost and performance may be dependent on the gain and on the noise, making the choice of sensor locations of great concern.

The rest of this paper is organized as follows: in Section 2 the details of the proposed method will be presented and discussed. Section 3 presents the performance of the proposed method and compares it with classical Kriging methods.

2. PROPOSED METHOD

This section details the proposed method to choose the best sensor locations for source extraction.

2.1. Linear estimator for source extraction

Let us assume that the observation $y(\mathbf{x}, t)$ at location \mathbf{x} and time t is a mixture of a source of interest $s(t)$ and a spatially correlated additive noise $n(\mathbf{x}, t)$ as

$$y(\mathbf{x}, t) = a(\mathbf{x})s(t) + n(\mathbf{x}, t), \quad (1)$$

where $a(\mathbf{x})$ is the spatial gain of the signal of interest $s(t)$ at location \mathbf{x} . The vector $\mathbf{x} \in \mathbb{R}^D$ represents the coordinates of the sensor in the D -dimensional space. Typically, $D \in \{1, 2, 3\}$, *i.e.* if the sensor can be placed on a curve, on a surface or in 3D space, respectively.

Considering that K sensors have been placed at locations $\mathbf{X}_K = \{\mathbf{x}_k\}_{k \in \{1, \dots, K\}}$, the aim of linear source extraction is to design a vector $\mathbf{f} \in \mathbb{R}^K$ to estimate the source by

$$\hat{s}(t) = \mathbf{f}^T \mathbf{y}(\mathbf{X}_K, t) = \mathbf{f}^T \mathbf{a}(\mathbf{X}_K) s(t) + \mathbf{f}^T \mathbf{n}(\mathbf{X}_K, t), \quad (2)$$

where $\mathbf{y}(\mathbf{X}_K, t) = [y(\mathbf{x}_1, t), \dots, y(\mathbf{x}_K, t)]^T$, $\mathbf{a}(\mathbf{X}_K) = [a(\mathbf{x}_1), \dots, a(\mathbf{x}_K)]^T$ and $\mathbf{n}(\mathbf{X}_K, t) = [n(\mathbf{x}_1, t), \dots, n(\mathbf{x}_K, t)]^T$. Classically, a criterion to choose the best \mathbf{f} is the output signal-to-noise ratio (SNR) defined by

$$SNR(\mathbf{f}) = \mathbb{E}[(\mathbf{f}^T \mathbf{a}_K s(t))^2] / \mathbb{E}[(\mathbf{f}^T \mathbf{n}_K(t))^2], \quad (3)$$

where for the sake of simplicity \mathbf{w}_K stands for $\mathbf{w}(\mathbf{X}_K)$ for any spatial function $\mathbf{w}(\cdot)$. Assuming that the signal time samples are temporally zero-mean, independent and identically distributed (iid) and denoting $\sigma_s^2 = \mathbb{E}[s(t)^2]$ and $R_K^n = \mathbb{E}[\mathbf{n}_K(t) \mathbf{n}_K^T(t)]$, then the SNR (3) resumes to $SNR(\mathbf{f}) = \sigma_s^2 (\mathbf{f}^T \mathbf{a}_K \mathbf{a}_K^T \mathbf{f}) / (\mathbf{f}^T R_K^n \mathbf{f})$. Maximizing it to express the best extraction vector \mathbf{f}^* leads to¹

$$\mathbf{f}^* = (R_K^n)^{-1} \mathbf{a}_K, \quad (4)$$

and the achieved output SNR is given by

$$SNR(\mathbf{f}^*) = \sigma_s^2 \mathbf{a}_K^T (R_K^n)^{-1} \mathbf{a}_K. \quad (5)$$

2.2. Optimal sensor location for source extraction

If M sensors are used, the aim of optimal sensor location for source extraction is thus to find the best location \mathbf{X}_M^* so that the output SNR (3) is maximum. According to (5), this problem resumes to choose \mathbf{X}_M such that

$$J(\mathbf{X}_M) = \mathbf{a}_M^T (R_M^n)^{-1} \mathbf{a}_M \quad (6)$$

is maximum: $\mathbf{X}_M^* = \arg \max_{\mathbf{X}_M} J(\mathbf{X}_M)$.

¹The scaling factor is here not tackled since in source extraction, the main goal is to enhance the signal of interest. Additional prior information on the signal amplitude can then be used to properly scale the extracted source.

A difficulty arises from this scheme: the optimal extraction vector \mathbf{f}_M^* (4) needs a perfect knowledge of the spatial gain $a(\mathbf{x})$ of the source of interest and of the spatial covariance of the noise to express the criterion (6). To overcome this, we assume that R_M^n can be modelled with a covariance kernel function $k^n(\mathbf{x}, \mathbf{x}')$ [13] that has been estimated almost perfectly (for instance from previous recording without the source of interest). On the contrary, the spatial gain $a(\mathbf{x})$ of the source of interest is imperfectly known and is modelled as a stochastic Gaussian process:

$$\hat{a}(\mathbf{x}) \sim \mathcal{GP}(m^a(\mathbf{x}), k^a(\mathbf{x}, \mathbf{x}')), \quad (7)$$

where $m^a(\mathbf{x})$ is the mean function and $k^a(\mathbf{x}, \mathbf{x}')$ is the covariance function. From this modelling, $m^a(\cdot)$ is the main behaviour of the spatial gain and $k^a(\cdot, \cdot)$ represents the uncertainty that we have on it. In practice, these two functions can be expressed either from some prior knowledge of the physical recorded phenomenon, either from previous estimation, such as the output of independent component analysis applied on previously recorded data. Consequently, in practice the criterion to optimize is defined as the mean output SNR

$$\hat{J}(\mathbf{X}_M) = \mathbb{E}[\hat{\mathbf{a}}_M^T (R_M^n)^{-1} \hat{\mathbf{a}}_M], \quad (8)$$

where $\hat{\mathbf{a}}_M$ is an estimation of \mathbf{a}_M . Using uncertainty model (7), \hat{J} resumes to

$$\hat{J}(\mathbf{X}_M) = (\mathbf{m}_M^a)^T (R_M^n)^{-1} \mathbf{m}_M^a + \text{Tr}((R_M^n)^{-1} R_M^a), \quad (9)$$

where $R_M^a \in \mathbb{R}^{M \times M}$ is the covariance matrix whose (i, j) th element is $k^a(\mathbf{x}_i, \mathbf{x}_j)$ and $\text{Tr}(\cdot)$ is the trace operator. The optimal sensor locations are thus obtained as

$$\hat{\mathbf{X}}_M = \arg \max_{\mathbf{X}_M} \hat{J}(\mathbf{X}_M). \quad (10)$$

However, directly maximizing (9) can lead to a high computational cost since it needs to place M sensors in a D -dimensional space simultaneously (*i.e.* it is an optimization problem of size $M \times D$). To avoid this, one can use a greedy approach that selects the M sensors by sequentially selecting $N < M$ sensors at a time. Assuming that K sensors have already been placed, to choose the locations of the N following ones, criterion (8) will be recast as

$$\hat{J}(\mathbf{X}_N | \mathbf{X}_K) = \mathbb{E}[\hat{\mathbf{a}}_{K+N}^T (R_{K+N}^n)^{-1} \hat{\mathbf{a}}_{K+N} | \mathbf{X}_K], \quad (11)$$

where $K+N$ means $\{\mathbf{X}_N \cup \mathbf{X}_K\}$ and thus $\hat{\mathbf{a}}_{K+N} \in \mathbb{R}^{K+N}$ can be divided as $\hat{\mathbf{a}}_{K+N} = [\hat{\mathbf{a}}_K^T, \hat{\mathbf{a}}_N^T]^T$. The conditional mean (11) means that $\hat{\mathbf{a}}_K$ and the upper diagonal block of R_{K+N}^n are independent of the new sensor locations \mathbf{X}_N .

Once the sensor locations $\hat{\mathbf{X}}_M$ are obtained, one can extract the source of interest (2) using the separation vector

$$\hat{\mathbf{f}}_M = (R_M^n)^{-1} \mathbf{m}_M^a. \quad (12)$$

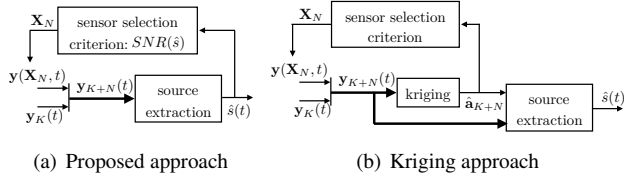


Fig. 1. Comparison between Kriging and the proposed sensor selection scheme for source extraction.

It is important to note that the proposed approach gives a direct criterion for sensor placement for source extraction (Fig. 1(a)). This is in contrast to standard sensor placement using Kriging which, in the case of source extraction, corresponds to a two-step procedure: first, standard Kriging will select the sensor locations so that information on the spatial gain of the source is maximized and then using the selected sensor positions and the predictions of their gains, one can extract the source of interest from the mixtures (Fig. 1(b)).

2.3. Performance analysis

Assuming that the estimation of $a(\mathbf{x})$ by the mean function $m(\mathbf{x})$ can be written as $m^a(\mathbf{x}) = a(\mathbf{x}) + b(\mathbf{x})$, where $b(\mathbf{x})$ would represent the estimation error, it is worth noting that the optimized criterion (9) can be expressed as

$$\hat{J}(\mathbf{X}_M) = J(\mathbf{X}_M) + 2\mathbf{a}_M^T (R_M^n)^{-1} \mathbf{b}_M^a + (\mathbf{b}_M^a)^T (R_M^n)^{-1} \mathbf{b}_M^a + \text{Tr}((R_M^n)^{-1} R_M^a). \quad (13)$$

In other words, the criterion \hat{J} is equal to the true criterion J (knowing perfectly $a(\mathbf{x})$) plus a term depending on the estimation error \mathbf{b}_M^a and on the uncertainty R_M^a on the spatial gain. Consequently, this shows that depending on the particular realization of error term, the optimal selected sensor locations $\hat{\mathbf{X}}_M$ are not necessarily the true optimal sensor locations \mathbf{X}_M^* . The larger the error and uncertainty are, the more different the estimated SNR \hat{J} to the optimal SNR J is.

For separation vector (12), the achieved SNR is given by

$$SNR(\hat{\mathbf{f}}_M) = \sigma_s^2 \frac{(\mathbf{z}_M^T \mathbf{m}_M^a)^2 + \mathbf{z}_M^T R_M^a \mathbf{z}_M}{(\mathbf{m}_M^a)^T (R_M^n)^{-1} \mathbf{m}_M^a + \text{Tr}((R_M^n)^{-1} R_M^a)}$$

with $\mathbf{z}_M = (R_M^n)^{-1} \mathbf{a}_M$, which can be rewritten as

$$SNR(\hat{\mathbf{f}}_M) = SNR(\mathbf{f}_M^*) \alpha / \beta, \quad (14)$$

where

$$\beta = 1 + [2\mathbf{z}_M^T \mathbf{b}_M^a + (\mathbf{b}_M^a)^T (R_M^n)^{-1} \mathbf{b}_M^a + \text{Tr}((R_M^n)^{-1} R_M^a)] / (\mathbf{z}_M^T \mathbf{a}_M),$$

$$\alpha = (1 + (\mathbf{z}_M^T \mathbf{b}_M^a) / (\mathbf{z}_M^T \mathbf{a}_M))^2 + (\mathbf{z}_M^T R_M^a \mathbf{z}_M) / (\mathbf{z}_M^T \mathbf{a}_M)^2.$$

This means that the achieved SNR, $SNR(\hat{\mathbf{f}})$, is equal to the oracle one, $SNR(\mathbf{f}^*)$, up to a multiplicative term that depends both on the error \mathbf{b}_M^a of the estimation on \mathbf{a}_M and

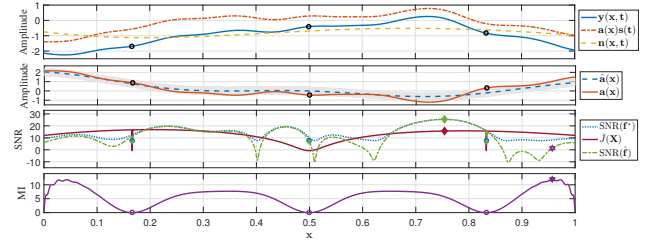


Fig. 2. Illustration of the proposed method to select optimal sensor location. From top to bottom: example of recordings $y(x, t) = a(x)s(t) + n(x, t)$ for a given t . 2nd plot: actual gain $a(x)$ and its estimation $\hat{a}(x)$. Grey shadow is the uncertainty. 3rd plot: Oracle SNR (5), estimated SNR (9) and achieved SNR (14). Bottom: MI criterion [10]. In all plots, the circles represent the three pre-selected sensors. The diamonds indicate the maximum of the proposed criterion and the stars of the MI criterion.

on the uncertainty R_M^a of $a(\mathbf{x})$. This multiplicative term is lower than one (by the definition of \mathbf{f}^*) indicating a loss of performance due to the error of estimation and uncertainty compared to the oracle SNR, $SNR(\mathbf{f}^*)$.

3. NUMERICAL EXPERIMENTS

The proposed method to choose the location of the sensors (Fig. 1(a)) will be compared to Kriging methods (Fig. 1(b)) based on entropy [8,9], referred to as ‘Entropy’ and on mutual information [10], referred to as ‘MI’. The methods will be compared according to the achieved output SNR (14) relying on the extraction vector $\hat{\mathbf{f}}_M$ that depends on the chosen sensor locations. The optimal SNR (5) will be denoted oracle SNR since it assumes a perfect knowledge on the gain $a(\mathbf{x})$.

3.1. Numerical setup

The data are generated synthetically, and D , the dimension of the space, is equal to 1. The range of x is normalized leading to $x \in [0, 1]$ in a grid of size 200 along with three initial sensors to be located arbitrary at $K = \{33, 100, 167\}$. Both signals $a(x)$ and $n(x)$ are produced from Gaussian processes (GP) $\mathcal{GP}(m(x), C(x, x'))$, with a square exponential covariance function $C(x, x') = \sigma^2 \exp(-(x - x')^2 / (2\rho^2))$. The mean function $m^n(x)$ for the noise is 0 and $m^a(x)$ for the gain $a(x)$ is randomly generated by a GP with a zero-mean function and $\rho = 0.2$. The length scale ρ_a (resp. ρ_n) related to $a(x)$ (resp. to $n(x)$) will be changed as described in the next sections.

3.2. Accuracy of the proposed criterion

In the first part of the simulations, the accuracy of the proposed method is evaluated and presented in Fig.2. To generate $a(x)$, the smoothness and power parameters defining the covariance function are $\rho_a = 0.05$ and $\sigma_a = 0.5$, respectively. 20000 time samples of $s(t)$ are randomly inde-

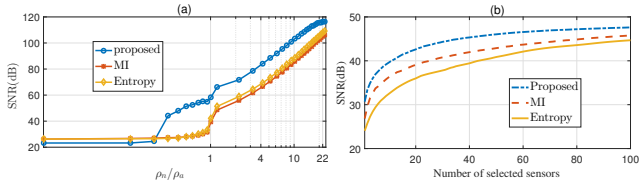


Fig. 3. (a): The dependency of the performance to the smoothness ratio between noise and spatial gain $\frac{\rho_n}{\rho_a}$. (b): Comparing the performance of proposed method with entropy and MI according to the improvement of SNR versus the number of selected sensor using greedy method.

pendently drawn from a zero mean normal distribution with $\sigma_s = 1$. The 20000 taps of noise $n(x, t)$ are generated from the GP with $\rho_n = 0.5$ and σ_n so that the average SNR on each sensor is of -5dB .

To optimize (9), a grid search method is used: 1001 linearly spaced possible sensor locations were tested between 0 and 1, and 3 sensors are pre-located at 0.166, 0.5 and 0.833, respectively. The results are shown in Fig. 2. In the third sub-figure, it is seen that the proposed criterion $\hat{J}(\mathbf{x})$ (*i.e.* the estimated SNR from imperfect knowledge on $a(x)$) plotted in red has roughly the same behaviour than the oracle SNR (5) which requires perfect knowledge of the gain. The differences are explained by (13) and come from the estimation error $b(x)$ and the uncertainty. Interestingly, the maximum value of \hat{J} (9) is located in a range of high value for the actual SNR (14) leading thus to an output SNR of about 25dB with 4 sensors while the output SNR with only the 3 pre-located sensors is equal to 9dB.

Moreover, it is worth noting that the achieved SNR with 4 sensors, $SNR(\hat{\mathbf{f}})$, in green, can be lower than the output SNR with only the 3 pre-located sensors (*e.g.*, for $x \geq 0.85$). This means that adding a sensor will not necessarily leads to a better SNR but can deteriorate the quality of the extraction. This counter intuitive result can be explained from the imperfect estimation $\hat{a}(x)$. In this case, based on an inadequate model, the extraction vector based on $K + 1$ sensors $\hat{\mathbf{f}}_{K+1}$ leads to a worse estimation $\hat{s}(t)$ than using only K sensors using $\hat{\mathbf{f}}_K$.

Finally, in the bottom sub-figure in Fig. 2, we can clearly see that the MI criterion leads to a far from optimal choice of sensor location with respect to SNR.

3.3. Comparison with state of the art

In this part, firstly, the sensitivity of the methods to the smoothness of the noise relative to the smoothness of the spatial gains are compared. In Fig.3(a) the actual SNRs are plotted according to different ratios ρ_n/ρ_a . In this experiment, all the simulation parameters are similar to the previous experiment (Section 3.2) except that the average SNR on each sensor is set to be 3dB. Also, ρ_a is set 0.05. For each ratio ρ_n/ρ_a , 50 data sets were randomly generated. From this figure, it can be concluded that the proposed method performs better than the other methods in different smoothness

situations. It is noticeable that for the very small amounts of ρ_n/ρ_a , the performance of none of the methods provide significant improvement. For instance, consider the point where $\rho_n/\rho_a = 3$; the resulted SNR after adding one new sensor using the proposed method is about 80dB , while the value of SNRs using MI and entropy are around 65dB, respectively. Note that depending on the ratio ρ_n/ρ_a , the optimal location of the new sensor varies.

In the second part of the simulations (Fig. 3(b)), the efficiency of the proposed method to choose the sensor locations is tested. The sensor locations were chosen with a greedy approach: at each iteration, a single new sensor is added to the previous set of sensors, starting with an empty set. Three different criteria are compared to select the new sensor location: the proposed method, MI and entropy. This figure shows the actual output SNR (14) for each size of the set of sensors. In this part, ρ_n/ρ_a is set to be 0.01, and 20 data sets were randomly generated. The performance of our method is significantly better than the other two methods, specially when the number of sensors is very limited. For instance, if the maximum number of 10 sensors can be used, the results achieved by the proposed method is about 11dB better than entropy and also 6dB better than MI. To reach a SNR of 40dB, the proposed criterion needs about 10 sensors, while more than 25 are required by the MI method and about 45 with the entropy one, showing the efficiency of the proposed approach to choose relevant sensor locations for source extraction.

4. CONCLUSIONS AND PERSPECTIVES

In this paper we have studied the problem of optimal sensor placement for signal extraction from noisy underdetermined measurements collected by a limited number of sensors. The average output SNR of the linearly extracted signal has been proposed as a quality criterion to select sensor locations. Such a criterion includes the uncertainty on the spatial gain of the source to be extracted, providing a suitable solution for the optimal sensor placement problem. Numerical simulations have shown the superior efficiency and accuracy of the proposed method in the source extraction problem when compared to classical sensor placement criteria such as entropy and mutual information.

For future works, in addition to a more in-depth analysis of the influence of modeling parameters, several other aspects can be studied: first, we can obtain an analytic characterization of the regions where placing new sensors deteriorate the extraction. Second, we can propose an extension of this approach to the case where an imperfectly known noise covariance kernel is considered. Finally, the proposed method is a model based approach, requiring knowledge on the spatial gains and noise, the possibility to tackle the optimal sensor placement problem in a data driven approach, for example using independent component analysis, can also be studied.

5. REFERENCES

- [1] J. W. Berry, L. Fleischer, W. E. Hart, C. A. Phillips, and J. P. Watson, "Sensor placement in municipal water networks," *Journal of Water Resources Planning and Management*, vol. 131, no. 3, May 2005.
- [2] M. Meo and G. Zumpano, "On the optimal sensor placement techniques for a bridge structure," *Engineering Structures*, vol. 27, no. 10, pp. 1488–1497, August 2005.
- [3] H. J. Hermens, B. Frenriks, C. Disselhorst-Klugb, and G. Rau, "Development of recommendations for SEMG sensors and sensor placement procedures," *Journal of Electromyography and Kinesiology*, vol. 10, no. 5, pp. 361–374, October 2000.
- [4] S. Y. Chen and Y. F. Li, "Automatic sensor placement for model-based robot vision," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 34, no. 1, pp. 393–408, February 2004.
- [5] C. K. Daniel, "Sensor placement for on-orbit modal identification and correlation of large space structures," *Journal of Guidance, Control, and Dynamics*, vol. 14, no. 2, pp. 251–259, 1991.
- [6] S. Martinez and F. Bullo, "Optimal sensor placement and motion coordination for target tracking," *Automatica*, vol. 42, no. 4, pp. 661–668, April 2006.
- [7] A. Krause, J. Leskovec, C. Guestrin, J. VanBriesen, and C. Faloutsos, "Efficient sensor placement optimization for securing large water distribution networks," *Journal of Water Resources Planning and Management*, vol. 134, no. 6, November 2008.
- [8] M. C. Shewry and H. P. Wynn, "Maximum entropy sampling," *Journal of Applied Statistics*, vol. 14, no. 2, pp. 165–170, 1987.
- [9] N. Cressie, "The origins of kriging," *Mathematical Geology*, vol. 22, no. 3, pp. 239–252, 1990.
- [10] A. Krause, A. Singh, and C. Guestrin, "Near-optimal sensor placements in gaussian processes: Theory, efficient algorithms and empirical studies," *Journal of Machine Learning Research*, vol. 9, pp. 235–284, 2008.
- [11] P. Comon and C. Jutten, Eds., *Handbook of Blind Source Separation Independent Component Analysis and Applications*, Academic Press, 2010.
- [12] J.-F. Cardoso, "Blind signal separation: statistical principles," *Proceedings of the IEEE*, vol. 86, no. 10, pp. 2009–2025, October 1998.
- [13] C. E. Rasmussen and C. K. I. Williams, *Gaussian processes for machine learning*, vol. 1, MIT press Cambridge, 2006.