
Charles C. Castello and Jeffrey Fan
Department of ECE
Florida International University
Miami, FL, USA

Asad Davari
Department of ECE
WVU Institute of Technology
Montgomery, WV, USA

Ruei-Xi Chen
Dept. of Comp. Sci. and Info. Eng.
St. John’s University
Taipei, Taiwan

Abstract — This paper presents a novel strategy in determining an optimal sensor placement scheme for environmental monitoring using Wireless Sensor Networks (WSN). This is accomplished by minimizing the variance of spatial analysis based on randomly chosen points representing the sensor locations. These points are assigned randomly generated measurements based on a specified distribution. Spatial analysis is employed using Geostatistical Analysis (classical variography and ordinary point kriging) and optimization occurs with Monte Carlo Analysis. A simple example of measuring mercury in soil is illustrated in finding the optimal sensor placement using WSNs. Studied variables include the number of sensor locations, variances, and Monte Carlo repetitions.

Keywords — Optimal Sensor Placement; Wireless Sensor Network; WSN; Environmental Monitoring; Geostatistical Analysis; Classical Variography; Ordinary Point Kriging; Monte Carlo Analysis

I. INTRODUCTION

Wireless Sensor Networks (WSN) have become an important component in today’s society due to the numerous applications such as defense, manufacturing, security, and weather forecasting [1]-[3]. Research dealing with the improvement of WSNs ranges from energy efficient routing protocols [4], energy efficient Medium Access Control (MAC) protocols [5][6], and improved sensitivity in sensor technology [7][8]. Another research topic dealing with WSNs is sensor placement. This is an extremely important area of research due to the difficulty of determining optimal coverage with a limited amount of sensor nodes. The complexity of determining optimal sensor placement is caused by the properties of the environment, the types of obstacles in the environment, and the sensing phenomena [9]. This research focuses on the optimal sensor placement in WSN by using Geostatistical Analysis which utilizes classical variography and ordinary point kriging and Monte Carlo Analysis.

A grid map is created based on area size and grid division. Values for the measured phenomena are randomly generated for each grid location based on the characteristics of said phenomena (e.g. mean, standard deviation, and probability distribution). The locations of N sensor nodes in the WSN are randomly generated on the grid map and spatial analysis is performed based on the location’s respective phenomena values and using Geostatistical Analysis. The mean variance is calculated based on the spatial analysis results and then stored. This process occurs for M repetitions in Monte Carlo Analysis and the sensor placement configuration where the minimum mean variation produced is the optimal sensor placement. As M repetitions are increased, the minimum mean variation and error decrease due to the Law of Large Numbers.

There has been much research dealing with optimal sensor deployment in WSNs. Schoellhammer et al. [9] evaluates two different approaches to sensor placement. Spatial-temporal statistical analysis is utilized which combines spline-based modeling, principal component analysis, and data partitioning. An Integer Linear Programming (ILP) formulation is used to prove the optimal solution to the network deployment problem. Research work in [10] investigates the relationship between the network lifetime and sensor coverage. Optimal sensor placement is accomplished based on sensing range where maximizing the network lifetime is also taken into consideration. Krause et al. [11] utilizes non-parametric probabilistic models called Gaussian Processes (GPs) for the spatial phenomena of interest and for the spatial variability of link qualities which allows for the estimation of predictive power and communication cost of unsensed locations. A novel, polynomial-time, data-driven algorithm, pSPIEL, is also presented, which selects the Sensor Placements at Informative and cost-Effective Locations. Lastly, Armaou et al. [12] studies optimal sensor placement in the presence of disturbance by using linear parabolic partial differential equations. From the literature review dealing with optimal sensor placement, there were no other research works that use Geostatistical Analysis and the Monte Carlo technique in determining an optimal sensor placement scheme which adds to not only the sensor placement research area, but also optimal estimation.

The remainder of this research paper is organized as follows: Section II reviews Geostatistical Analysis methodologies being used in the proposed technique, Section III describes the proposed optimal sensor placement technique, Section IV describes the experimental setup,
Section V examines generated results, and Section VI concludes with final thoughts and future work.

II. GEOSTATISTICAL ANALYSIS

There have been many studies in environmental sciences that use geostatistical analysis to help understand phenomena on a spatial plane. This type of analysis could provide valuable information for improved scientific understanding, risk assessment of contaminants or pollutants, and decision support. Examples of such work include using geostatistics to assess mercury in soils around a coal-fired power plant in Baoji, China [13], analyzing spatial correlation of nitrogen dioxide (NO₂) concentrations in Milan, Italy [14], studying spatial distribution of soil lead in the mining site of Silvermines, Ireland [15], and development of a spectrophotometer in generating detailed soil maps for rice paddy fields using geostatistical analysis [16].

The geostatistical methods used in this paper will now be described which include variography and kriging, briefly summarized from [17]-[18]. The theoretical basis of both techniques can be found in greater detail in [19]-[21].

A. Classical Variogram

In geostatistical analysis, the variogram is used to describe the spatial dependency between referenced observations within the analyzed plane where the true variogram is generally unknown. Therefore, an estimation of the variogram can be calculated through known observations. This is accomplished by first determining the experimental variogram, given by the semivariance:

\[ \gamma(h) = 0.5 \sum_{i=1}^{N(h)} (z_{x} - z_{x+h})^2 \]  

(1)

where \( z_x \) is the observed value at point \( x \) and \( z_{x+h} \) is an observed value at another point within distance \( h \). \( \gamma(h) \) is also known as a semivariogram or variogram and the distance between observations is known as lag distance.

The next step is to summarize the experimental variogram with a variogram estimator, which determines the central tendency and is similar to descriptive statistics derived from univariate observations [17]. The variogram estimator is calculated by Equation (2) where \( N(h) \) represents the amount of pairs within lag interval \( h \).

\[ \gamma_k(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} (z_{x} - z_{x+h})^2 \]  

(2)

Lastly, a variogram model or parametric curve is fitted to the variogram estimator. This is similar to frequency distribution fitting where the frequency distribution is modeled by a distribution type and its parameters [17]. The three commonly used models in variography are: spherical, exponential, and linear.

B. Ordinary Point Kriging

In order to interpolate observations on a regular grid, ordinary point kriging is used, which is considered the most commonly utilized method. Ordinary point kriging uses weighted averages and neighboring observations to predict unobserved points:

\[ \hat{z}_{x_0} = \sum_{i=1}^{N} \lambda_i z_{x_i} \]  

(3)

where \( \lambda_i \) are estimated weight values. In order to guarantee that estimates are unbiased, the sum should equal one.

\[ \sum_{i=1}^{N} \lambda_i = 1 \]  

(4)

Equation (5) shows the average estimation error must equal zero where \( \hat{z}_{x_0} \) is the unknown value.

\[ E(\hat{z}_{x_0} - z_{x_0}) = 0 \]  

(5)

Calculating the mean-square error using Equations (3) through (5) in terms of the variogram and using a Lagrange multiplier \( v \) for optimization yields a linear kriging system of \( N+1 \) equations and \( N+1 \) unknowns which is calculated by:

\[ \sum_{i=1}^{N} \lambda_i \gamma(x_i, x_j) + v \gamma(x_i, x_0) \]  

(6)

where \( \lambda_i \) is the weight for the \( i \)th data point, \( \gamma(x_i, x_j) \) is the variogram between data points \( x_i \) and \( x_j \), and \( \gamma(x_i, x_0) \) is the variogram between the data point and unobserved point. The kriging variance is calculated by:

\[ \sigma^2(x_0) = \sum_{i=1}^{N} \lambda_i \gamma(x_i, x_0) + v \gamma(x_i, x_0) \]  

(7)

III. OPTIMAL PLACEMENT THEORY

An overview of the proposed optimal placement technique is shown in Figure 1. The first step of the process is to define key variables such as the area dimensions \((L \times W)\), grid division of said area \((X \times Y)\), the nodal density \((N)\), and the number of repetitions in the Monte Carlo Analysis \((M)\). The characteristics of the phenomena being monitored must also be specified such as the mean and standard deviation values. A grid map is generated based on the defined dimensions and data is randomly produced for each grid point which is established from the defined statistical characteristics of the phenomena. \( N \) random grid points are chosen from the grid map and spatial analysis is performed based on those location’s respective values. Variography is first used in order to determine spatial dependency of referenced observations and kriging, to interpolate across space for predicted values between observed locations [23].

The mean variance of the kriging results is calculated and stored as the optimization variable and steps 4 through 7 are repeated \( M \) times where the minimum mean variance is found. The \( N \) sensor locations that generated the minimum
mean variance is therefore the most optimal sensor placement strategy based on the \( M \) repetitions.

IV. EXPERIMENTAL SETUP

The sensor placement strategy discussed previously will be applied to the sensing of Total Mercury (THg) outside Oak Ridge Reservation’s Y-12 Security Complex at Oak Ridge, TN, where mercury contamination is considered a major concern. Mean (\( \sigma \)) and standard deviation (\( \mu \)) values of THg at this location were taken from Liu et al. [22] where \( \sigma = 191.9 \) ng/g and \( \mu = 25.55 \) ng/g which are based on redoximorphic concentrations. These two values will be used to randomly generate THg concentrations based on assuming a Gaussian distribution during the optimal sensor placement procedure.

The experiment is setup using a 40m x 40m area with a grid division of 10 x 10 which is shown in Figure 2. Varying numbers of sensors were studied (\( N = 10, 25, 50, \) and 75) with Monte Carlo repetitions of \( M = 1, 5, 10, 50, 100, 500, 1k, 5k, 10k, 20k, 30k, 40k, \) and 50k. Based on the Law of Large Numbers, when \( M = 50k \), error is equal to 0.4472% where convergence occurs at 1/\( M \) .

V. RESULTS

The resulting optimal sensor placement scheme when \( N = 50 \) and \( M = 50k \) is shown in Figure 3 where the kriging and variance results are shown in Figures 4 and 5 respectively.

Figure 1 - Optimal Sensor Placement Procedure

Figure 2 - Grid Points Map of Potential Sensor Locations

Figure 3 - Optimal Sensor Placement when N=50 and M=50k

Figure 4 – Kriging Results when N=50 and M=50k
Simulations varying the amount of sensors and the Monte Carlo repetitions were performed where \( N = 10, 25, 50, \) and \( 75 \) and \( M = 1, 5, 10, 50, 100, 500, 1k, 5k, 10k, 20k, 30k, 40k, \) and \( 50k. \) Resulting minimum variances are shown in Figure 6 where the a decrease in the minimum mean variance trend occurs as the number of Monte Carlo repetitions increase, which is to be expected due to the Law of Large Numbers. The respective values of these graphs can be viewed in Table 1.

<table>
<thead>
<tr>
<th>( M )</th>
<th>( M ) Variance ((N=10))</th>
<th>( M ) Variance ((N=25))</th>
<th>( M ) Variance ((N=50))</th>
<th>( M ) Variance ((N=75))</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>446.91</td>
<td>318.12</td>
<td>218.74</td>
<td>157.41</td>
</tr>
<tr>
<td>5</td>
<td>440.30</td>
<td>313.23</td>
<td>209.26</td>
<td>155.47</td>
</tr>
<tr>
<td>10</td>
<td>430.59</td>
<td>305.78</td>
<td>206.32</td>
<td>155.29</td>
</tr>
<tr>
<td>50</td>
<td>421.63</td>
<td>304.60</td>
<td>203.07</td>
<td>154.40</td>
</tr>
<tr>
<td>100</td>
<td>413.20</td>
<td>297.56</td>
<td>201.84</td>
<td>153.90</td>
</tr>
<tr>
<td>500</td>
<td>409.52</td>
<td>293.07</td>
<td>201.31</td>
<td>152.88</td>
</tr>
<tr>
<td>1k</td>
<td>411.84</td>
<td>289.21</td>
<td>198.99</td>
<td>152.38</td>
</tr>
<tr>
<td>5k</td>
<td>408.53</td>
<td>281.10</td>
<td>197.46</td>
<td>151.03</td>
</tr>
<tr>
<td>10k</td>
<td>407.45</td>
<td>290.90</td>
<td>196.78</td>
<td>149.56</td>
</tr>
<tr>
<td>20k</td>
<td>407.45</td>
<td>288.56</td>
<td>195.76</td>
<td>150.39</td>
</tr>
<tr>
<td>30k</td>
<td>407.45</td>
<td>288.43</td>
<td>195.88</td>
<td>150.60</td>
</tr>
<tr>
<td>40k</td>
<td>409.23</td>
<td>288.21</td>
<td>195.57</td>
<td>150.60</td>
</tr>
<tr>
<td>50k</td>
<td>405.86</td>
<td>286.52</td>
<td>194.51</td>
<td>150.09</td>
</tr>
</tbody>
</table>

### VI. Conclusion

An optimal placement strategy using Geostatistical Analysis and Monte Carlo theory has been developed. Simulations have been processed using a THg example by extracting statistical characteristics from previous research. It was shown through the simulations that the minimum mean variation decreases as the Monte Carlo repetitions increase which is shown in the \( N = 10, 25, 50, \) and \( 75 \) simulations. The ultimate goal in determining the optimal sensor placement configuration is finding the sensor placement with the minimum mean variance.

Potential future work for the optimal sensor placement strategy includes taking into consideration other aspects of WSN such as communication costs. Obstacles can also be considered when dealing with the determination of the optimal sensor placement configuration. Other spatial analysis techniques can also be investigated to potentially improve the accuracy of the proposed method. This technique can be expanded from 2-D analysis to 3-D analysis based on application need and Lastly, further inquiries can be made into modifying the proposed sensor placement scheme for visual sensors in applications such as video surveillance, target tracking and identification, traffic monitoring, and smart homes.

### ACKNOWLEDGMENT

This project was supported by the National Aeronautics and Space Administration through the University of Central Florida’s Florida Space Grant Consortium.

### REFERENCES


