*Research Article*

# A Novel Bayesian Method for Calculating Circular Error Probability with Systematic-Biased Prior Information

**Bowen Liu** [iD]**, Xiaojun Duan** [iD]**, and Liang Yan**

*College of Liberal Arts and Sciences, National University of Defense Technology, Changsha, Hunan 410000, China*

Correspondence should be addressed to Xiaojun Duan; xjduan@nudt.edu.cn

Circular Error Probability (CEP) is defined as the radius of a circle where the probability of an impact point being inside is 50%, which is also widely used as a measure of the guidance weapon systems' precision. In order to achieve a fusion of various test information, Bayesian methods and improved Bayesian methods have been extensively studied in calculating the CEP. Nevertheless, these methods could fail when there exists unknown systematic bias in the prior information. Therefore, a novel method called Bayesian estimation based on representative points (BERP) with an optimization procedure for determining the optimal number of representative points is proposed in this paper. Explicit theoretical analyses demonstrate that the BERP outperforms the classical Bayesian methods when fusing the slightly biased prior information and also give the bound of the systematic bias for stopping using the heavily biased prior information. Moreover, when the systematic bias is within the bound, simulation results indicate that our method is credible and outperforms the classical Bayesian method in calculating the CEP of guidance weapon systems.

## 1. Introduction

Performance evaluation of complex equipment, such as the guidance weapon systems, is very important before application. In the assessment of impact accuracy of the guidance weapon systems, CEP is the most commonly used as a measure, which can integrate the precision and dispersion to assess the impact accuracy [1–3]. Usually, the tighter the pattern of impact point errors, the smaller the CEP we can get; i.e., there is higher impact accuracy of guidance weapon systems. Let $(X, Z)$ be the impact point errors of the projectile, where $X$ and $Z$ are downrange and cross-range misses, respectively; then the CEP can be defined as follows [4, 5]:

$$\mathbb{P}\left(X^2 + Z^2 \leq \text{CEP}^2\right) = \iint_{x^2 + z^2 \leq \text{CEP}^2} f(x, z) \, dx \, dz \tag{1}$$
$$= 0.5$$

It is shown in the above equation that the CEP can be calculated by numerical integrations when giving the probability density function $f(x, z)$. In practice, $(X, Z)$ are usually

assumed to follow the bivariate normal distribution, i.e., $(X, Z) \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are the mean and the covariance, respectively. Then the problem of calculating CEP rests on estimating $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, whose accuracy will determine the precision of the CEP calculation.

In the performance evaluation of guidance weapon systems, the impact points used to calculate the CEP are collected from various tests. The realistic tests of guidance weapon systems are usually extremely expensive and time-consuming, so generally the sample size of impact points is very small. Moreover, the calculation of CEP based on the small sized data would lead to unreliable results, so it is reasonable to introduce the prior information to refine the results. Therefore, Bayesian estimation is widely used to achieve the fusion of the prior information and the realistic test information [6, 7] and hence to increase the reliability and the accuracy of the estimation of $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, as well as the CEP. In fact, the most important prior information is provided by the substitute tests in the development processes. The performance evaluation is a sequential process, so the data collected from substitute tests are usually regarded as prior samples. In order to use the prior information more

reasonably, several strategies are introduced in bringing the credibility of the prior information into Bayesian estimation, such as data compatibility tests, the information divergence, and the theory of fuzzy operators [8, 9]. In the CEP calculation, Huang comes up with a measure of credibility from physics resources of data [10], which has a good estimation accuracy if we have a clear understanding of the physical background. Duan suggested a method with all prior information normalized into one test sample [11], which can reduce the deviation when prior information got distorted, but more theoretical discussion is needed for this method when considering the information loss and fusion efficiency of the normalization.

However, the substitute tests for the guidance weapon systems may be systematically biased compared with realistic tests. Since the pattern of systematic bias is unknown, we cannot estimate the bias but can only give a rough range of it. Simulation results demonstrate that the systematically biased prior information will cause serious impact on the mean estimation $\boldsymbol{\mu}$ when applying classical Bayesian estimation directly. The estimation of the parameter $\boldsymbol{\Sigma}$ is slightly affected by the unknown systematic bias because of the same guidance system of these tests. The improved methods considering the normalization of prior information could reduce the estimate bias of $\boldsymbol{\mu}$, but this will cause much loss of prior information and inevitably leads to an unreliable calculation of CEP. One possible way to solve the problem is by choosing appropriate samples, such as representative points to generate new prior information.

Representative points (RPs), also known as Principal Points [12], are a group of points that could represent a distribution with the least mean squared error (MSE). RPs' theory is brought up in 1990s and now widely applied in clustering analyses [13], statistical simulations [14], image processing [15], and so on. In this paper, we resample samples from the systematically biased data to get the RPs which are regarded as the new prior information, expecting to reduce the estimate bias of $\boldsymbol{\mu}$ substantially. This new estimation method is called Bayesian estimation based on representative points (BERP). Meanwhile, we propose an optimization procedure which balances the effects of estimate bias and information loss to determine the optimal number of RPs. In addition, two theorems are proposed to prove that the estimate bias and MSE of $\boldsymbol{\mu}$ with RPs are smaller than those with the raw data. Furthermore, we also analyze the bound of the systematic bias for stopping using the heavily biased prior information. Within the bound, both the simulations and authentic experiments show that our BERP outperforms the classical Bayesian methods in estimating the parameters. In the performance evaluation of guidance weapon systems, via BERP it is better to choose RPs from the raw prior samples as new prior information when calculating the CEP. On the whole, by using the theory of RPs, our works enrich the Bayesian methods especially on the prior information fusion patterns theoretically and provide a possible way to solve the engineering problem in performance evaluation of guidance weapon systems.

The rest of this paper is organized as follows. Section 2 introduces the classical Bayesian estimation applied in CEP calculation. Section 3 describes some notations and preliminaries of RPs, the process of BERP, and the optimization procedure for determining the optimal number of RPs. In Section 4, we propose two theorems to compare the estimation performances by BERP and classical Bayesian estimation and analyze the bound of systematic bias for stopping using the biased prior information. Section 5 shows three numerical experiments about our new method. The conclusion is given in Section 6.

## 2. Bayesian Estimation in CEP Calculation

*2.1. Calculation of CEP.* Suppose that $(X, Z)$ follows the bivariate normal distribution $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where

$$
\begin{aligned}
\boldsymbol{\mu} &= \left(\mu_x, \mu_z\right)^T, \\
\boldsymbol{\Sigma} &= \begin{pmatrix} \sigma_x^2 & \rho\sigma_x\sigma_z \\ \rho\sigma_x\sigma_z & \sigma_z^2 \end{pmatrix}
\end{aligned}
\tag{2}
$$

The parameters $\sigma_x$, $\sigma_z$ are standard deviations of impact point errors for downrange direction and cross-range direction, respectively; $\mu_x$, $\mu_z$ are means of impact point errors for each direction; $\rho$ ($0 \leq |\rho| < 1$) is the correlation coefficient of $X$ and $Z$. In most cases, $X$ and $Z$ are independent of each other. Nevertheless, even if $\rho \neq 0$, we can use the orthogonal transformation to achieve the decorrelation of $X, Z$. So we assume that $\rho = 0$ in the rest of the paper. Under the assumption that $\rho = 0$, the CEP $R$ satisfies the equation

$$
\iint_{x^2+z^2 \leqslant R^2} \frac{1}{2\pi\sigma_x\sigma_z} \exp\left\{ -\frac{\left(x-\mu_x\right)^2}{2\sigma_x^2} - \frac{\left(z-\mu_z\right)^2}{2\sigma_z^2} \right\} dx dz = 0.5
\tag{3}
$$

After estimating the parameters $\mu_x$, $\mu_z$, $\sigma_x$, and $\sigma_z$, the CEP $R$ can be calculated by numerical integrations. More details about calculating CEP are given in [16].

*2.2. Classical Bayesian Estimation.* As described in the Introduction, the performance evaluation of guidance weapon systems is a sequential process; the distribution parameters would change when fusing test data from different stage. The realistic tests are conducted to refine the previous evaluation results based on the substitute tests. Therefore, the data collected from substitute tests are usually regarded as prior samples.

Because the procedures of estimating the parameters $(\mu_x, \sigma_x)$ and $(\mu_z, \sigma_z)$ are similar, we take the downrange direction of impact point errors as an example to introduce the classical Bayesian estimation, where $X$ follows the normal distribution $N(\mu_x, \sigma_x^2)$. For convenience, we drop the subscripts of $(\mu_x, \sigma_x^2)$ as $(\mu, \sigma^2)$ and let $\pi_1(\mu, \sigma^2)$ be the joint prior distribution of $(\mu, \sigma^2)$. In Bayesian theory, the conjugate prior distributions of $\mu$ and $\sigma^2$ are normal distribution $N(\mu_1, \sigma^2/k_1)$ and inverse Gamma distribution

$IGa(\alpha_1, \beta_1)$, respectively. As for the distribution parameters $\mu_1$, $k_1$, $\alpha_1$, and $\beta_1$, they are determined by (7). The probability density function of $IGa(\alpha_1, \beta_1)$ is

$$f\left(\sigma^2\right) = \frac{\beta_1^{\alpha_1}}{\Gamma\left(\alpha_1\right)} \left(\frac{1}{\sigma^2}\right)^{\alpha_1+1} e^{-\beta_1/\sigma^2} \tag{4}$$

where $\Gamma(\cdot)$ means the Gamma function. Moreover, the joint prior distribution $\pi_1(\mu, \sigma^2)$ is the normal-inverse Gamma distribution, so we have

$$\begin{aligned}
\pi_1\left(\mu, \sigma^2\right) &= \pi_1\left(\mu \mid \sigma^2\right) \cdot \pi_1\left(\sigma^2\right) \\
&\propto N\left(\mu_1, \frac{\sigma^2}{k_1}\right) \cdot IGa\left(\alpha_1, \beta_1\right)
\end{aligned} \tag{5}$$

Suppose the prior samples for downrange direction are $\mathbf{X}^{(1)} = \{x_1^{(1)}, x_2^{(1)} \cdots, x_{n_1}^{(1)}\}$, $n_1$ is the size of the samples; let

$$\begin{aligned}
\overline{X}^{(1)} &= \frac{1}{n_1} \sum_{i=1}^{n_1} x_i^{(1)} \\
S_1^2 &= \frac{1}{n_1} \sum_{i=1}^{n_1} \left(x_i^{(1)} - \overline{X}^{(1)}\right)^2
\end{aligned} \tag{6}$$

Then the estimates of the parameters of the joint prior distribution are

$$\begin{aligned}
\alpha_1 &= \frac{1}{2} \sum_{i=1}^{n_1} \left(x_i^{(1)} - \overline{X}^{(1)}\right)^2 = \frac{n_1 S_1^2}{2} \\
\beta_1 &= \frac{1}{2}\left(n_1 - 1\right) \\
\mu_1 &= \overline{X}^{(1)} \\
k_1 &= n_1
\end{aligned} \tag{7}$$

Similarly, when the realistic test samples $\mathbf{X}^{(2)} = \{x_1^{(2)}, x_2^{(2)}, \cdots, x_{n_2}^{(2)}\}$ are obtained, $n_2$ is the size of the samples; let

$$\begin{aligned}
\overline{X}^{(2)} &= \frac{1}{n_2} \sum_{i=1}^{n_2} x_i^{(2)} \\
S_2^2 &= \frac{1}{n_2} \sum_{i=1}^{n_2} \left(x_i^{(2)} - \overline{X}^{(2)}\right)^2
\end{aligned} \tag{8}$$

Because of the property of conjugate prior distribution, the posterior distribution of $(\mu, \sigma^2)$ is also a normal-inverse Gamma distribution:

$$\pi_2\left(\mu, \sigma^2 \mid \overline{X}^{(2)}, S_2^2\right) \propto N\left(\mu_2, \frac{\sigma^2}{k_2}\right) \cdot IGa\left(\alpha_2, \beta_2\right) \tag{9}$$

where $\mu_2$, $k_2$, $\alpha_2$, and $\beta_2$ are the parameters of the normal-inverse Gamma distribution, and the estimates of the parameters are

$$\begin{aligned}
\alpha_2 &= \alpha_1 + \frac{n_2 S_2^2}{2} + \frac{n_1 n_2 \left(\overline{X}^{(2)} - \mu_1\right)^2}{2\left(n_1 + n_2\right)} \\
\beta_2 &= \beta_1 + \frac{n_2}{2} \\
\mu_2 &= \frac{n_1 \mu_1 + n_2 \overline{X}^{(2)}}{n_1 + n_2} \\
k_2 &= n_1 + n_2
\end{aligned} \tag{10}$$

So the estimates of $(\mu, \sigma^2)$ by classical Bayesian estimation are

$$\begin{aligned}
\widehat{\mu}_{\text{Bayes}} &= \mu_2 \\
\widehat{\sigma}_{\text{Bayes}}^2 &= \frac{\alpha_2}{\beta_2 - 1}
\end{aligned} \tag{11}$$

As for the cross-range direction of impact point errors, the estimates of $\mu_z$ and $\sigma_z$ can also be calculated by (11). After obtaining the estimates of $\mu_x$, $\mu_z$, $\sigma_x$, and $\sigma_z$, we can calculate the CEP of impact points by (3).

## 3. Bayesian Estimation Based on Representative Points

In this section, we propose the novel method Bayesian estimation based on representative points and the procedure of determining the optimal number of RPs. RPs can not only optimally represent the distribution of prior information in terms of MSE principle, but also have smaller sample size compared with raw prior samples. Therefore, RPs can retain the useful information of prior samples and reduce the estimate bias of $\mu_x$, $\mu_z$. If we search RPs as new prior information, we may get more accurate and reliable CEP of guidance weapon systems.

*3.1. Methods for Searching Representative Points.* In this subsection, we will give a brief introduction of RPs and methods for searching RPs. Assume that $\mathbf{X} \in R^p$ is a $p$ dimensional random vector, and the probability density function of $\mathbf{X}$ is $f(x)$. Define the mean squared error for a set of points $\{r_1, r_2, \cdots, r_k\} \subseteq R^p$ of the random vector $\mathbf{X}$ as follows:

$$\begin{aligned}
\text{MSE}\left(r_1, r_2, \cdots, r_k\right) &= \mathbb{E}\left(\min_{i=1,\cdots,k} \|x - r_i\|^2\right) \\
&= \int_{R^p} \min_{i=1,\cdots,k} \|x - r_i\|^2 f(x)\, dx
\end{aligned} \tag{12}$$

where $\|\cdot\|$ stands for $L_2$-norm. The vectors $\{r_1, \cdots r_k\} \in R^p$ are called $k$ representative points of a random vector $\mathbf{X}$ if

$$\text{MSE}\left(r_1, r_2, \cdots, r_k\right) \leq \text{MSE}\left(x_1, x_2, \cdots, x_k\right) \tag{13}$$

for all sets $\{x_1, x_2, \cdots, x_k\} \subseteq R^p$.

From the definition above, it is obvious that when $k = 1$, the single RP is the mean of $\mathbf{X}$. Searching the RPs equals doing the optimal grouping [17]; it is difficult to derive the concrete RPs theoretically even if the number of RPs is given. Flury has proved that there is no theoretical derivation of RPs when $k > 2$ [18]. So some approximation algorithms have been proposed to search RPs including k-means methods [19], parametric k-means methods [20], and nonparametric methods [21]. The k-means methods are searching the clustering centers as the RPs. Moreover, the parametric k-means methods resample large samples from a specific distribution whose parameters are estimated by maximum likelihood, then searching the RPs from the resampled samples by k-means methods. The main idea of nonparametric methods is to build the empirical distribution function of $\mathbf{X}$ and resample large samples from this empirical function, after which the RPs are chosen from the resampled samples. In most cases, nonparametric methods have better performance to represent a distribution in terms of MSE than the k-means methods and parametric k-means methods. The main steps of the nonparametric method introduced in [21] are shown as follows:

(i) **Step1.** For original samples $\mathbf{X}^{(0)} = \{x_1, x_2, \cdots, x_n\}$, use k-means algorithm to obtain $k$ points $\mathbf{P}^{(1)} = \{P_1^{(1)}, P_2^{(1)}, \cdots, P_k^{(1)}\}$ from $\mathbf{X}^{(0)}$ as an initial solution.

(ii) **Step2.** Use the kernel estimation method to estimate the density function of original samples, denoted as $\widehat{f}(x)$.

(iii) **Step3.** Use the Randomized Likelihood Sampling method to generate $N(N \gg n)$ samples $\mathbf{Z} = \{z_1, z_2, \cdots, z_N\}$ from the density function $\widehat{f}(x)$ as the training data.

(iv) **Step4.** Based on the $N$ training data $\mathbf{Z}$ and the $k$ starting points $\mathbf{P}^{(1)}$, use k-means algorithm to obtain $k$ RPs $\mathbf{P}^{(2)} = \{P_1^{(2)}, P_2^{(2)} \cdots, P_k^{(2)}\}$.

Following the four steps above, we can get the RPs from prior samples. After that, we can use the RPs as new prior information to estimate the population parameters. More details about BERP will be introduced in the next subsection.

*3.2. The Procedure of BERP.* In this subsection, we will introduce the procedure of BERP which is similar to the classical Bayesian estimation in Section 2.2. We also take the downrange direction of impact point errors as an example, suppose the original prior samples are $\mathbf{X}^{(1)} = \{x_1^{(1)}, x_2^{(1)}, \cdots, x_{n_1}^{(1)}\}$. Choose $n_r$ RPs from $\mathbf{X}^{(1)}$ by the nonparametric method introduced in Section 3.1 and denote $\mathbf{X}^{(rp)} = \{x_1^{(rp)}, x_2^{(rp)}, \cdots, x_{n_r}^{(rp)}\}$, and let

$$
\begin{aligned}
\overline{X}^{(rp)} &= \frac{1}{n_r} \sum_{i=1}^{n_r} x_i^{(rp)} \\
S_{rp}^2 &= \frac{1}{n_r} \sum_{i=1}^{n_r} \left( x_i^{(rp)} - \overline{X}^{(rp)} \right)^2
\end{aligned} \tag{14}
$$

Similar to Section 2.2, let $\alpha_{rp}$, $\beta_{rp}$, $\mu_{rp}$, and $k_{rp}$ be the parameters of the joint prior distribution. By searching RPs as new prior information, the estimates of these parameters are

$$
\begin{aligned}
\alpha_{rp} &= \frac{n_r S_{rp}^2}{2} \\
\beta_{rp} &= \frac{(n_r - 1)}{2} \\
\mu_{rp} &= \overline{X}^{(rp)} \\
k_{rp} &= n_r
\end{aligned} \tag{15}
$$

Moreover, let $\mu_{2r}$, $k_{2r}$, $\alpha_{2r}$, and $\beta_{2r}$ be the parameters of posterior distribution. So the estimates of these parameters are

$$
\begin{aligned}
\alpha_{2r} &= \alpha_{rp} + \frac{n_2 S_2^2}{2} + \frac{n_r n_2 \left( \overline{X}^{(2)} - \mu_{rp} \right)^2}{2 (n_r + n_2)} \\
\beta_{2r} &= \beta_{rp} + \frac{n_2}{2} \\
\mu_{2r} &= \frac{n_r \mu_{rp} + n_2 \overline{X}^{(2)}}{n_r + n_2} \\
k_{2r} &= n_r + n_2
\end{aligned} \tag{16}
$$

where $\overline{X}^{(2)}$ and $S_2^2$ are calculated by (8). So the estimates of $(\mu, \sigma^2)$ by BERP are

$$
\begin{aligned}
\widehat{\mu}_{\mathrm{BERP}} &= \mu_{2r} \\
\widehat{\sigma}_{\mathrm{BERP}}^2 &= \frac{\alpha_{2r}}{\beta_{2r} - 1}
\end{aligned} \tag{17}
$$

If we know the exact number of RPs, it is easy to estimate the parameters $\mu_x$, $\mu_z$, $\sigma_x$, and $\sigma_z$ by (17). However, it is difficult to determine the optimal number of RPs because there are no theoretical methods about this. Therefore, we propose an optimization procedure to determine the optimal number of RPs when there exists unknown systematic bias in prior samples.

*3.3. Optimal Number of Representative Points.* The approach to determining the optimal number of RPs varies with the background of practical problem. On the one hand, the RPs are closer to the original prior samples as the number grows. Therefore, the more the RPs chosen, the larger bias they may bring to the estimate of $\boldsymbol{\mu}$, which will also be validated in Theorem 1, Section 4. On the other hand, similar to the methods of normalizing prior information, choosing small number of RPs will cause great information loss. Therefore, we consider two factors when determining the optimal number of RPs: the estimate bias and the information loss. We still take the downrange direction of the impact point errors as an example to describe the optimization procedure. $B_{n_r}$ and $L_{n_r}$ stand for the estimate bias and the information

Define $N$ : Predetermined maximum number of RPs    $n_r$ : The current number of RPs
    $n_r^*$ : Optimal number of RPs    $B_{n_r}$ : Estimate bias    $L_{n_r}$ : Information loss
**for** $1 \leqslant n_r \leqslant N$ **do**
    Choose $n_r$ RPs from the prior samples and calculate $B_{n_r}$ and $L_{n_r}$
    Update $n_r = n_r + 1$
**end for**
    $n_r^* = \arg\min(B_{n_r} + L_{n_r})$
    **return** $n_r^*$

ALGORITHM 1: The optimization procedure.

loss, respectively, where $n_r$ stands for the number of RPs. The objective function $F_{n_r}$ is to achieve a balance between $B_{n_r}$ and $L_{n_r}$. So we have

$$F_{n_r} = B_{n_r} + L_{n_r} \tag{18}$$

In fact, it is hard to quantify the estimate bias $B_{n_r}$ without the true values of the population parameters. But there is a principle in the performance evaluation of guidance weapon systems that every realistic test sample should be used. So we can use the mean of realistic test samples to approximate the true value of parameter $\mu$, denoted as $\mu_{\text{true}}$. The estimate of $\mu$ by BERP is denoted as $\mu_{\text{BERP}}$. So the approximated $B_{n_r}$ is

$$B_{n_r} = \left|\mu_{\text{BERP}} - \mu_{\text{true}}\right| \tag{19}$$

As for $L_{n_r}$, the information loss could be estimated by Cox (1957) [17]:

$$L_{n_r} = \frac{E\left[X - \xi_{n_r}(X)\right]^2}{\sigma^2} = \sum_{i=1}^{n_r} \frac{\sigma_i^2}{\sigma^2} \tag{20}$$

where $\sigma^2$ is the variance of all prior samples; $\sigma_i^2$ is the variance of samples in $i^{th}$ class which is classified in Step 4 for searching RPs. As the number of RPs increases, the estimate bias would increase and the information loss would decrease. The optimal number of RPs is determined by the minimal value of the objective function

$$\min_{n_r} \left[\left|\mu_{\text{BERP}} - \mu_{\text{true}}\right| + \sum_{i=1}^{n_r} \frac{\sigma_i^2}{\sigma^2}\right] \tag{21}$$

Algorithm 1 describes the optimization procedure of determining the optimal number of RPs.

## 4. Theoretical Analysis of BERP

Estimate bias and MSE are the common measures to evaluate the quality of an estimator. So we will use them to analyse the theoretical performance of BERP and classical Bayesian estimation in this section. Suppose the posterior distribution of parameter $\theta$ is $f(\theta \mid \mathbf{X})$, $\overline{\theta}$ is the posterior expectation of $\theta$,

and $\theta_{\text{true}}$ is the true value of $\theta$. The estimate bias and MSE of the estimator $\widehat{\theta}$ are

$$\text{Bias}\left(\widehat{\theta}, \theta_{\text{true}} \mid X\right) = \overline{\theta} - \theta_{\text{true}} \tag{22}$$

$$\begin{aligned}
\text{MSE}\left(\widehat{\theta} \mid X\right) &= \text{E}^{\widehat{\theta}\mid X}\left(\widehat{\theta} - \theta_{\text{true}}\right)^2 \\
&= \text{E}^{\widehat{\theta}\mid X}\left(\widehat{\theta} - \overline{\theta} + \overline{\theta} - \theta_{\text{true}}\right)^2 \\
&= \text{Var}\left(\widehat{\theta} \mid X\right) + \left(\overline{\theta} - \theta_{\text{true}}\right)^2 \\
&= \text{Var}\left(\widehat{\theta} \mid X\right) \\
&\quad + \left(\text{Bias}\left(\widehat{\theta}, \theta_{\text{true}} \mid X\right)\right)^2
\end{aligned} \tag{23}$$

where $\text{Var}(\widehat{\theta} \mid X)$ is the variance of $\widehat{\theta}$.

Let realistic test samples follow the normal distribution $N(\mu, \sigma^2)$, and the prior samples follow the normal distribution $N(\mu + \eta, \sigma^2)$, where $\eta$ is the systematic bias. Let $n_1$, $n_2$ be the sample sizes of prior samples and realistic test samples, respectively. Suppose the posterior estimate of $\mu$ by classical Bayesian estimation is $\widehat{\mu}_{\text{Bayes}}$, $\overline{\mu}_{\text{Bayes}}$ is its posterior expectation. The posterior estimate of $\mu$ by BERP is $\widehat{\mu}_{\text{BERP}}$, $\overline{\mu}_{\text{BERP}}$ is its posterior expectation, and $n_r$ is the number of RPs. Let $\mu_{\text{true}}$ be the true value of parameter $\mu$. There are two theorems to compare the estimate bias and MSE of the two estimators $\widehat{\mu}_{\text{Bayes}}$ and $\widehat{\mu}_{\text{BERP}}$.

**Theorem 1.** *In the case that there exists systematic bias in prior samples, one has*

$$Bias\left(\widehat{\mu}_{BERP}, \mu_{true} \mid X\right) \leq Bias\left(\widehat{\mu}_{Bayes}, \mu_{true} \mid X\right) \tag{24}$$

*Proof.* From (11) and (17), the estimate bias of $\widehat{\mu}_{\text{BERP}}$ and $\widehat{\mu}_{\text{Bayes}}$ is

$$\text{Bias}\left(\widehat{\mu}_{\text{BERP}}, \mu_{\text{true}} \mid X\right) = \overline{\mu}_{\text{BERP}} - \mu_{\text{true}} = \frac{n_r \eta}{n_r + n_2} \tag{25}$$

$$\text{Bias}\left(\widehat{\mu}_{\text{Bayes}}, \mu_{\text{true}} \mid X\right) = \overline{\mu}_{\text{Bayes}} - \mu_{\text{true}} = \frac{n_1 \eta}{n_1 + n_2} \tag{26}$$

The estimate bias of $\mu_{BERP}$ increases with the number of RPs. Since the size of prior samples is much larger than the size of RPs, we have $n_1 \gg n_r$. Therefore there is

$$\frac{n_r}{n_r + n_2} < \frac{n_1}{n_1 + n_2} \tag{27}$$

So

$$\text{Bias}\left(\widehat{\mu}_{\text{BERP}}, \mu_{\text{true}} \mid X\right) \le \text{Bias}\left(\widehat{\mu}_{\text{Bayes}}, \mu_{\text{true}} \mid X\right) \qquad (28)$$

$\square$

**Theorem 2.** *In the case that $\eta^2 \geqslant \sigma^2/n_2$, one has*

$$MSE\left(\widehat{\mu}_{BERP} \mid X\right) \le MSE\left(\widehat{\mu}_{Bayes} \mid X\right) \qquad (29)$$

*Proof.* The MSE of $\widehat{\mu}_{\text{Bayes}}$ and $\widehat{\mu}_{\text{BERP}}$ are

$$\text{MSE}\left(\widehat{\mu}_{\text{Bayes}} \mid X\right) = \text{Var}\left(\widehat{\mu}_{\text{Bayes}} \mid X\right) \\ + \left(\text{Bias}\left(\widehat{\mu}_{\text{Bayes}}, \mu_{\text{true}} \mid X\right)\right)^2 \qquad (30)$$

$$\text{MSE}\left(\widehat{\mu}_{\text{BERP}} \mid X\right) = \text{Var}\left(\widehat{\mu}_{\text{BERP}} \mid X\right) \\ + \left(\text{Bias}\left(\widehat{\mu}_{\text{BERP}}, \mu_{\text{true}} \mid X\right)\right)^2 \qquad (31)$$

where

$$\text{Var}\left(\widehat{\mu}_{\text{Bayes}} \mid X\right) = \frac{\alpha_2}{k_2\left(\beta_2 - 1\right)}$$

$$= \frac{k_2 \cdot \left(n_1 \cdot S_1^2 + n_2 \cdot S_2^2\right) + n_1 \cdot n_2 \cdot \left(\overline{X}^{(2)} - \overline{X}^{(1)}\right)^2}{k_2^2 \cdot \left(k_2 - 3\right)}$$

$$\text{Var}\left(\widehat{\mu}_{\text{BERP}} \mid X\right) = \frac{\alpha_{2r}}{k_{2r}\left(\beta_{2r} - 1\right)}$$

$$= \frac{k_{2r} \cdot \left(n_r \cdot S_{rp}^2 + n_2 \cdot S_2^2\right) + n_r \cdot n_2 \cdot \left(\overline{X}^{(2)} - \overline{X}^{(rp)}\right)^2}{k_{2r}^2 \cdot \left(k_{2r} - 3\right)} \qquad (32)$$

Because prior samples and realistic test samples both follow the normal distribution $N(\mu, \sigma^2)$ and there exists systematic bias $\eta$ in prior samples, we have the approximated results:

$$S_1^2 \approx S_2^2 \approx S_{rp}^2 \approx \sigma^2 \\ \left(\overline{X}^{(2)} - \overline{X}^{(1)}\right)^2 \approx \left(\overline{X}^{(2)} - \overline{X}^{(rp)}\right)^2 \approx \eta^2 \qquad (33)$$

Let $\lambda_1 = n_1 + n_2$ and $\lambda_r = n_r + n_2$, and $\Delta_1$ is determined as follows:

$$\Delta_1 = \frac{\lambda_1^2 \lambda_r^2}{\left(\lambda_1 - 3\right)\left(\lambda_r - 3\right)\left(n_1\lambda_r + n_r\lambda_1\right) - n_1 n_r \left(\lambda_1 + \lambda_r + n_2 - 3\right) + n_2^3 - 3n_2^2} \qquad (34)$$

when $\text{MSE}(\widehat{\mu}_{\text{BERP}} \mid X) \le \text{MSE}(\widehat{\mu}_{\text{Bayes}} \mid X)$, if and only if $\eta^2 \geqslant \sigma^2/n_2 \cdot \Delta_1$. Because $n_1 \gg n_2, n_1 \gg n_r$, we can get $\Delta_1 < 1$. To sum up, when $\eta^2 \ge \sigma^2/n_2$, there is

$$\text{MSE}\left(\widehat{\mu}_{\text{BERP}} \mid X\right) \le \text{MSE}\left(\widehat{\mu}_{\text{Bayes}} \mid X\right) \qquad (35)$$

$\square$

The estimate bias of $\widehat{\mu}_{\text{BERP}}$ is smaller than that of $\widehat{\mu}_{\text{Bayes}}$ when there exists unknown systematic bias in prior samples. Moreover, the MSE of $\widehat{\mu}_{\text{BERP}}$ is also smaller than that of $\widehat{\mu}_{\text{Bayes}}$ when $\eta^2 \ge \sigma^2/n_2$. In most cases, $\sigma^2/n_2$ is smaller than $\eta^2$. Therefore, it can be concluded that BERP has better accuracy for estimating the parameter $\mu$ of normal distribution than classical Bayesian estimation when there exists unknown systematic bias in prior samples.

However, when the systematic bias is beyond a certain bound, it may be better to stop fusing the biased prior samples even if they may provide some useful information. Suppose $\widehat{\mu}_{MLE}$ is estimated by maximum likelihood estimation without using the prior information, it is obviously an unbiased estimate. The MSE of $\widehat{\mu}_{MLE}$ is

$$\text{MSE}\left(\widehat{\mu}_{\text{MLE}} \mid X\right) = \text{Var}\left(\widehat{\mu}_{\text{MLE}} \mid X\right) = \frac{\sigma^2}{n_2} \qquad (36)$$

when $\text{MSE}(\widehat{\mu}_{\text{MLE}} \mid X) \le \text{MSE}(\widehat{\mu}_{\text{BERP}} \mid X)$, if and only if $\eta^2 \geqslant \sigma^2/n_2 \cdot \Delta_2$, where

$$\Delta_2 = \frac{\left(n_r - 3\right)\left(n_r + n_2\right)^2}{n_r n_2 + n_r^2 \left(n_2 + n_r - 3\right)} \qquad (37)$$

Moreover, compared with (34), there is $\Delta_1 < \Delta_2$, which means that when $\eta^2 \geqslant \sigma^2/n_2 \cdot \Delta_2$, $\text{MSE}(\widehat{\mu}_{\text{MLE}} \mid X)$ is also smaller than $\text{MSE}(\widehat{\mu}_{\text{Bayes}} \mid X)$. Therefore, if $\eta^2$ is larger than $\sigma^2/n_2 \cdot \Delta_2$, we should stop using the prior information; $\sqrt{\sigma^2/n_2 \cdot \Delta_2}$ is the bound of the systematic bias.

## 5. Numerical Experiments

In this section, three numerical experiments are provided to validate that BERP can help to get more accurate and reliable calculation of CEP when the systematic bias is within the bound. The first one is to show the performance of optimization procedure when determining the optimal number of RPs; the second one is to compare the estimation accuracy of BERP with that of classical Bayesian estimation; the third one is to compare the calculation of CEP based on BERP with classical Bayesian estimation.

*Example 1.* In this example, we will analyze the optimization procedure for determining the optimal number of RPs. Let $n_1 = 200$ be the sample size of the prior samples and $n_2 = 5$ be the sample size of realistic test samples. The prior samples and the realistic test samples are generated from normal distributions $N(\mu + \eta, \sigma^2)$ and $N(\mu, \sigma^2)$, respectively, where $\mu = 0, \eta^2 = 10, \sigma^2 = 50$. The RPs are searched by
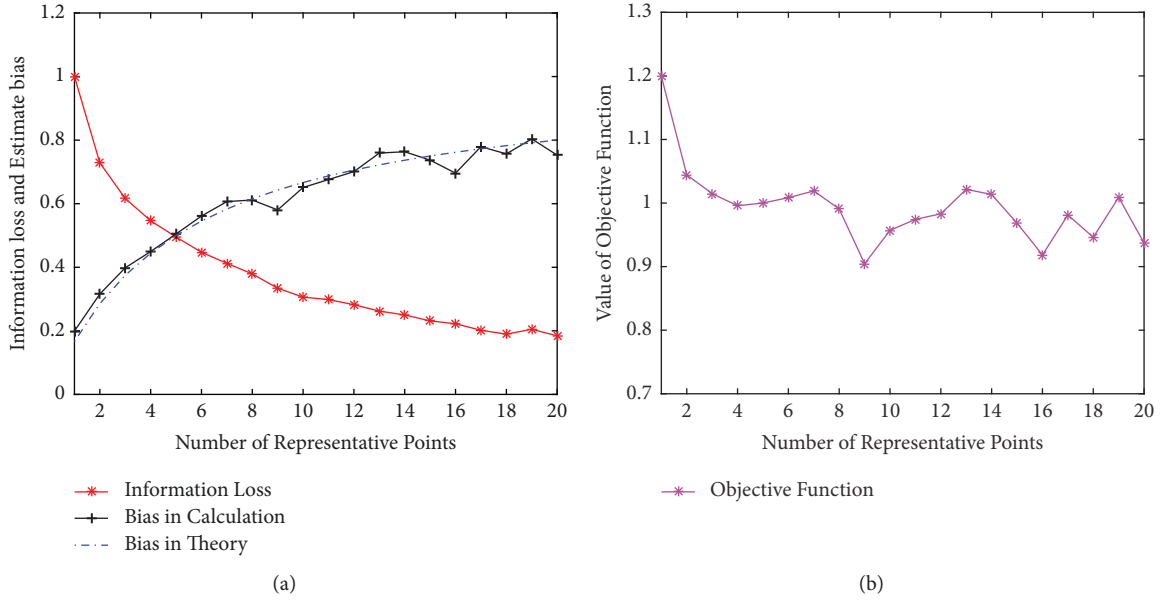
FIGURE 1: Performance for different representative points.

the nonparametric method. In order to reduce the influence of random factors on simulation result, resample samples 100 times under the same circumstance and get the average results. Figure 1 shows the variation of the information loss, the estimate bias, and the objective function for different numbers of RPs, where the estimate bias and information loss are both calculated after normalization.

As displayed in Figure 1, the estimate bias approximated by (19) is very close to the theoretical one, which means that using $B_{n_r}$ to quantify the estimate bias is reasonable. In addition, when the number of RPs increases, the information loss decreases while the estimate bias increases, so the objective function can balance the two factors well when determining the optimal number of RPs. From Figure 1(b), when the number of RPs is 9, the objective function reaches the minimal value, so the optimal number of RPs is 9 in this example.

*Example 2.* In this example, we will compare the estimation performance of the two methods. Let $n_1 = 200$ be the sample size of the prior samples, $n_2 = 5$ be the sample size of the realistic test samples, and $N = 20$ be the predetermined maximum number of RPs. The prior samples and realistic test samples are generated from normal distributions $N(\mu + \eta, \sigma^2)$ and $N(\mu, \sigma^2)$, respectively, where $\mu = 0$ and $\eta$ is the systematic bias in prior samples. We set different values for $\eta$ in the simulations to investigate the estimation performance when estimating the parameter $\mu$. Similar to Example 1, we resample samples 100 times under the same circumstance and get the average results in order to reduce the influence of random factors. Table 1 shows the simulation results.

From Table 1, we can summarize two conclusions. (1) Within the bound of the systematic bias shown in Section 4, if there exists slight systematic bias in prior samples, $\widehat{\mu}_{\text{BERP}}$ is much more closer to the true value of the parameter $\mu$ than $\widehat{\mu}_{\text{Bayes}}$, and MSE($\widehat{\mu}_{\text{BERP}}$) is also smaller than MSE($\widehat{\mu}_{\text{Bayes}}$) in

most cases. Therefore, BERP has higher accuracy to estimate the parameter $\mu$ than classical Bayesian estimation when there exists slight systematic bias in prior information. (2) If there is no systematic bias in prior samples, $\widehat{\mu}_{\text{BERP}}$ is very close to $\widehat{\mu}_{\text{Bayes}}$, and MSE($\widehat{\mu}_{\text{BERP}}$) is a little larger than MSE($\widehat{\mu}_{\text{Bayes}}$). There is no obvious difference between the estimation accuracies of the two methods. Moreover, without the systematic bias in prior samples, the optimal number of RPs is close to the predetermined maximum number 20. In this case, BERP is degenerated into the classical Bayesian estimation to some extent.

*Example 3.* When there is systematic bias in prior information, the CEP calculation based on BERP and classical Bayesian estimation is discussed in this example. We simulate $n_1 = 200$ prior samples $(\mathbf{X}^{(1)}, \mathbf{Z}^{(1)})^T$ from the bivariate normal distribution $N(\boldsymbol{\mu} + \boldsymbol{\eta}, \boldsymbol{\Sigma})$ and $n_2 = 5$ realistic test samples $(\mathbf{X}^{(2)}, \mathbf{Z}^{(2)})^T$ from the bivariate normal distribution $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\eta} = (\sqrt{10}, \sqrt{10})^T, \boldsymbol{\mu} = (0, 0)^T$, and $\boldsymbol{\Sigma} = \text{diag}(40, 40)$. The parameters $\mu_x$, $\mu_z$, $\sigma_x$, and $\sigma_z$ are estimated by BERP and classical Bayesian estimation, respectively. Based on the estimates of the population parameters, we use the numerical integration to get the CEP. Figure 2 shows the simulation results about CEP calculation.

As shown in Figure 2, the true CEP is $\text{CEP}_{\text{true}} = 7.4466$, the CEP calculation based on BERP is $\text{CEP}_{\text{BERP}} = 7.6532$, and the CEP calculation based on classical Bayesian estimation is $\text{CEP}_{\text{Bayes}} = 8.4402$. It is easy to conclude that the CEP calculation based on BERP is much closer to the true CEP when the systematic bias is within the bound. In addition, the CEP calculation without using the prior information is defined as $\text{CEP}_{\text{noprior}}$. If the prior information were effectively ignored, the $\text{CEP}_{\text{noprior}}$ would be unreliable and unstable because of the large MSE of the estimation of $\mu$ without

TABLE 1: Comparison of the simulation result of the two methods.

| $\sigma^2$ | $\eta^2$ | $\widehat{\mu}_{\text{Bayes}}$ | $\widehat{\mu}_{\text{BERP}}$ | $\text{MSE}(\widehat{\mu}_{\text{Bayes}})$ | $\text{MSE}(\widehat{\mu}_{\text{BERP}})$ | Opt $N_{Rps}$ |
|---|---|---|---|---|---|---|
| 40 | 0 | **0.0071** | 0.0419 | **0.1985** | 1.6958 | 19 |
| 50 | 0 | **-0.0185** | -0.0720 | **0.2482** | 2.1204 | 19 |
| 60 | 0 | **-0.0178** | -0.0796 | **0.2979** | 2.5442 | 19 |
| 40 | 5 | 2.1657 | **0.1930** | 4.9576 | **4.8515** | 14 |
| 50 | 5 | 2.1885 | **0.3104** | **5.0081** | 5.4469 | 13 |
| 60 | 5 | 2.1661 | **0.3347** | **5.0574** | 6.0391 | 13 |
| 40 | 10 | 3.0963 | **0.8185** | 9.7162 | **7.032** | 9 |
| 50 | 10 | 3.0684 | **1.0232** | 9.7688 | **7.7469** | 9 |
| 60 | 10 | 3.0920 | **1.0394** | 9.8186 | **8.4485** | 8 |
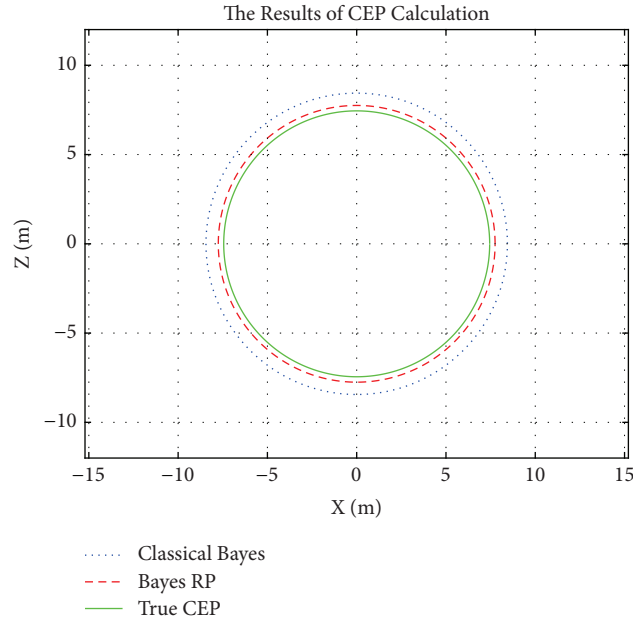


FIGURE 2: CEP calculated by different methods.

using prior information. Therefore, using BERP to estimate the parameters $\mu_x$, $\mu_z$ will get more accurate calculation of CEP than classical Bayesian estimation when there is slight systematic bias in prior information. In addition, CEP calculation by BERP also outperforms the method with the prior information ignored.

## 6. Conclusions

In this paper, we have investigated the methods for performance evaluation of guidance weapon systems. Because of the small sample size of the realistic test data, we would fuse the prior information in the evaluation. However, by classical Bayesian estimation, the unknown systematic bias in prior information may cause large deviation for CEP calculation. For purpose of addressing it, a novel Bayesian method called BERP is proposed in this paper, and the corresponding optimization procedure is designed. In addition, we also give the bound of systematic bias for stopping using the heavily biased prior information.

Within the bound of the systematic bias, theoretical analysis and simulation results prove that our new method has smaller estimate bias and MSE for estimating the mean of normal distribution than classical Bayesian estimation when there exists slight systematic bias in prior information. As for CEP calculation, the simulation results also validate that the CEP calculated by BERP is more accurate and reliable than the CEP calculated by classical Bayesian estimation. It can be concluded that a more accurate and reliable estimation of the CEP can be obtained via the BERP when the unknown systematic bias is within the bound.

There is no obvious difference of the estimation accuracy between the two methods; BERP also has a good estimation performance when there is no systematic bias. Therefore, in order to get accurate and reliable evaluation results of guidance weapon systems, it is better to calculate the CEP via BERP than classical Bayesian estimation if the systematic bias is within the bound. In contrast, if the systematic bias is beyond the bound, we should stop fusing the biased prior information and evaluate the performance

only by realistic test samples even if the sample size is small.

## Data Availability

The simulation data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

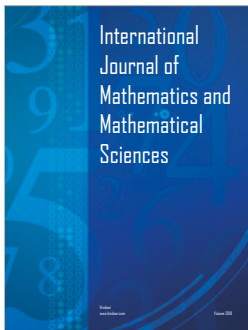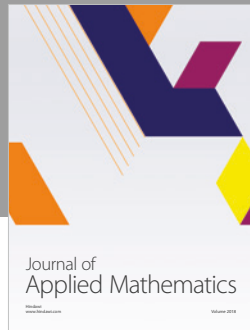The authors declare that they have no conflicts of interest.

## Authors' Contributions

All authors have contributed to the study and preparation of the article. Bowen Liu and Xiaojun Duan conceived the idea, derived equations, and did analysis. Bowen Liu and Liang Yan finished the programming work and wrote the paper. All authors have read and approved the final manuscript.

## Acknowledgments

## References

[1] J. C. Spall and J. L. Maryak, "A feasible Bayesian estimator of quantiles for projectile accuracy from non-iid data," *Journal of the American Statistical Association*, vol. 87, no. 419, pp. 676–681, 1992.

[2] J. T. Gillis, "Computation Of The Circular Error Probability Integral," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 27, no. 6, pp. 906–910, 1991.

[3] E. Williams C, "A comparison of circular error probable estimators for small samples," *DTIC Document*, 1997.

[4] Shengdi Zhang, Xiaojun Duan, Chang Li, Xiaojun Peng, and Qiang Zhang, "CEP Calculation Based on Weighted Bayesian Mixture Model," *Mathematical Problems in Engineering*, vol. 2017, pp. 1–7, 2017.

[5] J. Zhang and W. An, "Assessing circular error probable when the errors are elliptical normal," *Journal of Statistical Computation & Simulation*, vol. 82, no. 4, pp. 565–586, 2012.

[6] I. Bloch, "Information combination operators for data fusion: a comparative review with classification," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 26, no. 1, pp. 52–67, 1996.

[7] L. Ming and L. Fang, "Fusion Evaluation of Ammunition Based on Bayes Method," in *Recent Developments in Mechatronics and Intelligent Robotics*, vol. 691 of *Advances in Intelligent Systems and Computing*, pp. 115–119, Springer International Publishing, Cham, 2018.

[8] X.-J. Duan and G. Wang, "Weighted Bayesian fusion evaluation basing on composite equivalency model," *Guofang Keji Daxue Xuebao/Journal of National University of Defense Technology*, vol. 30, no. 3, pp. 90–94, 2008.

[9] X.-J. Duan and H.-Y. Huang, "Weighted Bayesian fusion evaluation based on information divergence of prior sample," *Binggong Xuebao/Acta Armamentarii*, vol. 28, no. 10, pp. 1276–1280, 2007.

[10] H. Huang, X. Duan, and Z. Wang, "Novel posterior-weighted Bayesian estimation method considering the credibility of the prior information," *Hangkong Xuebao/Acta Aeronautica et Astronautica Sinica*, vol. 29, no. 5, pp. 1245–1251, 2008.

[11] X. J. Duan and Y. Du X, "The Deduction of One Type of Bayes Formula and Its Application in Small Sample Evaluation," *Journal of Hunan University of Technology*, vol. 24, pp. 1–3, 2010.

[12] B. A. Flury, "Principal points," *Biometrika*, vol. 77, no. 1, pp. 33–41, 1990.

[13] S. Lühr and M. Lazarescu, "Incremental clustering of dynamic data streams using connectivity based representative points," *Data & Knowledge Engineering*, vol. 68, no. 1, pp. 1–27, 2009.

[14] K. Fang, M. Zhou, and W. Wang, "Applications of the representative points in statistical simulations," *Science China Mathematics*, vol. 57, no. 12, pp. 2609–2620, 2014.

[15] A. D. Sappa and M. Devy, "Fast range image segmentation by an edge detection strategy," in *Proceedings of the 3rd International Conference on 3-D Digital Imaging and Modeling, 3DIM 2001*, pp. 292–299, Canada, June 2001.

[16] Y. Wang, G. Yang, D. Yan, Y. Wang, and X. Song, "Comprehensive assessment algorithm for calculating CEP of positioning accuracy," *Measurement*, vol. 47, no. 1, pp. 255–263, 2014.

[17] D. R. Cox, "Note on Grouping," *Journal of the American Statistical Association*, vol. 52, no. 280, pp. 543–547, 1957.

[18] B. D. Flury, "Estimation of principal points," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, vol. 42, no. 1, pp. 139–151, 1993.

[19] J. A. Hartigan, *Clustering Algorithms*, John Wiley & Sons, 1979.

[20] T. Tarpey, "A parametric $k$-means algorithm," *Computational Statistics*, vol. 22, no. 1, pp. 71–89, 2007.

[21] Z. Qi, Y. Zhou, and K. Fang, "Representative points for location-biased datasets," *Communications in Statistics - Simulation and Computation*, pp. 1–14, 2017.