Received: 13 February 2020 | Revised: 26 June 2020 | Accepted: 8 September 2020

DOI: 10.1002/qre.2771

SPECIAL ISSUE ARTICLE

WILEY

Circular prediction regions for miss distance models under heteroskedasticity

Thomas H. Johnson[1] | **John T. Haman**[1] | **Heather Wojton**[1] | **Laura Freeman**[2]

[1] Institute for Defense Analyses, Alexandria, Virginia

[2] Virginia Tech Hume Center, Arlington, Virginia

Correspondence
Thomas H. Johnson, Institute for Defense Analyses, Alexandria, VA.
Email: thjohnso@ida.org

Abstract

Circular prediction regions are used in ballistic testing to express the uncertainty in shot accuracy. We compare two modeling approaches for estimating circular prediction regions for the miss distance of a ballistic projectile. The miss distance response variable is bivariate normal and has a mean and variance that can change with one or more experimental factors. The first approach fits a heteroskedastic linear model using restricted maximum likelihood, and uses the Kenward-Roger statistic to estimate circular prediction regions. The second approach fits an analogous Bayesian model with unrestricted likelihood modifications, and computes circular prediction regions by sampling from the posterior predictive distribution. The two approaches are applied to an example problem, and are compared using simulation.

KEYWORDS

circular error probable, department of defense, design of experiments, statistical modeling

1 | INTRODUCTION

In defense testing, miss distance measurements play an important role in weapon system evaluations. Weapon accuracy, whether it be for a hand gun, mortar, or artillery system, is one of the largest contributors to the weapon's effectiveness and lethality. A prudent weapon system evaluation leverages a designed experiment to provide an efficient characterization of accuracy as a function of anticipated combat conditions.

Miss distance is typically a univariate or bivariate response variable. A bivariate evaluation measures the x- and y-component distances between the projectile aim point and the impact point. A univariate evaluation collapses these measurements to a radial distance.

Bivariate evaluations contain more information and have practical advantages. Consider a notional artillery test that reveals that a projectile overshoots its target when it is exposed to certain combat conditions. Collapsing to a radial miss distance retains the magnitude of this miss, but not the direction. The directional information can assist in the diagnosis of the issue, improve target damage prediction, and help minimize collateral damage.

Historically, weapon accuracy evaluations estimate circular error probable (CEP) that assumes that the x- and y-component miss distances follow a bivariate normal distribution, the x- and y-component means are independent, and the x- and y-component variances are identical. That is,

$$\begin{pmatrix} x \\ y \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix} \right).$$

CEP is then estimated as the radius of the circular contour line representing the pth quantile of the joint distribution.[1–3]

Much of the original work on weapon accuracy evaluations did not use a designed experiment. For instance, Moranda,[3] Elder,[1] and Williams[4] collectively review more than 10 different CEP estimators. In each case, the estimator assumes that the data are collected under a fixed condition, resulting in a single CEP estimate.

Recently, model-based procedures have been used to account for a change in CEP with experimental factors. Zimmer and Casey[5] use a model-based approach to estimate CEPs and circular tolerance regions for an Army ballistic system test that includes firing charge, fuze mode, firing angle, and range to target as factors in the experimental design. Their approach accommodates heteroskedasticity in a normal linear model to capture a change in the CEP estimate.

Circular tolerance regions replace CEP estimates in some weapon accuracy studies to convey uncertainty while simplifying the presentation of results. A p content, $1-\alpha$ coverage, circular tolerance region contains a proportion p or more of the future miss distances, with confidence level $1-\alpha$ (eg, Ref. [6, p. 294]). Hall and Sheldon,[7] Didonato,[8] and Zhang and An[9] used tolerance regions to assess weapon accuracy.

Circular prediction regions serve a similar purpose as circular tolerance regions. A p content circular prediction region contains a proportion p of the future miss distances, on average (eg, Ref. [6, p. 295]). This is the two-dimensional version of a prediction interval. Computations of circular prediction regions are readily accessible in the literature for a variety of different models.

In this paper, we explore two approaches for constructing circular prediction regions using normal linear models that accommodate heteroskedasticity. The first is a frequentist approach that fits a heterokcedastic linear model using a restricted maximum likelihood (REML) algorithm proposed by Smyth.[10] The model permits the mean and variance to change with the experimental factors, and employs the Kenward-Roger statistic[11] to compute circular prediction regions that are centered at the mean.

The contributions of our first approach are as follows. Smyth's model and fitting procedure are intended for a univariate response variable. We extend its use to accommodate circular prediction region estimation for the bivariate miss distance response variable using a unique formulation of the model. Additionally, we have not seen the Kenward-Roger statistic used with Smyth's REML algorithm, or for estimating circular prediction regions.

The second approach fits an analogous Bayesian model, but with an unrestricted likelihood, and computes circular prediction regions by sampling from the posterior predictive distribution of the bivariate response, given the posterior distribution of all model parameters. We assume noninformative, flat priors for the mean and variance coefficients. We have not seen this approach used in a weapon accuracy study.

In the next section, we begin with a motivating example. Sections 3 and 4 detail the frequentist and Bayesian approaches, respectively. Section 5 applies each approach to the motivating example. Section 6 conducts a simulation study to investigate the empirical coverage rates of the two approaches, and Section 7 concludes with commentary about the advantages and disadvantages of each approach.

2 | MOTIVATING EXAMPLE

A new weapon is being developed and the testers would like to compare its accuracy to the old weapon. Each weapon is assessed across varying ammunition types and ranges from the target. The ammunition type is a categorical factor that has two levels: A or B, while range is continuous but is fixed at two levels: near (1000 ft) or far (2000 ft).

The experiment is a balanced 2^3 full factorial with five replicates per unique design point. That is, the experiment has $2^3 = 8$ unique design points, each with five replicates, resulting in 40 shots (observations) that are executed according to a completely randomized run order. Each shot results in a pair of miss distances: a horizontal (x) and a vertical (y) component. Table 1 presents the miss distance data for this example.

As a first attempt at an analysis, suppose we assume that the miss distance exhibits constant variance and fit the bivariate linear model

$$\begin{pmatrix} \underset{n\times 1}{\boldsymbol{x}} \\ \underset{n\times 1}{\boldsymbol{y}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{M} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{M} \end{pmatrix} \begin{pmatrix} \rho \\ \zeta \end{pmatrix} + \epsilon, \tag{1}$$

where

- $\underset{n\times 1}{\boldsymbol{x}}$ and $\underset{n\times 1}{\boldsymbol{y}}$ are the component miss distances,
- n is the number of observations (40 in this example),

TABLE 1 Example miss distance pairs, grouped by the three factors: distance, weapon type, and ammunition type. Each row represents a unique design point

Range	Weapon	Ammo type	Miss distance (x,y) in feet
Far	New	A	(16,−17), (23,−15), (15,2), (4,5), (9,2)
Near	New	A	(2,0), (4,−2), (−2,−4), (5,5), (2,3)
Far	Old	A	(−16,2), (−1,−2), (−8,−26), (−14,−11), (−4,2)
Near	Old	A	(0,0), (−2,2), (−2,−1), (1,−2), (−1,1)
Far	New	B	(19,6), (2,6), (4,0), (3,−7), (10,−10)
Near	New	B	(2,3), (0,7), (4,2), (0,3), (1,3)
Far	Old	B	(8,−6), (−5,0), (3,−5), (−8,−4), (4,−18)
Near	Old	B	(−1,8), (−2,4), (2,3), (−1,3), (2,3)

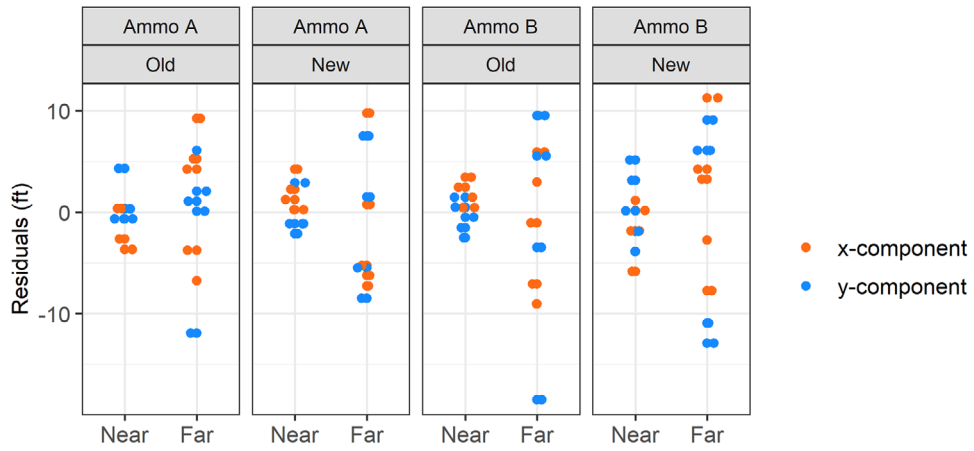

FIGURE 1 Residuals from linear model fit

- $\underset{n\times q}{\boldsymbol{M}}$ is the model matrix for the mean for each component,
- q is the number of coefficients for the mean for each component,
- $\underset{q\times 1}{\boldsymbol{\rho}}$ and $\underset{q\times 1}{\boldsymbol{\zeta}}$ are the coefficients for the mean for each component, and
- $\underset{2n\times 1}{\boldsymbol{\epsilon}} \sim N(\mathbf{0}, \sigma^2\boldsymbol{I})$ are the residuals.

A variety of diagnostic plots are available for checking the constant variance assumption in this model. One of these, which appears in Figure 1, shows the model fit residuals ($\boldsymbol{\epsilon}$) by factor-level setting. This plot reveals that the residual variability for the range-to-target "Far" setting is larger than "Near," indicating a violation of the constant variance assumption.

A simpler argument against this naive modeling approach is that the constant variance assumption is in contradiction with our desire to model a change in circular prediction regions. Constant variance will cause the radii of the circular prediction regions that are estimated at the design points to be identical. A different approach is needed to capture a change in miss distance variability.

3 | FREQUENTIST APPROACH

Our first approach is a frequentist method for constructing model-based circular prediction regions. This section focuses on theory, and is divided into three subsections: model formulation, model reduction, and circular prediction regions.

3.1 | Model formulation

The bivariate heteroskedastic linear model is

$$\boldsymbol{Y} \sim \mathcal{N}(\boldsymbol{A}\boldsymbol{\beta}, \boldsymbol{\Sigma}), \boldsymbol{\Sigma} = \text{diag}(\exp(\boldsymbol{\Omega}\boldsymbol{\gamma})), \tag{2}$$

where

- $\underset{2n\times 1}{\boldsymbol{Y}} = \begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{pmatrix}$ is the vector of miss distances,
- $\underset{2n\times 2q}{\boldsymbol{A}} = \begin{pmatrix} \boldsymbol{M} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{M} \end{pmatrix}$ is the model matrix for the mean,
- $\underset{2q\times 1}{\boldsymbol{\beta}} = \begin{pmatrix} \boldsymbol{\rho} \\ \boldsymbol{\zeta} \end{pmatrix}$ is the coefficient vector for the mean,
- $\underset{2n\times r}{\boldsymbol{\Omega}} = \begin{pmatrix} \boldsymbol{Z} \\ \boldsymbol{Z} \end{pmatrix}$ is the model matrix for the variance,
- $\underset{n\times r}{\boldsymbol{Z}}$ is the model matrix for the variance for each direction,
- $\underset{r\times 1}{\boldsymbol{\gamma}}$ is the coefficient vector for the variance, and
- r is the number of coefficients for the variance,

where the diag(·) function returns a square matrix with its argument along the diagonal and zeros elsewhere, and the exp(·) function operates element-wise on the entries of the $2n \times 1$ vector $\boldsymbol{\Omega}\boldsymbol{\gamma}$.

This model is sometimes referred to as a loglinear variance model (eg, JMP[12]), a double generalize linear model (eg, Montgomery[13]), or a heteroskedastic linear model (eg, Verbyla[14]). But in those references, the model focuses on a single response variable.

Here, we use the following model formulation to accommodate a bivariate response. First, we concatenate the response vectors, $\boldsymbol{x}$ and $\boldsymbol{y}$, into a single vector, $\boldsymbol{Y}$. Second, we create a block diagonal matrix, $\boldsymbol{A}$, for the model matrix for the mean, which allows the x- and y-component mean coefficient vectors, $\boldsymbol{\rho}$ and $\boldsymbol{\zeta}$, to be independently estimated. Third, we stack the model matrices for the variance to create a single matrix, $\boldsymbol{\Omega}$, which constrains the estimated x- and y-component variance coefficient vectors to be identical and denoted as $\boldsymbol{\gamma}$. The bivariate formulation of this model constrains individual prediction regions to be circular and centered at the mean.

The model is fit using an REML estimation technique proposed by Verbyla[14] and later generalized by Smyth.[10] The technique uses the method of scoring to determine the estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, which are denoted as $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\gamma}}$. We use the function `remlscore` of the R[15] package `statmod`[10] to fit the model.

3.2 | Frequentist model reduction

Model reduction involves a series of hypothesis tests to determine which x- and y-component coefficient pairs are included in the model. Let ρ_j and ζ_j, respectively, denote the jth component of $\boldsymbol{\rho}$ and $\boldsymbol{\zeta}$, where $j = 2, 3, 4, \ldots, q$. The null and alternative hypotheses for the jth test are

$$H_0 : \rho_j = 0 \text{ and } \zeta_j = 0, \tag{3}$$

$$H_a : \rho_j \neq 0 \text{ or } \zeta_j \neq 0, \tag{4}$$

and the Kenward-Roger[11] test statistic is

$$\hat{F} = \frac{1}{l}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)' \boldsymbol{L}\left(\boldsymbol{L}'\hat{\Phi}\boldsymbol{L}\right)^{-1}\boldsymbol{L}'\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right), \quad \text{where } \lambda\hat{F} \sim \mathcal{F}_{l,m}. \tag{5}$$

Here, $\boldsymbol{L}$ (size $l \times 2q$) specifies the linear combination of coefficients, $\hat{\boldsymbol{\Phi}}$ (size $2q \times 2q$) is the variance-covariance matrix of $\hat{\boldsymbol{\beta}}$ that is returned from the `remlscore` fitting function, and l is the number of simultaneous combinations of

coefficients that are involved in the hypothesis test, which is equal to 2. $\mathcal{F}_{l,m}$ denotes the F distribution with l numerator and m denominator degrees of freedom, and λ is a scaling constant. λ and m are calculated using Equations (A2) and (A3).

The test statistic and P-value for the jth hypothesis test are computed as follows. Let $\boldsymbol{B}_j$ denote a vector of size $1 \times q$ with its jth component equal to 1 and all other components equal to 0. To compute the test statistic for the jth hypothesis test, in Equation (5), substitute in the estimates for $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\Phi}}$, and set $l = 2$, $\boldsymbol{\beta} = \mathbf{0}$, and

$$\underset{2\times 2q}{\boldsymbol{L}} = \begin{pmatrix} \boldsymbol{B}_j & \mathbf{0} \\ \mathbf{0} & \boldsymbol{B}_j \end{pmatrix}. \tag{6}$$

The P-value for the jth test is equal to $1 - \mathcal{F}_{\hat{F}/\lambda,l,m}$, where $\hat{F}$ was computed in Equation (5). The null hypothesis is rejected and the coefficient pair is significant if the P-value is less than the prespecified significance level, α.

We use a backward selection technique to reduce the model. The approach starts with a full model. In a sequential process, it removes an insignificant coefficient pair from the model and recalculates the P-values for the remaining pairs. This process continues until only significant coefficient pairs remain.

3.3 | Circular prediction region

A p content circular prediction region estimated at the design point corresponding to the ith row of $\boldsymbol{M}$, denoted as $\boldsymbol{M}_i$, is computed as follows. Let

$$\underset{2\times 2q}{\boldsymbol{L}} = \begin{pmatrix} \boldsymbol{M}_i & \mathbf{0} \\ \mathbf{0} & \boldsymbol{M}_i \end{pmatrix}, \tag{7}$$

and define $\boldsymbol{G}$, which serves a similar purpose as $\boldsymbol{L}$ but for the variance coefficients, as

$$\underset{2\times r}{\boldsymbol{G}} = \begin{pmatrix} \boldsymbol{Z}_i \\ \boldsymbol{Z}_i \end{pmatrix}, \tag{8}$$

where $\boldsymbol{Z}_i$ denotes the ith row of $\boldsymbol{Z}$. Then, the equation for the radius of the circular prediction region that is centered at $\boldsymbol{L}\hat{\boldsymbol{\beta}}$ is

$$\sqrt{l\lambda\left(\operatorname{diag}(\exp(\boldsymbol{G}\hat{\boldsymbol{\gamma}})) + \boldsymbol{L}'\hat{\boldsymbol{\Phi}}\boldsymbol{L}\right)\mathcal{F}^{-1}_{p,l,m}}, \tag{9}$$

where $\mathcal{F}^{-1}_{p,l,m}$ returns the $p \times 100$th quantile of the cumulative F distribution with l numerator and m denominator degrees of freedom. Note that Equation (9) returns a 2×2 diagonal matrix. The diagonal terms are identical and are equal to the radius of the circular prediction region. Also, in this equation, $l = 2$, $\hat{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\Phi}}$ are obtained from the model fit, and λ and m are calculated using Equations (A2) and (A3).

4 | BAYESIAN APPROACH

4.1 | Bayesian model formulation

We may define a Bayesian approach that "mirrors" the frequentist approach. The Bayesian model has the same data likelihood as before, and places prior distributions on parameters $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$. Let $\theta = (\boldsymbol{\beta}', \boldsymbol{\gamma}')'$. We may write the data likelihood as

$$\boldsymbol{Y}|\theta \sim \mathcal{N}(\boldsymbol{A}\boldsymbol{\beta}, \boldsymbol{\Sigma}), \quad \boldsymbol{\Sigma} = \operatorname{diag}(\exp(\boldsymbol{\Omega}\boldsymbol{\gamma})). \tag{10}$$

In our computational exploratory analysis of this model, we have found empirically that independent noninformative priors for $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ provide acceptable results. That is, we set

$$\beta_i \sim \mathcal{U}(-\infty, +\infty), \quad i = 1, \ldots, q, \tag{11}$$

and

$$\gamma_j \sim \mathcal{U}(-\infty, +\infty), \quad j = 1, \ldots, r, \tag{12}$$

where $\mathcal{U}$ is the improper uniform distribution. Thus, the joint posterior distribution, $p(\boldsymbol{\theta}|\boldsymbol{Y})$, is proportional to the likelihood of the data. As we are principally interested in parameter and interval estimation, the improper prior is acceptable for our task. In our exploratory analysis of this model, we found that these improper prior distributions lead to a proper posterior distribution.

4.2 | Bayesian model reduction

Numerous information criterion statistics are available to aide in model comparison, but none are universally appealing. A cross-validation estimate of out-of-sample prediction error may be permissible for our problem because it can be used for comparing models with noninformative prior distributions.

We use the leave-one-out cross validation (LOOCV) statistic that appears in Equation (13) to compare models. This technique repeatedly partitions the data into n subsamples of size $n - 1$. Models on the n *training* data sets are fit, and predictions are made on the single, remaining y_i removed from each training data set.

$$\text{LOOCV} = \sum_{i=1}^{n} \log p_{(-i)}(y_i|\boldsymbol{\theta}_{\text{post}}), \tag{13}$$

where the distribution $p_{(-i)}(y_i|\boldsymbol{\theta}_{\text{post}})$ is the posterior predictive distribution of the training data without the ith observation evaluated at y_i. If the difference between the LOOCV of the full and reduced model is minimal (less than two standard errors from 0), we favor the reduced model. Details on LOOCV for Bayesian models are supplied by Gelman et al[16] and Vehtari et al.[17]

4.3 | Bayesian circular prediction interval

We compute circular prediction regions from the posterior predictive distribution of the Bayesian model. We can sample from the posterior predictive distribution,

$$p(\boldsymbol{y}^{rep}|\boldsymbol{Y}) = \int_{\boldsymbol{\theta}} p(\boldsymbol{y}^{rep}|\boldsymbol{\theta}) p(\boldsymbol{\theta}|\boldsymbol{Y}) d\boldsymbol{\theta},$$

using the statistical software, Stan,[18] by simulating draws from the data likelihood, $p(\boldsymbol{y}^{rep}|\boldsymbol{\theta})$, given draws from the posterior, $p(\boldsymbol{\theta}|\boldsymbol{Y})$. The resulting distribution will be centered at the posterior mode of the linear predictor, $\boldsymbol{A\beta}$, but have a larger variance to account for the inherent uncertainty involved in new data.

To form the bivariate prediction circles, we consider the joint distribution of $\boldsymbol{y}^{rep} = (x^{rep}, y^{rep})$ where x^{rep} and y^{rep} are draws from the posterior predictive distribution $p(\boldsymbol{y}^{rep}|\boldsymbol{\theta})$. Conditional on the specific rows of $\boldsymbol{M}$ and $\boldsymbol{Z}$ of import, we generate (say) 10 000 draws of $\boldsymbol{y}^{rep}$, yielding a bivariate sample.

A highest density region (HDR) containing $1 - \alpha$ of the joint posterior probability may be calculated using Hyndman's density quantile algorithm.[19] Hyndman's method depends on a density estimate, which may be formed nonparametrically using a kernel density estimate.

Kernel density estimates require selection of a bandwidth and kernel, but these "parameters" are unimportant as we can sample such large quantities of data from the posterior distribution that variation in the density estimates depending on the kernel and bandwidth quickly becomes negligible.

TABLE 2 ANOVA table for full model

Term	$\hat{\rho}$	$\hat{\zeta}$	$\hat{F}$	m	P-val
Intercept	1.95	−1.38			
Range	1.25	−3.42	7.11	47.76	.00
Weapon	4.20	0.97	9.95	47.76	.00
Ammo	−0.40	−1.43	1.17	47.76	.32
Range by Weapon	3.10	1.02	5.70	47.76	.01
Range by Ammo	−0.40	0.42	0.18	47.76	.83
Weapon by Ammo	0.62	0.18	0.90	45.81	.41

The contour of the $1-\alpha$ HDR is approximately circular in our problem. To "mirror" results from our frequentist approach, we constrain the prediction region to be exactly circular by computing the mean Euclidean distance from the posterior mode to points on the estimated $1-\alpha$ contour to form the radius of the circle. The posterior mode and this radius define the circular prediction region.

5 | APPLICATION TO THE EXAMPLE

Each approach is now applied to the example data from Section 2. We first formulate the full models (nonreduced) for the mean and variance, and then construct the corresponding model matrices.

The full model for the mean is second order, and has one intercept, three main effect coefficients, and three two-factor interaction coefficients. We code the factors using a typical "sum to zero contrast" (see the `contr.sum` function in the `stats` package[20] in R). In the model matrix, this coding scheme assigns a "−1" and "+1" to the first and second levels of a two-level factor. As a result, the full model coefficient vectors for the mean, $\boldsymbol{\rho}$ and $\boldsymbol{\zeta}$, are each of size 7×1, and the combined vector, $\boldsymbol{\beta}$, is of size 14×1.

We assume a structure for the variance formula. Visual inspection of the residuals in Figure 1 suggests a variance formula that includes an intercept term and a main effect for the range-to-target factor. Thus, $\boldsymbol{\gamma}$ is size 2×1.

5.1 | Frequentist approach

The model is fit using the `remlscore` function of the R[15] package `statmod`. The inputs to the function are $\boldsymbol{Y}$ (size 80×1), $\boldsymbol{A}$ (size 80×14), and $\boldsymbol{\Omega}$ (size 80×2). The function outputs the estimated coefficients for the mean $\hat{\boldsymbol{\beta}}$, which is composed of the estimated x- and y-component coefficients, $\hat{\boldsymbol{\rho}}$, and $\hat{\boldsymbol{\zeta}}$. The function also outputs the estimated variance coefficients, $\hat{\boldsymbol{\gamma}}$, in addition to $\hat{\boldsymbol{\Phi}}$ and $\hat{\boldsymbol{W}}$, which are the estimated covariance matrices for $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\gamma}}$, respectively.

Model reduction begins with a hypothesis test on each pair of coefficients in the full model. These results are presented in Table 2. The backward selection procedure sequentially removes insignificant terms from the model. As a result, the coefficient pairs for Ammo, Range by Ammo, and Weapon by Ammo are removed from the model.

Table 2 shows that Range, Weapon, and the interaction between Range and Weapon are significant at the $\alpha = .05$ significance level. The positive and large x-component coefficient for Weapon (equal to 4.20) indicates that the mean miss distance shifts right as Weapon changes from old to new. This trend is magnified as Range changes from near to far, as indicated by the positive and large x-component Range by Weapon interaction coefficient (equal to 3.10). The negative y-component coefficient for Range (equal to −3.42) indicates that the mean miss distance shifts downward as Range changes from near to far. Additionally, the intercept coefficients show that the overall mean miss distance is biased downward and to the right of the aim point (the origin).

Circular prediction regions are constructed using the reduced model and Equation (9). We set $P = .5$, which implies that a prediction region contains 50% of the sampled miss distances, on average. Figure 2 displays the circular prediction region for each unique mean that remains in the reduced model. The solid line shows the prediction region using the frequentist approach, while the dashed line corresponds to the Bayesian approach.

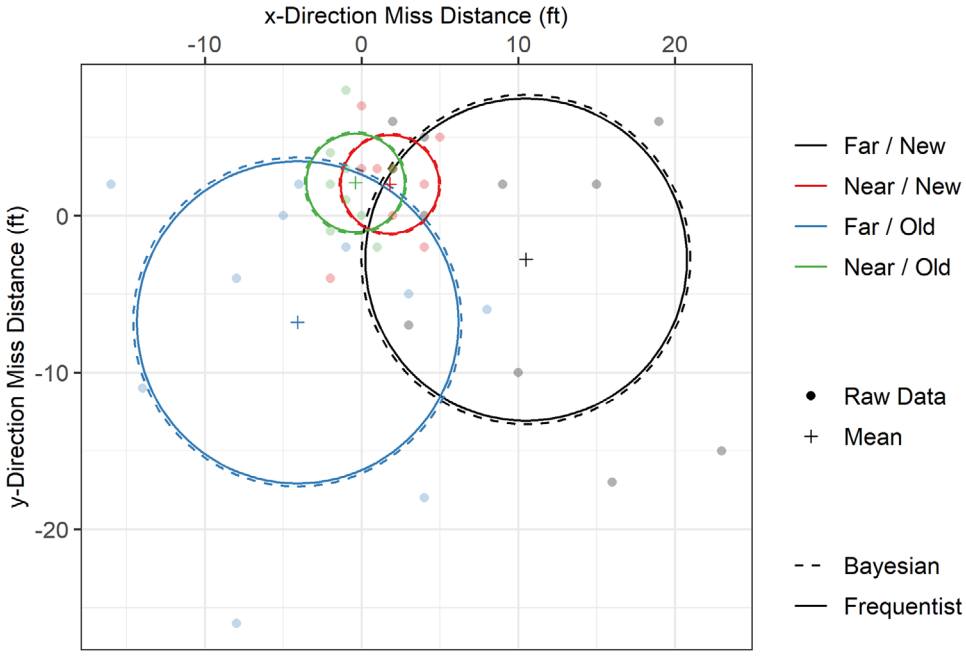

FIGURE 2 Circular prediction regions estimated from reduced model fit

TABLE 3 Posterior means, standard deviations, and 90% credible intervals for full model parameters

Term	Mean	Std. Dev.	90% Interval
$\rho_{\text{Intercept}}$	1.95	1.00	(0.31, 3.59)
ρ_{Range}	1.25	1.00	(−0.40, 2.89)
ρ_{Weapon}	4.20	1.00	(2.56, 5.83)
ρ_{Ammo}	−0.40	1.00	(−2.04, 1.24)
$\rho_{\text{Range:Weapon}}$	3.10	0.99	(1.46, 4.73)
$\rho_{\text{Range:Ammo}}$	−0.40	0.99	(−2.04, 1.24)
$\rho_{\text{Weapon:Ammo}}$	0.63	0.50	(−0.18, 1.46)
$\zeta_{\text{Intercept}}$	−1.37	1.00	(−3.02, 0.27)
ζ_{Range}	−3.42	1.00	(−5.06, −1.78)
ζ_{Weapon}	0.98	1.00	(−0.66, 2.61)
ζ_{Ammo}	−1.43	1.00	(−3.06, 0.22)
$\zeta_{\text{Range:Weapon}}$	1.03	1.00	(−0.61, 2.66)
$\zeta_{\text{Range:Ammo}}$	0.42	1.00	(−1.21, 2.06)
$\zeta_{\text{Weapon:Ammo}}$	0.18	0.50	(−0.64, 0.99)
σ_{Far}	8.54	1.08	(6.97, 10.48)
σ_{Near}	2.28	0.31	(1.85, 2.81)

5.2 | Bayesian approach

Applying the Bayesian approach outlined in Section 4, we sample from the joint posterior distribution of $\boldsymbol{\theta}|\boldsymbol{Y}$ using the statistical software Stan.[18] We use Stan partly because improper priors are allowed (as long as the resultant posterior distribution is proper) and partly because the built-in Markov chain Monte Carlo algorithm is relatively fast and reliable. For the presently considered example data, we sample 100 000 times from four Markov chains. Half of the samples are discarded for burn-in.

First, in Table 3, we examine the credible intervals for pairs (ρ_i, ζ_i) that contain 0. We identify that credible intervals for Ammo, Weapon by Ammo, and Range by Ammo contain 0 in both ρ and ζ. These variables are candidates for removal from the full model. A reduced model is considered that is composed of variables Weapon, Range, and Range by Weapon.

TABLE 4 Simulation study scenarios

Scenario	Experiment	Variance formula	Mean coefficients	Variance coefficients
1	Design 1	Formula 1	Pattern 1	Pattern 3
2	Design 2	Formula 1	Pattern 1	Pattern 4
3	Design 1	Formula 2	Pattern 1	Pattern 4
4	Design 2	Formula 2	Pattern 1	Pattern 3
5	Design 1	Formula 1	Pattern 2	Pattern 4
6	Design 2	Formula 1	Pattern 2	Pattern 3
7	Design 1	Formula 2	Pattern 2	Pattern 3
8	Design 2	Formula 2	Pattern 2	Pattern 4

We use LOOCV to estimate the out-of-sample predictive fit of the full model and the reduced model. For the full model, LOOCV $= -240.3$, and for the reduced model, LOOCV $= -241.7$. The difference is -1.4, with standard error 3.8, which shows that there is not a practical difference between the models, from a predictive point of view. We proceed in our Bayesian analysis with the reduced model.

In contrast to the frequentist approach, the Bayesian approach admits uncertainty intervals for all parameters of interest. For example, we provide credible intervals for σ_{Near} and σ_{Far}. The disjointedness of these two credible intervals provides evidence that the structure of $\boldsymbol{\Sigma}$ is appropriate in that a reduction to a constant variance $\boldsymbol{\Sigma}$ would be unreasonable.

Due to the noninformative prior distributions on θ, we do not observe any shrinkage in the posterior parameter estimates. Comparing Tables 2 and 3, we see that the posterior means are similar to frequentist parameter estimates.

The computed circular prediction regions are also very similar. For the frequentist approach, the radii of the four circular prediction regions in Figure 2 are 10.27, 10.27, 3.14, and 3.14 ft. The corresponding radii for the Bayesian approach are 10.59, 10.54, 3.21, and 3.23 ft. In this particular, a Bayesian approach with flat priors produced prediction regions with modestly larger radii.

6 | SIMULATION STUDY COMPARISON

The simulation study compares empirical frequentist coverage rates of circular prediction regions to their intended coverage rates. This study serves as a limited verification of each approach. The general steps of the simulation are to

1. assume values for the true coefficients,
2. generate a sample data set from these true coefficients,
3. fit the model to these data,
4. calculate the circular prediction interval,
5. generate an additional data point and record whether it falls within the circular prediction interval,
6. repeat steps two through five many times, and
7. compare the intended coverage rate to the proportion of times the additional data point fell within the circular prediction region.

The comparison is made across eight different scenarios. The scenarios are representative of typical defense tests, and differ by their experimental design and model formulas. The eight scenarios appear in Table 4.

The scenarios comprise two experimental designs. Design 1 is a full factorial experiment that has three two-level factors. Design 2 is a full factorial experiment that has two two-level factors and one three-level factor. Each is replicated 10 times, which provides 80 and 120 samples for Designs 1 and 2, respectively.

In all scenarios, the mean formula consists of main effects and two-factor interactions. The study includes two formulas for the variance. Formula 1 includes an intercept and one main effect for a two-level factor. Formula 2 includes an intercept and two main effects for two two-level factors.

TABLE 5 Coefficient patterns for simulation study

	Vector	Intercept setting	First setting	Second setting
Pattern 1	ρ	2	−3	1
	ζ	1	−1	−2
Pattern 2	ρ	3	2	1
	ζ	−2	−2	−1
Pattern 3	γ	3	−0.5	0.5
Pattern 4	γ	2	-1	1.5

TABLE 6 Simulated circular prediction interval coverage rates. Standard errors in parentheses

	Frequentist approach			Bayesian approach		
Scenario	Minimum	Mean	Maximum	Minimum	Mean	Maximum
1	0.883	0.895	0.906	0.888	0.900	0.911
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
2	0.887	0.896	0.910	0.892	0.900	0.910
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
3	0.885	0.893	0.904	0.881	0.896	0.911
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
4	0.889	0.898	0.908	0.897	0.904	0.914
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
5	0.886	0.896	0.910	0.891	0.900	0.912
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
6	0.892	0.900	0.909	0.897	0.904	0.912
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
7	0.889	0.897	0.908	0.898	0.907	0.917
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
8	0.876	0.892	0.908	0.873	0.893	0.915
	(0.006)	(0.006)	(0.007)	(0.006)	(0.006)	(0.007)

True coefficients (ρ, ζ, or γ) are set according to a specified pattern. The pattern sets the intercept equal to a specified value, and then in an alternating sequence, sets the remaining coefficients in the vector equal to a first or second setting. As shown in Table 5, Patterns 1 and 2 correspond to mean coefficients, and Patterns 3 and 4 correspond to variance coefficients.

To illustrate for Scenario 1, the x- and y-component mean coefficient vectors are of size 7×1 (one intercept, three main effects, three two-factor interactions); thus, Pattern 1 yields

$$\rho = \begin{bmatrix} 2 & -3 & 1 & -3 & 1 & -3 & 1 \end{bmatrix}, \tag{14}$$

$$\zeta = \begin{bmatrix} 1 & -1 & -2 & -1 & -2 & -1 & -2 \end{bmatrix}. \tag{15}$$

Moreover, for Scenario 1, given Formula 1 and Pattern 3, the variance coefficient vector is $\gamma = \begin{bmatrix} 3 & -.5 \end{bmatrix}$.

The outputs of the simulation are the simulated coverage rates. We set the target coverage rate equal to 90% ($P = .9$), as opposed to the 50% coverage rate we used in the example, to reduce the standard error in the simulated output.

The Bayesian fitting procedure is quite expensive, and thus, we only repeat steps 2 through 5 (see beginning of this section) 2500 times. Each Bayesian model fit uses four chains with 6000 iterations, and 2000 of which are for burn-in.

A simulated coverage rate is calculated for each circular prediction region. The number of circular prediction regions for a given scenario is determined by the experimental design. Design 1 has eight circular prediction regions, while Design 2 has 12 circular prediction regions.

The simulation results in Table 6 show the minimum, mean, and maximum simulated coverage rates among the 8 or 12 circular prediction regions for each scenario. The corresponding minimum, mean, and maximum standard error of the simulated coverage rates is shown in parentheses.

The simulation results show that for each of the scenarios, the approaches produce simulated frequentist coverage rates that are close to the intended rate (equal to 0.90). All mean coverage rates are within two standard errors of the intended coverage rate. The simulation study appears to have verified the implementation of the approaches for these limited numbers of scenarios.

As a closer look, we fit a linear model and conducted an analysis of variance using the simulated mean coverage rate as the response variable to screen for influential main effects in the simulation study. Though not shown, results indicated that the frequentist approach had a significantly lower mean simulated coverage rate (= 0.896) than the Bayesian approach (= 0.901). Additionally, Pattern 3 (= 0.900) for the variance formula had a significantly higher mean simulated coverage rate than Pattern 4 (= 0.896). Experiment, variance formula, and mean coefficients were not significant.

7 | DISCUSSION

We presented two approaches for estimating circular prediction regions that produced similar results. In the example problem, the conclusions from each approach were the same: the miss distance of the new weapon was biased to the right, which was exacerbated at long range, and gravity appears to have caused the projectile to drop at long range.

Further, in the example, both methods produced statistical intervals that were similar in size, a fact that could be reassuring to traditional practitioners of frequentist methods. The results of this application provide anecdotal evidence that in the absence of prior information, either statistical methodology is permissible.

The primary difference between results comes from the interpretation. The frequentist approach produces an interval that will contain future shots with probability $1 - \alpha$, where probability is typically interpreted in the sense of long-run frequencies. The Bayesian approach produces an interval that will contain future shots with conditional probability $1 - \alpha$, representing a subjective "degree of belief." Both produce an expected coverage $1 - \alpha$, but "expected" is in two different senses.

Our simulations showed that the Bayesian approach provided the intended coverage rate in *both* senses. Though the simulation was setup to verify the coverage rates according to long-run frequencies, we included the Bayesian approach in it anyhow. It was somewhat unsurprising to find that, under flat prior distributions, the Bayesian approach produced the intended frequentist coverage rate. Indeed, in simpler normal models (eg, Ref. [21, p. 301]), Bayesian prediction intervals have an identical closed-form equation as their frequentist counterpart.

A practical difference between approaches involved the difficulty of their implementation. Fitting the model and estimating prediction regions using the Bayesian approach required far fewer lines of code, and leveraged existing functionality from the Stan package[18] in R. Conversely, for the frequentist approach, we had to manually implement the Kenward-Roger statistic, leading to much more complexity.

The example problem considered a perfectly balanced factorial experiment, but in practice, it is not uncommon to encounter an unbalanced design. In such a case, the miss distance models will likely suffer from the same issues of multicollinearity as typical linear models. As multicollinearity worsens, deciphering significant factor effects from one another may become difficult. As an additional note about the experimental design, the example only employed categorical factors, but the models can readily accommodate continuous factors as well.

Future work may consider more complex variance structures for the bivariate miss distance model. By allowing the x- and y-component variances to differ and the x- and y-component covariance to be nonzero, the prediction regions could take the shape of a rotated ellipse. This flexibility may lead to a better fitting model in certain cases. Other work could investigate the possibility of constructing circular tolerance regions, perhaps using the Kenward Roger's statistic or an alternative test statistic.

ORCID

Thomas H. Johnson https://orcid.org/0000-0002-6782-3883

REFERENCES

1. Elder RL. *An Examination of Circular Error Probable Approximation Techniques*. Wright-Patterson Air Force Base, Ohio: Air Force Institute of Technology, Wright-Patterson School of Engineering; 1986.
2. Harter HL. Circular error probabilities. *J Am Stat Assoc*. 1960;55:723-731.
3. Moranda PB. Comparison of estimates of circular probable error. *J Am Stat Assoc*. 1959;54:794-800.

4. Williams CE. *A Comparison of Circular Error Probable Estimators for Small Samples*. Wright-Patterson Air Force Base, Ohio: Air Force Institute of Technology, Wright-Patterson School of Engineering; 1997.
5. Zimmer Z, Turner C. Regression based circular error probable: an application to ballistic systems. *Commun Stat Case Stud Data Anal Appl*. 2017;3:48-54.
6. Krishnamoorthy K, Mathew T. *Statistical Tolerance Regions: Theory, Applications, and Computation*. Hoboken, NJ: John Wiley & Sons; 2009.
7. Hall IJ, Sheldon DD. Improved bivariate normal tolerance regions with some applications. *J Qual Technol*. 1979;11:13-19.
8. DiDonato A. *Computation of the Circular Error Probable (CEP) and Confidence Intervals in Bombing Tests*. VA: Naval Surface Warfare Center Dahlgren; 2007.
9. Zhang J, Weilian A. Assessing circular error probable when the errors are elliptical normal. *J Stat Comput Simul*. 2012;82:565-586.
10. Smyth GK. An efficient algorithm for REML in heteroscedastic regression. *J Comput Graph Stat*. 2002;11:836-847.
11. Kenward MG, Roger JH. Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics*. 1997;53:983-997.
12. JMP Version 13, SAS Institute Inc., Cary, NC; 1989-2018.
13. Myers RH, Montgomery DC, Geoffrey VG, Robinson TJ. *Generalized Linear Models: With Applications in Engineering and the Sciences*. New York, NY: John Wiley & Sons; 2012.
14. Verbyla AP. Modelling variance heterogeneity: residual maximum likelihood and diagnostics. *J R Stat Soc Series B Stat Methodol*. 1993;55:493-508.
15. R Core Team. *R: A Language and Environment for Statistical Computing*. Austria: R Foundation for Statistical ComputingVienna; 2013. ISBN 3-900051-07-0.
16. Gelman A, Stern HS, Carlin JB, Dunson DB, Vehtari A, Rubin DB. *Bayesian Data Analysis*. Boca Raton: Chapman and Hall/CRC; 2013.
17. Vehtari A, Gelman A, Gabry J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat Comput*. 2017;27:1413-1432.
18. Carpenter B, Gelman A, Hoffman MD, et al. Stan: A probabilistic programming language. *J Stat Softw*. 2017;76:1-32.
19. Hyndman RJ. Computing and graphing highest density regions. *Am Stat*. 1996;50:120-126.
20. R Core Team. *R: A Language and Environment for Statistical Computing*. Austria: R Foundation for Statistical Computing Vienna; 2013. ISBN 3-900051-07-0.
21. Bolstad WM, Curran JM. *Introduction to Bayesian statistics*. Hoboken, NJ: John Wiley & Sons; 2016.

AUTHOR BIOGRAPHIES

Thomas H. Johnson is a Research Staff Member at the Institute for Defense Analyses in Alexandria, Virginia. He obtained an MS and PhD in Aerospace Engineering from Old Dominion University, and a BS in Aerospace Engineering from Boston University. His research interests include experimental design and regression modeling for a wide variety of defense problems that address system reliability, effectiveness, lethality, and vulnerability.

John T. Haman is a statistician at the Institute for Defense Analyses. He is a member of the Test Science team and is interested in statistical methodologies for operational evaluation. Currently, John is supporting electronic warfare systems for the Navy and Air Force. He obtained an MS and PhD in Statistics from Bowling Green State University, and a BS in Mathematics from Truman State University.

Heather Wojton is a Research Staff Member in the Operational Evaluation Division at the Institute for Defense Analyses. In her role, she leads an interdisciplinary team that facilitates data-driven decision-making within the DoD by advancing and applying statistical, behavioral, and data science methodologies to evaluate military programs. Currently, Dr. Wojton and her team, including Dr. Johnson and Dr. Haman, are advancing methods for M&S validation, testing AI-enabled systems and human-machine teams, and evaluating skill transfer in immersive virtual training environments.

Laura Freeman a Research Associate Professor and the Director of the Intelligent Systems Lab at the Virginia Tech Hume Center. Her research leverages experimental methods for conducting research that brings together cyberphysical systems, data science, artificial intelligence, and machine learning to address critical challenges in national security. She develops new methods for test and evaluation focusing on emerging system technology. She is also the Assistant Dean for Research in the National Capital Region, in that capacity, she works to shape research directions and collaborations in across the College of Science in the National Capital Region. Before working for Virginia Tech, Dr. Freeman was the Assistant Director of the Operational Evaluation Division at the Institute for Defense Analyses.

How to cite this article: Johnson TH, Haman JT, Wojton H, Freeman L. Circular prediction regions for miss distance models under heteroskedasticity. *Qual Reliab Engng Int*. 2021;37:2991–3003.
https://doi.org/10.1002/qre.2771

APPENDIX A: KENWARD-ROGER APPROXIMATION

Kenward and Roger approximate the distribution of F as

$$F \sim \lambda \mathcal{F}_{l,m}, \tag{A1}$$

where $\mathcal{F}_{l,m}$ is the F-distribution with l numerator degrees of freedom and m denominator degrees of freedom. The parameters λ and m are calculated after the data are collected, and are dependent on that data. The calculation of these parameters is somewhat involved. For completeness and because the "remlscore" R function does not implement the Kenward-Roger test, we include the equations below:

$$\lambda = \frac{m}{E[F](m-2)}, \tag{A2}$$

$$m = 4 + \frac{l+2}{lg-1}. \tag{A3}$$

The calculation of λ and m depends on the following equations:

$$g = \frac{\mathrm{Var}[F]}{2E[F]^2}, \tag{A4}$$

$$E[F] = 1 + \frac{A_2}{l}, \tag{A5}$$

$$\mathrm{Var}[F] = \frac{2}{l}(1+B), \tag{A6}$$

$$B = \frac{1}{2l}(A_1 + 6A_2), \tag{A7}$$

$$A_1 = \sum_{a=1}^{r}\sum_{b=1}^{r} W_{ab}\,\mathrm{tr}(\boldsymbol{\Theta\Phi P}_a\boldsymbol{\Phi})\mathrm{tr}(\boldsymbol{\Theta\Phi P}_b\boldsymbol{\Phi}), \tag{A8}$$

$$A_2 = \sum_{a=1}^{r}\sum_{b=1}^{r} W_{ab}\,\mathrm{tr}(\boldsymbol{\Theta\Phi P}_a\boldsymbol{\Phi\Theta\Phi P}_b\boldsymbol{\Phi}), \tag{A9}$$

$$\boldsymbol{\Theta} = \boldsymbol{L}\left(\boldsymbol{L}'\boldsymbol{\Phi L}\right)^{-1}\boldsymbol{L}', \tag{A10}$$

$$\boldsymbol{P}_c = \boldsymbol{A}'\frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \boldsymbol{\gamma}_c}\boldsymbol{A}, \tag{A11}$$

where W_{ab} is the element corresponding to the ath row and bth column of the variance-covariance matrix of $\hat{\boldsymbol{\gamma}}$. The `remlscore` function outputs this variance-covariance matrix. γ_c is the cth element of $\boldsymbol{\gamma} = [\gamma_1, \ldots, \gamma_r]'$.