Using a Generalized Linear Mixed Model Framework to Account for Spatial Variability in a Comparison of Orchard Sprayer Efficacy

William J. Price  
*University of Idaho*

Bahman Shafii  
*University of Idaho*, bshafii@uidaho.edu

Don Morishita  
*University of Idaho*

Follow this and additional works at: [http://newprairiepress.org/agstatconference](http://newprairiepress.org/agstatconference)

Part of the [Agriculture Commons](https://www.scholarworks.kstate.edu/agriculture_commons) and the [Applied Statistics Commons](https://www.scholarworks.kstate.edu/applied_statistics_commons)

This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License](https://creativecommons.org/licenses/by-nc-nd/4.0/).

**Recommended Citation**


This Event is brought to you for free and open access by the Conferences at New Prairie Press. It has been accepted for inclusion in Conference on Applied Statistics in Agriculture by an authorized administrator of New Prairie Press. For more information, please contact [cads@k-state.edu](mailto:cads@k-state.edu).
Using a Generalized Linear Mixed Model Framework to Account for Spatial Variability in a Comparison of Orchard Sprayer Efficacy

William J. Price and Bahman Shafii

Statistical Programs, College of Agricultural and Life Sciences, University of Idaho, Moscow, Idaho

Don Morishita

Department of Plant, Soil, and Entomological Sciences, University of Idaho

Uniform application of pesticides in vineyard and orchard systems can be difficult to achieve due to variability in the density and structure of the crop canopy. Depending on the equipment used and environmental conditions, applications can result in poor spray coverage, spray drift, and wasted spray which, in turn, are manifested as a combination of poor pesticide efficacy, economic losses and potential environmental problems for the grower. A study was therefore designed and carried out to test new sprayer equipment aimed at addressing these issues. Statistically, the study presented a unique replicated three dimensional spatial design which captured response variability (coverage) both within and across trees in an orchard setting. Application of a generalized linear mixed model framework allowed comparison of sprayer designs in terms of their application efficiencies while accounting for the intra- and inter-tree correlation of the coverage response. Examples demonstrating various models and their associated correlation structures are given and the resulting interpretations discussed.

I. INTRODUCTION

Table and wine grapes have a large economic value in the United States. For example, their estimated value in 2014 was 5.8 billion dollars (Crop Values 2014 Summary February 2015 USDA, National Agricultural Statistics Service). Management of this crop is, therefore, important for growers, producers, and consumers.

A primary concern in grape production is protection of the crop from insects and diseases. Insect and related pests such as aphids, mites, and borers can lead to plant damage and disease transmission. Various bacterial and fungal infections are additional, and often related, problems in certain climates. Management of these issues can involve considerable costs and use of material. In California, for example, application of up to 12 lbs of active ingredient per acre each year has been previously noted (California Pesticide Use Reporting Database).

Such high levels of pest control make the efficiency of pesticide application important. Efficient application can directly reduce costs to the grower while providing higher efficacy in pest control. It also has the broader benefits of limiting off target pesticide exposure, thereby minimizing environmental contamination.

One means of addressing these concerns is the modification of foliar sprayer designs. An efficient vineyard spraying system would apply a minimum amount of pesticide material to only those areas where they are needed, avoiding off target application, drift and waste. The study considered here
was carried out to evaluate a new sprayer that was designed with these goals in mind. Two additional spray systems, commonly used in vineyard settings, were also included in the study design for comparison. Although this study examined several aspects of spray coverage, both on and off the target vines, only a subset of the vine foliar data will be considered here as it presents a unique and interesting application of Generalized Linear Mixed Model (GLMM) methods.

II. METHODS

Source and description of the data

Three vineyard spraying systems were evaluated: Air Blast (AB), Cima Gearmore (CG), and an experimental precision application system, GenZ (GZ). The Air Blast sprayer is a common system used in orchard and vineyard settings. It consists of multiple spray nozzles arranged in a radial pattern in combination with a large fan (Figure 1a). It disperses spray droplets in a large omni-directional cloud covering both plant and ground material. A second common sprayer design, Cima Gearmore, applies spray solution in a more directional nature using high pressure venturi nozzles producing a fog of fine droplets (Figure 1b). The final design is an experimental system intended to direct the spray material only to the plant. The system “wraps” directed spray nozzles around a vineyard row with the intention of reducing spray drift and off target exposure (Figure 1c).

Figure 1. Three vineyard sprayer types: a) Air blast (AB), b) Cima Gearmore (CG), and c) GenZ (GZ).

The three sprayer types were laid out in a vineyard setting under a randomized complete block design having four blocks. Within the blocks, each sprayer plot had three data collection areas (PVC pipe
“trees” representing vines) and several drift PVC “trees” placed downwind of the sprayed target vineyard row (Figure 2). Within each PVC “tree”, multiple spray collection points were arranged with the intention of capturing spatial variability within a vine as well as the top & bottom sides of each position (Figures 3a and 3b). Water sensitive cards were placed facing the top and bottom of each position. Following application and drying, the cards were collected and optically scanned to determine the proportion of sprayer coverage (Figure 3c). For the analyses used here, only the top facing cards will be considered.

**Figure 2.** One block of the vineyard experimental layout. Each sprayer type was represented by three target trees (red) and four drift trees (white).

**Figure 3.** PVC trees, placed in the vineyard row canopy (a) were used to collect the data. Each tree had twelve canopy positions (b) with water sensitive cards (c).
Prior to analysis, the replicate PVC “trees” in each plot were averaged for each of the 12 card positions in order to mitigate within plot variability while maintaining the spatial structure. The final data form, therefore, contained one set of 12 points for each sprayer type block combination. The collective card positions can be considered as 12 sub-samples within a plot or, alternatively, they can be assigned spatial coordinates consisting of row and column positions. Analyses using both of these possibilities will be examined below.

**Statistical methods**

The response, proportion coverage \( Z_{ijxy} \), is a natural candidate for a beta distribution:

\[
Z_{ijxy} \sim \text{Beta}(p_{ijxy}) \quad \text{for } 0 < Z_{ijxy} < 1.0
\]  

(1)

where \( i = 1 \) to 3 sprayer types, \( j=1 \) to 4 blocks at row and column \((x, y)\), and \( p_{ijxy} \) is the proportion coverage of the \( i^{th} \) sprayer at row-column position \((x, y)\). Within the generalized linear mixed model (GLMM) framework, the link function for \( Z_{ijxy} \) is the generalized logit:

\[
\eta_{ijxy} = \ln(Z_{ijxy}/(1-Z_{ijxy}))
\]  

(2)

Three model scenarios were considered, similar to those proposed by Stroup (2013):

1) Randomized Complete Block (RCB) with sub-sampling, 2) Spatially arranged with a common variance-covariance matrix (Common), and 3) Spatial arranged with unique variance-covariance structures. The later option was investigated under three scenarios for defining the variance-covariance structure: a parametric spatial model (Spatial), a Toeplitz banded structure (Toep), and a non-parametric splined surface (Spline).

The RCB model is given as:

\[
\eta_{ijxy} = \mu + \tau_i + \beta_j + t_{ij} + e_{ijxy}
\]  

(3)

where \( \mu \) is an overall mean, \( \tau_i \) is the effect of sprayer \( i \), \( \beta_j \) is a random effect from block \( j \), \( t_{ij} \) is a random error term and \( e_{ijxy} \) is a random term due to sub-sampling. This model ignores any potential spatial structuring in the data.

The second scenario, spatial structure with a common variance-covariance matrix is defined as:

\[
\eta_{ijxy} = \mu + \tau_i + \omega_{ijxy}
\]  

(4)

where \( \mu \) is an overall mean, \( \tau_i \) is the effect of sprayer \( i \), \( \omega_{ijxy} \) is a random effect from position \((x, y)\) and the covariance between two positions, \( \text{cov}(Z_{ijxy}, Z_{ix'y'}) \), is specified through a parametric spatial model such as the spherical, Gaussian, or exponential and is assumed common across all three sprayer types.

The final scenario, spatial structure with a variance-covariance structure unique to each sprayer type, has a model given by:

\[
\eta_{ijxy} = \mu + \tau_i + \omega_{ijxy}
\]  

(5)

where \( \mu \) is an overall mean, \( \tau_i \) is the effect of sprayer \( i \), \( \omega_{ijxy} \) is a random effect from position \((x, y)\) and the covariance between two positions, \( \text{cov}(Z_{ijxy}, Z_{ix'y'}) \), can again be specified as a parametric form and assumed to be unique for each of the three sprayer types.
While there are many possible parametric model forms that could be assessed in eqs (4) and (5), this work will only present the spherical model as little difference was seen in the results from other model forms. This spherical spatial model is given by:

\[ \text{Cov}(Z_a, Z_b) = \sigma^2 \{ 1 - (3\delta_{ab}/2\alpha) + 0.5(\delta_{ab}/\alpha)^3 \} \]  

(6)

where \( \delta_{ab} \) is the distance between points \( Z_a \) and \( Z_b \). Eq (5) is applicable for \( \delta_{ab} \leq \alpha \) and set to 0 otherwise.

Alternatively, nonparametric spatial forms can also be used for the variance-covariance structure. One such structure is the banded Toeplitz matrix:

\[ \eta_{ijxy} = \mu + \tau_i + \nu_{ijxy} \]  

(7)

where \( \mu \) is an overall mean, \( \tau_i \) is the effect of sprayer \( i \), \( \nu_{ijxy} \) is again a random effect from position \((x, y)\) where the covariance between two positions, \( \text{cov}(Z_{ijxy}, Z_{ijx'y'}) \), is now a banded covariance matrix across the positions \((x, y)\) specified separately for each of the three sprayer types.

A last nonparametric alternative is a splined surface given as:

\[ \eta_{ijxy} = \mu + \tau_i + s_{ij} + \phi_{ijxy} \]  

(8)

where \( \mu \) is an overall mean, \( \tau_i \) is the effect of sprayer \( i \), \( s_{ij} \) is a continuous fixed splined effect computed over the spatial coordinates, and \( \phi_{ijxy} \) are the random effects due to the position \((x, y)\) and the covariance between two positions, \( \text{cov}(Z_{ijxy}, Z_{ijx'y'}) \), is based on Euclidian distances of spline knots and random effects (Rupert, Wand & Carroll 2003).

**Additional statistical methods**

As a means of initial data exploration, the spatial structure within each sprayer type was examined through sample variogram estimation defined as:

\[ \gamma(h)_{ijxy} = (1/|N(h)|) \sum |Z_{ijxy} - Z_{ijx'y'}|^2 \]  

(9)

where \(|N(h)|\) is the number of data pairs in the spatial grid separated by a distance (lag) \( h \), and \( \sum |Z_{ijxy} - Z_{ijx'y'}|^2 \) is the sum, over all lags, of the squared data pair differences. Plots of the sample variogram against lag \( h \) were used for assessing presence and shape of potential spatial structures.

**Model assessment and estimation**

All models were fit assuming a GLMM model with a beta likelihood and a logit link function estimated with the LaPlace optimization algorithm. Models were subsequently assessed and compared relative to their respective residual magnitudes and structures, parameter estimates and accuracy, and measures of fit: AIC, AICC, BIC. All statistical modeling and graphics were carried out using SAS 64bit, Version 9.4 (SAS, 2012).

**III. EMPIRICAL RESULTS**

Initially, the average proportion of coverage at each card position for each sprayer type was determined. These values were then plotted and color coded according to their respective positions and
values (Figure 4a). From these plots, some general patterns of coverage were evident: The AB sprayer had higher coverage on the canopy sides and lower values on the canopy interior. The CG sprayer resulted in high spray coverage on one side of the canopy, while the GZ sprayer gave higher values on both the canopy sides as well as the canopy top (Figure 4b).

Figure 4. Average percent coverage (a) of each sprayer type (red= high coverage, green=low coverage). Each sprayer had a unique coverage pattern (b).

Applying eq(9) to compute sample variograms for each sprayer gave differing patterns as shown in Figure 5. Here the AB sprayer shows a decreasing variance function with increasing distance between points. This is most likely due to the limited spatial resolution of the data in combination with the double sided coverage pattern of the AB sprayer where more distant positions are more similar than those close together. Both the CG and GZ sprayers result in nearly linear increases in variability with increasing distance between positions. In all cases, the limited size of the spatial grid (3 x 4) prevents the detection of a plateau point where variability between card positions stabilizes. Overall, it is clear that spatial variability exists within the canopy and that its degree and pattern varies across sprayers. The limited resolution, however, may make modeling of the spatial structure difficult.
Model Results

Analysis of variance results are given in Table 1. In all model scenarios, the effect of sprayer was highly significant with the exception of the splined variance-covariance model (p=0.0424). For all cases, the Air Blast sprayer had the highest mean levels of coverage at approximately 70% (Table 2 and Figure 6). The experimental sprayer, GZ, had about half as much coverage at 35% and the CG sprayer resulted in the lowest coverage around 19%.

Figure 5. Sample variograms for the three sprayer types given their general coverage patterns.

<table>
<thead>
<tr>
<th>Model</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>RCB</td>
<td>67.54</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Common</td>
<td>49.9</td>
<td>0.0002</td>
</tr>
<tr>
<td>Spatial</td>
<td>90.14</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Toep</td>
<td>59.02</td>
<td>0.0002</td>
</tr>
<tr>
<td>Spline</td>
<td>59.02</td>
<td>0.0424</td>
</tr>
</tbody>
</table>

Table 1. Analysis of variance results. Significance of the Sprayer type effect for the five model scenarios: Randomized complete block (RCB), Common spatial model (Common), Unique spatial model (Spatial), Banded Toeplitz (Toep), and the splined surface (Spline).
Table 2. Estimated means, standard errors, 95% confidence limits, and fit statistics for the three sprayer types: Air Blast (AB), Cima Gearmore (CG), and GenZ (GZ) under five model scenarios: Randomized complete block (RCB), Common spatial model (Common), Unique spatial model (Spatial), Banded Toeplitz (Toep), and the splined surface (Spline).

<table>
<thead>
<tr>
<th>Model</th>
<th>Sprayer Type</th>
<th>Mean</th>
<th>Standard</th>
<th>Lower</th>
<th>Upper</th>
<th>Fit Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AB</td>
<td>0.7068</td>
<td>0.0279</td>
<td>0.6342</td>
<td>0.7702</td>
<td>AIC -113.35</td>
</tr>
<tr>
<td></td>
<td>CG</td>
<td>0.1846</td>
<td>0.0221</td>
<td>0.1365</td>
<td>0.2448</td>
<td>AICC -112.74</td>
</tr>
<tr>
<td></td>
<td>GZ</td>
<td>0.3502</td>
<td>0.0298</td>
<td>0.2812</td>
<td>0.4261</td>
<td>BIC -117.04</td>
</tr>
<tr>
<td></td>
<td>AB</td>
<td>0.6912</td>
<td>0.0310</td>
<td>0.6106</td>
<td>0.7616</td>
<td>AIC -107.90</td>
</tr>
<tr>
<td></td>
<td>CG</td>
<td>0.1993</td>
<td>0.0248</td>
<td>0.1454</td>
<td>0.2670</td>
<td>AICC -107.29</td>
</tr>
<tr>
<td></td>
<td>GZ</td>
<td>0.3749</td>
<td>0.0333</td>
<td>0.2976</td>
<td>0.4591</td>
<td>BIC -90.08</td>
</tr>
<tr>
<td></td>
<td>AB</td>
<td>0.7198</td>
<td>0.0282</td>
<td>0.6460</td>
<td>0.7833</td>
<td>AIC -124.74</td>
</tr>
<tr>
<td></td>
<td>CG</td>
<td>0.1440</td>
<td>0.0155</td>
<td>0.1101</td>
<td>0.1861</td>
<td>AICC -123.08</td>
</tr>
<tr>
<td></td>
<td>GZ</td>
<td>0.3483</td>
<td>0.0348</td>
<td>0.2688</td>
<td>0.4374</td>
<td>BIC -95.04</td>
</tr>
<tr>
<td></td>
<td>AB</td>
<td>0.7090</td>
<td>0.0338</td>
<td>0.6199</td>
<td>0.7845</td>
<td>AIC -123.53</td>
</tr>
<tr>
<td></td>
<td>CG</td>
<td>0.1752</td>
<td>0.0241</td>
<td>0.1237</td>
<td>0.2421</td>
<td>AICC -119.78</td>
</tr>
<tr>
<td></td>
<td>GZ</td>
<td>0.3496</td>
<td>0.0891</td>
<td>0.1708</td>
<td>0.5838</td>
<td>BIC -99.78</td>
</tr>
<tr>
<td></td>
<td>AB</td>
<td>0.6356</td>
<td>0.1653</td>
<td>0.2333</td>
<td>0.9091</td>
<td>AIC -215.41</td>
</tr>
<tr>
<td></td>
<td>CG</td>
<td>0.0924</td>
<td>0.0490</td>
<td>0.0238</td>
<td>0.2982</td>
<td>AICC -213.41</td>
</tr>
<tr>
<td></td>
<td>GZ</td>
<td>0.3261</td>
<td>0.0198</td>
<td>0.2796</td>
<td>0.3762</td>
<td>BIC -237.41</td>
</tr>
</tbody>
</table>

The RCB model had estimated means with moderate precision and a fit value of AICC=-112.7. The common spatial model variance-covariance structure gave comparable significance, means and standard errors, however, it had the worst fit (AICC=-107.3). Given the differences seen in the sprayer coverage patterns and semivariograms (Figure 5), this is not unexpected. Allowing the variance-covariance structure to vary between sprayer types (Spatial model) does, in fact, improve the fit (AICC = -123.08) while maintaining the previous levels of significance, mean estimation and precision. The Spatial model also out performed the base RCB model in terms of fit. Hence, accounting for the spatial variability is important.

The Spatial model approximates the spatial structure through explicitly defined parametric modeling, such as the spherical model used here. The banded Toeplitz structure (Toep) was an alternative non-parametric means of achieving this variance-covariance approximation. The results for the Toep model had essentially equivalent levels of fit (AICC=-119.8) and mean estimation, but with slightly less precision on the means. This may be due to the explicit spatial model having stronger assumptions regarding the form of the spatial structure.

The final model (Spline) was a second non-parametric approximation to the variance-covariance structure and it gave results that differed most from the other scenarios. Although it gave the best
measure of fit (AICC = -213.4), the significance level was much lower (p = 0.0424) and the mean values were smaller in magnitude, and, in some cases, had much higher standard errors. This model showed a high level of instability during estimation, requiring care in model algorithm selection to achieve convergence. Furthermore, the authors were not able to completely define the spline surface uniquely for each sprayer type. While the model does provide a good AICC value, it was deemed as unreliable for these other reasons.

Figure 6. Estimated LSmeans and 95% confidence limits of proportion coverage from five generalized linear mixed model scenarios: a) Randomized complete block (RCB), b) A common spatial model for variance-covariance (Common), c) A unique spatial model for variance-covariance (Spatial), d) a banded Toeplitz variance-covariance for each sprayer (Toep), and e) a splined variance-covariance surface for each sprayer (Spline).
Residual structure for the three sprayer types and five model scenarios are summarized in Figure 7. The overall distributional structure of the residuals is similar across the five models in each sprayer type. The Toeplitz banded variance-covariance model had residuals that were slightly smaller in magnitude, while the Spline method often had a larger residual dispersion. Of the three sprayer types, the experimental GenZ sprayer had a relatively symmetric residual structure and somewhat more variability, while the Air Blast and Cima Gearmore sprayers had skewed residual distributions.

Figure 7. Residual structures of the five model scenarios and the three sprayer types: a) Air Blast, b) Cima Gearmore, and c) GenZ.

IV. CONCLUDING REMARKS

The generalized linear mixed model framework provided a good means to leverage and accommodate the response distribution (beta) as well as the spatial nature of this data set. Accounting for spatial variability improved model fit over the base RCB model, which ignored the spatial structure.
Among the spatial model options, assuming a common structure was a poor choice resulting in reduced model fit. This was not surprising as the data displayed unique spatial structure for the three sprayer types. Utilizing this difference in structure, the Spatial model, assuming unique spherical spatial models for each sprayer, gave a good fit, significance, and mean estimation. These qualities were also seen with the non-parametric banded Toeplitz option, although with slightly less precision. In cases where the resolution of the spatial grid is limited, such as in this data, such non-parametric options may be a good choice, especially if there is question about which parametric spatial model to assume. The alternative non-parametric option, Spline, performed poorly due to problems during estimation, lack of control over unique specification for each sprayer type, and loss of significance and mean precision. This option may hold promise, however, if these conditions can be resolve.

V. REFERENCES


