SPMODEL: SPATIAL MODELING IN R – SUPPLEMENTARY MATERIAL

A PREPRINT

Michael Dumelle [*]
United States
Environmental Protection Agency
200 SW 35th St, Corvallis, OR, 97333
Dumelle.Michael@epa.gov

Matt Higham
Department of Math, Computer Science, and Statistics
St. Lawrence University
23 Romoda Drive, Canton, NY, 13617
mhigham@stlawu.edu

Jay M. Ver Hoef
National Oceanic and Atmospheric Administration
Alaska Fisheries Science Center
Marine Mammal Laboratory, Seattle, WA, 98115
jay.verhoef@noaa.gov

April 26, 2022

Abstract

Enter the text of your abstract here.

***K*eywords** Spatial covariance · Linear Model · Autoregressive model

1 Covariance Functions

2 Estimation

2.1 Likelihood-based Estimation

Minus twice a profiled Gaussian log-likelihood, denoted $-2l(\boldsymbol{\theta}|\mathbf{y})$ is given by

$$-2l(\boldsymbol{\theta}|\mathbf{y}) = \ln|\boldsymbol{\Sigma}| + (\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}})^\intercal \boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}) + n \ln 2\pi, \tag{1}$$

where $\tilde{\boldsymbol{\beta}} = (\mathbf{X}^\intercal \boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^\intercal \boldsymbol{\Sigma}^{-1}\mathbf{y}$. Minimizing Equation 1 yields $\hat{\boldsymbol{\theta}}_{ml}$, the maximum likelihood estimates for $\boldsymbol{\theta}$. Then a closed for solution exists for $\hat{\boldsymbol{\beta}}_{ml}$, the maximum likelihood estimates for $\boldsymbol{\beta}$: $\hat{\boldsymbol{\beta}}_{ml} = \tilde{\boldsymbol{\beta}}_{ml}$, where $\tilde{\boldsymbol{\beta}}_{ml}$ is $\tilde{\boldsymbol{\beta}}$ evaluated at $\hat{\boldsymbol{\theta}}_{ml}$. Unfortunately $\hat{\boldsymbol{\theta}}_{ml}$ can be badly biased for $\boldsymbol{\theta}$ (especially for small sample sizes), which impacts the estimation of $\boldsymbol{\beta}$ (Patterson and Thompson 1971). This bias occurs due to the simultaneous estimation of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ To reduce this bias, restricted maximum likelihood estimation (REML) emerged (Patterson and Thompson 1971; Harville 1977; Wolfinger, Tobias, and Sall 1994). It can be shown that integrating $\boldsymbol{\beta}$ out of a Gaussian likelihood yields the restricted Gaussian likelihood used in REML estimation. Minus twice a restricted Gaussian log-likelihood, denoted $-2l_R(\boldsymbol{\theta}|\mathbf{y})$ is given by

$$-2l_R(\boldsymbol{\theta}|\mathbf{y}) = -2l(\boldsymbol{\theta}|\mathbf{y}) + \ln|\mathbf{X}^\intercal \boldsymbol{\Sigma}^{-1}\mathbf{X}| - p \ln 2\pi, \tag{2}$$

where p equals the dimension of $\boldsymbol{\beta}$. Minimizing Equation 2 yields $\hat{\boldsymbol{\theta}}_{reml}$, the restricted maximum likelihood estimates for $\boldsymbol{\theta}$. Then a closed for solution exists for $\hat{\boldsymbol{\beta}}_{reml}$, the restricted maximum likelihood estimates for $\boldsymbol{\beta}$: $\hat{\boldsymbol{\beta}}_{reml} = \tilde{\boldsymbol{\beta}}_{reml}$, where $\tilde{\boldsymbol{\beta}}_{reml}$ is $\tilde{\boldsymbol{\beta}}$ evaluated at $\hat{\boldsymbol{\theta}}_{reml}$.

[*]Corresponding Author

Generally the overall variance, σ^2, can be profiled out of Equation 1 and Equation 2. This reduces the number of parameters requiring optimization by one, which can dramatically reduce estimation time. For example, profiling σ^2 out of Equation 1 yields

$$-2l^*(\boldsymbol{\theta}^*|\mathbf{y}) = \ln|\boldsymbol{\Sigma}^*| + n\ln[(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}})^\intercal \boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}})] + n + n\ln 2\pi/n. \tag{3}$$

After finding $\hat{\boldsymbol{\theta}}^*_{ml}$ a closed form solution for $\hat{\sigma}^2_{ml}$ exists: $\hat{\sigma}^2_{ml} = [(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}})^\intercal \boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}})]/n$. Then $\hat{\boldsymbol{\theta}}^*_{ml}$ is combined with $\hat{\sigma}^2_{ml}$ to yield $\hat{\boldsymbol{\theta}}_{ml}$ and subsequently $\hat{\boldsymbol{\beta}}_{ml}$. A similar result holds for REML estimation. Profiling σ^2 out of Equation 2 yields

$$-2l^*_R(\boldsymbol{\theta}^*|\mathbf{y}) = \ln|\boldsymbol{\Sigma}^*| + (n-p)\ln[(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}})^\intercal \boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}})] + (n-p) + (n-p)\ln 2\pi/(n-p). \tag{4}$$

After finding $\hat{\boldsymbol{\theta}}^*_{reml}$ a closed form solution for $\hat{\sigma}^2_{reml}$ exists: $\hat{\sigma}^2_{reml} = [(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}})^\intercal \boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}})]/(n-p)$. Then $\hat{\boldsymbol{\theta}}^*_{reml}$ is combined with $\hat{\sigma}^2_{reml}$ to yield $\hat{\boldsymbol{\theta}}_{reml}$ and subsequently $\hat{\boldsymbol{\beta}}_{reml}$.

Both ML and REML estimation rely on the an $n \times n$ covariance matrix inverse. Inverting an $n \times n$ matrix is an enormous computational demand that scales cubically with the sample size. For this reason, ML and REML have historically been unfeasible to implement in their standard form with data larger than a few thousand observations. This motivates the use for the big data approaches outlined in Section (INSERT SECTION).

It is worth noting that the inverses themselves are not strictly needed for estimation (or prediction), though at least their square root is needed. In spmodel, calculating this square root requires a Cholesky decomposition, which still scales cubically with the sample size. Computing the Cholesky decomposition, however, is far more computationally efficient than computing the inverse. To see why only the Cholesky decomposition is needed, recall that the Cholesky decomposition of the covariance matrix $\boldsymbol{\Sigma}$ is $\mathbf{C}\mathbf{C}^\intercal$, where $\mathbf{C}$ is a lower triangular matrix (so $\mathbf{C}\mathbf{C}^\intercal = \boldsymbol{\Sigma}$). In the ML and REML likelihoods, $\boldsymbol{\Sigma}^{-1}$ is not needed on its own, only $\mathbf{X}^\intercal\boldsymbol{\Sigma}^{-1}\mathbf{X}$ and $\mathbf{X}^\intercal\boldsymbol{\Sigma}^{-1}\mathbf{y}$ are needed. We can rewrite the $\mathbf{X}^\intercal\boldsymbol{\Sigma}^{-1}\mathbf{X}$ as $\mathbf{X}^\intercal(\mathbf{C}^\intercal)^{-1}\mathbf{C}^{-1}\mathbf{X} = (\mathbf{C}^{-1}\mathbf{X})^\intercal\mathbf{C}^{-1}\mathbf{X}$. Then $\mathbf{C}^{-1}\mathbf{X}$ is efficiently solved by noticing that $\mathbf{C}^{-1}\mathbf{X} = \mathbf{A}$ for some matrix $\mathbf{A}$ implies $\mathbf{X} = \mathbf{C}\mathbf{A}$. This system can be efficiently solved for $\mathbf{A}$ using linear forward solves (forward substitution). Then $\mathbf{X}^\intercal\boldsymbol{\Sigma}^{-1}\mathbf{X} = \mathbf{A}^\intercal\mathbf{A}$. A similar approach is used to solve $\mathbf{X}^\intercal\boldsymbol{\Sigma}^{-1}\mathbf{y}$. Still, using Cholesky decomositions is unfeasible for sample sizes larger than a few thousand observations.

2.2 Semivariogram-based Estimation

An alternative approach to likelihood-based estimation is semivariogram-based estimation. The semivariogram of a constant-mean process $\mathbf{y}$ is the expectation of the squared half-difference between two observations h distance units apart. More formally, the semivariogram is denoted $\gamma(h)$ and defined as

$$\gamma(h) = \mathrm{E}(y_i - y_j)^2/2, \tag{5}$$

where $||y_i - y_j||_2 = h$ (the Euclidean distance). When the process $\mathbf{y}$ is second-order stationary, the semivariogram and covariance function are intimately connected: $\gamma(h) = \mathrm{Cov}(0) - \mathrm{Cov}(h)$, where $\mathrm{Cov}(0)$ is the covariance function evaluated at 0 (which is the overall variance, σ^2) and $\mathrm{Cov}(h)$is the covariance function evaluated at h. Both semivariogram approaches described next are more computationally efficient than ML or REML because their major computational burden (calculations based on pairs) scale the squared sample size (i.e., not the cubed sample size).

2.2.1 Weighted Least Squares

The empirical semivariogram is a moment-based estimate of the semivariogram denoted by $\hat{\gamma}(h)$ and defined as

$$\hat{\gamma}(h) = \frac{1}{2|N(h)|}\sum_{N(h)}(y_i - y_j)^2, \tag{6}$$

where $N(h)$ is the set of observations in $\mathbf{y}$ that are h units apart (distance classes) and $|N(h)|$ is the cardinality of $N(h)$ (Cressie 1993). Often the set $N(h)$ contains observations that are $h \pm \alpha$ apart – this approach is known as "binning" the empirical semivariogram. Typically, only certain h considered when constructing Equation 6 – a commonly used cutoff is to ignore h larger than half the maximum distance in the domain. One criticism of the empirical semivariogram is that distance bins and cutoffs tend to be artitrarily chosen (i.e., not chosen according to some statistical criteria).

w_i Name	w_i Form	`weight =`
Cressie	$\lvert N(h)\rvert/\gamma(h)_i^2$	`"cressie"`
Cressie (Denominator) Root	$\lvert N(h)\rvert/\gamma(h)_i$	`"cressie-droot"`
Cressie No Pairs	$1/\gamma(h)_i^2$	`"cressie-nopairs"`
Cressie (Denominator) Root No Pairs	$1/\gamma(h)_i$	`"cressie-droot-nopairs"`
Pairs	$\lvert N(h)\rvert$	`"pairs"`
Pairs Inverse Distance	$\lvert N(h)\rvert/h^2$	`"pairs-invd"`
Pairs Inverse (Root) Distance	$\lvert N(h)\rvert/h$	`"pairs-invsd"`
Ordinary Least Squares	1	`ols`

Table 1: spmodel table weights

Equation (6) is viewed as the average squared half-distance between two observations in $\mathbf{y}$. Cressie (1985) proposed estimating $\boldsymbol{\theta}$ by minimizing an objective function that involves γh and $\hat{\gamma}(h)$ and is based on a weighted least squares criterion. This criterion is defined as

$$\sum_i w_i[\hat{\gamma}(h)_i - \gamma(h)_i]^2, \tag{7}$$

where w_i, $\hat{\gamma}(h)_i$, and $\gamma(h)_i$ are the weights, empirical semivariogram, and semivariogram for the ith distance class. Minimizing Equation (7) yields $\hat{\boldsymbol{\theta}}_{wls}$, the semivariogram weighted least squares estimates of $\boldsymbol{\theta}$. After estimating $\boldsymbol{\theta}$, $\boldsymbol{\beta}$ estimates are constructed using generalized least squares: $\hat{\boldsymbol{\beta}}_{wls} = (\mathbf{X}^\intercal\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X})^{-1}\mathbf{X}^\intercal\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{y}$, where $\hat{\boldsymbol{\Sigma}}^{-1}$ is $\boldsymbol{\Sigma}$ evaluated at $\hat{\boldsymbol{\theta}}_{wls}$.

Cressie (1985) recommends setting the w_i in Equation (7) as $w_i = |N(h)|/\gamma(h)_i^2$, which gives more weights to distance classes with more observations ($|N(h)|$) and semivariances at shorter distances ($1/\gamma(h)_i^2$). The default in spmodel is to use these w_i – the type of w_i is changed via the `weights` argument to `splm()`. Table 2.2.1 contains all w_i available in spmodel.

Additionally, the number of $N(h)$ classes and maximum distance for h are specified by passing the `bins` and `cutoff` arguments to `splm()` (these arguments are passed via `...` to `esv()`). The default value for `bins` is 15 and the default value for the maximum h is half the maximum distance of the spatial domain's bounding box.

Recall that the semivariogram is defined for a constant-mean process. Generally, $\mathbf{y}$ does not necessarily have a constant mean. So the empirical semivariogram and $\hat{\boldsymbol{\theta}}_{wls}$ are typically constructed using the residuals from an ordinary least squares regression of $\mathbf{y}$ on $\mathbf{X}$ – these residuals are assumed to have mean zero.

After

2.2.2 Composite Likelihood

The composite likelihood approach involves constructing a likelihood based on conditional or marginal events for which log-likelihoods are available and then adding together these individual components. Composite likelihoods are attractive because they behave very similar to likelihoods but are easier to handle, both from a theoretical and a computational perspective. Curriero and Lele (1999) derive a particular composite likelihood for estimating semivariogram parameters. The negative log of this composite likelihood, denoted $\mathrm{CL}(h)$, is given by

$$\mathrm{CL}(h) = \sum_{i=1}^{n-1}\sum_{j>i}\left(\frac{(y_i - y_j)^2}{2\gamma(h)} + \ln(\gamma(h))\right) \tag{8}$$

where $\gamma(h)$ is the semivariogram (that depends on parameter vector $\boldsymbol{\theta}$). Minimizing Equation 8 yields $\hat{\boldsymbol{\theta}}_{cl}$, the semivariogram composite likelihood estimates of $\boldsymbol{\theta}$. After estimating $\boldsymbol{\theta}$, $\boldsymbol{\beta}$ estimates are constructed using generalized least squares: $\hat{\boldsymbol{\beta}}_{cl} = (\mathbf{X}^\intercal\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X})^{-1}\mathbf{X}^\intercal\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{y}$, where $\hat{\boldsymbol{\Sigma}}^{-1}$ is $\boldsymbol{\Sigma}$ evaluated at $\hat{\boldsymbol{\theta}}_{cl}$.

An advantage of the composite likelihood approach to semivariogram estimation is that it does not require arbitrarily specifying empircial semivariogram bins and cutoffs. It does tend to be more computationally demanding than the weighted least squares, however, as the composite likelihood is constructed from $\binom{n}{2}$ pairs for a sample size n, whereas the weighted least squares approach only requires calculating $\binom{|N(h)|}{2}$ pairs for each distance bin $N(h)$. As with the weighted least squares approach, Equation 8 requires constant-mean process, so typically the residuals from an ordinary least squares regression of $\mathbf{y}$ on $\mathbf{X}$ are used to estimate $\boldsymbol{\theta}$.

3 Hypothesis Testing

3.1 The General Linear Hypothesis Test

Hypothesis tests for each element in $\hat{\boldsymbol{\beta}}$ are available in `summary()`.

```
R> spmod <- splm(y ~ x + group, exdata, "exponential", xcoord, ycoord)
R> summary(spmod)


Call:
splm(formula = y ~ x + group, data = exdata, spcov_type = "exponential",
    xcoord = xcoord, ycoord = ycoord)

Residuals:
     Min       1Q   Median       3Q      Max 
-1.91090 -0.38965  0.07808  0.56560  2.29718 

Coefficients (fixed):
            Estimate Std. Error z value Pr(>|z|)  
(Intercept) -0.65261    0.38911  -1.677   0.0935 .
x           -0.07809    0.07357  -1.061   0.2885  
group2       0.29110    0.20997   1.386   0.1656  
group3       0.37734    0.20857   1.809   0.0704 .
group4       0.32637    0.21531   1.516   0.1296  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Pseudo R-squared: 0.04864

Coefficients (spatial covariance):
    de     ie  range 
0.6330 0.3791 1.2350 

Spatial covariance type: exponential
```

Test statistics are the ratio of the estimate to its standard error. These tests are useful for continuous predictors, but for categorical predictors, `summary()` only provides test statistics for each level, not all levels simultaneously. To find a test statistic for a categorical variable, there are two approaches. The first is to use `anova()`.

```
R> anova(spmod)

Analysis of Variance Table

Response: y
                Df     X2 Pr(>X2)
(Intercept) 1.0000 2.8129  0.0935
x           1.0000 1.1267  0.2885
group       3.0000 3.8742  0.2754
```

Test statistics from `anova()` are formed using the general linear hypothesis test. Let $\mathbf{L}$ be an $l \times p$ contrast matrix and l_0 be an $l \times 1$ vector. The null hypothesis is that $\mathbf{L}\hat{\boldsymbol{\beta}} = l_0$ and the alternative hypothesis is that $\mathbf{L}\hat{\boldsymbol{\beta}} \neq l_0$. Usually, l_0 is the zero vector (and in spmodel, this is assumed). The test statistic, denoted X^2, is given by

$$X^2 = [(\mathbf{L}\hat{\boldsymbol{\beta}} - l_0)^\intercal (\mathbf{L}(\mathbf{X}^\intercal \hat{\boldsymbol{\Sigma}} \mathbf{X})^{-1} \mathbf{L}^\intercal)^{-1} (\mathbf{L}\hat{\boldsymbol{\beta}} - l_0)] / rank(\mathbf{L}) \tag{9}$$

It is notoriously difficult to determine appropriate p-values for linear mixed models based on the general linear hypothesis test. lme4, for example, does not report p-values by default. There are three reasons we focus on next that explain why obtaining p-values is so challenging.

- The first (and often most important) challenge is that when estimating $\boldsymbol{\theta}$, it is usually not clear what the null distribution of the test statistic is. In certain cases such as ordinary least squares regression or certain experimental designs (e.g., blocked design, split plot design, etc.), Equation 9 is F-distributed (the test statistic is known) with known numerator and denominator degrees of freedom. But outside of these well-studied cases, no general results exist.
- The second challenge is that the standard error of the test statistic does not account for the uncertainty in $\hat{\boldsymbol{\theta}}$. For some approaches to addressing this problem, see Kackar and Harville (1984), Prasad and Rao (1990), Harville and Jeske (1992), and Kenward and Roger (1997).
- The third challenge is in determining denominator degrees of freedom. Again, in certain cases, these are known – but this is not true in general. For some approaches to addressing this problem, see Satterthwaite (1946), Schluchter and Elashoff (1990), Hrong-Tai Fai and Cornelius (1996), Kenward and Roger (1997), Littell et al. (2006), Pinheiro and Bates (2006), and Kenward and Roger (2009).

For these reasons, spmodel assumes a large-sample, Chi-squared approximation when calculating p-values. This approach addresses the three points above by assuming that with a large enough sample size:

- The numerator in Equation 9 tends to be asymptotically Chi-squared (under certain conditions) with $rank(\mathbf{L})$ degrees of freedom.
- The uncertainty from estimating $\hat{\boldsymbol{\theta}}$ is small enough to be safely ignored.

Because the approximation is asymptotic, degree of freedom adjustments can be ignored (it is also worth noting that an F distribution with infinite denominator degrees of freedom is a scaled Chi-squared distribution). A takeaway here is that this asymptotic approximation implies these p-values are likely unreliable with small samples.

A second approach to determining p-values is a likelihood ratio test for nested models. Let $l(\hat{\boldsymbol{\theta}}_0, \hat{\boldsymbol{\beta}}_0|\mathbf{y})$ be the log-likelihood from some reduced model and $l(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\beta}}_1|\mathbf{y})$ be the log-likelihood from some full model. When the reduced model is nested in the full model (i.e., the reduced model can be obtained by fixing some parameters of the full model), $[-2l(\hat{\boldsymbol{\theta}}_0, \hat{\boldsymbol{\beta}}_0|\mathbf{y})] - [-2l(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\beta}}_1|\mathbf{y})]$ is asymptotically Chi-squared with degrees of freedom equal to the difference in estimated parameters between the full and reduced model. To see whether there is evidence of spatial covariance, run

```
R> spmod_r <- splm(y ~ x, exdata, "none")
R> spmod_f <- splm(y ~ x, exdata, "exponential", xcoord, ycoord)
R> anova(spmod_r, spmod_f)

Likelihood Ratio Test

Response: y
           Df     X2 Pr(>X2)
1 vs 2  2.000 20.139       0
```

This output suggests evidence of spatial covariance. Because the likelihood relies on minimized likelihoods, they are only defined for ML or REML estimation. Furthermore, for REML estimation, likelihood ratio tests can only be used to compare nested models whose fixed effect structure does not change. This is because the REML likelihood (2) depends on the fixed effects through $\ln|\mathbf{X}^\intercal\boldsymbol{\Sigma}^{-1}\mathbf{X}|$. To use likelihood ratio tests for assessing the importance of fixed effects, parameters must be estimated using REML.

3.2 Contrasts

4 Random Effects

4.1 BLUPs

5 The Local list

References

Cressie, Noel. 1985. "Fitting Variogram Models by Weighted Least Squares." *Journal of the International Association for Mathematical Geology* 17 (5): 563–86.

———. 1993. *Statistics for Spatial Data*. John Wiley & Sons.

Curriero, Frank C, and Subhash Lele. 1999. "A Composite Likelihood Approach to Semivariogram Estimation." *Journal of Agricultural, Biological, and Environmental Statistics*, 9–28.

Harville, David A. 1977. "Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems." *Journal of the American Statistical Association* 72 (358): 320–38.

Harville, David A, and Daniel R Jeske. 1992. "Mean Squared Error of Estimation or Prediction Under a General Linear Model." *Journal of the American Statistical Association* 87 (419): 724–31.

Hrong-Tai Fai, Alex, and Paul L Cornelius. 1996. "Approximate F-Tests of Multiple Degree of Freedom Hypotheses in Generalized Least Squares Analyses of Unbalanced Split-Plot Experiments." *Journal of Statistical Computation and Simulation* 54 (4): 363–78.

Kackar, Raghu N, and David A Harville. 1984. "Approximations for Standard Errors of Estimators of Fixed and Random Effects in Mixed Linear Models." *Journal of the American Statistical Association* 79 (388): 853–62.

Kenward, Michael G, and James H Roger. 1997. "Small Sample Inference for Fixed Effects from Restricted Maximum Likelihood." *Biometrics*, 983–97.

———. 2009. "An Improved Approximation to the Precision of Fixed Effects from Restricted Maximum Likelihood." *Computational Statistics & Data Analysis* 53 (7): 2583–95.

Littell, Ramon C, George A Milliken, Walter W Stroup, Russell D Wolfinger, and Schabenberber Oliver. 2006. *SAS for Mixed Models*. SAS publishing.

Patterson, H Desmond, and Robin Thompson. 1971. "Recovery of Inter-Block Information When Block Sizes Are Unequal." *Biometrika* 58 (3): 545–54.

Pinheiro, José, and Douglas Bates. 2006. *Mixed-Effects Models in S and S-Plus*. Springer science & business media.

Prasad, NG Narasimha, and Jon NK Rao. 1990. "The Estimation of the Mean Squared Error of Small-Area Estimators." *Journal of the American Statistical Association* 85 (409): 163–71.

Satterthwaite, Franklin E. 1946. "An Approximate Distribution of Estimates of Variance Components." *Biometrics Bulletin* 2 (6): 110–14.

Schluchter, Mark D, and Janet T Elashoff. 1990. "Small-Sample Adjustments to Tests with Unbalanced Repeated Measures Assuming Several Covariance Structures." *Journal of Statistical Computation and Simulation* 37 (1-2): 69–87.

Wolfinger, Russ, Randy Tobias, and John Sall. 1994. "Computing Gaussian Likelihoods and Their Derivatives for General Linear Mixed Models." *SIAM Journal on Scientific Computing* 15 (6): 1294–1310.