MMANOVA: A general multilevel framework for multivariate analysis of variance

Steven Geinitz  
geinitz@math.uzh.ch  
University of Zurich

Reinhard Furrer  
furrer@math.uzh.ch  
University of Zurich

Stephan R. Sain  
ssain@ucar.edu  
National Center for Atmospheric Research

Abstract Classical analysis of variance requires that model terms be labeled as fixed or random and typically culminate by comparing variability from each batch (factor) to variability from errors; without a standard methodology to assess the magnitude of a batch’s variability, to compare variability between batches, nor to consider the uncertainty in this assessment. In this paper we support recent work, placing ANOVA into a general multilevel framework, then refine this through batch level model specifications, and develop it further by extension to the multivariate case. Adopting a Bayesian multilevel model parametrization, with improper batch level prior densities, we derive a method that facilitates comparison across all sources of variability. Whereas classical multivariate ANOVA often utilizes a single covariance criterion, e.g. determinant for Wilks’ lambda distribution, the method allows arbitrary covariance criteria to be employed. The proposed method also addresses computation. By introducing implicit batch level constraints, which yield improper priors, the full posterior is efficiently factored, thus alleviating computational demands. For a large class of models, the partitioning mitigates, or even obviates the need for methods such as MCMC. The method is illustrated with simulated examples and an application focusing on climate projections with global climate models.

Keywords: Bayesian inference; Constraints; Mixed model; Variance components
1 Introduction

Identifying and comparing variability among several factors is a fundamental task of statistical analysis. From initial exploratory steps, to model testing, analysis of variance plays a vital role in the practice of statistics. Gelman (2005) has outlined a general ANOVA methodology that fits a wide range of models and summarizes results in a manner that facilitates interpretation across different sources of variation. Gelman and Hill (2006) elaborate further, describing the methodology in terms of a multilevel model. One important contribution of this approach is in providing summaries that are more constructive than conclusions based on hypothesis tests. This framework is seeing usage by researchers from diverse fields e.g. ecology (Qian and Shen, 2007), genetics (Leinonen et al., 2008), and climate (Sain et al., 2011).

This paper presents a method that adopts the multilevel approach towards ANOVA, then extends it to multivariate settings, yielding a multilevel multivariate analysis of variance (MMANOVA) methodology. The strategy of initially treating all sources of variability in similar regard is naturally a part of the new method. We address known issues of applying constraints in variance analyses (Nelder, 1977, 1994, 1999) by constraining batch levels so that prior distributions are improper. The extension is seen as a valuable contribution, since multivariate ANOVA can further obfuscate model specification, interpretation, and computation. Kaufman and Sain (2010) offer an analysis of variance procedure that aids in model specification and interpretation, but requires computationally demanding MCMC steps. Recent methods involving approximations, e.g. integrated nested Laplace approximations (INLA) (Lindgren et al., 2011; Rue et al., 2009), have allowed for MCMC to be eliminated in many cases, significantly reducing computational demands. While the range of problems for which such methods are applicable is wide, the focus is not typically on variance parameters. Thus, the contribution of our work is in promoting a general ANOVA methodology. We accomplish this by supporting recent work with the same goal, by refining the model specification, and by extending this to multivariate cases.
In Section 2 we review the ANOVA formulation of Gelman (2005) and summarize some details of the approach. We then formally extend the idea to the multivariate case, discussing technical and computational aspects. Section 3 provides a demonstration of the method through a simulation example and through a climatological application using data from future climate projections given by several atmosphere-ocean general circulation models and global emissions scenarios. In Section 4 we discuss extensions and further computational benefits that are possible when the method is applied to common high-dimensional problems.

2 Analysis of Variance

Analysis of variance is widely accepted as a means of partitioning variability in a manner which allows it to be attributed to various factors. An important initial step in the analysis is considering each factor of the model to be fixed or random. This step, necessary in the classical setting, raises enough issues that statisticians have been obliged to address the “mixed models controversy” (Lencina et al., 2005; Voss, 1999). One might conclude that a consensus has still not been reached, given that John Nelder deemed it necessary to reiterate the requisite points of constraints and marginality over such a long period of time, beginning with Nelder (1977) and most recently Nelder (2008).

The rest of the section outlines a recent attempt by Gelman (2005) towards a more universal ANOVA methodology, then refines and extends the approach to a multivariate context.

2.1 Multilevel ANOVA

A fundamental contribution of the hierarchical regression approach to ANOVA employed by Gelman (2005) has been to indiscriminately consider all components in a model as random, thereby facilitating comparison across all sources of variability. The terminology is useful in supporting the indiscriminate nature of the method. The word batch is applied to all terms in the model, e.g. overall mean, factors, nested terms, interactions,
etc. The nature of the variability from a batch is further distinguished. The distinction traditionally made by random and fixed effects is instead addressed by considering a batch’s super and finite population variance. We now summarize the recent shift in ANOVA as a methodology in terms of a univariate linear model.

2.1.1 Model Parametrization

Following the notation of Gelman (2005), observations $Y_i, i = 1, \ldots, n$ are stated in terms of the additive decomposition

$$Y_i = \sum_{b=0}^{B} \beta^{(b)}_j,$$

or the alternative regression formulation

$$Y_i = \sum_{b=0}^{B} \sum_{j=1}^{n_b} x_{i,j}^{(b)} \beta^{(b)}_j,$$

with $\beta^{(b)}_j$ denoting individual batch levels and $x_{i,j}$ denoting explanatory variables. The regression formulation could be used for the additive decomposition, with explanatory variables set to either 0 or 1. Batch indices $b = 0$ and $b = B$ will often correspond to an overall mean, $\mu$, and to measurement errors, $\epsilon_i$, respectively, so that $n_0 = 1$, $n_B = n$. An individual batch is referenced by $b = 0, \ldots, B$ and consists of $n_b$ levels. Individual levels of a batch are denoted as $\beta^{(b)}_1, \ldots, \beta^{(b)}_{n_b}$ with $j^b_1, \ldots, j^b_{n_b}$ replicating the levels so that each observation is associated with exactly one batch level. We acknowledge that the additional level of subscripts and superscripts may seem contrived to some, although it is necessary for the general case. In practice the number of batches in the model is reasonable so that this can be avoided, as done in Section 3. Additional sub or superscripts can often be dropped. For example, a batch in its entirety is denoted by $\beta^{(b)} = \{\beta^{(b)}_1, \ldots, \beta^{(b)}_{n_b}\}$.

Given a batch, and an assumed distribution on model errors, a conventional fixed effects analysis often corresponds to the test $H_0: \beta^{(b)}_j = 0$ for $j = 1, \ldots, n_b$. While for a random effects model, assuming the $n_b$ levels of each batch $b$ to be modeled as Gaussian

$$\beta^{(b)}_j \mid \sigma^2_b \sim N(0, \sigma^2_b), \quad j = 1, \ldots, n_b,$$

(3)
a test for significant batch variation would be $H_0: \sigma_b^2 = 0$. Alternatively, the proposed methodology identifies two representations of variation of a given batch. The superpopulation variance, $\sigma_b^2$, corresponds to the variance of all potential, possibly infinitely many, levels of a batch. The finite-population variance represents variability of the specific set of batch levels that have been realized. Super and finite-population variances can be roughly related to the random effect variance component estimate, and the fixed effect within-group sum of squares, respectively. As an example, consider batch $b$ and its vector of batch levels, $\beta^{(b)} = (\beta^{(b)}_1, \ldots, \beta^{(b)}_{n_b})^T$ with $c_b$ constraints. Then the degrees of freedom are $\nu_b = n_b - c_b$, and the finite-population variance $s_b^2$ is $s_b^2 = \frac{1}{\nu_b} \beta^{(b)}^T (I_{n_b} - C_b^T (C_b C_b^T)^{-1} C_b) \beta^{(b)}$, where $I_{n_b}$ is the $n_b \times n_b$ identity and $C_b$ is the $c_b \times n_b$ constraint matrix such that $C_b \beta^{(b)} = 0$. Variance component estimation is made by decomposing the variance of the batch level estimates, $V_b = \text{Var}(\hat{\beta}^{(b)}) = \text{Var}(\{\text{E}(\hat{\beta}^{(b)} | \beta^{(b)})\}) + \text{E}\{\text{Var}(\hat{\beta}^{(b)} | \beta^{(b)})\}$, into the sum of the superpopulation variation, $\sigma_b^2$, plus the variability of the batch level estimations, $V_{\text{estimation}}$. The chosen estimate of the superpopulation variance, in this case the method-of-moments estimator, is then $\hat{\sigma}_b^2 = \hat{V}_b - \hat{V}_{\text{estimation}}$, where $\hat{V}_b = \frac{1}{\nu_b} \sum_{j=1}^{n_b} (\hat{\beta}_j^{(b)})^2$, and $\hat{V}_{\text{estimation}} = \sum_{k \in I(b)} \frac{n_k}{n_b} \hat{\sigma}_k^2$ includes superpopulation variances from other batches that enter into variability of the batch level estimates, indicated by the set $I(b)$. At a minimum, $\hat{V}_{\text{estimation}}$ will include $\hat{\sigma}_B^2$, the estimated error variance. For a large class of multilevel, hierarchical models, this strategy allows for all terms to be treated as random, and for their variabilities to be assessed.

2.1.2 Confirmatory Procedures

In regards to more inferential procedures, either a frequentist or Bayesian direction can be taken. In the frequentist case, an inverse-chi-square distribution, $\chi^2_{\nu_b}$, is employed to assess uncertainty in the superpopulation variance $\sigma_b^2$; since $\frac{1}{\nu_b} s_b^2 / \sigma_b^2$ is chi-square distributed with $\nu_b$ degrees of freedom. For batches with $\sum_{k \in I(b)} \frac{n_k}{n_b} \hat{\sigma}_k^2$, including more than only the error variance $\hat{\sigma}_B^2$, then a linear combination of inverse-chi-square distributions, $\sum_i m_i \chi^2_{\nu_i}$, is required, as described at the end of the previous section. As Gelman (2005)
states, these linear combinations may be dealt with directly, although simulation is often more straightforward. The simulation, described therein, is carried out by: 1) Obtain $R$ simulated raw variances, $V_b$, for each of the $B$ batches in the model with a random variable that is proportional to $\chi^2$ and corresponding degrees of freedom; 2) Calculate superpopulation variances using $\tilde{\sigma}_b^2 = \max(0, V_b - V_{\text{estimation}})$; 3) Simulate batch levels using newly generated superpopulation variances; 4) Calculate sample variances of each batch. This procedure then yields a (posterior) sample of superpopulation variances, batch levels, and finite-population variances, corresponding to the final three steps.

A strict Bayesian approach requires additional prior specifications, but yields posterior distributions of the superpopulation variances. However, this distinction, between the two schools of thought, can be seen as purely semantic. As Gelman (2005) states, “given $\sigma_b^2$, the parameters $\beta_j^{(b)}$ have a multivariate normal distribution (in Bayesian terms, a conditional posterior distribution; in classical terms, a predictive distribution)”. Thus, assessing uncertainty in the finite-population variances, $s_b^2$, is the same. In both cases, either batch levels themselves are simulated, or the distribution of $s_b^2$ is approximated with an appropriate chi-square random variable.

It should be clear that the uncertainty surrounding superpopulation variance parameters will typically be greater than that for finite-population variances. Intuitively, this is because superpopulation variances describe variability of levels that have not yet been realized.

### 2.2 Multilevel Multivariate ANOVA

Typical multivariate analysis of variance strategies rely on the distribution of the determinant of sums of squares matrices, i.e. Wilks’ lambda distribution (Mardia et al., 1979, p. 335), and culminate in $p$-value related conclusions. Hence, it does not easily facilitate inclusion of random effects, and thus no comparison across these effects. Using a Bayesian approach, we now derive a general multivariate methodology that seeks to provide results similar to those of Section 2.1.1. The method partitions variability by
batch in an efficient manner and further factors the posterior by batch into a batch’s superpopulation covariance posterior, and batch levels posteriors. In addition, we handle the issue of constraints in a way that does not commit what Nelder (1994) has called one of the false steps of linear models. Rather, the constraints are implicit, yielding improper batch level prior distributions. Another point which must be mentioned is that of matrix parameter estimation. Although the common limitations of covariance matrix estimation and modeling are issues that must be dealt with, it is important to first focus on, and refine multivariate ANOVA for familiar settings. In Section 4 we discuss issues, modifications, and implications of the method when higher dimensional data is used.

2.2.1 Multivariate Model Parametrization

Consider \(d\)-dimensional multivariate observations such that batch levels are vectors and batch variances are covariance matrices. Namely, (1)–(3) now contain vectors \(Y_i, \beta_j(b)\), matrices \(X_{i,j}\), and covariance matrices are \(\Sigma_b\), all of appropriate dimension. We adopt the strategy from the previous section, in indiscriminately considering all terms as potentially possessing variability. The multivariate analogue of (1) with Gaussian errors is

\[
Y_i \mid \{ \beta_j(b) \}_b, \Sigma_e \sim \mathcal{N}_d \left( \sum_{b=0}^{B-1} \beta_j(b), \Sigma_e \right), \quad i = 1, \ldots, n, \tag{4}
\]

and the remainder of the Bayesian model specification is given by

\[
\beta_j(b) \mid \Sigma_b \sim \mathcal{N}_d(\beta_j(b), \Sigma_b), \quad j = 1, \ldots, n_b, \tag{5}
\]

\[
U_b = \Sigma_b + \frac{n_b}{n} \Sigma_e \mid \Sigma_e \sim W^{-1}(\Psi_b, \kappa_b), \tag{6}
\]

\[
\Sigma_e \sim W^{-1}(\Psi_e, \kappa_e), \tag{7}
\]

for \(b = 0, \ldots, B - 1\) and with the inverse-Wishart distribution denoted by \(W^{-1}\). Batch indices \(b = 0\) and \(b = B\) respectively correspond to the intercept, or overall mean term, \(\mu\), and to measurement errors, \(e_i\). For notational convenience we will refer to \(\Sigma_\mu\) and \(\Sigma_e\), rather than \(\Sigma_0\) and \(\Sigma_B\). Typically zero-mean batch level priors are assumed; that is, \(\beta_0(b) = 0\). Setting \(\Psi_b = 0, \kappa_b = 0\) yields the noninformative prior \(p(U_b) \propto |U_b|^{-(d+1)/2}\).
Because inverse-Wishart support is given by the set of all positive definite matrices, (6) is referred to as a constrained inverse-Wishart distribution, since $U_b - \frac{n_b}{n} \Sigma_e$ is required to be positive definite. This covariance parametrization, $U_b = \Sigma_b + \frac{n_b}{n} \Sigma_e$, has previously been utilized in the context of multivariate random effects by Everson and Morris (2000), for which they develop efficient methods of sampling. Also, when only error terms contribute to the variance of batch level estimates, $U_b$ is analogous to $V_b$ of Section 2.1.1 with $I(b) = \{B\}$.

The choice of inverse-Wishart covariance priors, (6) and (7), has been made to balance the complexity of model specification with implementation and computation. However, given the considerable amount of research of covariance priors, there are other options available. Daniels (1999) and Daniels and Kass (2001) examine covariance priors that emphasize uniform shrinkage of the eigenvalues. Although informative covariance priors may be necessary in many cases, the usage of such priors will impact the computational demands required.

2.2.2 Posterior Distributions

Without additional specification, (4)–(7) yield an inadequate posterior. In an MCMC setting this may manifest itself by failure to converge, due to drifting in the parameter space. From a classical point of view, estimating the set of all batch levels, $\{\beta_j(b)\}$ would require additional constraints. The inclusion of similar constraints in the Bayesian model allows for a closed form of the posterior, as well as for factorization between batches. Degrees of freedom for each batch are then accounted for in the corresponding batch covariance posterior. This parametrization is also beneficial in terms of computation since batches are conditionally independent of one another. Using a vectorized form of the model, $Y = (Y_1^T, \ldots, Y_n^T)^T \in \mathbb{R}^{nd}$ and $\beta(b) = (\beta_1(b)^T, \ldots, \beta_n(b)^T)^T \in \mathbb{R}^{nb_d}$, is convenient for the development. The constraint $C_b \beta(b) = 0 \in \mathbb{R}^{c_b d}$, where there are $c_b$ constraints, combined with (4), (5), is now
\[
Y \mid \{\beta^{(b)}\}_{b}, \mathbf{C}_b \beta^{(b)} = 0, \Sigma_{\epsilon} \sim \mathcal{N}_d \left( \sum_{b=0}^{B-1} \beta^{(b)}, \mathbf{I}_n \otimes \Sigma_{\epsilon} \right), \quad (8)
\]

\[
\beta^{(b)} \mid \mathbf{C}_b \beta^{(b)} = 0, \Sigma_b \sim \mathcal{N}_{nb, 0}(\mathbf{1}_{nb} \otimes \beta^{(b)}_0, \bar{\Omega}_b), \quad (9)
\]

with \( \otimes \) denoting the Kronecker product. The rank-deficient \( \bar{\Omega}_b \), due to the constraint, causes (9) to be improper. To derive this improper distribution begin with the unconstrained and vectorized form of (5), which has covariance \( \Omega_b = \mathbf{I}_{nb} \otimes \Sigma_b \). The density is then stated through a decomposition of the precision, \( Q_b = \Gamma \Lambda \Gamma^T \otimes \Sigma_b^{-1} \). Assuming \( c_b \) constraints, the rank deficiency is introduced by removing the corresponding number of eigenvalues from the diagonal matrix \( \Lambda \), e.g. \( \bar{\Lambda} = \text{diag}(0, \ldots, 0, \lambda_{c_b+1}, \ldots, \lambda_{nb}) \), leading to \( \bar{Q}_b = \Gamma \bar{\Lambda} \Gamma^T \otimes \Sigma_b^{-1} \). This method of addressing linear constraints is useful for other general improper distributions and intrinsic Gaussian Markov random fields, as illustrated by Rue and Held (2005). Let \( |Q|_* \) be the pseudo determinant of a singular matrix, that is, the product of its non-zero eigenvalues. In the case of the identity matrix being used as the first matrix term in the Kronecker product, the eigenvalues are one, thus densities of the likelihood, (8), and of batch level priors, (9) are

\[
p(Y \mid \{\beta^{(b)}_j\}_b, \Sigma_{\epsilon}) \propto |\Sigma_{\epsilon}|^{-n/2} \exp \left( -\frac{1}{2} \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^T \Sigma_{\epsilon}^{-1} (Y_i - \hat{Y}_i) \right) \]

\[
+ \left( -\frac{1}{2} \sum_{b=0}^{B-1} \sum_{j=1}^{nb} (\beta^{(b)}_j - \hat{\beta}^{(b)}_j)^T \frac{\mathbf{1}_{nb}}{n_b} \Sigma_{\epsilon}^{-1} (\beta^{(b)}_j - \hat{\beta}^{(b)}_j) \right),
\]

\[
p(\beta^{(b)} \mid \Sigma_b) \propto |\bar{Q}_b|_*^{1/2} \exp \left( -\frac{1}{2} (\beta^{(b)} - 1 \otimes \beta^{(b)}_0)^T \bar{Q}_b (\beta^{(b)} - 1 \otimes \beta^{(b)}_0) \right) \]

\[
\propto |\Sigma_b|^{-(nb-c_b)/2} \exp \left( -\frac{1}{2} \sum_{j=c_b+1}^{nb} (\beta^{(b)}_j - \beta^{(b)}_0)^T \Sigma_b^{-1} (\beta^{(b)}_j - \beta^{(b)}_0) \right),
\]

where \( \hat{\cdot} \) denotes least-squares estimates. For orthogonal batches, the full posterior from these densities and from batch covariance prior densities, can be conveniently factored into

\[
p(\Sigma_{\epsilon}, \{\Sigma_b, \beta^{(b)}_j\}_{b=0}^{B-1} \mid Y) = p(\Sigma_{\epsilon} \mid Y) \prod_{b=0}^{B-1} p(\Sigma_b, \beta^{(b)}_j \mid Y, \Sigma_{\epsilon}). \quad (10)
\]
Each joint batch density, \( p(\Sigma_b, \beta^{(b)} \mid Y, \Sigma_c) \), is then factored further. Using known matrix identities, e.g. \((A^{-1} + B^{-1})^{-1} = A(A + B)^{-1}B\), the identity
\[
(x - s)^T U^{-1}(x - s) + (x - t)^T V^{-1}(x - t) = \text{tr} \left( (U + V)^{-1}(s - t)(s - t)^T \right) + (x - P^{-1}m)^T P(x - P^{-1}m),
\]
is derived, where \( P = U^{-1} + V^{-1}, m = U^{-1}s + V^{-1}t \). Accounting for model constraints, together with (11), the batch superpopulation posterior and the batch levels are found through the decomposition of quadratic forms of batch levels and least squares estimates
\[
\text{tr} \left[ (\Sigma_\beta + \frac{n_b}{n} \Sigma_c)^{-1} B_\beta \right] + \sum_{j=1}^{c_n} (\beta_j^{(b)} - \hat{\beta}_j^{(b)})^T n_\beta^{-1} \Sigma_c^{-1} (\beta_j^{(b)} - \hat{\beta}_j^{(b)}) + \sum_{j=cn+1}^{n_b} (\beta_j^{(b)} - P_b^{-1} m_j^{(b)})^T P_b (\beta_j^{(b)} - P_b^{-1} m_j^{(b)}),
\]
where \( P_b = \Sigma_b^{-1} + \frac{n_b}{n} \Sigma_c^{-1}, m_j^{(b)} = \frac{n_b}{n} \Sigma_c^{-1} \beta_j^{(b)} + \Sigma_b^{-1} \beta_0^{(b)} \), and \( \text{tr}(\cdot) \) denoting the trace operator. Additionally, \( B_\beta = \sum_{j=cn+1}^{n_b} (\beta_j^{(b)} - \beta_0^{(b)})(\beta_j^{(b)} - \beta_0^{(b)})^T \), is analogous to a matrix sums of squares of the unconstrained batch level estimates that has been adjusted by the prior mean. The full joint posterior is then factored as
\[
p(\Sigma_c, \{ \Sigma_b, \beta^{(b)} \}_{b=0}^{B-1} \mid Y) = p(\Sigma_c \mid Y) \prod_{b=0}^{B-1} p(\Sigma_b \mid Y, \Sigma_c) p(\beta^{(b)} \mid Y, \Sigma_c, \Sigma_b),
\]
where the product denotes batch posterior independence, and thus no need for computationally intensive MCMC procedures. The corresponding distributions of (12) are
\[
\Sigma_c \mid Y \sim W^{-1} \left( \Psi_c + \sum_{i=1}^{n} (Y_i - \hat{Y}_i)(Y_i - \hat{Y}_i)^T, \kappa_c + n - \sum_{b=0}^{B-1} n_b \right),
\]
\[
\Sigma_b + \frac{n_b}{n} \Sigma_c \mid Y, \Sigma_c \sim W^{-1} \left( \Psi_b + B_b, \kappa_b + n_b - c_b \right),
\]
\[
\beta_j^{(b)} \mid Y, \Sigma_c, \Sigma_b \sim \begin{cases} N_d \left( \hat{\beta}_j^{(b)}, \frac{n_b}{n} \Sigma_c \right) & j = 1, \ldots, c_b, \\ N_d \left( P_b^{-1} m_j^{(b)}, P_b^{-1} \right) & j = c_b + 1, \ldots, n_b. \end{cases}
\]
Batch levels (15), which reflect both free and constrained parameter estimates, can then be sampled and adjusted accordingly to obtain posteriors for finite-population covariances, \( S_b \). Recall finite-population parameters focus on observed levels of a batch, not on all potential unobserved batch levels.
In all cases thus far covariances $\Sigma_b$ are assumed to be of full rank. Hence, improper covariance posteriors will be due only to an insufficient number of observed levels of the batch. Díaz-García et al. (1997) offer a comprehensive look at all possible cases of improper Wishart distributions and following their terminology this would be classified as a pseudo-inverse-Wishart. Uhlig (1994) as well as Srivastava (2003) consider sampling with a pseudo-singular-Wishart distribution. Extending their work to inverse-Wishart distributions is one method for dealing with moderate discrepancies in the number of observed levels, $n_b < d$. Addressing cases in which $n_b \ll d$ is discussed in Section 4. Even in the remaining case, $n_b \geq d$, simulation from a posterior is not always efficient. Because support of the inverse-Wishart posterior requires positive definiteness in two respects, $\Sigma_\beta > 0$, and $\Sigma_\beta - \frac{n_b}{n}\Sigma_\epsilon > 0$, the usual method of rejection sampling from an inverse-Wishart distribution is not always practical. Everson and Morris (2000) describe a more computationally efficient method to maintain positive definiteness through a Cholesky decomposition and maintaining positive eigenvalues while the sample realization is generated.

Multivariate sources of variability do not always yield a single, clear criterion that indicates the greatest contributor to overall variability. For scalar variance components, $\sigma_b^2 > \sigma_\epsilon^2$ is clearly interpreted, however, due to the partial ordering of positive definite matrices, the analogous statement on covariance matrices is not useful. In other words, there is not a single, obvious comparison that can be made to determine which of two covariances are “greater”. Depending on the setting, there may exist an adequate scalar that sufficiently summarizes covariance characteristics. For the volume of ellipsoidal contours the determinant, $|\Sigma|$, achieves this, while in other cases the sum of all entries, $1^T \Sigma 1$, or the sum of the marginal variances, $\text{tr}(\Sigma)$, may be appropriate. Scalar criteria with corresponding uncertainty intervals then allow multivariate sources of variability to
be directly compared. Mardia et al. (1979) employed the determinant and trace, which correspond to the product and sum of eigenvalues; referring to them as the generalized variance and total variance, respectively. We have amended the terminology, since inclusion of the sum of all matrix elements requires further delineation, by denoting $\mathbf{1}^T \Sigma \mathbf{1}$ as the total variance, and $\text{tr}(\Sigma)$ as the total marginal variance.

Effectively relaying results of analysis of variance is one of the motivating factors that Gelman (2005) cites. The classical table of $p$-values does not yield any indication of batches with the largest variances, nor are the required assumptions on other batches clear. Nelder (1999) discusses many of the issues of over-reliance upon the $p$-value, and its ineffectiveness as a tool for communicating results. Uncertainty intervals are the default choice for presenting results. Visual plots are convenient since they facilitate simultaneous comparison of the relative variability contributions, their magnitudes, and the magnitude of the uncertainty in the estimates. For direct comparison of batches $b$ and $b'$, statements of the form $P\left(g(\Sigma_b) > g(\Sigma_{b'}) \mid \mathbf{Y}\right)$, utilizing an arbitrary matrix criterion $g(\cdot)$, may also be found.

### 3 Examples

This section covers two examples of the outlined methodology to carry out confirmatory procedures on multivariate data. The first is a toy example in which output and summaries are given in order to provide further insight. The second example utilizes global averages of temperature and precipitation predictions using 13 atmosphere-ocean general circulation models and 3 global greenhouse gas emissions scenarios that have been identified by the Intergovernmental Panel on Climate Change.

#### 3.1 Simulation

The process $\mathbf{Y}_{ij} = \mathbf{\mu} + \mathbf{\alpha}_i + \mathbf{\epsilon}_{ij} \in \mathbb{R}^d$, i.e. $B = 2$, $d = 3$, is used to illustrate the method of Section 2.2. To generate the three-dimensional observations, parameters $\Sigma_{\alpha}, \Sigma_{\epsilon}$, and $\mathbf{\mu}$ are fixed. An individual simulation is then performed by generating $\mathbf{\alpha}_i$ and $\mathbf{\epsilon}_{ij}, i =$
1, …, \( n_\alpha \), \( j = 1, \ldots, n_\epsilon \), from mean-zero multivariate Gaussian distributions with their respective fixed covariances. The mean term \( \mu \) is added to generated data resulting in \( n = n_\alpha n_\epsilon \) observations.

Using (13)–(15), posterior distributions for these covariance criteria are obtained for three distinct simulation scenarios. The three scenarios can be explained by the ambiguous description that variability introduced by batch \( \alpha \) is greater than (case 1), less than (case 2), or comparable to (case 3) variability introduced by error batch \( \epsilon \). More specifically, covariance matrices are decomposed into a vector of marginal standard deviations \( s \) and a correlation matrix \( R \), e.g. \( \Sigma_\alpha = \text{diag}(s_\alpha) R_\alpha \text{diag}(s_\alpha) \). For all simulation correlation matrices are fixed. Batch levels \( \alpha_i \) have the unique correlation structure \( (R_\alpha)_{1,2} = 0.3, (R_\alpha)_{1,3} = 0.1, (R_\alpha)_{2,3} = 0.5 \). Errors \( \epsilon_{ij} \) have the correlation matrix, \( R_\epsilon \), with autocorrelation structure \( (R_\epsilon)_{i,j} = \rho^{|i-j|} \), and \( \rho = 0.2 \). Error marginal variances are additionally held constant at 1 over all simulations, \( \Sigma_\epsilon = R_\epsilon \), so that only \( s_\alpha \) is distinct for each case.

The objective of the analysis is to assess the relative variability introduced by batch \( \alpha \) and batch \( \epsilon \), as well as the uncertainty in the assessment. Further, this is to be done in an appropriate multivariate context. Figure 1 displays the results of two simulation runs under the first scenario (case 1) using the determinant. In one simulation, the number of batch level realizations are \( n_\alpha = 5, n_\epsilon = 3 \) and in the second \( n_\alpha = 8, n_\epsilon = 5 \), which is to say less vs. more data. The left-most graph of Figure 1 displays uncertainty intervals, with narrow lines denoting 95% quantiles, thicker lines 50%, and a vertical tick placed at the median. The upper set of intervals, which are intuitively wider, correspond to less data, while the lower set of intervals correspond to more. By vertical comparison of the uncertainty intervals, we see that for \( n_\alpha = 5, n_\epsilon = 3 \) all intervals overlap, and hence no distinction can be made between the sources of variability. For \( n_\alpha = 8, n_\epsilon = 5 \) however, there is no overlap of the uncertainty intervals, suggesting that both superpopulation variability, and the finite-population variability are greater than error variability. Additionally, Figure 1 offers a diagnostic look at covariance posteriors. The center figure
displays ellipses from first two principal components corresponding to 2.5%, 50%, and 97.5% determinant-ordered percentiles of the posterior distributions, specific values of which can be seen on endpoints of the uncertainty intervals. The right-most graph of Figure 1, which shows all 1000 ellipses from each posterior distribution, offers a look at the size, shape, and orientation, for an overall comparison of the uncertainty in the batch covariances. Size renders an idea of the magnitude of the marginal variances. The disparity in size between the ellipses corresponding to $\mathbf{S}_\alpha$ and $\mathbf{\Sigma}_\epsilon$ suggest that the marginal variances of the former are greater than those of the latter. Through shape, one may glean some insight into batch covariance dependence. Lastly, the orientation, or varying orientation, suggests the uncertainty of the dependence, e.g. as the orientation of the ellipses corresponding to $\mathbf{\Sigma}_\alpha$ fluctuate greatly, there is not much that can be said about its dependence structure.

For comparison, consider a frequentist approach to a simplified form of the problem. Beginning with $\mathbf{X}_i \sim \mathcal{N}_d(0, \mathbf{\Sigma}), i = 1, \ldots, n$, it is known that $\mathbf{U} = \sum_i (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^T \sim W_d(\mathbf{\Sigma}, n-1)$, from which we derive the distributions

$$|\mathbf{U}| \sim |\mathbf{\Sigma}| \cdot \prod_{i=1}^d \chi_{n-1-d+i}^2,$$

$$\mathbf{1}^T \mathbf{U} \mathbf{1} \sim \mathbf{1}^T \mathbf{\Sigma} \mathbf{1} \cdot \chi_{n-1}^2,$$

$$\text{tr}(\mathbf{U}) \sim \sum_{i=1}^d \lambda_i \cdot \chi_{n-1}^2, \quad \lambda_1, \ldots, \lambda_d = \text{eig}(\mathbf{\Sigma}).$$

The first two offer pivots and thus allow for closed form expressions that yield confidence intervals for values of interest $|\mathbf{\Sigma}|$, $\mathbf{1}^T \mathbf{\Sigma} \mathbf{1}$. The confidence interval for $\text{tr}(\mathbf{\Sigma})$ is based on the normal approximation that matches the first two moments, $E\{\text{tr}(\mathbf{U})\} = (n-1)\text{tr}(\mathbf{\Sigma})$, and $\text{Var}\{\text{tr}(\mathbf{U})\} = 2(n-1)\text{tr}(\mathbf{\Sigma}^2)$. This approximation has been chosen in the spirit of moment matching approximations used by Imhof (1961) as applied to quadratic forms of random vectors. Note that these confidence intervals assume that the $\alpha_i$ are directly observed, which is not the case for our proposed method. Rather, the classical methods shown are included only for comparison.

To gain insight into the coverage success and uncertainty interval widths, we have
carried out \( S = 100 \) simulations over different values of \( n_\alpha \) and each of the different scenarios of variability sources. For all simulations the number of replicates at each level is fixed at \( n_\epsilon = 15 \). Results in all 3 cases were relatively similar with respect to coverage, noting however that uncertainty interval widths increase as the magnitude of variability increases. Thus, only the scenario in which variability sources are comparable (case 3) has been shown (Figure 2). Coverage and interval widths for \( \Sigma_\alpha \) and \( \Sigma_{\text{Freq}} \) should be compared as they correspond to the same true, unknown covariance. Despite the fact that the methodology does not assume realizations of \( \alpha \) to be observed directly, but rather indirectly through \( Y \), results are comparable to the frequentist approach outlined.

Figure 1: Simulation for two distinct sample size pairs, \( n_\alpha = 5, n_\epsilon = 3 \), and \( n_\alpha = 8, n_\epsilon = 5 \). In both, batch \( \alpha \) varies greater than the errors, with \( s_\alpha = (\sqrt{2}, \sqrt{2}, \sqrt{3})^T \). Uncertainty intervals with thin narrow lines denoting 95\% uncertainty level endpoints, thicker middle portion denoting 50\% endpoints, and vertical tick at median. Larger red vertical ticks denote true parameter values. Higher-positioned, wider intervals correspond to smaller batch sample sizes, while lower-positioned, narrower intervals correspond to larger batch sample sizes (left). For second set of sample size pairs, ellipses from first two principal components are displayed for 2.5\%, 50\%, and 97.5\% determinant-ordered percentiles of posterior distributions when 1000 posterior realizations have been drawn (center). All 1000 posterior ellipses (right).
Figure 2: For $n_\alpha = 5, \ldots, 100$, coverage is estimated with $S = 100$ simulations in which the sources of variability are comparable (case 3). Coverage estimates for $\Sigma_\alpha$, $S_\alpha$, and $\Sigma_\epsilon$, where grey line denotes 95% nominal coverage (top). Average uncertainty interval widths on log scale (bottom).

3.2 Application

In this example our methodology is applied to a bivariate dataset of global temperature (Celsius) and precipitation (mm/day) for 9 decadal averages of boreal summer months, June, July, August, during the remaining century. The first batch in the model consists of 13 levels, each representing a single atmosphere-ocean general circulation model (AOGCM) developed by several international climate research institutions as part of the CMIP3 project (Meehl et al., 2000) in the framework of the Fourth Assessment Report (AR4) for the Intergovernmental Panel on Climate Change (IPCC). The second batch covers 3 greenhouse gas emissions scenarios that have been defined by the Special Report on Emissions Scenarios (SRES), which are identified as A1B, A2, and B1 (Nakićenović and Swart, 2000). One fundamental objective of the analysis is then to compare how these factors contribute to overall variability of global climate averages, how they relate to one another, and what the uncertainty of this assessment is.

Bias and dependence among climate models is an issue that has more recently begun to be examined further, beginning with Tebaldi and Knutti (2007), Jun et al. (2008), Knutti et al. (2010), and references therein. Despite this, we adopt the statistical assumption
that has traditionally been used when with working with sets of AOGCMs, which is
to assume that they are independently drawn from a common process representative of
true climate characteristics. Using this assumption our approach can be seen as a useful
exploratory tool, and may be further adopted to address contrasts of batch levels, and
thus to identify similar batch levels. Preliminary analysis steps have suggested the model

\[ Y_{ijt} = \mu_0 + \alpha_{0,i} + \beta_{0,j} + \gamma_{ij} + x_{1,t}\mu_1 + x_{1,t}\alpha_{1,i} + x_{1,t}\beta_{1,j} + x_{2,t}\mu_2 + \epsilon_{ijt}, \tag{16} \]

where \( i = 1, \ldots, n_\alpha = 13, \ a \in 1, \ldots, n_\beta = 3, \ t = 1, \ldots, n_t = 9, \ n = n_\alpha n_\beta n_t, \) and \( d = 2. \)

Time covariate \( x_1 \) is centered such that \( x_{1,t} = -4, \ldots, 4, \) and \( x_2 \) is transformed to be
orthogonal to other predictors in the model. Batches of interest are AOGCM, \( \alpha, \) and
SRES, \( \beta, \) and their interaction, \( \gamma. \) The first two are further specified as a constant effect,
\( \alpha_0, \beta_0, \) as well as with respect to time, \( \alpha_1, \beta_1. \)

Posterior distributions of batches \( \alpha_0, \beta_0, \) and \( \gamma \) are derived from \( (14) \) and \( (15). \)

Batches \( \alpha_1 \) and \( \beta_1 \) differ slightly as they correspond to the regression model formulation.

Multivariate batch levels associated with a covariate would, in general, be multiplied by
a matrix, e.g. \( X_{1,t}. \) Using a matrix covariate, superpopulation and batch level posteriors
of batch \( \alpha_1 \) are then

\[
\Sigma_{\alpha_1} + V_{\alpha_1} | Y, \Sigma_\epsilon \sim W^{-1} \left( \sum_{i=c_{\alpha_1}+1}^{n_\alpha} \hat{\alpha}_{1,i} \hat{\alpha}^T_{1,i} + \epsilon_{\alpha_1} \right), \tag{17} \\
\alpha_{1,i} | Y, \Sigma_\epsilon, \Sigma_{\alpha_1} \sim \begin{cases} 
N_d \left( \hat{\alpha}_{1,i}, \frac{m_{\alpha}}{n} \Sigma_\epsilon \right) & i = 1, \ldots, c_{\alpha_1}, \\
N_d \left( P^{-1}_{\alpha_1} m_{\alpha_1,i}, P^{-1}_{\alpha_1} \right) & i = c_{\alpha_1} + 1, \ldots, n_\alpha,
\end{cases} \tag{18} 
\]

where \( V_{\alpha_1} = \frac{1}{n_\beta} (\sum_{t=1}^{n_t} X_{1,t}^T \Sigma_\epsilon^{-1} X_{1,t})^{-1}, \) \( P_{\alpha_1} = \Sigma_{\alpha_1}^{-1} + V_{\alpha_1}^{-1}, \) and \( m_{\alpha_1,i} = V_{\alpha_1}^{-1} \hat{\alpha}_{1,i}. \) For
model \( (16) \) the covariate matrix is \( X_{1,t} = \text{diag}(x_{1,t}), \) and thus \( V_{\alpha_1} = (n_\beta \sum_{t=1}^{n_t} x_{1,t}^2)^{-1} \Sigma_\epsilon. \)

The posterior of batch \( \beta_1 \) is found similarly.

Figure 3 suggests that AOGCM is the most distinguishing feature. Figures 4 and
5 confirm this assessment since \( \alpha_0 \) is seen as the most significant source of variability
among all batches. Comparison of Figures 4, 5 also illustrate the additional uncertainty
of superpopulation parameters over their finite-population counterparts. Superpopulation
covariance criteria uncertainty intervals are wide because they account for uncertainty in unobserved batch levels, particularly in the case when a small number of batch levels have been observed. Finite-population covariance uncertainty intervals are generally smaller, because they are concerned with variability of only the batch levels that have been realized.

Figure 3: Global temperature (top) in degrees Celsius and precipitation (bottom) as mm/day over $n_t = 9$ decadal intervals for $n_\alpha = 13$ AOGCMs and $n_\beta = 3$ emissions scenarios.

Figure 4: Superpopulation covariance uncertainty intervals using determinant, total variance, and total marginal variance criteria shown on a log scale. Nominal coverages of 0.95 (thin lines) and 0.50 (thick lines), and the median (vertical line) are denoted using quantiles of the corresponding posterior distributions.

Posterior predictive distributions, often used to perform model checking and diagnos-
Figure 5: Finite-population covariance uncertainty intervals using determinant, total variance, and total marginal variance criteria shown on a log scale. Nominal coverages of 0.95 (thin lines) and 0.50 (thick lines), and the median (vertical line) are denoted using quantiles of the corresponding posterior distributions.

Figure 6 examines posterior density $p(\tilde{Y}_{ijt} - \tilde{Y}_{ij1} | \{Y_{ijt}\})$, the difference between posterior predictive distributions at the final, $t = n_t$, and initial, $t = 1$, decades. Thus, the focus is on temporal batches, $\alpha_{1i}', \beta_{1j}'$. Indices $i', j'$ signify new, unobserved batch levels. Linear and quadratic terms, $\mu_1, \mu_2$ are included, although additional variability from these terms has been disregarded. The left-most panel, in which every batch contributes a new batch level realization, shows a large degree of variation. The center panel assumes that the AOGCM observed in the original data is to be used, thus variability from these specific batch level posteriors is included. For SRES a novel batch level is assumed, thus a realization utilizing the SRES superpopulation posteriors is included. A subset of three observed AOGCM levels has been displayed, selected so as to best represent the range and relative distances of their peaks. However, the high degree of variation introduced by the new emissions scenario level makes even these distributions nearly indistinguishable. In the right-most plot, using observed emissions batch levels, $n_\beta = 3$, the additional variability introduced comes primarily from the new AOGCM
batch level to be observed. The posterior predictive plots are particularly useful for determining whether the variability from each batch is due to the magnitude of the batch variability itself, or due to uncertainty in the assessment itself.

Figure 6: Posterior predictive distribution, \( p(Y_{i'j'n_t} - Y_{i'j'1}|\{Y_{ijt}\}) \), where both \( i', j' \) represent new, unobserved AOGCM and SRES batch levels (left). Posterior predictive \( p(Y_{ij'n_t} - Y_{ij'1}|\{Y_{ijt}\}) \) with observed AOGCM batch levels, \( i = 3, 9, 12 \) (solid, dashed, dotted) and unobserved SRES batch level \( \beta_{1,j'} \) (center). Posterior predictive \( p(Y_{i'jn_t} - Y_{i'j1}|\{Y_{ijt}\}) \) with unobserved AOGCM batch level \( \alpha_{1,i'} \) and all SRES levels, \( j = 1, 2, 3 \) (solid, dashed, dotted), that have been observed (right). Density contours correspond to quantiles 0.05, 0.25, 0.75. Horizontal and vertical axes denote precipitation in mm/day and temperature in degrees Celsius, respectively.

4 Discussion

The first contribution of this paper has been in extending recent philosophical shifts in the treatment of analysis of variance to multivariate settings. New analysis of variance approaches allow appropriate parameters, e.g. super or finite population, to be used to answer the correct research question, while at the same time providing coherent model definition, implementation, and interpretation. This same flexibility has been extended to multivariate cases; in that the researcher can guide covariance criteria choices, rather than the method determining the criterion. The second contribution has been in providing a foundation for computational efficiency, which is necessary for dimension scalability.
Using improper batch level priors we have shown that it is possible to minimize dependencies between batch covariances. In many cases this reduces, or eliminates, the need for complex and computationally demanding analyses.

Further extensions to the methodology must explicitly address increasing dimensionality. For moderately sized dimensions $d$, relative to number of observations, improper inverse-Wishart distributions, and/or priors that impose particular dependence structures, are possible options. For cases in which $d$ is very large, stricter covariance assumptions may be employed. In the spatial context, properties such as stationarity allow covariance parameter space to be reduced, e.g. range, sill, and nugget in a spatial covariance function. Because simultaneous estimation of such parameters is nontrivial, some parameters are often assumed, or estimated empirically in earlier analysis steps, as in Furrer et al. (2007). In other cases, so as to maintain computational feasibility, sparsity restrictions are placed on covariances (Cressie and Johannesson 2008; Furrer et al. 2006; Stein 2008). For many such scenarios a covariance is decomposed into a correlation matrix and a scalar variance parameter. Our method is then carried out with the inverse-Wishart posterior density transformed through a spectral decomposition of the correlation matrix, thus allowing for efficient posterior sampling for cases in which $d \gg n$.

This extension offers an alternative to geostatistical model analyses that have previously relied on computationally intensive MCMC methods, and is the focus of current research. Other difficulties encountered are unbalanced designs and linearly dependent predictors. MCMC may be utilized for sets of dependent batch levels. Development for these cases is another area of current research.

**Acknowledgments**

We acknowledge the modeling groups, the Program for Climate Model Diagnosis and Intercomparison (PCMDI) and the WCRP’s Working Group on Coupled Modeling (WGCM) for their roles in making available the WCRP CMIP3 multi-model dataset. Support of this dataset is provided by the Office of Science, U.S. Department of Energy.
References

Cressie, N. and Johannesson, G. (2008). Fixed rank kriging for very large spatial data
sets. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **70**, 209–226.

Daniels, M. J. (1999). A prior for the variance in hierarchical models. *Canad. J. Statist.*, **27**, 567–578.

Daniels, M. J. and Kass, R. E. (2001). Shrinkage estimators for covariance matrices. *Biometrics*, **57**, 1173–1184.

Díaz-García, J. A., Gutierrez Jáimez, R., and Mardia, K. V. (1997). Wishart and pseudo-
Wishart distributions and some applications to shape theory. *J. Multivariate Anal.*, **63**, 73–87.

Everson, P. J. and Morris, C. N. (2000). Simulation from Wishart distributions with
eigenvalue constraints. *J. Comput. Graph. Statist.*, **9**, 380–389.

Furrer, R., Genton, M. G., and Nychka, D. (2006). Covariance tapering for interpolation
of large spatial datasets. *J. Comput. Graph. Statist.*, **15**, 502–523.

Furrer, R., Sain, S. R., Nychka, D. W., and Meehl, G. A. (2007). Multivariate Bayesian
analysis of atmosphere-ocean general circulation models. *Environ. Ecol. Stat.*, **14**, 249–266.

Gelman, A. (2005). Analysis of variance: Why it is more important than ever. *Ann. Statist.*, **33**, 1–31.

Gelman, A. and Hill, J. (2006). *Data Analysis Using Regression and Multi-
level/Hierarchical Models*. Cambridge University Press, 1st edition.

Imhof, J. P. (1961). Computing the distribution of quadratic forms in normal variables.
*Biometrika*, **48**, 419–426.
Jun, M., Knutti, R., and Nychka, D. W. (2008). Spatial analysis to quantify numerical model bias and dependence: How many climate models are there? *J. Amer. Statist. Assoc.*, **103**, 934–947.

Kaufman, C. G. and Sain, S. R. (2010). Bayesian functional ANOVA modeling using Gaussian process prior distributions. *Bayesian Anal.*, **5**, 847–874.

Knutti, R., Furrer, R., Tebaldi, C., Cermak, J., and Meehl, G. A. (2010). Challenges in combining projections from multiple climate models. *Journal of Climate*, **23**, 2739–2758.

Leinonen, T., O’Hara, R. B., Cano, J. M., and Merilä, J. (2008). Comparative studies of quantitative trait and neutral marker divergence: a meta-analysis. *Journal of Evolutionary Biology*, **21**, 1–17.

Lencina, V. B., Singer, J. M., and Stanek, E. J. (2005). Much ado about nothing: the mixed models controversy revisited. *International Statistical Review*, **73**, 9–20.

Lindgren, F., Rue, H., and Lindström, J. (2011). An explicit link between gaussian fields and gaussian markov random fields: the stochastic partial differential equation approach. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **73**, 423–498.

Mardia, K. V., Kent, J. T., and Bibby, J. M. (1979). *Multivariate Analysis*. Academic Press.

Meehl, G. A., Boer, G. J., Covey, C., Latif, M., and Stouffer, R. J. (2000). The coupled model intercomparison project (CMIP). *American Meteorological Society Bulletin*, **81**, 313–318.

Nakićenović, N. and Swart, R., editors (2000). *Special Report on Emission Scenarios*. Intergovernmental Panel on Climate Change, Cambridge University Press.

Nelder, J. A. (1977). A reformulation of linear models. *J. Roy. Statist. Soc. Ser. A*, **140**, 48–77.

23
Nelder, J. A. (1994). The statistics of linear models: back to basics. *Statist. Comput.*, 4, 221–234.

Nelder, J. A. (1999). From statistics to statistical science. *J. R. Stat. Soc. Ser. D The Statistician*, 48, 257–269.

Nelder, J. A. (2008). What is the mixed-models controversy? *International Statistical Review*, 76, 134–135.

Qian, S. S. and Shen, Z. (2007). Ecological applications of multilevel analysis of variance. *Ecology*, 88, 2489–2495.

Rue, H. and Held, L. (2005). *Gaussian Markov Random Fields: Theory and Applications*. Chapman & Hall, London.

Rue, H., Martino, S., and Chopin, N. (2009). Approximate Bayesian inference for latent gaussian models by using integrated nested laplace approximations. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 71, 1–35.

Sain, S. R., Nychka, D., and Mearns, L. (2011). Functional ANOVA and regional climate experiments: a statistical analysis of dynamic downscaling. *Environmetrics*, 22, 700–711.

Srivastava, M. S. (2003). Singular Wishart and multivariate beta distributions. *Ann. Statist.*, 31, 1537–1560.

Stein, M. L. (2008). A modeling approach for large spatial datasets. *J. Korean Statist. Soc.*, 37, 3–10.

Tebaldi, C. and Knutti, R. (2007). The use of the multi-model ensemble in probabilistic climate projections. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.*, 365, 2053–2075.

Uhlig, H. (1994). On singular Wishart and singular multivariate beta distributions. *Ann. Statist.*, 22, 395–405.
Voss, D. T. (1999). Resolving the mixed models controversy. *Amer. Statist.*, 53, 352–356.