

Bayesian Functional ANOVA Modeling Using Gaussian Process Prior Distributions

Cari G. Kaufman and Stephan R. Sain

Abstract

Functional analysis of variance (ANOVA) models partition a functional response according to the main effects and interactions of various factors. This article develops a general framework for functional ANOVA modeling from a Bayesian viewpoint, assigning Gaussian process prior distributions to each batch of functional effects. We discuss the choices to be made in specifying such a model, advocating the treatment of levels within a given factor as dependent but exchangeable quantities, and we suggest weakly informative prior distributions for higher level parameters that may be appropriate in many situations. We discuss computationally efficient strategies for posterior sampling using Markov Chain Monte Carlo algorithms, and we emphasize useful graphical summaries based on the posterior distribution of model-based analogues of traditional ANOVA decompositions of variance. We illustrate this process of model specification, posterior sampling, and graphical posterior summaries in two examples. The first considers the effect of geographic region on the temperature profiles at weather stations in Canada. The second example examines sources of variability in the output of regional climate models from a designed experiment.

Keywords: analysis of variance, climate models, functional data, variance components

Cari Kaufman is Assistant Professor, Statistics Department, University of California, Berkeley, CA 94720 (email: cgk@stat.berkeley.edu). Stephan Sain is Project Leader, Geophysical Statistics Project, National Center for Atmospheric Research, Boulder, CO 80305 (email: ssain@ucar.edu).

1 Introduction

Functional analysis of variance (ANOVA) models are appropriate when the data consist of functions that are expected to differ according to some set of factors (Ramsay and Silverman, 2005, Chapter 13). For example, our work is motivated by the need to compare sources of variability in the projections made by computer models of climate. In this case the categorical factors can be the choice of climate model or the choice of various input values to the model, and the response is naturally a function of space and time. However, functional ANOVA models have proven useful in analyzing data from a variety of other fields (see e.g. Brumback and Rice, 1998; Spitzner et al., 2003; Wang et al., 2003).

Functional ANOVA models partition the functional response according to the main effects and interactions of the factors. For example, consider two crossed factors, with levels denoted by i and j . Let $Y_{ijk}(x)$ denote an observation from replication k under levels i and j of the factors, evaluated at x . The model partitions the functional response according to

$$Y_{ijk}(x) = \mu(x) + \alpha_i(x) + \beta_j(x) + (\alpha\beta)_{ij}(x) + \epsilon_{ijk}(x), \quad (1)$$

for $i = 1, \dots, m_A$, $j = 1, \dots, m_B$, $k = 1, \dots, n_{ij}$, and $x \in \mathcal{X} \subset \mathfrak{R}^d$. Each of the terms on the right hand side is a function mapping into the same space as the observations, and these may be modeled in a variety of ways. For example, smoothing spline ANOVA models express the effects as linear combinations of some underlying basis functions, and the coefficients on these basis functions are then chosen to minimize a criterion balancing goodness of fit with a measure of smoothness of the fitted functions (see Gu, 2002, for an overview). The tradeoff between the two is governed by the choice of smoothing parameter, which can be made according to various risk estimates. The connection between fitted smoothing splines and the limiting Bayes rule under a particular sequence of prior distributions has long been recognized (Wahba, 1978), and this connection can be used to motivate model choices in a Bayesian analysis (Barry, 1996). However, this formulation is not always intuitive or appropriate for analyzing a particular set of functional data. In this paper, we propose a fully Bayesian framework for functional ANOVA modeling. We view the functional effects on the right hand side of (1) and similar models as unknown quantities about which we have

some, perhaps vague, prior beliefs, for example that they belong to a particular function space. We use Gaussian process distributions as priors over these function spaces, and we make inference about the effects by conditioning on the observations. Some advantages of this approach are

1. The model provides a natural framework for inference, including simultaneous credible sets for functions. We also obtain posterior distributions for model-based analogues of the usual ANOVA decompositions of variance. As these vary over the domain of the functions, they can be used to create graphical displays that give an immediate sense of how different sources of variability contribute to the functional response.
2. The covariance parameters of the Gaussian processes, which play a role similar to the smoothing parameters in spline models, are estimated along with the functions themselves, rather than imposing a fixed roughness penalty. This extra source of uncertainty is naturally incorporated into posterior inference.
3. The prior specification accommodates a wide class of functions, and prior knowledge about the functions can be incorporated if desired. The Gaussian process distributions are extendable to an arbitrary number of dimensions of the functional response.

Our work draws heavily on models for spatial data, in that the Gaussian process prior distributions assigned to the various effects have covariance functions commonly used in geostatistics. However, because the term “spatial ANOVA” often indicates treating spatial regions as categorical factors, and because our method can be generalized to any number of dimensions, we will refer to it as Gaussian process ANOVA. This model has many connections to existing methods, as ANOVA models are a special case of the linear model with normal errors, a central tool in statistics. We elaborate on some particular connections to Bayesian linear model theory.

1.1 Bayesian ANOVA and Random Effects Models

It has been common practice in Bayesian ANOVA models to treat the levels of a given factor as one would “random effects” in a classical linear model, that is, as conditionally *iid* random variables with mean zero and a common variance component (Lindley and Smith, 1972; Gelman, 2005). For example, in the one-way ANOVA model for a scalar response, the usual model is

$Y_{ij} = \mu + \alpha_i + \epsilon_{ij}$, with $\mu|\mu_0, \sigma_\mu^2 \sim N(\mu_0, \sigma_\mu^2)$, $\alpha_i|\sigma_\alpha^2 \stackrel{iid}{\sim} N(0, \sigma_\alpha^2)$, and $\epsilon_{ij}|\sigma_\epsilon^2 \stackrel{iid}{\sim} N(0, \sigma_\epsilon^2)$, for $i = 1, \dots, m, j = 1, \dots, n_i$. One rationale for this choice is that it clearly satisfies certain invariance properties that characterize our understanding of the ANOVA decomposition: the joint distribution of the responses Y_{ij} is unaltered by permuting replications within a given level, or by permuting the various levels within a factor (Dawid, 1977). However, without any constraints on the individual levels, the model is over-parameterized, leading to Bayesian nonidentifiability (Gelfand and Sahu, 1999). That is, the marginal distribution for $\delta = \mu - \sum_i \alpha_i/m$ is not updated by the likelihood. In theory this is not an issue provided the prior distribution is proper, and we simply ignore δ in posterior inference. However, in practice this nonidentifiability means that Markov Chain Monte Carlo (MCMC) algorithms may drift to extreme values in the overparameterized space, even as they remain stable in the lower-dimensional subspace identified by the likelihood. This creates the potential for numerical instability (Gelfand and Sahu, 1999).

This issue has been addressed using various reparameterizations of the original prior distribution, such as hierarchical centering (Gelfand et al., 1995) and centering by sweeping (Vines et al., 1996). Hierarchical centering reparameterizes the model above to $Y_{ij} = \eta_i + \epsilon_{ij}$, $\eta_i|\mu, \sigma_\alpha^2 \stackrel{iid}{\sim} N(\mu, \sigma_\alpha^2)$, while centering by sweeping uses the prior distribution obtained by marginalizing over δ . However, hierarchical centering cannot be carried out for all factors if there are two or more crossed factors in the model, and centering by sweeping suffers the drawback that the implied prior distribution for the reparameterized model can make posterior sampling difficult in practice. For this work, we prefer another approach, which is to condition on identifying constraints in the prior distribution itself; see Smith (1973) and Nobile and Green (2000) for other examples. This prior distribution still satisfies Dawid's (1977) invariance properties: though dependent, the levels remain exchangeable. The benefit of this approach in the functional ANOVA context is that the partitioning of variability is unambiguous, so that variances and correlation parameters for the Gaussian processes we assign to each factor are interpretable as the factor's magnitude of effect and extent over the domain. Without these constraints, the nonidentifiability in the effects carries over to the distributions for these higher-level parameters, and sampling via MCMC is in our experience extremely inefficient.

1.2 Bayesian Analysis of Variance Components

Gelman (2005) makes a useful distinction for Bayesian ANOVA models, contrasting the variance components for the distribution of levels, or “superpopulation variances,” and the variance components calculated *from* the observed levels, or “finite population variances.” For example, in the one-way model above, σ_α^2 is the superpopulation variance, while the finite population variance is

$$s_\alpha^2 = \frac{1}{m-1} \boldsymbol{\alpha}' \left[\mathbf{I} - \frac{1}{m} \mathbf{J} \right] \boldsymbol{\alpha} \quad (2)$$

where $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_m)'$, \mathbf{I} is the $m \times m$ diagonal matrix, and \mathbf{J} is the $m \times m$ matrix of ones. This is the model-based analogue of the mean square for between-group variability one would calculate in a traditional ANOVA model. Correspondingly, the mean square can be thought of simply as a point estimate of this quantity, while the Bayesian model provides a full posterior distribution. As in Gelman (2005), we do not consider the testing problem here, instead estimating and comparing the finite population variances in the spirit of exploratory data analysis.

The model we propose for the functional response contains a single superpopulation variance for each factor, corresponding to the marginal prior variance for that factor’s Gaussian process prior distribution. However, more interesting in this case are the finite population variances. We now have $s_\alpha^2(x)$, which is a functional parameter of interest (a function of the $\{\alpha_i(x)\}$). While our prior distributions are specified such that the finite population variance is constant over the domain, we can examine the posterior distribution of $s_\alpha^2(x)$ and other finite population variance components to examine the magnitude of the contribution of each factor over the domain of the function. We describe the calculation of these quantities in Section 2.4.

1.3 Outline

In the next section we propose a framework for Bayesian functional ANOVA models using Gaussian process prior distributions, starting with the two-way model in (1) and then moving on to a more general formulation. We make some suggestions regarding the choices involved in defining the model for a given application. We describe an MCMC algorithm used to fit the model and some ways to make posterior sampling more efficient. We then describe in more detail the calculation of the posterior distributions for the finite population variance components and some useful graphical

displays for comparing sources of variation in the functional response. Section 3 describes this process of model specification, posterior sampling, and creating graphical posterior summaries for two examples of functional data. The first is a simple one-way model for a one dimensional response, while the second is a two-way model with temporal trends for a spatial response. We conclude with an overview of the method and some potential areas for future model development.

2 Gaussian Process ANOVA Model

Suppose we observe the functional response at x_1, \dots, x_p . (For notational simplicity, we will assume throughout that the x values are the same for all combinations of levels, although the models we propose apply equally well to the more general case.) Starting with the two-way model as in (1), let the vector $\mathbf{Y}_{ijk} = (Y_{ijk}(x_1), \dots, Y_{ijk}(x_p))'$ represent the k^{th} response with factor A at level i and factor B at level j . We model \mathbf{Y}_{ijk} as a finite set of observations from an underlying smooth realization of a stochastic process Y_{ijk} defined on $\mathcal{X} \subseteq \mathfrak{R}^d$. Let $\mu_{ij}(x) = \mu(x) + \alpha_i(x) + \beta_j(x) + (\alpha\beta)_{ij}(x)$. Then the first stage of the model is

$$Y_{ijk} | \{\mu_{ij}\}, \sigma_\epsilon^2, \theta_\epsilon \stackrel{indep}{\sim} GP(\mu_{ij}, \sigma_\epsilon^2 R_{\theta_\epsilon})$$

for $i = 1, \dots, m_A$, $j = 1, \dots, m_B$, $k = 1, \dots, n_{ij}$, where the notation $GP(h, K)$ denotes a Gaussian process distribution with mean function h and covariance function K . Here we have separated the covariance function into the marginal variance σ_ϵ^2 , and R_{θ_ϵ} , a member of a particular class of correlation functions indexed by θ .

We now specify Gaussian process prior distributions for μ , $\{\alpha_i\}$, $\{\beta_j\}$, and $\{(\alpha\beta)_{ij}\}$, taking each batch of functions to be independent of the other batches and independent of the residuals a priori, and assigning each batch its own set of higher-level parameters. For a given set of q parametric regression functions $\{f_\ell\}$, we model

$$\mu | \{\phi_\ell\}, \sigma_\mu^2, \theta_\mu \sim GP \left(\sum_{\ell=1}^q \phi_\ell f_\ell, \sigma_\mu^2 R_{\theta_\mu} \right). \quad (3)$$

The prior distributions for the batches of functions $\{\alpha_i\}$, $\{\beta_j\}$, and $\{(\alpha\beta)_{ij}\}$ satisfy the constraints $\sum_i \alpha_i(x) = 0$, $\sum_j \beta_j(x) = 0$, $\sum_i (\alpha\beta)_{ij}(x) = 0$, and $\sum_j (\alpha\beta)_{ij}(x) = 0$ for all x . Specifically,

we define a distribution for $\{\alpha_i\}$ such that each α_i is marginally a mean zero Gaussian process, and

$$\text{Cov}(\alpha_i(x), \alpha_{i'}(x')) = \begin{cases} (1 - \frac{1}{m_A})\sigma_\alpha^2 R_{\theta_\alpha}(x, x') & i = i' \\ -\frac{1}{m_A}\sigma_\alpha^2 R_{\theta_\alpha}(x, x') & i \neq i' \end{cases} \quad (4)$$

We define the prior for $\{\beta_j\}$ in an analogous fashion, with parameters σ_β^2 and θ_β , and m_B levels rather than m_A . We also specify mean zero Gaussian process prior distributions for the interaction terms, with a slightly more complicated covariance structure imposed by the two sets of sum to zero constraints:

$$\text{Cov}((\alpha\beta)_{ij}(x), (\alpha\beta)_{i'j'}(x')) = \sigma_{\alpha\beta}^2 R_{\theta_{\alpha\beta}}(x, x') \begin{cases} (m_A - 1)(m_B - 1)/(m_A m_B) & i = i', j = j', \\ (1 - m_A)/(m_A m_B) & i = i', j \neq j' \\ (1 - m_B)/(m_A m_B) & i \neq i', j = j' \\ 1/(m_A m_B) & i \neq i', j \neq j' \end{cases}$$

The model specification is completed by the choice of regression functions in the mean for μ and the correlation functions, and prior distributions over the higher-level parameters. These choices will be application specific, although we make some general suggestions in section 2.2.

2.1 General Formulation

We now describe the Gaussian process ANOVA model for an arbitrary number of crossed and/or nested factors. Let each ‘‘batch’’ of functions, for example the levels within a particular factor or interaction, be denoted by subscript b , and let i index the observations in the dataset. Extending the notation of Gelman (2005) for ANOVA models to a functional response, write

$$Y_i(x) = \sum_{b=0}^B \beta_{j_i^b}^{(b)}(x),$$

where $\beta_1^{(b)}, \dots, \beta_{m_b}^{(b)}$ are the functions in batch b and j_i^b indicates the particular value of j corresponding to the i^{th} observation for this batch. Note that the sum includes both the grand mean, defining $\mu \equiv \beta^{(0)}$, and the error terms, defining $\epsilon_i \equiv \beta_{j_i^b}^{(B)} = Y_i - \sum_{b=0}^{B-1} \beta_{j_i^b}^{(b)}$. We assign to each batch of functions a joint Gaussian process distribution. For the grand mean and error terms, we

have

$$\begin{aligned} \beta^{(0)}|\{\phi_\ell\}, \sigma_0^2, \theta_0 &\sim GP(\sum_\ell \phi_\ell f_\ell, \sigma_0^2 R_{\theta_0}) \\ \beta_{j_i}^{(B)}|\sigma_B^2, \theta_B &\stackrel{iid}{\sim} GP(0, \sigma_B^2 R_{\theta_B}), \quad i = 1, \dots, n \end{aligned} \quad (5)$$

As in the two-way case, we assign Gaussian process distributions to the batches of functions representing the main effects and interactions, and we constrain them to sum to zero across their margins. For a particular batch b of functions, let $\mathbf{C}^{(b)}$ be a $m_b \times c_b$ matrix with linearly independent columns such that $\mathbf{C}^{(b)'}\boldsymbol{\beta}^{(b)}(x) = \mathbf{0} \forall x$, where $\boldsymbol{\beta}^{(b)}(x) = (\beta_1^{(b)}(x), \dots, \beta_{m_b}^{(b)}(x))'$. Then define $\mathbf{P}^{(b)} = \left[\mathbf{I}_{m_b} - \mathbf{C}^{(b)}(\mathbf{C}^{(b)'}\mathbf{C}^{(b)})^{-1}\mathbf{C}^{(b)'} \right]$. Finally, assign a mean zero multivariate Gaussian process distribution to $\beta_1^{(1)}, \dots, \beta_B^{(m_b)}$, with

$$Cov(\beta_j^{(b)}(x), \beta_{j'}^{(b)}(x)) = \mathbf{P}_{jj'}^{(b)}\sigma_b^2 R_{\theta_b}(x, x'). \quad (6)$$

Again, the model specification is completed by defining the regression functions, class of correlation functions, and prior distribution for the higher level mean and covariance parameters for each batch.

2.2 Model Specification

2.2.1 Mean and Covariance Functions

The mean structure in (5) is meant to capture obvious patterns in the common response; for example in Section 3.1 we model temperature profiles over the year as having an underlying sinusoidal pattern. Because we are interested in the finite population variances and not the superpopulation variances, it does not matter from a theoretical point of view how we specify the mean or what effect this has on the interpretation of σ_μ^2 . Often a single intercept value here will be adequate, taking $f \equiv 1$. From a practical point of view, ϕ and σ_μ^2 are only weakly identified in this model, so care is required in specifying the prior distributions for these parameters, as we describe below.

We suggest choosing the class of correlation functions R to be stationary and isotropic, and equating each θ with a single range parameter ρ . The covariance function for the error terms may also include a nugget term corresponding to measurement error variance. One can also use a more flexible class of covariance functions, for example to model nonstationarity. However, the benefits

of added model complexity should be weighed in terms of the effect on posterior inference for the quantities of interest: the factor levels and their finite population variances. The covariance functions affect these only insofar as they allow sharing of information across the function. A less flexible class of functions in this case may not be optimal, but it is unlikely to make a large difference in the inference provided the data are not very sparse over the domain of the functions.

2.2.2 Parameterization of the Levels

In some cases it may be easier to represent the dependence structure in (6) by reparameterizing the levels of a given factor as a linear combination of independent processes in a lower-dimensional subspace. For example, consider the main effect of factor A . Write $\alpha_i = \sum_{k=1}^{m_A-1} M_{ik} \alpha_k^*$, where M is a $m \times (m-1)$ matrix and $\alpha_1^*, \dots, \alpha_{m_A-1}^* \stackrel{iid}{\sim} GP(0, \sigma_\alpha^2 R_{\theta_\alpha})$. We require that $\sum_{i=1}^p \alpha_i(x) = 0 \forall x$ and the prior covariance in (4) holds, which is true provided

1. $\sum_{i=1}^m M_{ik} = 0, \quad k = 1, \dots, m-1$
2. $\sum_{k=1}^{m-1} M_{ik}^2 = (1 - 1/m), \quad i = 1, \dots, m$
3. $\sum_{k=1}^{m-1} M_{ik} M_{i'k} = -1/m, \quad i \neq i'$

These conditions can easily be satisfied by rescaling the columns of a matrix of Helmert contrasts. For example, when $m_A = 3$, start with the matrix whose rows are $\{(1, 0), (-1/2, 1), (-1/2, -1)\}$. The columns of this matrix already sum to zero; the idea is to multiply each column by some scalar such that the second two conditions hold. The multiplier for the first column is clearly $\sqrt{2/3}$, and some quick algebra shows the multiplier for the second column is $\sqrt{1/2}$. The updated matrix is $M = \{(\sqrt{2/3}, 0), (-\sqrt{2/3}/2, \sqrt{1/2}), (-\sqrt{2/3}/2, -\sqrt{1/2})\}$.

2.2.3 Prior Choices

Our guiding principle in choosing prior distributions for higher-level parameters is to include weakly constraining prior information where it exists, as a way of regularizing the inference. By “regularization” in this context, we mean that by assigning low prior probability to certain regions of parameter space, we can improve the interpretability of the model parameters, as well as preventing numerical instability due to sample paths drifting to extreme values in the MCMC algorithm. For

example, although the mean of μ and its variance σ_μ^2 are only weakly identified in the likelihood, prior information is often available about the mean, based on physical or other constraints. Likewise, the (σ^2, ρ) pairs will tend to be only weakly identified, which is related to the equivalence of the Gaussian processes for certain combinations of parameter values (Stein, 1999; Zhang, 2004). Standard noninformative priors as in Berger et al. (2001) can be computationally expensive, involving derivatives of each element of the correlation matrix. To reduce the computational burden and to regularize the inference, we suggest taking $p(\sigma^2) \propto 1/\sigma^2$, but choosing a proper prior distribution for the ρ parameters. This distribution should be chosen such that realizations of a Gaussian process with the corresponding correlation function cover the range of anticipated behavior. This specification can still be vague if desired, allowing both flat and wiggly realizations. We illustrate this choice in the examples.

2.3 Posterior Sampling

One can generate posterior samples from the Gaussian processes ANOVA model using a Gibbs sampler, using the Metropolis-Hastings algorithm to sample distributions not available in closed form. The Gaussian process distributions for the functions imply multivariate normal distributions for those functions evaluated at a finite set of points. One would typically sample the functions at the set of unique x values in the dataset, to facilitate computation of the likelihood, although additional x values may also be included. Boldface parameters in the following should be interpreted to mean the vector of evaluations for the corresponding process over the x values of interest.

The full conditional distributions for ϕ , $\boldsymbol{\mu}$, and the main effects and interactions are all multivariate normal. However, there is dependence in the prior distributions, induced by the sum-to-zero constraints. If the levels have been reparameterized as in Section 2.2.2, one can carry out sampling in the lower dimensional subspace and then transform back when making posterior inference. Deriving full conditional distributions for the reparameterized model is straightforward, as the prior distribution can be factored into independent components and the contributions from the likelihood simply involve contrasts of the observations. An alternative method is to keep the original, degenerate parameterization and to sample a batch of parameters as a block. For example, in the one-way model in which the $\{\alpha_i\}$ have prior covariance structure as in (4), we may factor the prior

distribution according to

$$\begin{aligned}
 p(\boldsymbol{\alpha}_1 | \sigma_\alpha^2, \rho_\alpha) &= MVN\left(0, \frac{m-1}{m} \sigma_\alpha^2 \boldsymbol{\Gamma}(\rho_\alpha)\right) \\
 p(\boldsymbol{\alpha}_i | \boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_{i-1}, \sigma_\alpha^2, \rho_\alpha) &= MVN\left(-\frac{\sum_{k=1}^{i-1} \boldsymbol{\alpha}_k}{m-i+1}, \frac{m-i}{m-i+1} \sigma_\alpha^2 \boldsymbol{\Gamma}(\rho_\alpha)\right), \quad i = 2, \dots, m,
 \end{aligned}$$

where $\boldsymbol{\Gamma}(\rho)$ denotes the $p \times p$ correlation matrix $\{R_\rho(x_i, x_j)\}$ and p is the number of x values at which the functions are being sampled. Note that the final distribution for α_m is degenerate, reflecting the sum to zero constraint. Letting “*Rest*” denote all parameters except for the collection $\{\alpha_i\}$, we generate $\{\alpha_i^*\}$ from $p(\{\alpha_i\} | \textit{Rest})$ by first generating α_1^* from $p(\alpha_1 | \textit{Rest})$, then iteratively sampling α_i^* from $p(\alpha_i | \alpha_1^*, \dots, \alpha_{i-1}^*, \textit{Rest})$ for $i = 2, \dots, m-1$. These distributions are also multivariate normal and are straightforward to derive using this factorization of the prior. The sample for α_m is simply $-\sum_{i=1}^{m-1} \alpha_i$. The first example in Section 3 uses this blocking strategy, while the second example uses the reparameterization as in Section 2.2.2.

Within each iteration it is also advisable to sample each (σ^2, ρ) pair as a block. Now letting “*Rest*” denote all parameters except σ^2 and ρ , first we sample ρ^* from $p(\rho | \textit{Rest})$ using a Metropolis-Hastings step, then we sample σ^{2*} from $p(\sigma^{2*} | \rho^*, \textit{Rest})$. These parameters tend to be highly correlated in posterior samples, and our experience has been that sampling them in this way dramatically improves the mixing of the MCMC samples. We have also found it helpful to randomize the order in which each parameter or block of parameters is updated within each iteration, as suggested in (Roberts and Sahu, 1997).

2.4 Graphical Posterior Summaries

The posterior samples can be used to create a variety of useful graphics to summarize various aspects of the posterior distribution. We focus on two graphical displays in particular, globally defined intervals of high posterior probability and plots of the finite population variances.

2.4.1 Global Credible Intervals

Bayesian “confidence intervals” have been used in spline smoothing for some time (Wahba, 1983; Gu and Wahba, 1993). The posterior samples can be used to generate similar intervals of high posterior

probability for each of the levels for a given factor, or for other quantities of interest. These intervals are not uniquely defined, but we suggest starting with point-wise intervals at each x value where the effects have been sampled, for example taking as the lower and upper bounds $(\ell(x), u(x))$ the 0.025 and 0.975 quantiles of the posterior distribution. This produces point-wise intervals. To calculate the simultaneous intervals, find ϵ such that $\hat{P}[g(x) \in (\ell(x) - \epsilon, u(x) + \epsilon) \forall x \in \mathcal{X}] = 0.95$, where g is the function of interest, \hat{P} is the proportion of posterior samples satisfying the criterion and \mathcal{X} is the set of x values at which posterior samples of the functions have been generated. Plots of $[\ell(x) - \epsilon, \ell(x) + \epsilon]$, linearly interpolating between $x \in \mathcal{X}$, then give a graphical summary of a high probability region for the entire function g . For sufficiently dense \mathcal{X} , this will be a good approximation to the true functional intervals.

2.4.2 Finite Population Variance Plots

Extending the notation of Gelman (2005) to functional effects, we define the finite population variances for the Gaussian process ANOVA model as

$$s_b^2(x) = \frac{1}{m_b - c_b} \boldsymbol{\beta}^{(b)}(x)' \mathbf{P}^{(b)} \boldsymbol{\beta}^{(b)}(x),$$

where $\boldsymbol{\beta}^{(b)}(x)$ and $\mathbf{P}^{(b)}$ are as defined in Section 2.1. Note that this definition also includes the error term, for which there are no constraints and $\mathbf{P}^{(b)}$ is simply diagonal. Each s_b^2 is the functional analogue of a mean square quantity in traditional ANOVA. One could consider carrying out a traditional ANOVA analysis at each x of interest in the domain and then simply plotting these mean squares. However, the Bayesian functional ANOVA has the advantage that s_b^2 is explicitly modeled as a function, for which we obtain posterior samples via the sampled effects, and so we can construct both point-wise and global intervals for these functions. We can also look at posterior probabilities for various relationships between the finite population variances over the functional domain, exploring the regions of the domain in which various factors are most important. We demonstrate this form of graphical inference in the examples.

3 Examples

We present two examples of Gaussian process ANOVA models. The first example is a simple one-way model for a one dimensional response, while the second is a two-way model with temporal trends for a spatial response. The models for each example were fit using the R language for statistical computing; data and R code for the examples are available online at <http://www.image.ucar.edu/~cgk>.

3.1 Example I: Temperature Profiles at Canadian Weather Stations

We consider the Canadian weather data introduced by Ramsay and Silverman (2005), which is available as part of the `fda` package in R. The data consist of monthly average temperatures for 35 Canadian weather stations. The stations are divided into four climate zones: Atlantic, Continental, Pacific, and Arctic. The data are shown in Figure 1. Ramsay and Silverman (2005) estimated the temperature profiles in each zone using a functional ANOVA model, representing the effects using Fourier basis functions and minimizing a penalized least squares criterion. They also calculated point-wise confidence intervals for the deviations of each profile from the average profile. Using a Gaussian process ANOVA model, we construct both pointwise and global credible intervals for the deviations from the average profile, and we study the posterior distribution of the finite population variances to determine the months in which the categorization by zone has the largest impact.

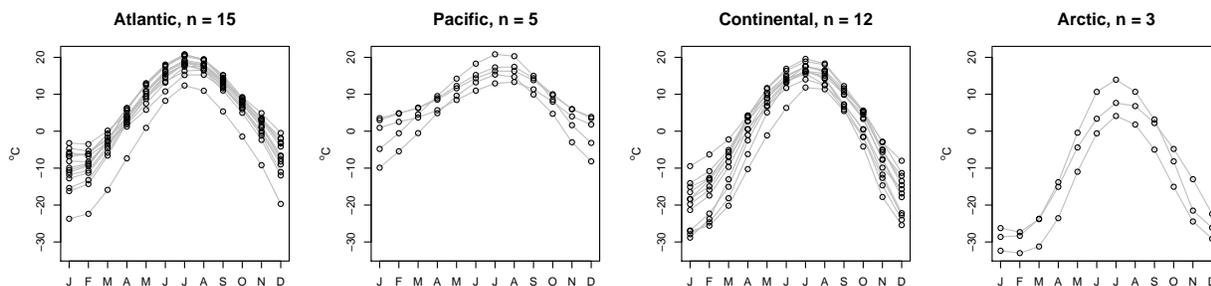


Figure 1: Average monthly temperature profiles from 35 Canadian weather stations.

3.1.1 Model Specification

We model the average temperature for a station j falling into zone i at time t as $Y_{ij}(t) = \mu(t) + \alpha_i(t) + \epsilon_{ij}(t)$, where $t \in [0, 1]$ represents fraction of the year. The function $\mu + \alpha_i$ represents the expected temperature profile for zone i , with α_i modeling deviations from the average profile μ .

As in the general formulation, we take the batches μ , $\{\alpha_i\}$, and $\{\epsilon_{ij}\}$ to be independent a priori. We specify distributions that reflect our belief that these functions are smooth and periodic by using Gaussian process distributions with periodic means and covariance functions. Specifically, let $d(t, t') = 2 \sin(\psi_{t,t'}/2)$, where $\psi_{t,t'}$ is the angle in radians between $2\pi t$ and $2\pi t'$. Define

$$R_{\rho,\nu}(t, t') = \frac{(d(t, t')/\rho)^\nu}{2^{\nu-1}\Gamma(\nu)} \mathcal{K}_\nu(d(t, t')/\rho), \quad (7)$$

which is the Matérn correlation function (Matérn, 1986) with parameters ρ and ν , evaluated at $d(t, t')$. Because the Matérn correlation function is positive definite in \mathbb{R}^2 , $R_{\rho,\nu}$ is a valid periodic correlation function on $[0, 1]$ (Yaglom, 1987, page 389). For this analysis, we fix $\nu = 2$ throughout, and we write $R_\rho(t, t')$ to denote $R_{\rho,2}(t, t')$ as in (7).

We specify $E[\mu(t)] = \phi_0 + \phi_1 \cos(2\pi t) + \phi_2 \sin(2\pi t)$ and $Cov(\mu(t), \mu(t')) = \sigma_\mu^2 R_{\rho_\mu}(t, t')$. We incorporate the constraint $\sum_{i=1}^4 \alpha_i(t) = 0$ into the prior distribution by taking $\alpha_1, \dots, \alpha_4$ to have a multivariate Gaussian process distribution with mean zero and covariance structure as in (4). Finally, we model ϵ_{ij} as independent mean zero Gaussian processes, each with covariance function $\sigma_\epsilon^2 R_{\rho_\epsilon}(t, t')$.

Considering the prior distributions for higher-level parameters, we take the elements of $\phi = (\phi_0, \phi_1, \phi_2)'$ to have independent normal distributions with mean zero. Following the principle of including weak prior information based on physical constraints, we take ϕ_0 to have a prior variance of 50, which implies low probability that ϕ_0 , the overall yearly mean, will fall outside the interval $[-15^\circ C, 15^\circ C]$. Likewise, the amplitude of the cosine curve, $\sqrt{\phi_1^2 + \phi_2^2}$, is unlikely to be outside $[-50^\circ C, 50^\circ C]$. Rather than specifying a joint prior for these values, which also control the phase shift of the mean curve, we assign them prior variance of $25^2 = 625$, meaning that a value greater than 50 is unlikely for either.

For the σ^2 parameters we use the standard non-informative prior distribution $p(\sigma^2) \propto 1/\sigma^2$. We take the ρ parameters to have independent Gamma(1, 1) prior distributions. This allows for a variety of correlation lengths. Figure 2 shows simulated realizations of mean zero Gaussian processes for a variety of plausible values of ρ under this distribution.

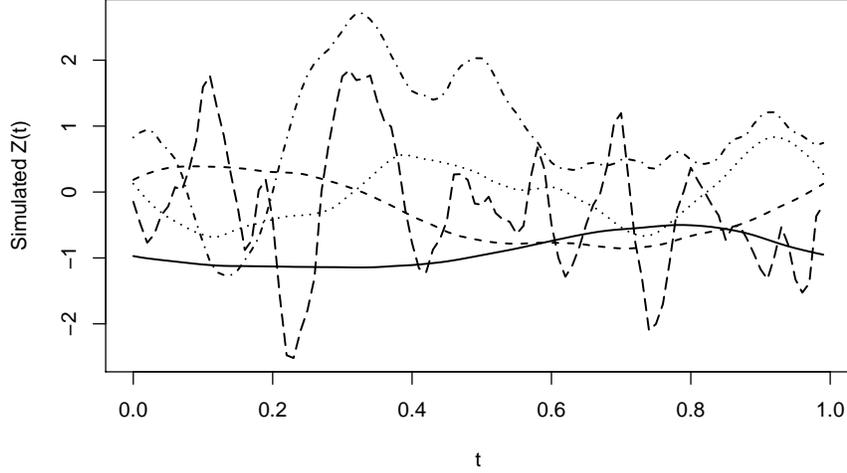


Figure 2: Simulated realizations from Gaussian process distributions with variance one and correlation functions $R_\rho(t, t')$, with ρ set to the 0.9 (—), 0.75 (---), 0.5 (···), 0.25 (-·-), and 0.1 (- -) quantiles of a Gamma(1, 1) distribution.

3.1.2 Posterior Sampling

We observe $\mathbf{Y}_{ij} \equiv (Y_{ij}(t_1), \dots, Y_{ij}(t_{12}))'$, for $i = 1, \dots, 4$, $j = 1, \dots, n_i$, and $t_k = (k - 0.5)/12$ for $k = 1, \dots, 12$. Based on the joint specification above, the likelihood is

$$\mathbf{Y}_{ij} | \boldsymbol{\mu}, \boldsymbol{\alpha}_i, \sigma_\epsilon^2, \rho_\epsilon \stackrel{indep}{\sim} MVN(\boldsymbol{\mu} + \boldsymbol{\alpha}_i, \sigma_\epsilon^2 \boldsymbol{\Gamma}(\rho_\epsilon)), \quad i = 1, \dots, 4, j = 1, \dots, n_i,$$

where the bold symbols $\boldsymbol{\mu}$ and $\boldsymbol{\alpha}_i$ indicate vectors of the corresponding Gaussian processes evaluated at t_1, \dots, t_k and the notation $\boldsymbol{\Gamma}(\rho)$ indicates the 12×12 correlation matrix $\{R_\rho(t_i, t_j)\}$. The Gaussian process prior distributions above also imply prior multivariate normal distributions for $\boldsymbol{\mu}$ and $\{\boldsymbol{\alpha}_i\}$.

The Gibbs sampler iterates between sampling ϕ and $\boldsymbol{\mu}$, sampling $\{\boldsymbol{\alpha}_i\}$ as a block as described in Section 2.3, and sampling each of (σ_μ^2, ρ_μ) , $(\sigma_\alpha^2, \rho_\alpha)$, and $(\sigma_\epsilon^2, \rho_\epsilon)$ as a block, also described in Section 2.3. We carried out 20,000 iterations of the Gibbs sampling algorithm, which took approximately 5 minutes on a laptop computer. We tried several values for each of the proposal variances in the Metropolis-Hastings steps for the ρ parameters until we achieved acceptance rates of twenty to thirty percent for each. The resulting sample paths for the parameters were such that retaining every twentieth iteration produced sample paths with no evidence of autocorrelation. The sample paths appeared to have a stationary distribution within only a few iterations, but to be conservative we discarded the first 5000 iterations for burn-in, giving 750 roughly independent samples.

3.1.3 Graphical Posterior Summaries

Figure 3 plots the point-wise and global intervals of high posterior probability, as described in Section 2.4, for the grand mean μ and the regional effects $\{\alpha_i\}$. As in Ramsay and Silverman (2005), we conclude that the temperature profile for the Atlantic region tends to be slightly warmer than the mean profile, the profile for the Pacific region is warmer than the mean profile during the winter months, the Continental profile is slightly colder during the winter, and the Arctic profile is always colder, particularly so in March. These inferences are somewhat heuristic, as we are interpreting the overall shape of the intervals. However, it would be straightforward to examine, for example, the posterior distribution for the month in which the Arctic profile differed the most from the mean profile, by simply calculating this quantity for each posterior sample and examining its empirical distribution. One could also compare specific contrasts of interest between the regions.

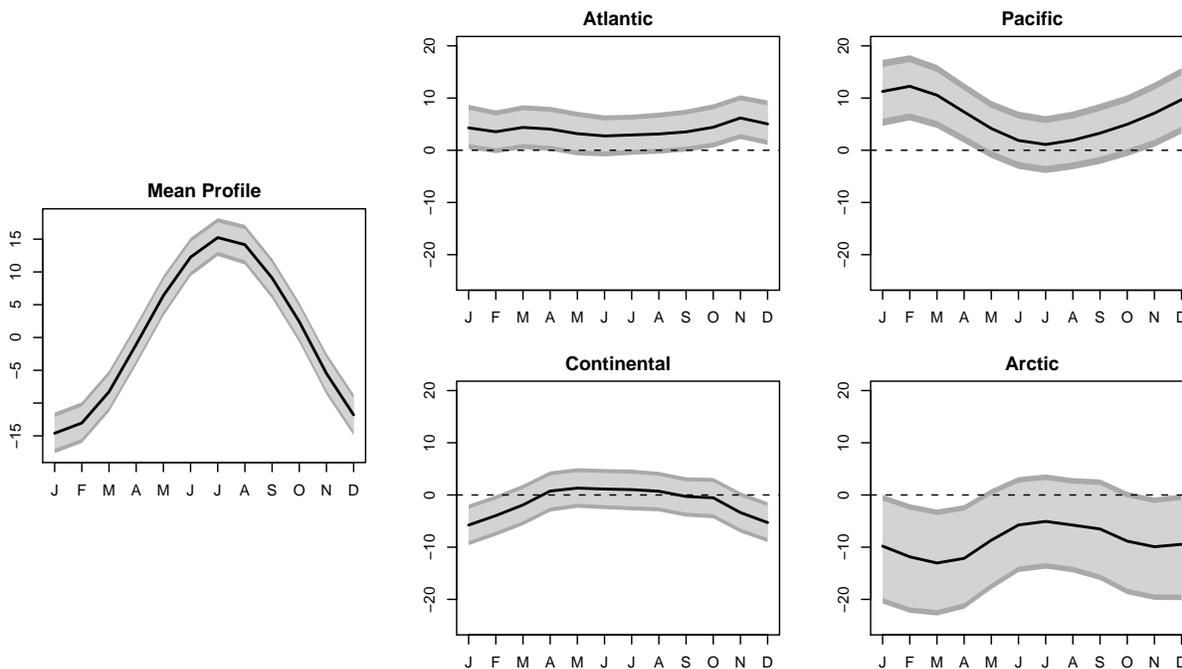


Figure 3: Posterior means (—), pointwise credible intervals (light gray shading) and global credible (union of light and dark gray shading) for the mean profile μ and the regional effects $\{\alpha_i\}$.

Figure 4 shows point-wise and global credible intervals for the finite population standard deviations $s_\alpha(x)$ and $s_\epsilon(x)$ and their ratio. First examine $s_\alpha(x)$, the finite population standard deviation for the effect of region. The posterior intervals indicate that the grouping by region has the largest

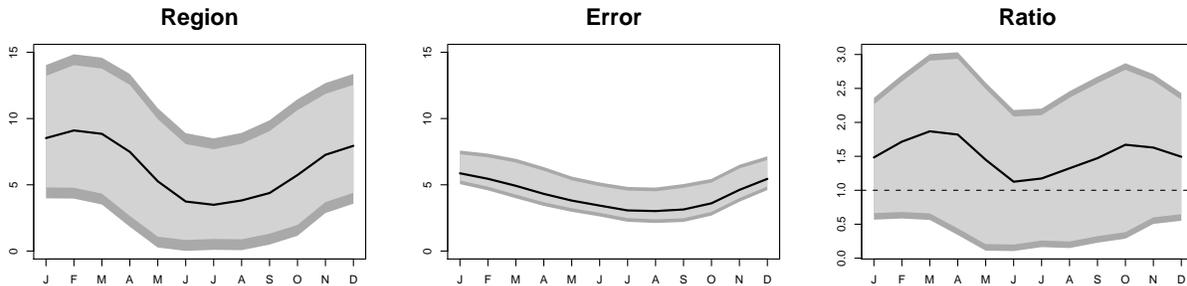


Figure 4: The first two panels show point-wise (light grey shading) and global (union of light and dark grey shading) 95% credible intervals for the finite population standard deviations $s_\alpha(x)$ (region) and $s_\epsilon(x)$ (error). The last panel gives the intervals for $s_\alpha(x)/s_\epsilon(x)$.

effect from late fall to early spring. Next looking at $s_\epsilon(x)$, the finite population standard deviation for the error, we see that the variability within regions tends to be smaller than the variability between regions, although the credible intervals overlap. The variability within regions also is highest during the winter and spring months. However, the posterior distribution of $s_\alpha(x)/s_\epsilon(x)$ indicates that, relative to the magnitude of the error, the effect of region is largest in early spring and fall, with a smaller effect during the winter. Looking again at the data in Figure 1, there does appear to be some variability in the winter months that is not accounted for by the current set of geographic regions. The Pacific and Continental regions in particular appear to contain subgroups of stations whose behavior is not captured by the overall profiles for those regions. Although this problem may be seen even in the data for this example, a larger dimensional functional response or a greater number of factors will make this model inadequacy much harder to diagnose, making the plots of the finite population variances valuable diagnostic tools.

3.2 Example II: Regional Climate Model Experiment

Regional climate models (RCMs) are used by climate scientists to model the evolution of the climate system over a limited area, using discretized versions of physical processes. These models address smaller spatial regions than do global climate models (GCMs), sometimes called general circulation models. However, the higher resolution in RCMs better captures the impact of local features such as lakes and mountains, as well as subgrid-scale atmospheric processes that are only approximated in GCMs. Due to their limited area, RCMs require boundary conditions, and these are often provided by the output of GCMs. This is also referred to as “downscaling” the GCM output using the RCM.

Climate scientists are interested in how much variability in the RCM output is attributable to the RCM itself, and how much is due simply to large-scale boundary conditions provided by the GCM.

The PRUDENCE project (Christensen et al., 2002) crossed the factors of RCM model choice and GCM boundary conditions in a designed experiment involving regional models over Europe from various climate research centers. We examine a subset of the data consisting of control runs (1961-1990) for two RCMs crossed with two GCMs, looking at output over the United Kingdom and Ireland. The two regional models we consider are HIRHAM, developed in collaboration between the Danish Meteorological Institute, the Royal Netherlands Meteorological Institute, and the Max Planck Institute for Meteorology, and RCAO, developed at the Rossby Centre at the Swedish Meteorological and Hydrological Institute. The two GCMs are ECHAM4, from the Max Planck Institute, and HadAm3H, from the Hadley Centre in the United Kingdom. Details regarding all the models and references concerning their development can be found at <http://prudence.dmi.dk/>.

Figure 5 shows average summer temperatures from 30 years of output in the four combinations of RCM and GCM. Note that there are similar large-scale patterns in the means for a particular set of GCM boundary conditions (compare between columns), and there are similar smaller-scale patterns for a particular choice of RCM (compare between rows). These observations suggest a decomposition of the mean temperature response into the effect of RCM, effect of GCM, and their possible interaction. The magnitude of these effects and their values over various regions can be used as a diagnostic tool. For example, if there is disagreement between models in a given region, then the model builders can focus their attention on that region. However, it is natural to compare the magnitude of the disagreement to the models’ “internal” variability, that is, the variability from year to year. We use the Gaussian process ANOVA model to quantify these sources of variability in the model output.

3.2.1 Model Specification

Let $Y_{ijt}(s)$ denote the output of RCM i with boundary conditions provided by GCM j , at time t and location s . We code the years 1961 to 1990 as $t_k = k - 15.5, k = 1, \dots, 30$, so that the model intercept corresponds to the midpoint of the time interval. Initial analyses showed a mild increasing trend in the data for all models, the magnitude of which varied little between models or locations.

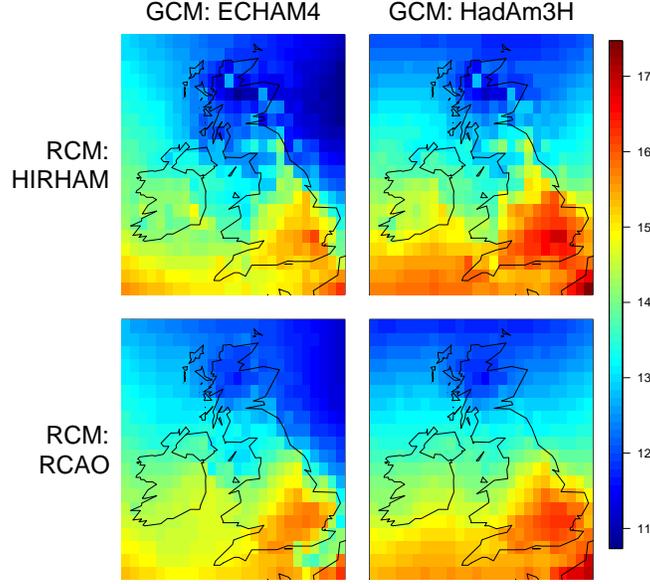


Figure 5: Average summer temperatures ($^{\circ}\text{C}$) in control runs (corresponding to 1961–1990) of the Prudence Project experiment, taken over the 30 years of model output.

Therefore, we used an expanded version of (1) with a single time effect γ :

$$Y_{ijt}(s) = \mu(s) + \alpha_i(s) + \beta_j(s) + (\alpha\beta)_{ij}(s) + \gamma t + \epsilon_{ijt}(s).$$

We take μ to have a single intercept parameter, with $\mu \sim GP(\mu_0, \sigma_{\mu}^2 R_{\rho_{\mu}})$. Here we take R_{ρ} for all processes to be the Matérn correlation function on \mathfrak{R}^2 with parameters ρ and $\nu = 2$. Because there are only two levels per factor, it is easy to reparameterize the effects to satisfy the sum to zero constraints. This simplifies the Gibbs sampling algorithm, as discussed in Section 2.3. Let $i = -1$ represent the RCM HIRHAM, and let $i = 1$ represent the RCM RCAO. Likewise, let $j = -1$ represent GCM ECHAM4, and let $j = 1$ represent GCM HadAm3H. Then let

$$\begin{aligned} \alpha_i &= i\alpha, & \alpha &\sim GP(0, \sigma_{\alpha}^2 R_{\rho_{\alpha}}) \\ \beta_j &= j\beta, & \beta &\sim GP(0, \sigma_{\beta}^2 R_{\rho_{\beta}}) \\ (\alpha\beta)_{ij} &= ij(\alpha\beta), & (\alpha\beta) &\sim GP(0, \sigma_{\alpha\beta}^2 R_{\rho_{\alpha\beta}}) \end{aligned}$$

The σ^2 values are rescaled compared to their definitions in Section 2.1, but this does not change the joint distribution. We interpret $\mu_{ijt} = \mu + i\alpha + j\beta + ij(\alpha\beta)_{ij} + \gamma t$ as the expected or climatological

temperature field under RCM i and GCM j at time t . Note that although the climate model output is deterministic, we can never know μ_{ijt} with certainty, due to fluctuations around μ_{ijt} from year to year within the model and the finite number of years we observe. The goal of this analysis is to carry out statistical inference for $\mu_{ijt}(s)$ and the elements of its ANOVA decomposition, given the observed model output. We assume that the observations are centered around μ_{ijt} (that is, that the models are in equilibrium), so we take ϵ_{ijt} to have mean zero, with $Y_{ijt}|\mu, \alpha, \beta, (\alpha\beta), \sigma_\epsilon^2, \rho_\epsilon \stackrel{iid}{\sim} GP(\mu + i\alpha + j\beta + ij(\alpha\beta), \sigma_\epsilon^2 R_{\rho_\epsilon})$.

We specify $\mu_0 \sim N(15, 6.25)$, assigning low probability to the event that μ_0 lies outside the interval $[10^\circ C, 20^\circ C]$. We again assign $p(\sigma^2) \propto 1/\sigma^2$ for all the variance parameters, and take $\rho \sim Gamma(1, 100)$ for all the range parameters. This allows for a variety of smoothness in the realizations, from functions which vary at the grid-scale to those that are essentially flat over the domain we are considering.

3.2.2 Posterior Sampling

Gibbs sampling for this example is straightforward, with normal full conditional distributions available for $\mu_0, \boldsymbol{\mu}, \boldsymbol{\alpha}, \boldsymbol{\beta}, (\boldsymbol{\alpha}\boldsymbol{\beta})$, and γ , with bold symbols indicating vectors of the corresponding functions evaluated at centers of the RCM grid boxes. The (σ^2, ρ) parameters can again be blocked, first sampling ρ from its distribution conditional on everything but σ^2 , then sampling σ^2 from its inverse gamma full conditional distribution. We generated ten thousand iterations, which took approximately 24 hours on 3.2 GHz dual processor compute node with 4 GB of memory. Based on visual inspection of the sample paths, we discarded the first 3000 samples for burn-in and retained every tenth sample thereafter, for a total of 700 roughly independent samples.

3.2.3 Graphical Posterior Summaries

Figure 6 shows the posterior means for $\boldsymbol{\mu}, \boldsymbol{\alpha}, \boldsymbol{\beta}$, and $(\boldsymbol{\alpha}\boldsymbol{\beta})$. There appears to be very little interaction between RCM and GCM. Most of the difference in the mean response is due to GCM, which imposes a large effect in terms of both magnitude and spatial extent. With boundary conditions provided by the HadAm3H GCM, the output tends to be warmer, particularly over the North Sea. The effect of RCM is smaller in magnitude for the majority of locations, and the effects are more localized. The RCM RCAO is cooler in the west and warmer in the east, although the direction of the effect

varies greatly around the coastline.

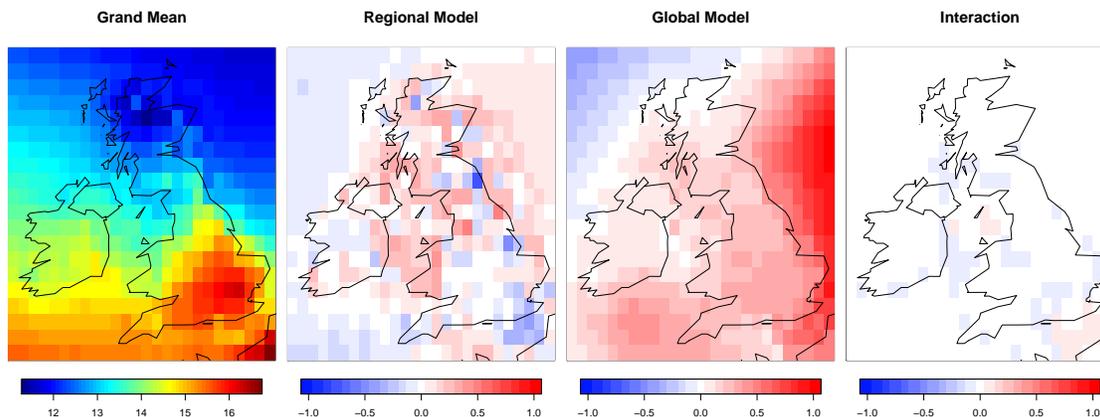


Figure 6: Posterior means of the grand mean μ , the main effect of regional model α , the main effect of global model β , and the interaction $(\alpha\beta)$. The effects shaded from blue to red are interpreted as deviations from the grand mean. Due to the ± 1 coding, the difference between levels are twice these values. The units for shading are $^{\circ}C$.

Figure 7 makes the comparison of the magnitude of the effects more explicit, plotting the posterior means for the finite population standard deviations corresponding to each effect. Because there are only two levels, the first three panels are directly related to the parameters whose means are plotted in Figure 6. Specifically, one can calculate under this parameterization that $s_{\alpha}^2(x) = 2\alpha^2(x)$, $s_{\beta}^2(x) = 2\beta^2(x)$, and $s_{\alpha\beta}^2(x) = 4(\alpha\beta)^2(x)$. The finite population standard deviation for the error or internal variability term, analogous to the error sum of squares, is large overall, with a mean that is larger over land than over oceans. It appears that the GCM is the largest source of variability for many locations, particularly in the North Sea.

However, these plots show only the posterior means of the finite population standard deviations; to make inference about their relative magnitudes, we need to take into account the full posterior distribution. One way of doing this is to plot the posterior probability of specific relationships between the finite population variances, as in Figure 8. The probability that the effect of regional model exceeds internal variability is large only for a few locations along the eastern coastline. The probability that the effect of global model exceeds internal variability is large mainly for locations in the North Sea, as well as a few locations to the northwest where the GCM HadAM3H tends to produce cooler temperatures. However, there are a number of locations for which the variability due to regional model exceeds that of global model with high probability. This indicates that although

both RCM and GCM have relatively small effects relative to the internal variability of the models, the choice of RCM does make a difference in the downscaling of the GCM for many locations.

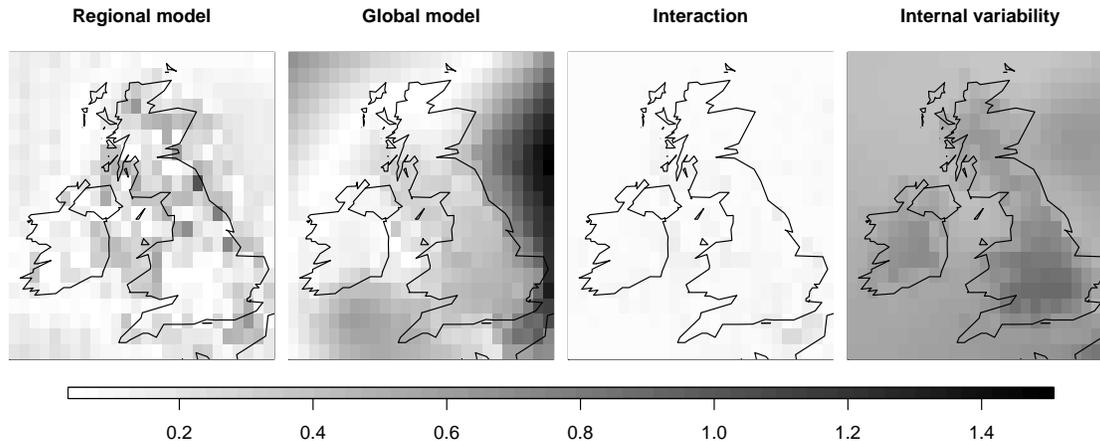


Figure 7: Posterior means of the finite population standard deviations for regional model (s_α), global model (s_β), interaction ($s_{\alpha\beta}$), and internal variability (s_ϵ). The units for shading are $^\circ C$.

4 Discussion

We have presented a general framework for Bayesian functional ANOVA modeling in arbitrary dimension with any number of crossed and/or nested factors. The model assigns Gaussian process prior distributions to each batch of functional parameters, corresponding to the levels of the various main effects and interactions. We impose prior dependence on each batch so that the functions satisfy identifiability constraints. These facilitate interpretation and numerical sampling of the unknown parameters. The posterior distributions of the finite population variances, which are model-based analogues of the traditional ANOVA decompositions of variance, can be compared graphically to analyze the contribution of each factor over various regions of the functional domain. In addition, we can obtain both point-wise and global credible intervals for any functional quantity of interest in the model. These intervals automatically incorporate uncertainty about the smoothness of the functions, by effectively integrating over prior uncertainty in the covariance parameters for each batch of Gaussian processes.

Our statement of the model in Section 2.1 was general, making very few assumptions about the form of the mean or covariance functions for the underlying Gaussian processes. We have

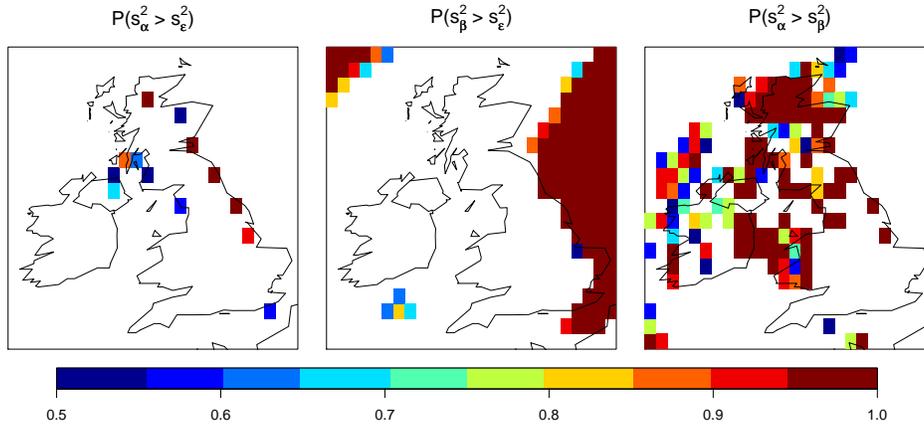


Figure 8: Shading indicates the posterior probabilities for relationships between the finite population variances. The first, $P(s_\alpha^2 > s_\epsilon^2)$, is the posterior probability that the effect of regional model is larger than the internal variability of the models. Likewise, $P(s_\beta^2 > s_\epsilon^2)$ is the posterior probability that the effect of global model is larger than internal variability. These are greater than 95% for only a fraction of the model locations. The final panel, $P(s_\alpha^2 > s_\beta^2)$, is the posterior probability that the effect of regional model is larger than the effect of global model. Probabilities less than 50% are not shaded.

used simple isotropic covariance functions of the sort often used in geostatistical models, although this is not a requirement of the model. The computational burden of using a strictly positive covariance function of this type will increase with the number of x values at which the response function is evaluated, as can be seen in the drastic difference in the time required to fit the model in Section 3.1, with 12 distinct x values, and Section 3.2, with 520 distinct x values. This is due to the computational difficulty of evaluating the determinant and inverse of the covariance matrices in the model, which grows as $O(p^3)$, where p is the number of distinct x values. To facilitate computation, it may be desirable to use a correlation with compact support (Gneiting, 2002), or to impose Markovian structure as in the generalized additive models of Fahrmeir and Lang (2001).

The model choices for the mean and covariance structure can and should be tailored to each particular data analysis, although we prefer simple prior choices over more complicated ones for the reasons discussed in Section 2.2.1. Note that although the prior structure may be simple, the process of conditioning on the observations introduces a variety of interesting nonstationarities, as illustrated in the plots of the posterior distributions for the finite population variances, which differ markedly across the domain of the functions.

Our focus has been on exploratory techniques and graphical summaries. However, it would be

possible to extend this framework to allow for more explicit testing of the possibility of null effects by allowing the distributions for the variance components σ^2 in each batch of Gaussian processes to have mass at zero, or to model the levels as a mixture distribution and determine their similarity by the posterior probability that they fall into the same component of the mixture. This would be a functional version of the ANOVA framework suggested by Nobile and Green (2000).

Acknowledgements

The data for the regional models example have been provided by Hayley Fowler (University of Newcastle, Newcastle upon Tyne) through the PRUDENCE data archive, funded by the EU through contract EVK2-CT2001-00132. The data are available to download from <http://prudence.dmi.dk/>. This research was supported by the North American Regional Climate Change Assessment Program under National Science Foundation grants ATM-03534131 and ATM-0534173, the Geophysical Statistics Project at the National Center for Atmospheric Research under National Science Foundation grant DMS-0355474, and the Statistical and Applied Mathematical Sciences Institute under National Science Foundation grant DMS-0112069.

References

- Barry, D. (1996). An empirical Bayes approach to growth curve analysis. *The Statistician*, 45:3–19.
- Berger, J., De Oliveira, V., and Sansó, B. (2001). Objective Bayesian analysis of spatially correlated data. *Journal of the American Statistical Association*, 96:1361–1374.
- Brumback, B. and Rice, J. (1998). Smoothing spline models for the analysis of nested and crossed samples of curves. *Journal of the American Statistical Association*, 93:961–976.
- Christensen, J., Carter, T., and Giorgi, F. (2002). PRUDENCE employs new methods to assess European climate change. *EOS, Transactions American Geophysical Union*, 83.
- Dawid, A. P. (1977). Invariant distributions and analysis of variance models. *Biometrika*, 64:291–297.

- Fahrmeir, L. and Lang, S. (2001). Bayesian inference for generalized additive mixed models based on Markov random field priors. *Applied Statistics*, 50:201–220.
- Gelfand, A. and Sahu, S. (1999). Identifiability, improper priors, and Gibbs sampling for generalized linear models. *Journal of the American Statistical Association*, 94:247–253.
- Gelfand, A., Sahu, S., and Carlin, B. (1995). Efficient parameterisations for normal linear mixed models. *Biometrika*, 82:479–488.
- Gelman, A. (2005). Analysis of variance - Why it is more important than ever. *The Annals of Statistics*, 33:1–53.
- Gneiting, T. (2002). Compactly supported correlation functions. *Journal of Multivariate Analysis*, 83:493–508.
- Gu, C. (2002). *Smoothing Spline ANOVA Models*. Springer.
- Gu, C. and Wahba, G. (1993). Smoothing spline ANOVA with component-wise Bayesian “confidence intervals”. *Journal of Computational and Graphical Statistics*, 2:97–117.
- Lindley, D. and Smith, A. (1972). Bayes estimates for the linear model. *Journal of the Royal Statistical Society, Series B (Methodological)*, 34:1–41.
- Matérn, B. (1986). *Spatial Variation*. Springer-Verlag, second edition.
- Nobile, A. and Green, P. (2000). Bayesian analysis of factorial experiments by mixture modelling. *Biometrika*, 87:15–35.
- Ramsay, J. and Silverman, B. (2005). *Functional Data Analysis*. Springer.
- Roberts, G. and Sahu, S. (1997). Updating schemes, correlation structure, blocking and parameterization for the gibbs sampler. *Journal of the Royal Statistical Society, Series B*, 59:291–317.
- Smith, A. (1973). Bayes estimates in one-way and two-way models. *Biometrika*, 60:319–329.
- Spitzner, D., Marron, J., and Essick, G. (2003). Mixed-model functional ANOVA for studying human tactile perception. *Journal of the American Statistical Association*, 98(98).

- Stein, M. (1999). *Interpolation of Spatial Data: Some Theory for Kriging*. Springer.
- Vines, S., Gilks, W., and Wild, P. (1996). Fitting Bayesian multiple random effects models. *Statistics and Computing*, 6:337–346.
- Wahba, G. (1978). Improper priors, spline smoothing and the problem of guarding against model errors in regression. *Journal of the Royal Statistical Society. Series B (Methodological)*, 40:364–372.
- Wahba, G. (1983). Bayesian “confidence intervals” for the cross-validated smoothing spline. *Journal of the Royal Statistical Society, Series B (Methodological)*, 45:133–150.
- Wang, Y., Ke, C., and Brown, M. (2003). Shape invariant modelling of circadian rhythms with random effects and smoothing spline ANOVA decompositions. *Biometrics*, 59:804–812.
- Yaglom, A. (1987). *Correlation theory of stationary and related random functions*. Springer-Verlag.
- Zhang, H. (2004). Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *Journal of the American Statistical Association*, 99:250–261.