Supplement of Geosci. Model Dev., 17, 1133–1151, 2024
https://doi.org/10.5194/gmd-17-1133-2024-supplement

Geoscientific Model Development
Open Access

Supplement of

Functional analysis of variance (ANOVA) for carbon flux estimates from remote sensing data

Jonathan Hobbs et al.

Correspondence to: Jonathan Hobbs (jonathan.m.hobbs@jpl.nasa.gov)

S1 Supplement

This supplement includes additional exposition of the statistical modeling assumptions in the functional analysis of variance (ANOVA) methodology. This is accompanied by an overview of the Markov chain Monte Carlo (MCMC) algorithm for Bayesian estimation of the parameters and functional ANOVA components.

S1.1 Spatial Statistical Model

The general functional ANOVA and corresponding notation follow Kaufman and Sain (2010),

$$y_{ijk}(\mathbf{s}) - y_{ijk}^{(0)}(\mathbf{s}) = \mu(\mathbf{s}) + \alpha_i(\mathbf{s}) + \beta_j(\mathbf{s}) + (\alpha\beta)_{ij}(\mathbf{s}) + \epsilon_{ijk}(\mathbf{s}), \tag{S1}$$

where $y_{ijk}(\mathbf{s})$ is a carbon flux estimate and $y_{ijk}^{(0)}(\mathbf{s})$ is the corresponding prior flux at spatial location $\mathbf{s}$. The prior flux fields $y_{ijk}^{(0)}(\mathbf{s})$ for the CMS-Flux inversions over Eurasia are shown in Figure S1. The prior flux fields for the OCO-2 MIP inversions are shown in Figure S2 (North America) and Figure S3 (Africa). In this study, the spatial fields are provided at N locations for each of the $n_\alpha \times n_\beta$ combinations of the experimental factors and for all n_ϵ replicates. For this collection of realizations, the model can be represented in a form with the N locations assembled into vectors,

$$\mathbf{y}_{ijk} - \mathbf{y}_{ijk}^{(0)} = \boldsymbol{\mu} + \boldsymbol{\alpha}_i + \boldsymbol{\beta}_j + (\boldsymbol{\alpha\beta})_{ij} + \boldsymbol{\epsilon}_{ijk} \tag{S2}$$

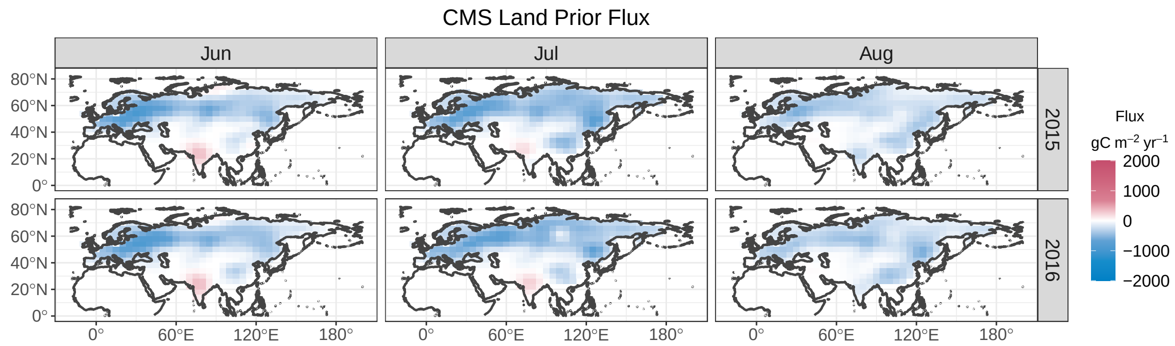

Figure S1. Prior fluxes $y_{ijk}^{(0)}(\mathbf{s})$ for the CMS-Flux inversions over Eurasia in the records of fused CO_2 investigation (units are gC m^{-2} yr^{-1}).

Further, for convenience, let $\boldsymbol{\Sigma}_\mu$ represent the covariance for the full spatial field $\boldsymbol{\mu}$, with analogous notation for the other ANOVA processes. Functional ANOVA models for spatial data often specify Gaussian process (GP) models with Matérn covariance structure as prior distributions for the ANOVA processes (Kaufman and Sain, 2010; Kang and Cressie, 2013). This approach provides a flexible representation of spatial correlation as a function of distance, but these GP models yield dense covariance matrices ($\boldsymbol{\Sigma}_\mu$), making the computations for MCMC expensive. Therefore, we use a Vecchia approximation (Katzfuss and Guinness, 2021) for each GP. This approximation facilitates computational demand for MCMC.

Since the $k = 1, \ldots, n_\epsilon$ replicates are consecutive months in the carbon flux examples, a more general treatment of the error terms $\boldsymbol{\epsilon}_{ijk}$ allows for temporal correlation. The joint $Nn_\epsilon \times Nn_\epsilon$ spatio-temporal covariance for the error vector for each treatment combination is defined as

$$\boldsymbol{\Sigma}_\epsilon = Cov(\boldsymbol{\epsilon}_{ij}),$$

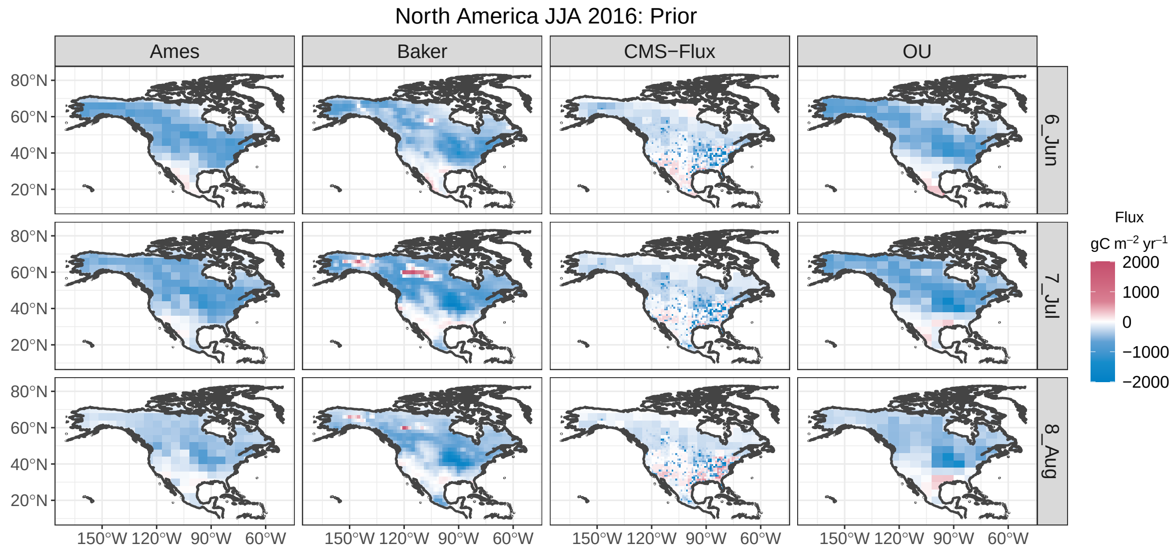

Figure S2. Prior fluxes $y_{ijk}^{(0)}(\mathbf{s})$ for the OCO-2 flux model intercomparison project over North America for JJA 2016 (units are gC m^{-2} yr^{-1}).

where the replicate error vectors are stacked,

$$\boldsymbol{\epsilon}_{ij} \equiv \begin{bmatrix} \boldsymbol{\epsilon}_{ij1} \\ \cdots \\ \boldsymbol{\epsilon}_{ijn_\epsilon} \end{bmatrix}.$$

Analogous definitions for stacked versions of the data vectors $\mathbf{y}_{ij}$ and flux priors $\mathbf{y}_{ij}^{(0)}$ are also used.

Each GP is parameterized by a standard deviation σ, a spatial range parameter λ, and a smoothness parameter ν. In addition, a temporal autocorrelation parameter ρ is included for the error process. These parameters are likely to vary across the GPs for the different ANOVA components, but they can be estimated from the data and can be included in the MCMC procedure. The fully Bayesian inference includes prior distributions for the GP parameters. In our implementation, we assume the same

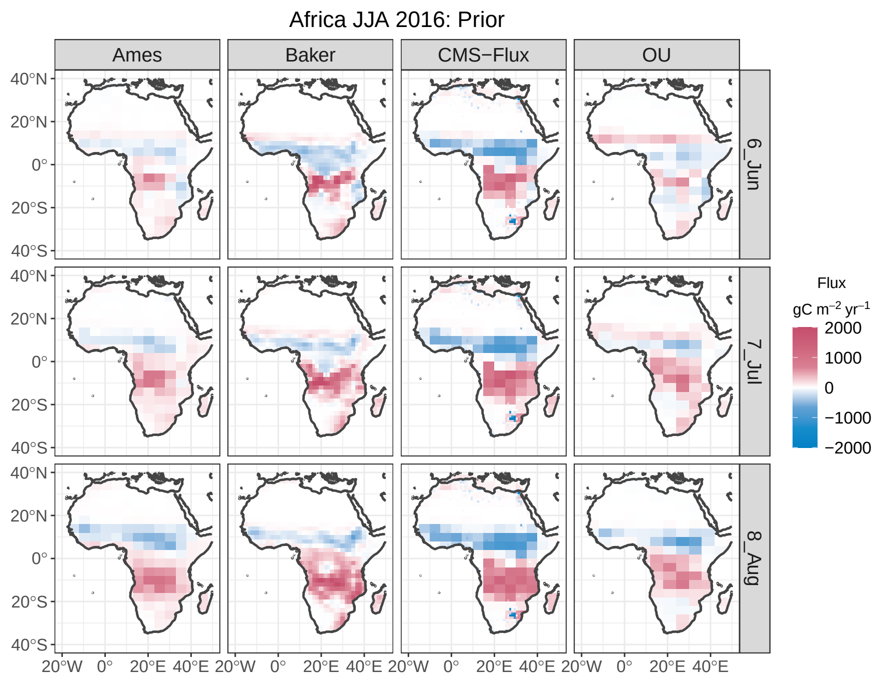

Figure S3. Prior fluxes $y_{ijk}^{(0)}(\mathbf{s})$ for the OCO-2 flux model intercomparison project over Africa for JJA 2016 (units are gC m^{-2} yr^{-1}).

functional form for a GP parameter prior across the ANOVA components. These forms are

$$
\begin{aligned}
\sigma_\epsilon &\sim \text{Half-normal}\,(r_\epsilon) \\
\pi(\sigma_\epsilon) &= \frac{\sqrt{2}}{r_\epsilon\sqrt{\pi}} \exp\left(-\frac{\sigma_\epsilon^2}{2r_\epsilon^2}\right),\ \sigma_\epsilon \geq 0 \\
\lambda_\epsilon &\sim \text{Lognormal}\,(v_\epsilon, w_\epsilon) \\
\pi(\lambda_\epsilon) &= \frac{1}{\lambda_\epsilon w_\epsilon \sqrt{2\pi}} \exp\left(-\frac{(\log\lambda_\epsilon - v_\epsilon)^2}{2w_\epsilon^2}\right),\ \lambda_\epsilon > 0 \\
\nu_\epsilon &\sim \text{Scaled Beta}\,(b_{1,\epsilon}, b_{2,\epsilon}, c_{1,\epsilon}, c_{2,\epsilon}) \\
\pi(\nu_\epsilon) &= \left(\frac{\nu_\epsilon - c_{1,\epsilon}}{c_{2,\epsilon} - c_{1,\epsilon}}\right)^{b_{1,\epsilon}-1} \left(\frac{c_{2,\epsilon} - \nu_\epsilon}{c_{2,\epsilon} - c_{1,\epsilon}}\right)^{b_{2,\epsilon}-1} \frac{\Gamma(b_{1,\epsilon} + b_{2,\epsilon})}{\Gamma(b_{1,\epsilon})\,\Gamma(b_{2,\epsilon})},\ c_{1,\epsilon} \leq \nu_\epsilon \leq c_{2,\epsilon} \\
\rho_\epsilon &\sim \text{Scaled Beta}\,(p_{1,\epsilon}, p_{2,\epsilon}, q_{1,\epsilon}, q_{2,\epsilon}) \\
\pi(\rho_\epsilon) &= \left(\frac{\rho_\epsilon - q_{1,\epsilon}}{q_{2,\epsilon} - q_{1,\epsilon}}\right)^{p_{1,\epsilon}-1} \left(\frac{q_{2,\epsilon} - \rho_\epsilon}{q_{2,\epsilon} - q_{1,\epsilon}}\right)^{p_{2,\epsilon}-1} \frac{\Gamma(p_{1,\epsilon} + p_{2,\epsilon})}{\Gamma(p_{1,\epsilon})\,\Gamma(p_{2,\epsilon})},\ q_{1,\epsilon} \leq \rho_\epsilon \leq q_{2,\epsilon}
\end{aligned}
$$

where the hyperparameters in parentheses are chosen to vary across ANOVA components. The half-normal form for the GP standard deviations is motivated by the literature on variance components in hierarchical models (Gelman, 2006). The scaled Beta distribution for ν constrains the support for the smoothness parameter and improves convergence of the MCMC sampler in our experience. The resulting prior distributions are shown in Figure S4.

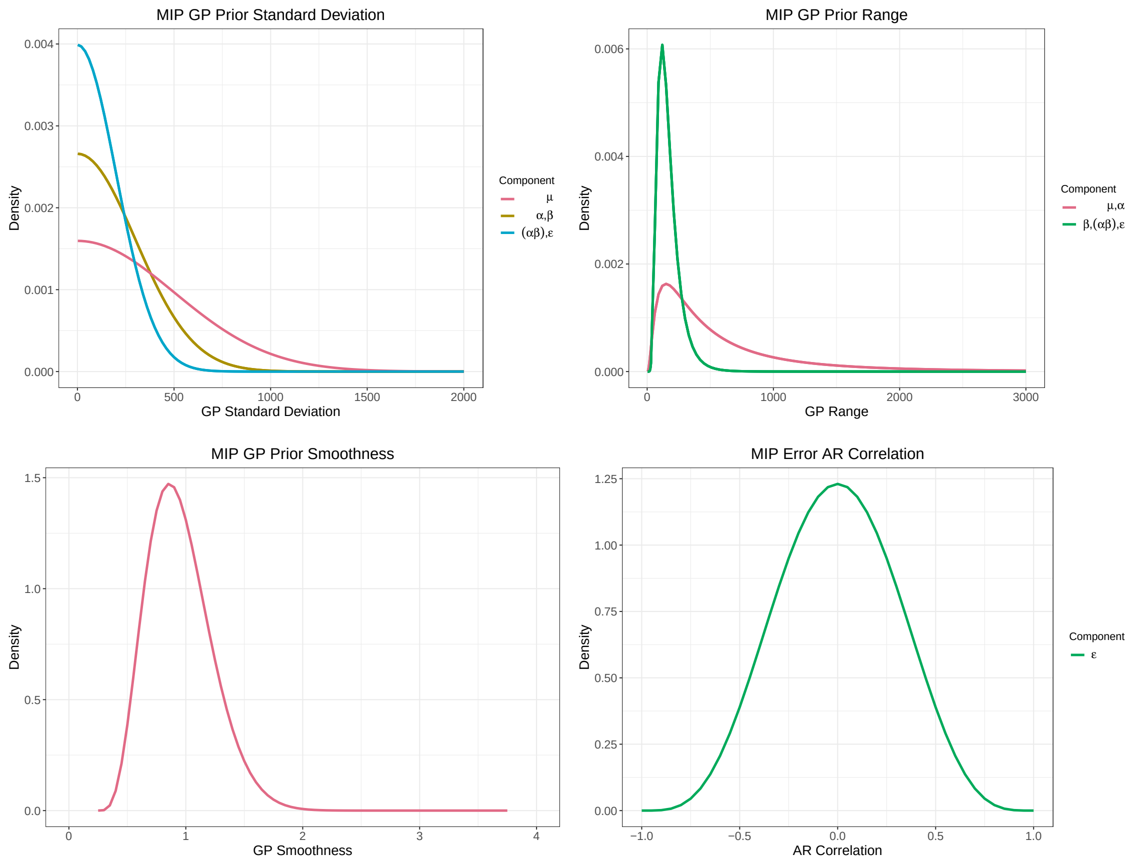

Figure S4. Prior densities for spatio-temporal covariance parameters. Top left panel depicts half-normal priors for GP standard deviation σ (units are gC m^{-2} yr^{-1}). Top right panel depicts lognormal priors for GP range λ (units are km). Bottom left panel depicts scaled Beta priors for GP smoothness ν. Bottom right panel depicts scaled Beta prior for the error process autoregressive (AR) correlation ρ_ϵ. ANOVA components have different prior parameters as indicated by colors in each panel.

S1.2 Parameterization via Contrasts

In classical ANOVA, the factorial model (S1) is over-parameterized. There are only $n_\alpha \times n_\beta$ unique group means for the various combinations of factors. This can be reconciled in the model by imposing constraints on the main effects and interactions. Kaufman and Sain (2010) extend this to functional ANOVA by imposing sum-to-zero constraints for each spatial location. For the first main effect, define $n_\alpha - 1$ contrast effects $\alpha^*_{i'}$. Then the original main effects can be related to the contrast effects through a $n_\alpha \times (n_\alpha - 1)$ contrast matrix $\mathbf{H}_\alpha$. The collection is a set of Helmert contrasts that are scaled such that $\mathbf{H}_\alpha^T \mathbf{H}_\alpha = \mathbf{I}$. Tables S1 and S2 provide the contrast matrices for the OCO-2 MIP examples where $n_\alpha = 4$ and $n_\beta = 2$.

S1.3 Gibbs Sampler

For many Bayesian hierarchical models, the posterior distribution is not readily available in closed form, but it can be sampled via MCMC. We implement a Gibbs sampler, which iterates conditional sampling among the stochastic nodes given all other

Table S1. Matrix of contrast coefficients, $\mathbf{H}_\alpha$, for the OCO-2 MIP functional ANOVA with four levels, $n_\alpha = 4$.

Level	Contrast		
	$i' = 1$	$i' = 2$	$i' = 3$
$i = 1$	-0.707107	-0.408248	-0.288675
$i = 2$	0.707107	-0.408248	-0.288675
$i = 3$	0	0.816496	-0.288675
$i = 4$	0	0	0.866025

Table S2. Matrix of contrast coefficients, $\mathbf{H}_\beta$, for the OCO-2 MIP functional ANOVA with four levels, $n_\beta = 2$.

Level	Contrast
	$j' = 1$
$j = 1$	-0.707107
$j = 2$	0.707107

elements and the data (Gelman et al., 2013). For the functional ANOVA model, this sampling alternates between updates for the various ANOVA component GPs and for the GP spatial dependence parameters and variances.

S1.3.1 Algorithm

1. Update for functional ANOVA spatial processes
   The ANOVA structure for a complete balanced design, with all levels of all factors observed, allows for convenient forms for conditional distributions that rely on marginal sums, sometimes termed "row" and "column" sums in classical ANOVA. Because of the zero-sum constraints, the MCMC updates are performed for the reduced rank GPs $\boldsymbol{\alpha}^*$, $\boldsymbol{\beta}^*$, and $(\boldsymbol{\alpha\beta}^*)$. In each case, the conditional posterior distribution, denoted e.g. $p(\boldsymbol{\mu}|\mathbf{y},\boldsymbol{\theta})$, given the data and GP model parameters $\boldsymbol{\theta}$, is a multivariate Gaussian distribution, with means and covariances defined below. At each iteration of the Gibbs sampler, these calculations are performed given the current values of the GP dependence parameters, and a draw from each of these multivariate Gaussian distributions is performed to update the ANOVA components.

   – For the mean field,

   $$p(\boldsymbol{\mu}|\mathbf{y}-\mathbf{y}^{(0)},\boldsymbol{\theta}) \sim \mathcal{N}\left(\mathbf{a}_\mu, \mathbf{S}_\mu\right)$$
   $$\mathbf{S}_\mu = \left(\boldsymbol{\Sigma}_\mu^{-1} + n_\alpha n_\beta \mathbf{D}_\epsilon^T \boldsymbol{\Sigma}_\epsilon^{-1}\mathbf{D}_\epsilon\right)^{-1}$$
   $$\mathbf{a}_\mu = \mathbf{S}_\mu \left(\mathbf{D}_\epsilon^T \boldsymbol{\Sigma}_\epsilon^{-1} \sum_{i=1}^{n_\alpha}\sum_{j=1}^{n_\beta}\left(\mathbf{y}_{ij} - \mathbf{y}_{ij}^{(0)}\right)\right)$$

   where $\mathbf{D}_\epsilon$ is a $Nn_\epsilon \times N$ sparse design matrix,

   $$\mathbf{D}_\epsilon = \mathbf{1}_{n_\epsilon} \otimes \mathbf{I}_N.$$

   – For the Factor A main effect(s), $\boldsymbol{\alpha}^*$, the following level-specific sums are computed for each $i = 1,\ldots,n_\alpha$,

   $$\mathbf{y}_{i\cdot\cdot} = \sum_{j=1}^{n_\beta}\left(\mathbf{y}_{ij} - \mathbf{y}_{ij}^{(0)}\right)$$

   Similarly, the contrast coefficients are applied to yield contrast sums,

   $$\mathbf{y}_{i'\cdot\cdot} = \sum_{i=1}^{n_\alpha} h_{\alpha(i,i')}\mathbf{y}_{i\cdot\cdot},$$

where $h_{\alpha(i,i')}$ is the row i, column i' element of $\mathbf{H}_\alpha$.
Then the conditional posterior distribution for the contrast effect $\boldsymbol{\alpha}^*_{i'}$

$$p(\boldsymbol{\alpha}^*_{i'}|\mathbf{y}-\mathbf{y}^{(0)},\boldsymbol{\theta}) \sim \mathcal{N}\left(\mathbf{a}_{\alpha,i'},\mathbf{S}_\alpha\right)$$
$$\mathbf{S}_\alpha = \left(\boldsymbol{\Sigma}_\alpha^{-1} + n_\alpha n_\beta \mathbf{D}_\epsilon^T \boldsymbol{\Sigma}_\epsilon^{-1}\mathbf{D}_\epsilon\right)^{-1}$$
$$\mathbf{a}_{\alpha,i'} = \mathbf{S}_\alpha\left(\mathbf{D}_\epsilon^T\boldsymbol{\Sigma}_\epsilon^{-1}\mathbf{y}_{i'\cdot\cdot}\right)$$

– For the Factor B main effect(s), $\boldsymbol{\beta}^*$, the following level-specific sums are computed for each $j = 1,\ldots,n_\beta$,

$$\mathbf{y}_{\cdot j\cdot} = \sum_{i=1}^{n_\alpha}\left(\mathbf{y}_{ij} - \mathbf{y}_{ij}^{(0)}\right)$$

The contrast coefficients are applied to yield contrast sums,

$$\mathbf{y}_{\cdot j'\cdot} = \sum_{j=1}^{n_\beta} h_{\beta(j,j')}\mathbf{y}_{\cdot j\cdot},$$

where $h_{\alpha(j,j')}$ is the row j, column j' element of $\mathbf{H}_\beta$.
Then the conditional posterior distribution for the contrast effect $\boldsymbol{\beta}^*_{j'}$

$$p(\boldsymbol{\beta}^*_{j'}|\mathbf{y}-\mathbf{y}^{(0)},\boldsymbol{\theta}) \sim \mathcal{N}\left(\mathbf{a}_{\beta,j'},\mathbf{S}_\beta\right)$$
$$\mathbf{S}_\beta = \left(\boldsymbol{\Sigma}_\beta^{-1} + n_\alpha n_\beta \mathbf{D}_\epsilon^T \boldsymbol{\Sigma}_\epsilon^{-1}\mathbf{D}_\epsilon\right)^{-1}$$
$$\mathbf{a}_{\beta,j'} = \mathbf{S}_\beta\left(\mathbf{D}_\epsilon^T\boldsymbol{\Sigma}_\epsilon^{-1}\mathbf{y}_{\cdot j'\cdot}\right)$$

– For the interaction effect(s), deviations for each treatment combination are retained,

$$\mathbf{y}_{ij\cdot} = \left(\mathbf{y}_{ij} - \mathbf{y}_{ij}^{(0)}\right)$$

In a similar fashion, the contrast coefficients are assembled,

$$\mathbf{y}_{i'j'\cdot} = \sum_{i=1}^{n_\alpha}\sum_{j=1}^{n_\beta} h_{\alpha(i,i')}h_{\beta(j,j')}\mathbf{y}_{ij\cdot}$$

Then the conditional posterior posterior distribution for the contrast effect $(\boldsymbol{\alpha\beta})^*_{i'j'}$,

$$p((\boldsymbol{\alpha\beta})^*_{i'j'}|\mathbf{y}-\mathbf{y}^{(0)},\boldsymbol{\theta}) \sim \mathcal{N}\left(\mathbf{a}_{(\alpha\beta),i'j'},\mathbf{S}_{(\alpha\beta)}\right)$$
$$\mathbf{S}_{(\alpha\beta)} = \left(\boldsymbol{\Sigma}_{(\alpha\beta)}^{-1} + n_\alpha n_\beta \mathbf{D}_\epsilon^T \boldsymbol{\Sigma}_\epsilon^{-1}\mathbf{D}_\epsilon\right)^{-1}$$
$$\mathbf{a}_{(\alpha\beta),i'j'} = \mathbf{S}_{(\alpha\beta)}\left(\mathbf{D}_\epsilon^T\boldsymbol{\Sigma}_\epsilon^{-1}\mathbf{y}_{i'j'\cdot}\right)$$

For each of these updates, if either $n_\alpha > 2$ or $n_\beta > 2$, the posterior means for multiple contrast effects can be solved simultaneously. The Vecchia approximation (Katzfuss and Guinness, 2021) enables efficient construction of the sparse Cholesky factors needed for simulating realizations from these multivariate distributions. These computations are implemented with the `GPvecchia` package in the R statistical computing environment (Katzfuss et al., 2020).

2. Update for spatio-temporal covariance parameters
   The spatial covariance parameters, e.g. $\boldsymbol{\theta}_\mu \equiv (\sigma_\mu, \lambda_\mu, \nu_\mu)$, for each ANOVA component are sampled with Metropolis-Hastings (MH) steps. The MH proposal distributions are multivariate Gaussian for the transformation $(\log(\sigma), \log(\lambda), \text{logit}(\nu))$. During a burn-in period for the Markov chain, the proposal covariance is adapted periodically (Gelman et al., 2013). These MCMC steps involve evaluation of the prior densities $\pi(\boldsymbol{\theta})$ and multivariate Gaussian likelihoods, e.g. $f(\boldsymbol{\mu}|\boldsymbol{\Sigma}_{\boldsymbol{\mu}})$, for the relevant GPs.

   – For the mean field, the conditional posterior includes the likelihood for the corresponding ANOVA GP,

     $$p\left(\sigma_\mu, \lambda_\mu, \nu_\mu | \boldsymbol{\mu}\right) \propto \pi\left(\sigma_\mu\right) \pi\left(\lambda_\mu\right) \pi\left(\nu_\mu\right) f\left(\boldsymbol{\mu} | \boldsymbol{\Sigma}_\mu\right)$$

   – For the main and interaction effects, the conditional posterior includes the likelihood for each of the contrast effects. For example, this posterior for the Factor A main effect is

     $$p\left(\sigma_\alpha, \lambda_\alpha, \nu_\alpha | \boldsymbol{\alpha}^*\right) \propto \pi\left(\sigma_\alpha\right) \pi\left(\lambda_\alpha\right) \pi\left(\nu_\alpha\right) \prod_{i'=1}^{n_\alpha - 1} f\left(\boldsymbol{\alpha}_{i'}^* | \boldsymbol{\Sigma}_\alpha\right)$$

   – For the ANOVA error term, the update includes the spatial covariance parameters, as well as the temporal autocorrelation. This results in a four-element MH proposal, including $\text{logit}(\rho)$. The conditional posterior combines the likelihood across all replicates,

     $$p\left(\sigma_\epsilon, \lambda_\epsilon, \nu_\epsilon, \rho_\epsilon | \boldsymbol{y} - \boldsymbol{y}^{(0)}\right) \propto \pi\left(\sigma_\epsilon\right) \pi\left(\lambda_\epsilon\right) \pi\left(\nu_\epsilon\right) \pi\left(\rho_\epsilon\right) \prod_{i=1}^{n_\alpha} \prod_{j=1}^{n_\beta} f\left(\boldsymbol{y}_{ij} - \boldsymbol{y}_{ij}^{(0)} | \boldsymbol{\mu}, \boldsymbol{\alpha}^*, \boldsymbol{\beta}^*, (\boldsymbol{\alpha\beta})^*, \boldsymbol{\Sigma}_\epsilon\right)$$

S1.3.2 Initialization

The full MCMC algorithm can successfully converge if the parameters and functional ANOVA states are initialized to reasonable values. In addition, experience indicates that fixing some GP parameters in the preliminary MCMC burn-in also aids convergence. The procedure outlined below has proved successful for the examples in this work as well as tests with simulated data.

1. Compute single-location ANOVA estimates. Location-specific ANOVA models (S1) are fit by ordinary least squares.

2. Perform grid-search for preliminary estimates of GP parameters. Given the initial ANOVA component estimates, the GP likelihood is evaluated over a coarse grid of values for the GP parameters $\boldsymbol{\theta}$ for all components (mean, main and interaction effects, noise).

3. Run burn-in for ANOVA elements only. The functional ANOVA states in the MCMC are initialized to modest random perturbations of the least-squares estimates, and a short MCMC is run with GP parameters fixed at the values obtained in the grid search. This MCMC is a Gibbs sampler only with no MH updates.

4. Run full adaptive MCMC. Starting from the final MCMC states from the previous step, the ANOVA components and GP parameters are sampled with the full Metropolis-within-Gibbs algorithm. The MH proposal covariances are updated adaptively in this stage. This burn-in was run for 20,000 iterations.

5. Run post burn-in sampling with fixed MH proposals. Suitable MH proposal covariances are fixed for the GP parameter updates, and the full MCMC is run for an additional 50,000 iterations. The posterior summaries are constructed from only this stage of the MCMC, where the chain has reached the stationary distribution.

References

Gelman, A.: Prior distributions for variance parameters in hierarchical models, Bayesian Anal., 1, 515–534, https://doi.org/10.1214/06-BA117A, 2006.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B.: Bayesian Data Analysis, Chapman & Hall/CRC, Boca Raton, FL, third edn., 2013.

Kang, E. L. and Cressie, N.: Bayesian Hierarchical ANOVA of Regional Climate-Change Projections from NARCCAP Phase II, Int. J. Appl. Earth Obs. Geoinf., 22, 3–15, https://doi.org/10.1016/j.jag.2011.12.007, 2013.

Katzfuss, M. and Guinness, J.: A general framework for Vecchia approximations of Gaussian processes, Stat. Sci., 36, 124–141, https://doi.org/10.1214/19-STS755, 2021.

Katzfuss, M., Jurek, M., Zilber, D., and Gong, W.: GPvecchia: Scalable Gaussian-Process Approximations, R package version 0.1.3, 2020.

Kaufman, C. G. and Sain, S. R.: Bayesian Functional ANOVA Modeling Using Gaussian Process Prior Distributions, Bayesian Anal., 5, 123–150, https://doi.org/10.1214/10-BA505, 2010.