

# My title

Justice Akuoko-Frimpong, Jonathan Ta      Author Two

## 1 Introduction

Spatial heterogeneity in environmental processes presents significant challenges for traditional regression modeling approaches. Consider air pollution monitoring across an urban-rural gradient: the relationship between particulate matter (PM2.5) concentrations and predictor variables like traffic density or industrial emissions often varies substantially across space. Global regression models that assume constant coefficients fail to capture these localized relationships, potentially leading to incorrect scientific conclusions and suboptimal policy decisions. Our work develops a computationally efficient Bayesian implementation of spatially varying coefficient (SVC) models, which address this limitation by allowing regression coefficients to change smoothly over space. Let  $s_i \in \mathbb{R}^2$  for each  $i = 1, \dots, n$  denote spatial locations where we observe response variables  $Y(\mathbf{s})$  and covariates  $\mathbf{X}(\mathbf{s}) = (X_1(\mathbf{s}), \dots, X_p(\mathbf{s}))^\top$ . The SVC model specification:

$$Y(\mathbf{s}) = \sum_{j=1}^p X_j(\mathbf{s})w_j(\mathbf{s}) + \epsilon(\mathbf{s}),$$

where  $w_r(\mathbf{s})$  are spatially varying coefficients modeled as Gaussian processes:

$$w_r \sim \text{GP}(0, C(s, s'))$$

$$C(s, s') = \sigma_r^2 \exp(-\phi_r^{-1} \|s - s'\|^2)$$

with  $\epsilon(\mathbf{s}) \sim \text{MVN}(0, \tau^2 I_n)$  representing measurement error. The key innovations in our implementation include:

## Computational Algorithm

We implement a Markov Chain Monte Carlo (MCMC) algorithm with three key innovations:

1. **Low rank Gaussian Process:** Using knots to reduce the  $O(n^3)$  computational complexity of Gaussian process models.
2. **Adaptive Metropolis-Hastings:** For updating spatial range parameters  $\phi_r$  with automatic tuning of proposal distributions:

$$\phi_r^{(new)} = \phi_r^{(current)} + \epsilon, \quad \epsilon \sim N(0, \sigma_{adapt}^2)$$

where  $\sigma_{adapt}^2$  adjusts based on acceptance rates during burn-in.

3. **Efficient linear algebra:** Utilizing Cholesky decompositions for all covariance matrix operations, with pre-computation of constant distance matrices:

## 2 Methods

### 2.1 SVC Model

Let  $s_i \in \mathbb{R}^2$  for each  $i = 1, \dots, n$  denote a spatial location for which we have collected data, and  $\mathbf{s} = (s_1, \dots, s_n)^\top$  be the vector of all such locations. Then  $Y(\mathbf{s})$  are univariate dependent variables and  $\mathbf{X}(\mathbf{s}) = (X_1(\mathbf{s}), \dots, X_p(\mathbf{s}))^\top$  are  $p \times 1$  vectors of covariates. A linear regression model with spatially varying coefficients assumes  $Y(\mathbf{s})$  are dependent on

$\mathbf{X}(\mathbf{s})$  as follows:

$$Y(\mathbf{s}) = \sum_{j=1}^p X_r(\mathbf{s})w_r(\mathbf{s}) + w_0(\mathbf{s}) + \epsilon(\mathbf{s}),$$

where  $w_r(\mathbf{s})$  are the regression coefficients corresponding to  $X_r(\mathbf{s})$ ,  $w_0(\mathbf{s})$  are spatial random effects, and  $\epsilon(\mathbf{s})$  are independently and identically distributed multivariate normal (MVN) measurement errors, i.e.

$$\epsilon(\mathbf{s}) \sim \text{MVN}(0, \tau^2 I_n).$$

For convenience, we will treat  $w_0(\mathbf{s})$  as the regression coefficients corresponding to  $X_0(\mathbf{s}) = \mathbf{1}_n$ , a vector of length  $n$  with every element equal to 1.

For each  $w_r(\mathbf{s})$ , we assign a Gaussian process (GP) prior with squared exponential covariance, i.e.

$$w_r \sim \text{GP}(0, C(s, s')), \text{ where}$$

$$C(s, s') = \sigma_r^2 \exp\{\phi_r^{-1} \|s - s'\|^2\}.$$

We refer to  $C(\cdot, \cdot)$  as the covariance function, which calculates the covariance between two locations  $s$  and  $s'$ . Squared exponential covariance is a commonly used covariance function due to useful properties such as infinite smoothness and gradual decrease in covariance with distance. The parameter  $\sigma_r^2$  is the spatial variance and  $\phi_r$  is the spatial decay, which indicates how quickly correlation decreases with the squared distance.

We assign inverse-gamma conjugate priors for the variance parameters  $\tau^2$  and each  $\sigma_r^2$ . In addition, each  $\phi_r$  is assigned a uniform prior. In summary,

$$\tau^2 \sim \text{Inv.Gamma}(\alpha_\tau, \beta_\tau),$$

$$\sigma_r^2 \sim \text{Inv.Gamma}(\alpha_r, \beta_r), \text{ and}$$

$$\phi_r \sim \text{Uniform}(l_r, u_r).$$

## 2.2 MCMC Algorithm

The primary function in the `svc` package, `svclm`, samples from the joint posterior distribution of  $w_r(\mathbf{s})$ ,  $\phi_r$ ,  $\sigma_r^2$ , and  $\tau^2$ , for  $r = 0, \dots, p$  using a Gibbs sampler. We will explain how each parameter is sampled at iteration  $t + 1$ .

### 2.2.1 Spatial Decay

The parameter  $\phi_r^{(t+1)}$  is updated via a random walk Metropolis algorithm. Because  $\phi_r^{(t)} \in (l_r, u_r)$ , each proposal is calculated as

$$\phi'_r = f^{-1}(f(\phi_r^{(t)}) + U) = l_r + \frac{u_r - l_r}{1 + \exp\left\{\log\left(\frac{u_r - l_r}{\phi_r^{(t)} - l_r} - 1\right) + U\right\}},$$

where  $U$  is sampled from a normal distribution with mean 0.

The initial standard deviation for  $U$  is specified using the "phi\_proposal\_sd\_start" argument or is 1 by default. Then the standard deviation is updated at each iteration via a Robust Adaptive Metropolis (RAM) algorithm (Vihola 2012).

### 2.2.2 Spatially Varying Coefficients

Let  $\tilde{Y}(\mathbf{s}) = Y(\mathbf{s}) - \sum_{j \neq r} X_j(\mathbf{s})w_j^{(t)}(\mathbf{s})$ . Then

$$\frac{\tilde{Y}}{X_r}(\mathbf{s}) | w_r^{(t)}(\mathbf{s}), \tau^2 \sim \text{MVN}\left(w_r^{(t)}(\mathbf{s}), \tau^{2(t)} \text{diag}\left(\frac{1}{X_r^2(\mathbf{s})}\right)\right) := \text{MVN}(\mu, \Sigma).$$

Then assuming the prior  $w_r(\mathbf{s}) \sim \text{MVN}\left(0, \sigma_r^{2(t)} C\left(\phi_r^{(t)}\right)\right) := \text{MVN}(0, \Sigma_0)$ , we sample  $w_r^{(t+1)}(\mathbf{s})$  from  $\text{MVN}(\mu_1, \Sigma_1)$ , where

$$\Sigma_1 = (\Sigma_0^{-1} + \Sigma^{-1}) \text{ and } \mu_1 = \Sigma_1 \Sigma^{-1} \frac{\tilde{Y}}{X_r}(\mathbf{s}).$$

### 2.2.3 Spatial Variance

We sample  $\sigma_r^{2(t+1)}$  from  $\text{Inv.Gamma}\left(\alpha_r + \frac{n}{2}, \beta_r + w_r^{(t)}(\mathbf{s})^\top C\left(\phi_r^{(t)}\right)^{-1} w_r^{(t)}(\mathbf{s})\right)$ .

#### 2.2.4 Nugget

We sample  $\tau^{2(t+1)}$  from  $\text{Inv.Gamma}\left(\alpha_t + \frac{n}{2}, \beta_t + \frac{1}{2} \left( Y(\mathbf{s}) - \sum_{j=0}^p X_r(\mathbf{s}) w_r^{(t)}(\mathbf{s}) \right)\right)$ .

### 3 Data Analysis

The real world data used in this section is a gridded satellite data with latitude, longitude, temperature, Normalized Difference Vegetation Index (NDVI); which measures vegetation health/greenness, and emissivity which represents a surface's efficiency in emitting thermal radiation. We used temperature as a univariate outcome and used emissivity and NDVI as covariates. The variables showed moderate spatial autocorrelation with low to moderate variance.

#### 3.1 Model fitting

### 4 Conclusion

Conclusion