A Multivariate Spatio-Temporal Fay-Herriot Model for Forest Carbon Pools Across the Contiguous US

Elliot Shannon, Andrew O. Finley, Paul May, Sudipto Banerjee

Department of Forestry
Department of Statistics and Probability
Michigan State University

The Joint Statistical Meetings, August 6, 2024
Portland, Oregon

Motivation

- The United Nations require annual reporting of greenhouse gas emissions from five sectors:
  - Energy
  - Industry
  - Agriculture
  - **Forestry**
  - Waste
- The quality of these reports relies on the data and methods used to make estimates.

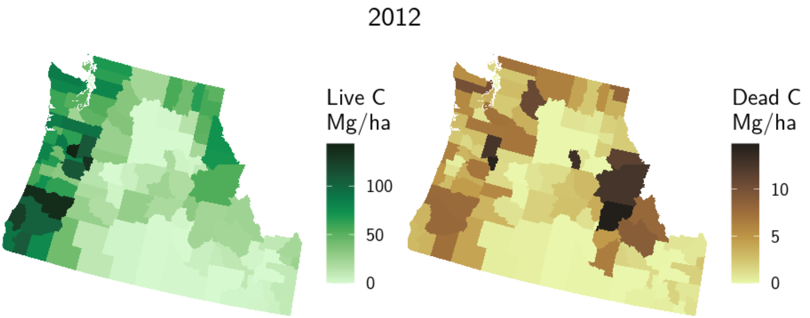
Source: Eggleston et al. (2006)

Motivation

- The Forest Inventory and Analysis (FIA) program of the US Forest Service measures forest carbon at inventory plots for different carbon "pools".
    - Live trees
    - Dead trees
    - Leaf litter
    - Soil
- Interested in status, trend, and change estimates at fine spatial and temporal scales (i.e., annual county-level estimates).

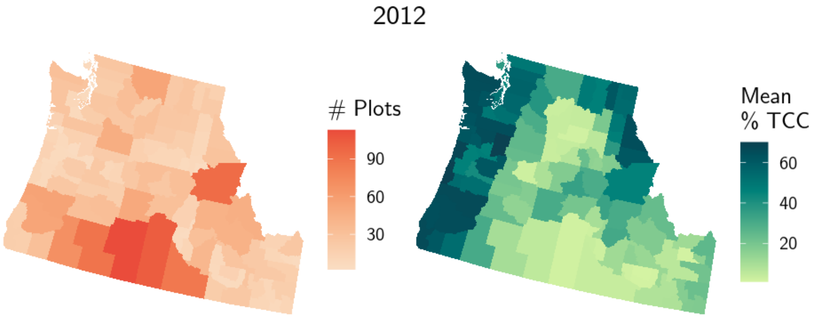
Data

- Carbon density measurements for each pool are collected on a rotating basis from inventory plots.
- Direct estimates are calculated at the county level for years 2008 – 2021.
- Due to sparseness of inventory plots and the small areas of interest, direct estimates may be missing for some counties/years.



Data

- Uncertainty in direct estimates from few inventory plots and/or measurement variation within a given county.
- Auxiliary data such as remotely sensed tree canopy cover (TCC) may be informative as a covariate.



Objectives

- Develop a dynamic, non-stationary, multivariate spatio-temporal model that:
    - accommodates spatial and temporal dependence *between* pools,
    - acknowledges dependence *among* pools, while allowing non-stationary correlations,
    - incorporate uncertainty stemming from direct estimates due to sample size and/or variability in measurements,
    - leverages available spatially and/or spatio-temporally varying predictors
    - provides improved carbon density estimates for each pool in all counties/years.

Direct Estimates

Let $y_{m,i,j,t}$ be the measured carbon density for the m^{th} carbon pool from the i^{th} plot in county j at time t, with $m = 1, \ldots, M$, $i = 1, \ldots, n_{m,j,t}$, $j = 1, \ldots, J$, and $t = 1, \ldots, T$.

The direct estimate of the mean carbon density of pool m in county j at time t is

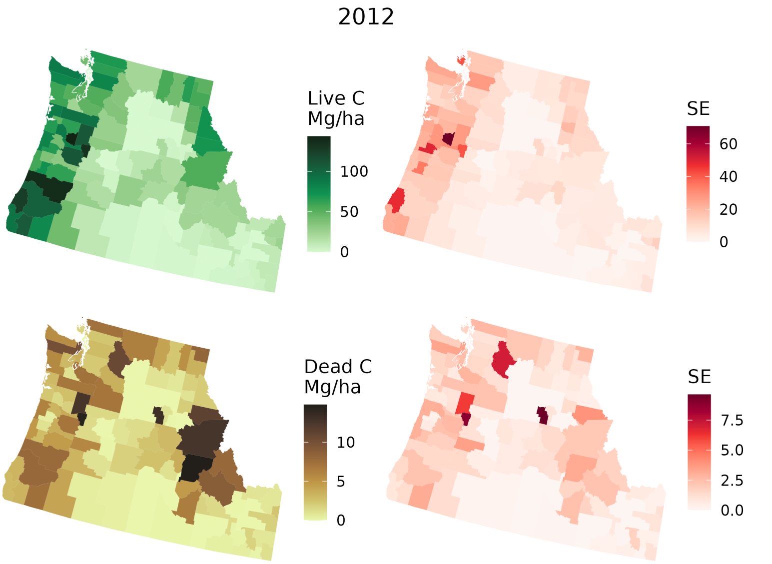
$$\hat{\mu}_{m,j,t} = \frac{1}{n_{m,j,t}} \sum_{i=1}^{n_{m,j,t}} y_{m,i,j,t} \tag{1}$$

with associated variance

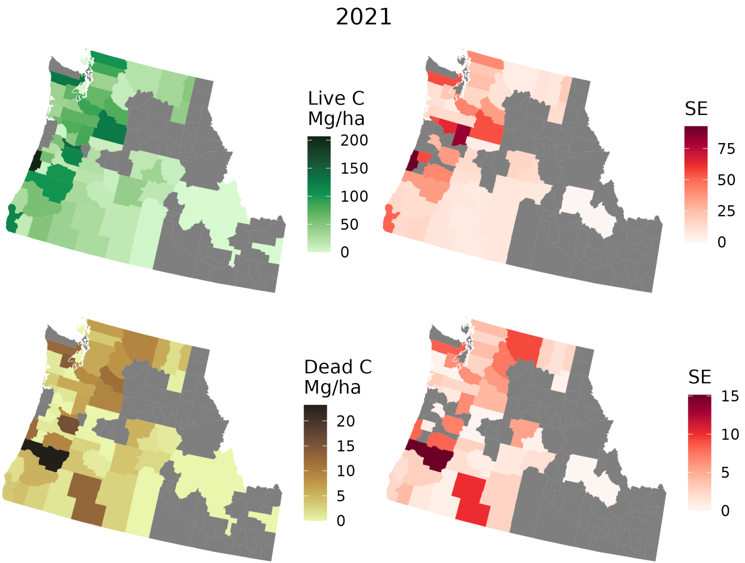
$$\hat{\sigma}^2_{m,j,t} = \frac{1}{n_{m,j,t}(n_{m,j,t} - 1)} \sum_{i=1}^{n_{m,j,t}} (y_{m,i,j,t} - \hat{\mu}_{m,j,t})^2 \tag{2}$$

Missingness occurs when $n_{m,j,t} = 0$, $n_{m,j,t} = 1$, or all measurements with a county/year are equal.

Direct Estimates



Direct Estimates



Proposed Model

For county j at time t, we are interested in the length M vector of latent means $\boldsymbol{\mu}_{j,t} = (\mu_{1,j,t}, \ldots, \mu_{M,j,t})^\top$.

We leverage direct estimates $\hat{\boldsymbol{\mu}}_{j,t} = (\hat{\mu}_{1,j,t}, \ldots, \hat{\mu}_{M,j,t})^\top$ and $\hat{\boldsymbol{\sigma}}^2_{j,t} = (\hat{\sigma}^2_{1,j,t}, \ldots, \hat{\sigma}^2_{M,j,t})^\top$, along with auxiliary data available for each county and time, $\mathbf{X}_{j,t}$.

The proposed Fay–Herriot (FH) model is

$$\hat{\boldsymbol{\mu}}_{j,t} = \boldsymbol{\mu}_{j,t} + \boldsymbol{\delta}_{j,t} \tag{3}$$

$$\boldsymbol{\mu}_{j,t} = \mathbf{X}_{j,t}\boldsymbol{\beta}_j + \mathbf{u}_{j,t} + \boldsymbol{\varepsilon}_{j,t} \tag{4}$$

Proposed Model

Model Components:

$$\hat{\boldsymbol{\mu}}_{j,t} = \boldsymbol{\mu}_{j,t} + \boldsymbol{\delta}_{j,t} \tag{3}$$

$$\boldsymbol{\mu}_{j,t} = \mathbf{X}_{j,t}\boldsymbol{\beta}_j + \mathbf{u}_{j,t} + \boldsymbol{\varepsilon}_{j,t} \tag{4}$$

- $\boldsymbol{\delta}_{j,t} \overset{\text{ind}}{\sim} MVN(\mathbf{0}, \boldsymbol{\Sigma}_{\delta,j,t})$ Direct estimate error term.
- $\mathbf{X}_{j,t}$ $M \times MP$ block diagonal matrix of P many predictor variables.
- $\boldsymbol{\beta}_j$ length MP vector of county-varying regression coefficients.
- $\mathbf{u}_{j,t}$ length M vector of county- and time-varying intercepts.
- $\boldsymbol{\varepsilon}_{j,t} \overset{\text{iid}}{\sim} MVN(\mathbf{0}, \boldsymbol{\Sigma}_{\varepsilon})$ Latent error term.

Spatio-temporal intercept

For $t = 1, \ldots, T$,

$$\mathbf{u}_{j,t} = \mathbf{u}_{j,t-1} + \mathbf{w}_{j,t}, \quad (\mathbf{u}_{j,0} \equiv 0) \tag{5}$$

$$\mathbf{w}_{j,t} = \mathbf{A}\mathbf{v}_{j,t} \tag{6}$$

$\mathbf{A}$ is the cholesky square root of the spatial random effects cross covariance matrix.
Elements of $\mathbf{v}_{j,t} = (v_{1,j,t}, \ldots, v_{M,j,t})^\top$ are modeled as

$$\begin{pmatrix} v_{m,1,t} \\ \vdots \\ v_{m,J,t} \end{pmatrix} \sim MVN\left(\mathbf{0}, \mathbf{Q}(\rho_{v,m,t})^{-1}\right), \quad m = 1, \ldots, M \tag{7}$$

where $\mathbf{Q}(\rho_{v,m,t})^{-1}$ is a CAR correlation matrix with spatial correlation parameter $\rho_{v,m,t}$ (Besag, 1974).

Spatially-varying coefficient

We have $\boldsymbol{\beta}_j = (\boldsymbol{\beta}_{1,1,j}, \ldots, \boldsymbol{\beta}_{M,P,j})^\top$.

The elements of $\boldsymbol{\beta}_j$ are again modeled using the CAR structure

$$\begin{pmatrix} \beta_{m,p,1} \\ \vdots \\ \beta_{m,p,J} \end{pmatrix} \sim MVN\left(\mathbf{0}, \tau^2_{\beta,m,p}\mathbf{Q}(\rho_{\beta,m,p})^{-1}\right), \quad \begin{matrix} m = 1, \ldots, M \\ p = 1, \ldots, P \end{matrix} \tag{8}$$

where $\tau^2_{\beta,m,p}\mathbf{Q}(\rho_{\beta,m,p})^{-1}$ is the CAR covariance matrix with scalar variance $\tau^2_{\beta,m,p}$ and spatial correlation parameter $\rho_{\beta,m,p}$.

Direct estimate variance

We consider $\mathbf{\Sigma}_{\delta,j,t} = \text{diag}\left(\sigma^2_{1,j,t}, \ldots, \sigma^2_{M,j,t}\right)$. where each $\sigma^2_{m,j,t}$ is given an

Inverse Gamma prior of the form

$$\sigma^2_{m,j,t} \sim IG\left(\frac{n_{m,j,t}}{2}, \frac{(n_{m,j,t}-1)\hat{\sigma}^2_{m,j,t}}{2}\right) \tag{9}$$

for $m = 1, \ldots, M$.

Proposed Model

Latent error:
A basic specification assumes $\mathbf{\Sigma}_\varepsilon = \text{diag}(\sigma^2_{\varepsilon,1}, \ldots, \sigma^2_{\varepsilon,M})$, with each $\sigma^2_{\varepsilon,m}$ following an Inverse Gamma prior.

Priors:
Finally, we specify vague Inverse Gamma priors for variance terms τ^2's, Uniform$(0, 1)$ priors on spatial dependence parameters ρ's, and an Inverse Wishart prior on $\mathbf{AA}^\top$.

Posterior samples are generated using Gibbs sampling for parameters with explicit full conditional distributions, and Metropolis steps for all other parameters.

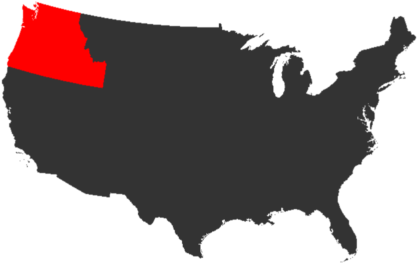
Preliminary Results Setting

- **Data**: Live and dead tree carbon density in Washington, Oregon, and Idaho from 2008 – 2021.

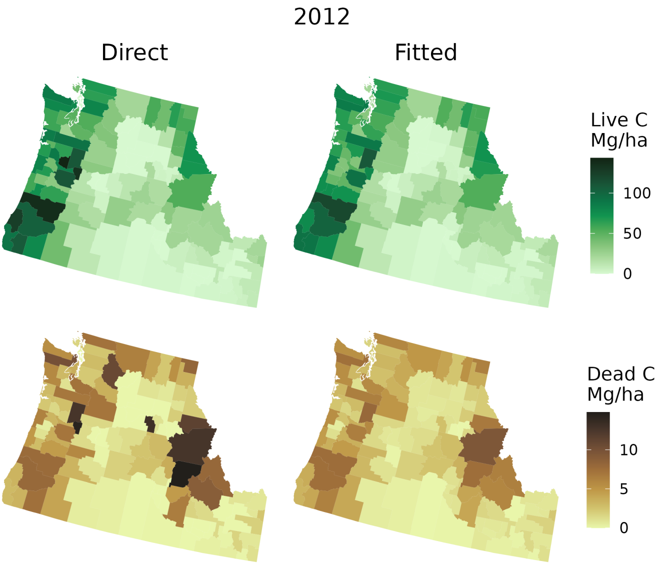
- We do not implement space varying coefficients β_j, but do incorporate dynamically evolving coefficients β_t.

$$\boldsymbol{\beta}_t = \boldsymbol{\beta}_{t-1} + \boldsymbol{\eta}_t \tag{10}$$

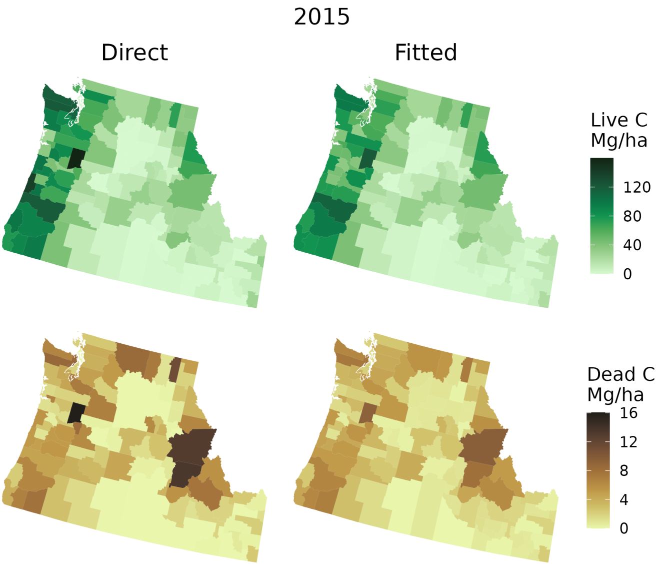
$$\boldsymbol{\eta}_t \sim MVN(0, \boldsymbol{\Sigma}_\eta) \tag{11}$$



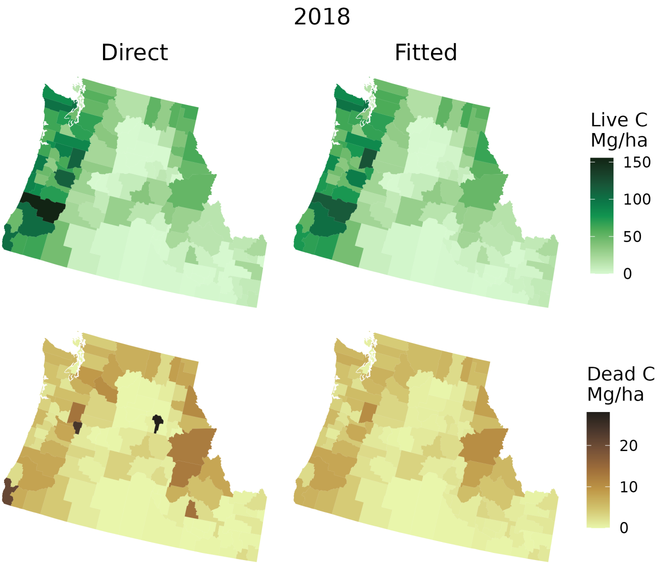
Preliminary Results



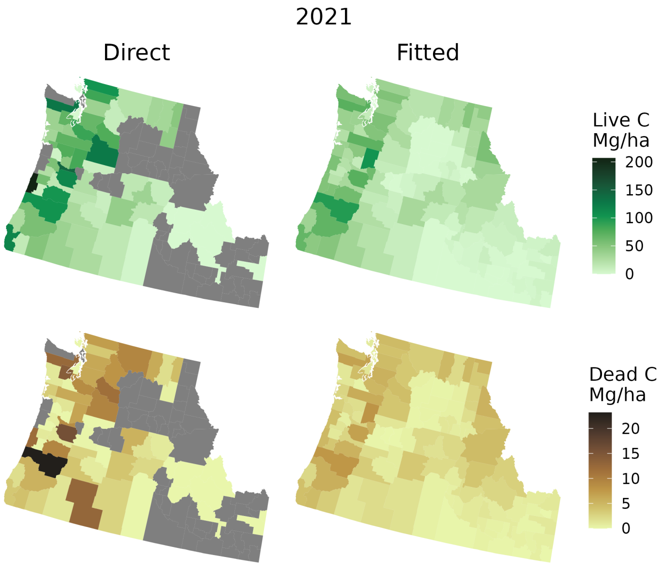
Preliminary Results



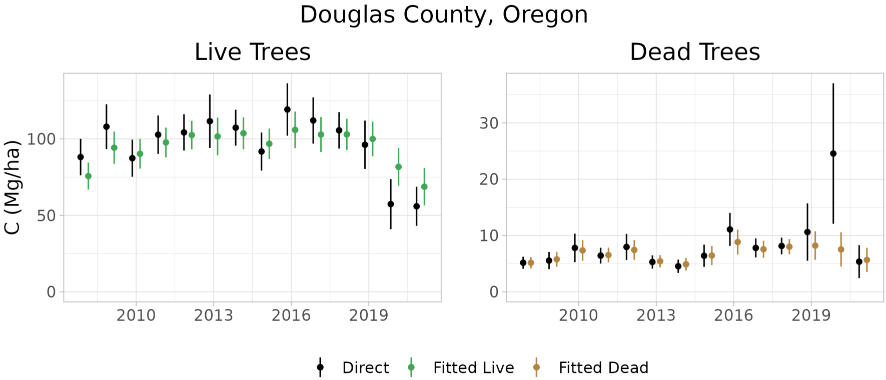
Preliminary Results



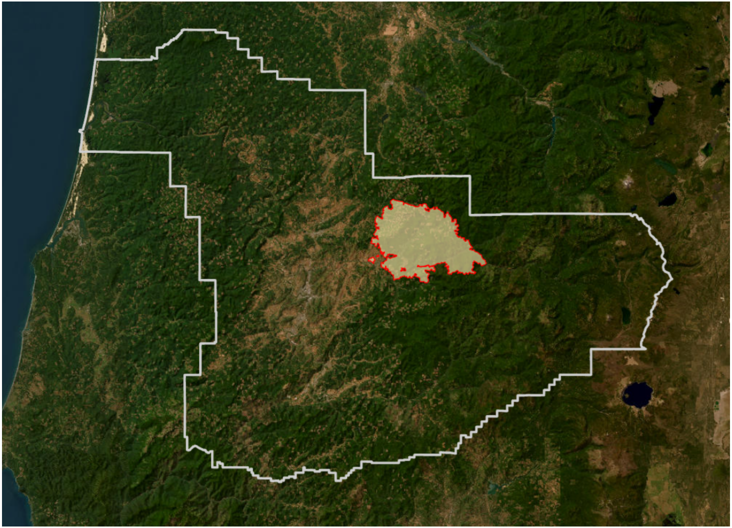
Preliminary Results



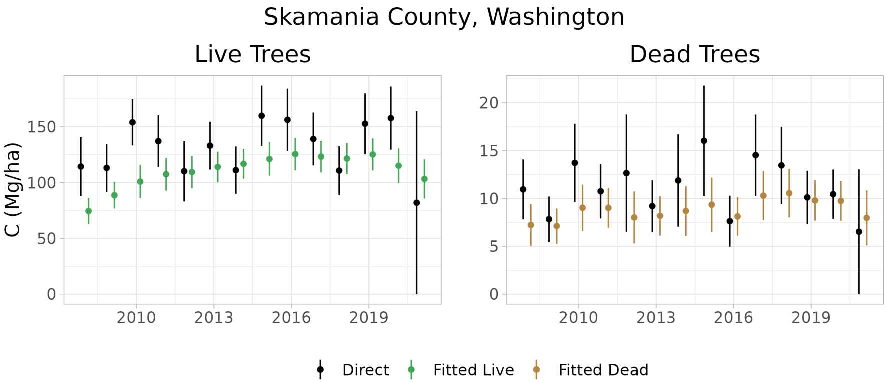
Preliminary Results



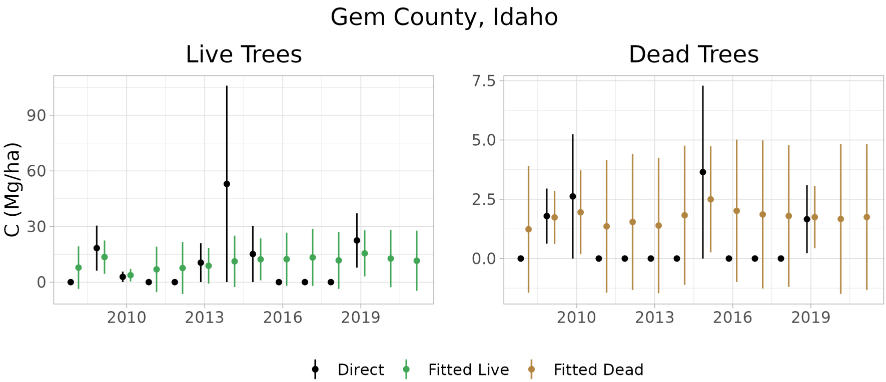
2020 Archie Creek Wildfire



Preliminary Results

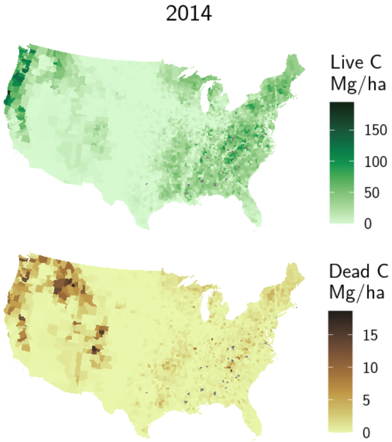


Preliminary Results



Next Steps

- Extend to entire CONUS
- Incorporate space-varying regression coefficient for TCC.
- Allow $\mathbf{A}$ to vary spatially $\mathbf{A}_j$ and, perhaps, temporally $\mathbf{A}_{j,t}$.
- Other carbon pools (down woody material, leaf litter, soil, etc.)



References

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society: Series B (Methodological)*, 36(2):192–225.

Bradley, J. R., Holan, S. H., and Wikle, C. K. (2015). Multivariate spatio-temporal models for high-dimensional areal data with application to Longitudinal Employer-Household Dynamics. *The Annals of Applied Statistics*, 9(4):1761 – 1791.

De Witte, D., Abad, A. A., Molenberghs, G., Verbeke, G., Sanchez, L., Mas-Bermejo, P., and Neyens, T. (2023). A multivariate spatio-temporal model for the incidence of imported covid-19 cases and covid-19 deaths in cuba. *Spatial and Spatio-temporal Epidemiology*, 45:100588.

Eggleston, H. S., Buendia, L., Miwa, K., Ngara, T., and Tanabe, K. (2006). 2006 ipcc guidelines for national greenhouse gas inventories.

Fay, R. E. and Herriot, R. A. (1979). Estimates of income for small places: An application of james-stein procedures to census data. *Journal of the American Statistical Association*, 74(366).

Guhaniyogi, R., Finley, A., Banerjee, S., and Kobe, R. (2013). Modeling complex spatial dependencies: Low-rank spatially varying cross-covariances with application to soil nutrient data. *Journal of Agricultural, Biological, and Environmental Statistics*, 18.

Jin, X., Banerjee, S., and Carlin, B. P. (2007). Order-free co-regionalized areal data models with application to multiple-disease mapping. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(5):817–838.

Jin, X., Carlin, B. P., and Banerjee, S. (2005). Generalized hierarchical multivariate car models for areal data. *Biometrics*, 61(4):950–961.