# A Double Fixed Rank Kriging Approach to Spatial Regression Models with Covariate Measurement Error

Nickson Ning, Francis Hui, Alan Welsh

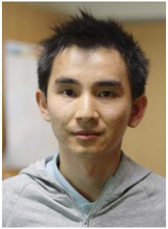
Research School of Finance, Actuarial Studies and Statistics

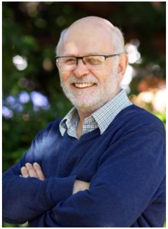
NATURAM PRIMUM COGNOSCERE RERUM

Australian National University

# Reference

Ning, X., Hui, F.K.C., Welsh, A. (2023). A Double Fixed Rank Kriging Approach to Spatial Regression Models with Covariate Measurement Error. *Environmetrics*, **34**, e2771.
https://doi.org/10.1002/env.2771.



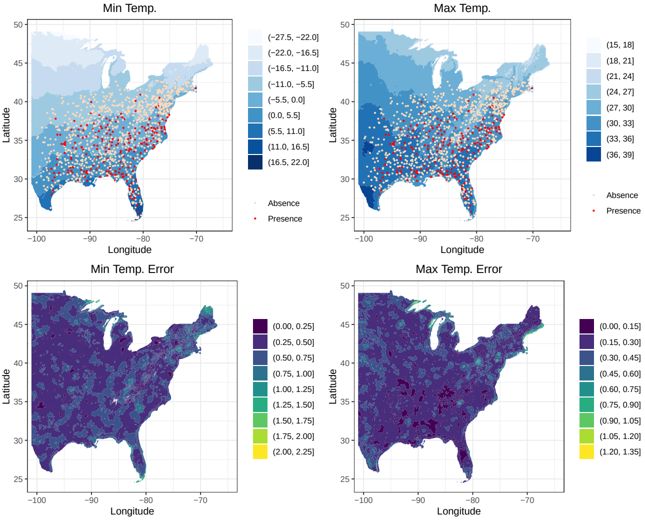


# Carolina Wren

# Carolina Wren

- Bird
- Cute
- Provides motivating data for our methodology

# Motivating data



- Many disciplines e.g., environmental monitoring, ecology, epidemiology collect and analyse spatially indexed data
- Often exhibit measurement error in the covariates

# Spatial GLMMs

For $i = 1, \ldots, n$, let $y(\boldsymbol{s}_i)$ denote a response measured at spatial location $\boldsymbol{s}_i \in \mathbb{D}$ in some spatial domain $\mathbb{D}$.

$$f(y(\boldsymbol{s}_i)|\vartheta(\boldsymbol{s}_i), \phi) = \exp\left\{[y(\boldsymbol{s}_i)\vartheta(\boldsymbol{s}_i) - a\{\vartheta(\boldsymbol{s}_i)\}]/\phi + c(y(\boldsymbol{s}_i), \phi)\right\}$$

$$g\{\mu(\boldsymbol{s}_i)\} = \beta_0 + \boldsymbol{x}(\boldsymbol{s}_i)^\top \boldsymbol{\beta} + H_\rho(\boldsymbol{s}_i)$$

$H_\rho(\boldsymbol{s})$ is an unknown spatial field capturing the (residual) spatial correlation in the responses.

Commonly used for spatially indexed data (Diggle et al., 1998; Lindgren et al., 2011; Datta et al., 2016).

# Measurement Error

For observation $i = 1, \ldots, n$ and covariate $j = 1, \ldots, J$,

$$w_j(\boldsymbol{s}_i) = x_j(\boldsymbol{s}_i) + \epsilon_{ij}, \quad (1)$$

$\epsilon_{ij} \overset{ind.}{\sim} N(0, \sigma_j^2)$.

Well-known that naively using $w_j(\boldsymbol{s}_i)$ instead of $x_j(\boldsymbol{s}_i)$ causes attenuation/bias towards zero (see e.g., Wald (1940)).

# Literature

- Many ways to model $H_\rho(\boldsymbol{s}_i)$ (Albert and McShane, 1995; Cressie and Johannesson, 2008)

- Classical measurement error model (1) for non-spatial data is also well-studied (Carroll et al., 2006; Fuller, 2009), e.g., SIMEX

- Both: Alexeeff et al. (2016); Zhang et al. (2016)

- No need to know/estimate $\sigma_j^2$: Huque et al. (2016)

# Double FRK

For covariate $j = 1, \ldots, J$, assume that $x_j(\boldsymbol{s}_i)$ are realisations from an underlying spatial field $G_j(\boldsymbol{s})$. Cressie and Johannesson (2008) Fixed-rank Kriging:

$$\begin{aligned} x_j(\boldsymbol{s}_i) = G_j(\boldsymbol{s}_i) = \boldsymbol{B}_j^\top(\boldsymbol{s}_i)\boldsymbol{\theta}_j; \quad \boldsymbol{\theta}_j \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_j) \\ H_\rho(\boldsymbol{s}_i) = \boldsymbol{B}_\rho^\top(\boldsymbol{s}_i)\boldsymbol{\theta}_\rho; \quad \boldsymbol{\theta}_\rho \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_\rho). \end{aligned} \tag{2}$$

# Double FRK cont'd

- **First step:** For each covariate $x_j(\boldsymbol{s}_i)$, fit a basis function model (GAM/LMM) to the error-contaminated covariates $w_j(\boldsymbol{s}_i)$.

$$w_j(\boldsymbol{s}_i) = \boldsymbol{B}_j^\top(\boldsymbol{s}_i)\boldsymbol{\theta}_j + \epsilon_{ij}.$$

  Using the model, get predictions $\hat{x}_j(\boldsymbol{s}_i) = \boldsymbol{B}_j^\top(\boldsymbol{s}_i)\hat{\boldsymbol{\theta}}_j$ for the underlying true covariates $x_j(\boldsymbol{s}_i)$.

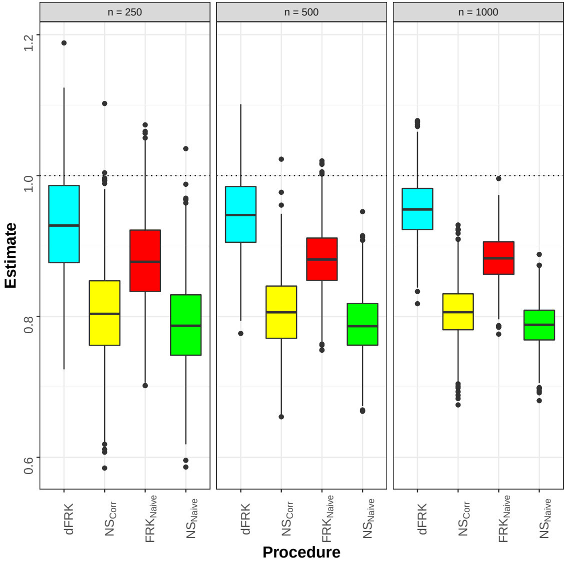
- **Second Step:** Fit the spatial GLMM (e.g., using a Laplace approximation), treating the $\hat{x}_j(\boldsymbol{s}_i)$'s as if they were the true underlying covariate values $x_j(\boldsymbol{s}_i)$.

$$g\{\mu(\boldsymbol{s}_i)\} = \beta_0 + \hat{\boldsymbol{x}}(\boldsymbol{s}_i)^\top\boldsymbol{\beta} + \boldsymbol{B}_\rho^\top(\boldsymbol{s}_i)\boldsymbol{\theta}_\rho.$$

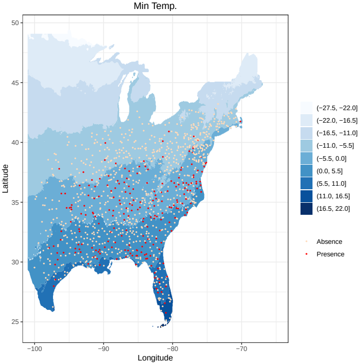
# Simulations

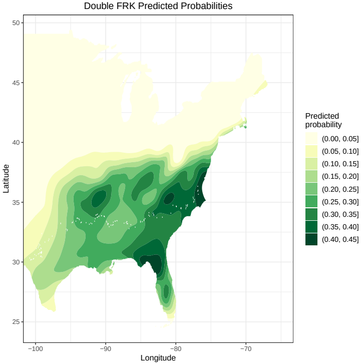
- Simulated from spatial GLMM and covariate measurement error model
- Compared slope estimates produced by different methods
- $dFRK$: adjust for both covariate measurement error and residual spatial autocorrelation. $NS_{corr}$: No residual spatial autocorrelation adjustment. $FRK_{naive}$: No measurement error adjustment. $NS_{naive}$: no adjustment for either.

# Simulations



# Motivating data revisited

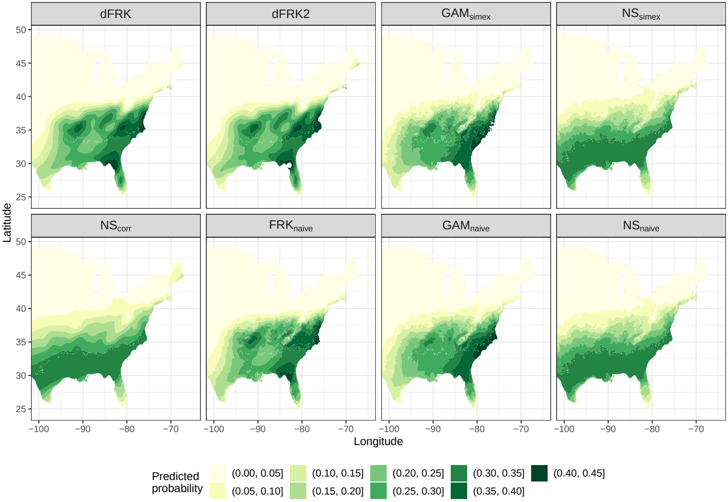


# Carolina Wren: Estimates

| Spatial | M. Error | Estimates | | | Standard Error | | |
|---|---|---|---|---|---|---|---|
| | | Intercept | Temp. | Temp.$^2$ | Intercept | Temp. | Temp.$^2$ |
| FRK | FRK | -0.463 | 0.148 | -0.0362 | 0.125 | 0.0247 | 0.0060 |
| FRK2 | FRK | -0.551 | 0.147 | -0.0327 | 0.116 | 0.0244 | 0.0054 |
| GAM | SIMEX | -0.607 | 0.192 | -0.0224 | 0.320 | 0.1444 | 0.0065 |
| None | SIMEX | -0.789 | 0.106 | -0.0189 | 0.099 | 0.0190 | 0.0033 |
| None | FRK | -0.793 | 0.106 | -0.0186 | 0.098 | 0.0189 | 0.0033 |
| FRK | None | -0.814 | 0.116 | -0.0250 | 0.330 | 0.0869 | 0.0106 |
| GAM | None | -0.769 | 0.119 | -0.0212 | 0.258 | 0.1086 | 0.0058 |
| None | None | -0.801 | 0.105 | -0.0182 | 0.098 | 0.0188 | 0.0032 |

# Carolina Wren: Out-of-sample Prediction



# Carolina Wren: Out-of-sample Prediction

| Spatial | M. Error | MSE | PD | AUC |
|---|---|---|---|---|
| FRK | FRK | **0.1555** | **0.4764** | **0.6839** |
| FRK2 | FRK2 | 0.1562 | 0.4792 | 0.6830 |
| GAM | SIMEX | 0.1614 | 0.4882 | 0.6425 |
| None | SIMEX | 0.1581 | 0.4823 | 0.6444 |
| None | FRK | 0.1576 | 0.4808 | 0.6478 |
| FRK | None | 0.1570 | 0.4787 | 0.6673 |
| GAM | None | 0.1604 | 0.4853 | 0.6483 |
| None | None | 0.1580 | 0.4817 | 0.6468 |

# Conclusion

- Extend work of Huque et al. (2016) to a generalised linear mixed model framework
- Deals with spatial correlation in responses and covariate measurement error simultaneously
- No need for validation/replication data to estimate $\sigma_j^2$: exploits the inherent spatial autocorrelation
- Key assumption: covariates are realisations of some smooth underlying spatial field

THE
END

THANKS!