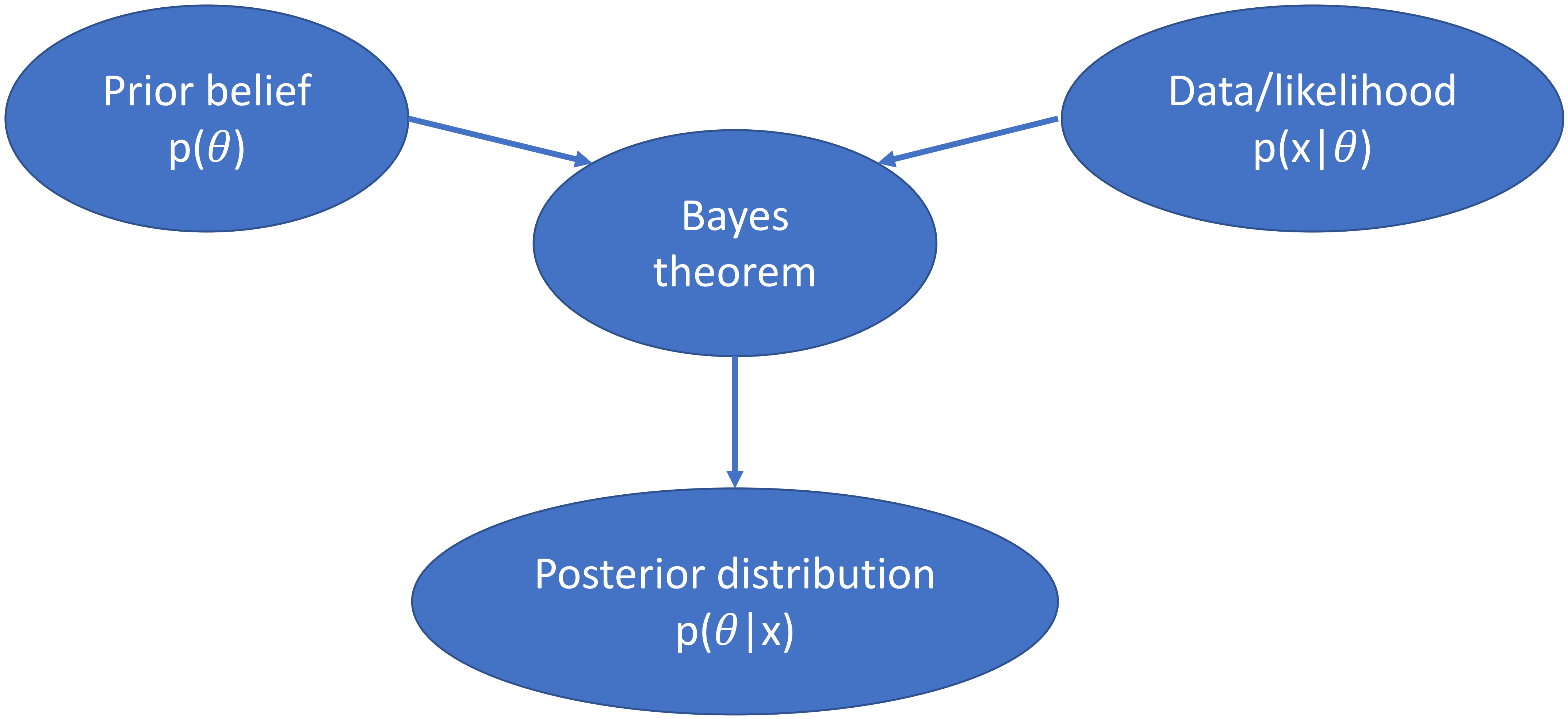
Empirical Bayesian spatial models using mgcv

Sophie Lee

PHID meeting, 2nd December, 2021

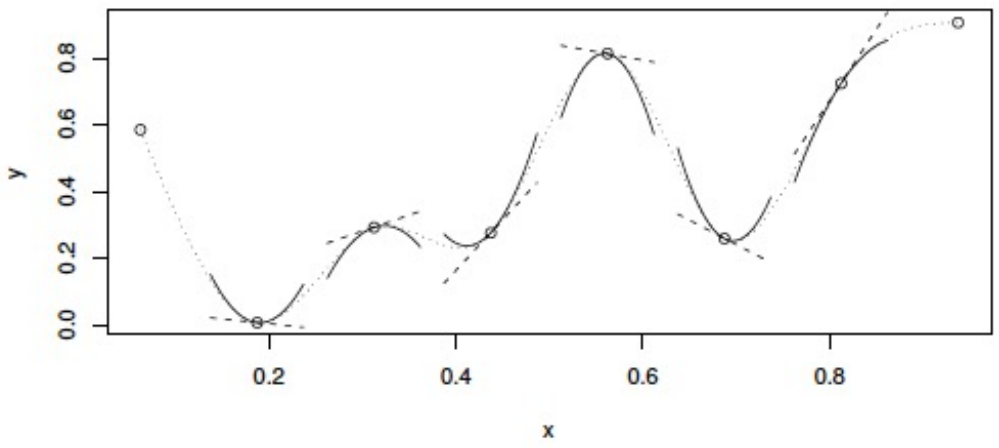
Empirical Bayesian vs fully Bayesian



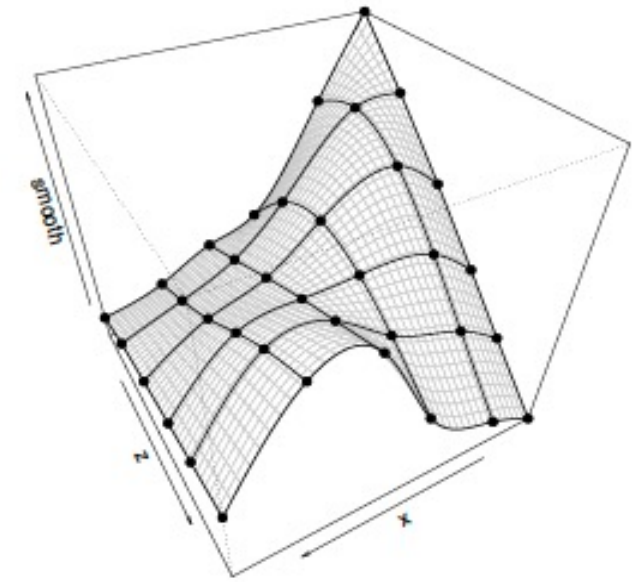
Generalised additive models

$$g(y_i) = \beta_0 + \sum_{j=1}^{n} \beta_j X_{ij} + \sum_{k=1}^{m} f_k(Z_i)$$

Cubic spline



Thin plate spline



Wood, 2017, CRC press

Bayesian framing of GAMs

- To avoid overfitting the data, REstricted Maximum Likelihood (REML) produces a smoothing penalty λ when fitting smooth functions

- Prior belief: $f_k(Z_i)$ is more smooth than wiggly

$$g(y_i) = \beta_0 + \sum_{j=1}^{n} \beta_j X_{ij} + \sum_{k=1}^{m} f_k(Z_i)$$

- λ is estimated using the data, therefore our approach is empirical rather than fully Bayesian

Bayesian framing of GAMs

$$g(y_i) = \beta_0 + \sum_{j=1}^{n} \beta_j X_{ij} + \sum_{k=1}^{m} f_k(Z_i)$$

- Smooth functions are estimated as a a linear combination of basis functions ($b_j(Z_i)$) and coefficients:

$$f_k(Z_i) = \sum_{j=n+1}^{p} b_j(Z_i)\beta_j$$

- We assume that βs have a zero mean Gaussian prior with precision proportional to the <u>smoothing penalty</u>

Bayesian framing of GAMs

- Continuing this Bayesian view, the posterior distribution of βs is:

$$\beta | y, \lambda \sim N(\hat{\beta}, V_{\beta})$$

- We can simulate from this distribution using coefficient estimates ($\hat{\beta}$) and the precision matrix (V_{β}) from mgcv model outputs

- From this we can produce credible intervals for the coefficient estimates and combine these with linear predictors to produce estimates of the outcome

Simulating from the posterior

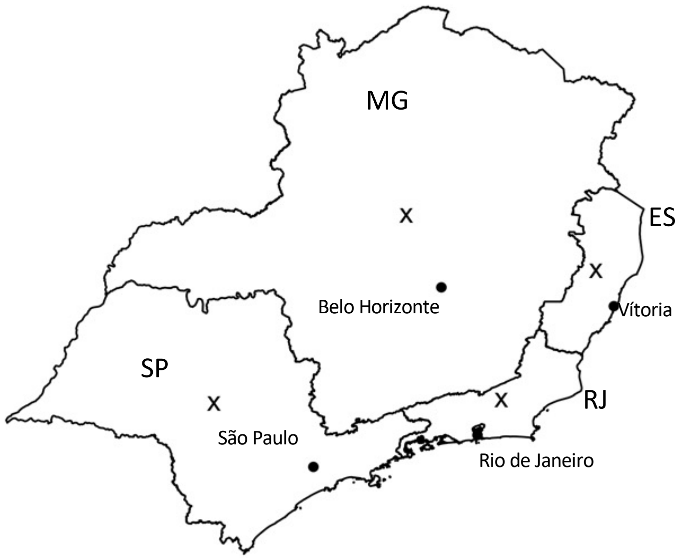
$$g(y_i) = \beta_0 + \sum_{j=1}^{n} \beta_j X_{ij} + \sum_{k=1}^{m} f_k(Z_i)$$
$$= \beta_0 + \sum_{j=1}^{n} \beta_j X_{ij} + \sum_{j=n+1}^{p} b_j(Z_i)\beta_j$$

- Smooth functions are linear combinations of basis functions and βs

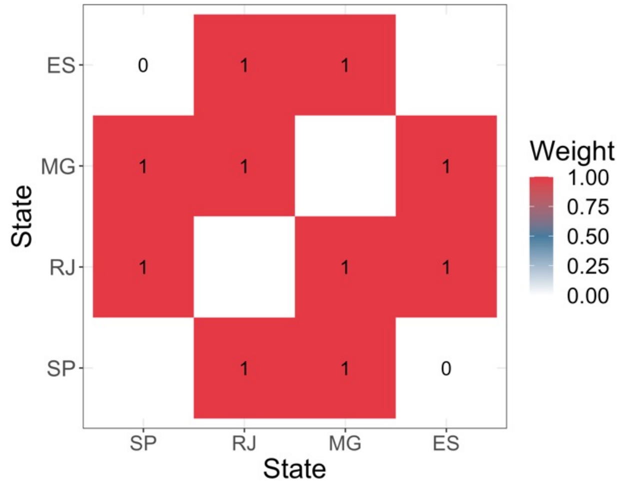
- Basis functions for each observation can be extracted and combined with simulated βs in the same way observed covariates (X_i) can be

- This produces simulations from the posterior of the response

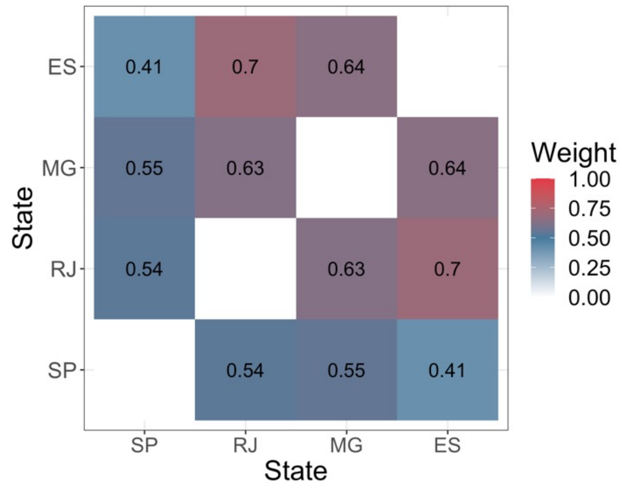
Fully Bayesian spatial models



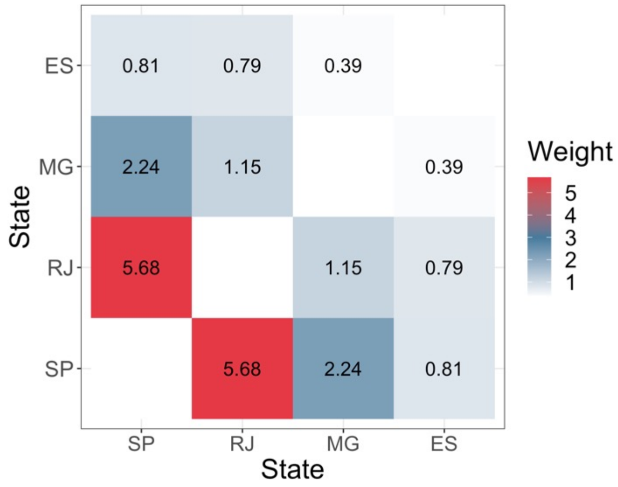
Neighbourhood-based:



Distance-based:

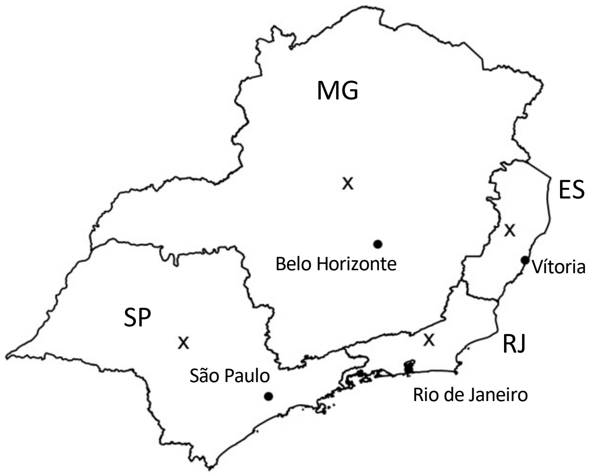


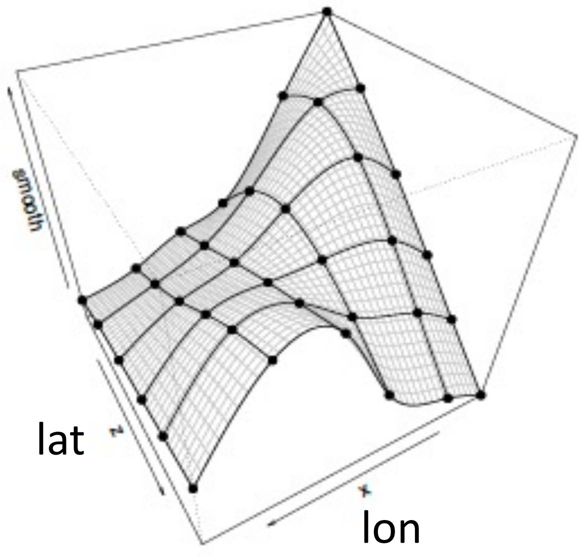
Human movement (air travel):



Lee S.A. et al., 2021, J. R. Soc. Interface

Empirical Bayesian spatial models





- Apply a thin plate spline to coordinates to produce a spatially smoothed plane
- Less rigid, context specific
- Under this approach, the smooth has the same structure (and interpretation) as random effects

Extra reading/references

- Wood, S.N., 2017. *Generalized additive models: an introduction with R*. CRC press.
- Fahrmeir, L., Kneib, T. and Lang, S., 2004. Penalized structured additive regression for space-time data: a Bayesian perspective. *Statistica Sinica*, pp.731-761.
- Wood, S.N., 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *73*(1), pp.3-36.
- Lee, S.A., Economou, T., de Castro Catão, R., Barcellos, C. and Lowe, R., (in press). The impact of climate suitability, urbanisation, and connectivity on the expansion of dengue in 21st century Brazil.