# 12.x Spatial Generalized Linear Models

Generalized linear models have become very popular in recent years. Many models and analyses for counts, proportions, binary data, etc., have been introduced through the years. However, Nelder and Wedderburn (1972) unified many of the approachs by using the classical linear model framework along with a link function, and Wedderburn (1974) introduced quasi-likelihood estimation methods. A popular textbook has been McCullagh and Nelder (1989).

Many types of data are binary, counts, or positive continuous. Early attempts to model such data relied on transformations to "near normal" so that classical linear model methods could be used. For example, a square root transformation was often used for count data. However, Nelder and Wedderburn (1972) introduced a natural extension to linear models using parameteric distributions such as the Poisson for counts, the Bernoulli for binary data, etc., called the generalized linear model (GLM, McCullagh and Nelder 1989), which have become very popular and generally preferred to data transformations. A natural extension of GLMs by allowing latent random effects in the linear mixed model, to create a class of generalized linear mixed model (GLMM, Breslow and Clayton 1993). The latent random effects are generally assumed to be independent and identically distributed from normal distribution. However, it is also possible to for the latent random effects to be spatially autocorrelated, leading to the spatial generalized linear model (SGLM, Gotway and Stroup 1997; Diggle et al. 1998), which we review here.

Historically, for areal data, such as the models in Chapter 7, there are equivalent models for discrete data, such as those that are binary or counts. These have been termed the autologistic, auto-Poisson models, autobinomial, and auto negative binomial, with obvious connections to their nonspatial distributions (Besag 1974; Cressie 1993). These models have not been very popular, as the conditional specification does not always lead to a recognizable likelihood, or, indeed, as for the auto-Poisson, that likelihood may not have a closed form under positive autocorrelation. We will not discuss these models further.

Another class of models are based on a hierarchical constuction, where the mean of any of the distributions in GLMs is allowed to vary by using spatial random effects in the mean structure. Here, there are three broad methods of analysis. The most obvious method is to take a Bayesian approach and compute the posterior distribution of all latent spatial variables and parameters. This has been extremely popular beginning with disease-mapping (Clayton and Kaldor 1987) and the introduction of the `WinBUGS` software (Lunn et al. 2000).

Another approach is the penalized quasi-likelihood models (Breslow and Clayton 1993; Wolfinger and O'Connell 1993). These models have been implemented in popular software such as the `glmmPQL` function in the `MASS` package in `R` and the `GLIMMIX` package in `SAS`.

## 12.x.1 Spatially structured dependence

A very general way to create spatially structured dependence for GLMMs is through a hierarchical construction. We will use the notation $[\mathbf{y}|\boldsymbol{\xi}]$ to denote any probability density function of the random variable $\mathbf{y}$ conditional on parameters, or other fixed variables, $\boldsymbol{\xi}$. We can have a joint distribution on the left side of the conditional bar, and multiple parameters and fixed values on the right, e.g., $[\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_k | \boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \ldots, \boldsymbol{\xi}_k]$. For example, let $[\mathbf{y}|\boldsymbol{\xi}]$ be a Poisson distribution with mean parameter $\boldsymbol{\xi}$. For the hierarchical construction of a SGLM, we condition on spatially-autocorrelated random effects $\mathbf{w}$, thus $[\mathbf{y}|\mathbf{w}]$, where $\mathbf{w}$ is generally considered to have multivariate normal distribution, which can be denoted $[\mathbf{w}|\boldsymbol{\theta}]$, where the vector $\boldsymbol{\theta}$ contains covariance parameters. For example, $\boldsymbol{\theta}$ often contains the partial sill, range, and nugget effect for geostatistical models. Hence, the joint distribution of the data $\mathbf{y}$ and the latent random effects $\mathbf{w}$ is $[\mathbf{y}, \mathbf{w}|\boldsymbol{\theta}] = [\mathbf{y}|\mathbf{w}][\mathbf{w}|\boldsymbol{\theta}]$.

The model for the data $\mathbf{y}$ can have more parameters than just the mean, in which case we write it $[\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\phi}]$, where it is parameterized so that $\mathrm{E}(\mathbf{y}) = \boldsymbol{\mu}$. If our linear model is $\boldsymbol{\eta}$, then generalized linear models establish a link function between $\boldsymbol{\mu}$ and $\boldsymbol{\eta}$, which we denote as $g(\boldsymbol{\mu}) = \mathbf{w}$, where $g(\cdot)$ is called the link function. For the Poisson example, $g(\cdot)$ is often the log function. Link functions are monotonic so that $g^{-1}(\cdot)$ is one-to-one with $g(\cdot)$. The log link makes sense for the Poisson example because $g^{-1}(\cdot)$ is the exponential function, and hence $\mathbf{w}$ is unconstrained. The negative binomial distribution can also be parameterized with a mean, and an extra parameter that allows for overdispersion, which we would write as $[\mathbf{y}|\boldsymbol{\mu}, \phi]$, where $\phi$ is the overdispersion parameter. Now, let us write

$$\mathbf{w} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e},$$

where this model is the same one defined in (1.1), where $\mathrm{var}(\mathbf{e}) = \boldsymbol{\Sigma}_{\boldsymbol{\theta}}$, and we use the subscript to show the dependence of $\boldsymbol{\Sigma}$ on $\boldsymbol{\theta}$. Thus, we use the notation $[\mathbf{w}|\mathbf{X}, \boldsymbol{\beta}, \boldsymbol{\theta}]$ to indicate the probability density function $N(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}})$. Then, a very general model can be constructed hierarchically as,

$$[\mathbf{y}, \mathbf{w}|\boldsymbol{\phi}, \boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{X}] = [\mathbf{y}|\mathbf{w}, \boldsymbol{\phi}][\mathbf{w}|\boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{X}]. \tag{12.1}$$

As a concrete example, suppose that $[\mathbf{y}|\mathbf{w}]$ is Poisson, and $[\mathbf{w}|\boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{X}]$ is multivariate normal, then the joint likelihood is

$$[\mathbf{y}, \mathbf{w}|\boldsymbol{\phi}, \boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{X}] = \left(\prod_{i=1}^{n} \frac{\exp(w_i)^{y_i}\exp(-\exp(w_i))}{y_i!}\right)\frac{\exp-[(\mathbf{w}-\mathbf{X}\boldsymbol{\beta})^T\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}(\mathbf{w}-\mathbf{X}\boldsymbol{\beta})]}{(2\pi)^{n/2}|\boldsymbol{\Sigma}_{\boldsymbol{\theta}}|^{1/2}},$$

and note the use of $\mu_i = g^{-1}(w_i) = \exp(w_i)$.

The joint distribution 12.1 is the basis for inference, either by

- putting prior distributions on $\boldsymbol{\phi}$, $\boldsymbol{\beta}$, and $\boldsymbol{\theta}$ and computing, or sampling from, the joint posterior distribution $[\mathbf{w}, \boldsymbol{\phi}, \boldsymbol{\beta}, \boldsymbol{\theta}|\mathbf{y}, \mathbf{X}]$ using any of a variety of Bayesian methods, or

- approximating (12.1) with a quasi-likehood and using iterative fitting algorithms for $\mathbf{w}, \boldsymbol{\phi}, \boldsymbol{\beta}, \boldsymbol{\theta}$, or

- integrating over $\mathbf{w}$ using a Laplace approximation, and integrating over $\boldsymbol{\beta}$ as in REML, and use maximum likelihood to estimate $\boldsymbol{\phi}, \boldsymbol{\theta}$ marginally, followed by GLS estimation of $\boldsymbol{\beta}$ and prediction for $\mathbf{w}$.

Below, we will give more details on the Laplace approximation and the marginal maximum likelihood approach.

### 12.x.2 Exploratory spatial data analysis

la te da

### 12.x.3 Parametric models for the mean structure

la te da

### 12.x.4 Parametric models for the covariance structure

la te da

### 12.x.5 Marginal MLE for Covariance Parameters

We would like to marginalize the distribution $[\mathbf{w},\mathbf{y}|\boldsymbol{\phi},\boldsymbol{\beta},\boldsymbol{\theta}] = [\mathbf{y}|\mathbf{w},\boldsymbol{\phi}][\mathbf{w}|\boldsymbol{\beta},\boldsymbol{\theta}]$ over both $\mathbf{w}$ and $\boldsymbol{\beta}$ to obtain a likelihood of the data and variance/covariance parameters: $[\mathbf{y}|\boldsymbol{\phi},\boldsymbol{\theta}]$. This is obtained by,

$$\int_{\mathbf{w}}\int_{\boldsymbol{\beta}}[\mathbf{w},\mathbf{y}|\boldsymbol{\phi},\boldsymbol{\beta},\boldsymbol{\theta}]d\boldsymbol{\beta}d\mathbf{w} = \int_{\mathbf{w}}[\mathbf{y}|\mathbf{w},\boldsymbol{\phi}]\int_{\boldsymbol{\beta}}[\mathbf{w}|\boldsymbol{\beta},\boldsymbol{\theta}]d\boldsymbol{\beta}d\mathbf{w}.$$

When $[\mathbf{w}|\boldsymbol{\beta},\boldsymbol{\theta}]$ is Gaussian, $\int_{\boldsymbol{\beta}}[\mathbf{w}|\boldsymbol{\beta},\boldsymbol{\theta}]d\boldsymbol{\beta}$ is the likelihood for restricted maximum likelihood estimation (REML) (see Exercise 8.4),

$$[\mathbf{w}|\boldsymbol{\theta}] \equiv \int_{\boldsymbol{\beta}}[\mathbf{w}|\boldsymbol{\beta},\boldsymbol{\theta}]d\boldsymbol{\beta} = \frac{1}{C_n}\exp[(\mathbf{w}-\mathbf{X}\hat{\boldsymbol{\beta}})^T\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}(\mathbf{w}-\mathbf{X}\hat{\boldsymbol{\beta}})]$$

where $C_n = \sqrt{2\pi^N|\boldsymbol{\Sigma}_{\boldsymbol{\theta}}||\mathbf{X}^T\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\mathbf{X}|}$ and

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\mathbf{w}. \tag{12.2}$$

Thus, we next need,

$$[\mathbf{y}|\boldsymbol{\phi},\boldsymbol{\theta}] = \int_{\mathbf{w}}[\mathbf{y}|\mathbf{w},\boldsymbol{\phi}][\mathbf{w}|\boldsymbol{\theta}]d\mathbf{w}$$

Let us denote $\ell(\mathbf{w};\mathbf{y},\boldsymbol{\phi},\boldsymbol{\theta}) = \log([\mathbf{y}|\mathbf{w},\boldsymbol{\phi}][\mathbf{w}|\boldsymbol{\theta}])$. Consider,

$$\int_{\mathbf{w}} e^{\ell(\mathbf{w};\mathbf{y},\boldsymbol{\phi},\boldsymbol{\theta})}d\mathbf{w}.$$

Let $\mathbf{g}$ be the gradient vector with $i$th element

$$\mathbf{g}_i = \frac{\partial\ell(\mathbf{w};\mathbf{y},\boldsymbol{\phi},\boldsymbol{\theta})}{\partial w_i}$$

and the Hessian matrix with $i,j$th element,

$$\mathbf{H}_{i,j} = \frac{\partial^2\ell(\mathbf{w};\mathbf{y},\boldsymbol{\phi},\boldsymbol{\theta})}{\partial w_i\partial w_j}$$

Using the multivariate Taylor series expansion around some point $\mathbf{a}$,

$$\int_{\mathbf{w}} e^{\ell(\mathbf{w};\mathbf{y},\boldsymbol{\phi},\boldsymbol{\theta})} d\mathbf{w} \approx \int_{\mathbf{w}} e^{\ell(\mathbf{a};\mathbf{y},\boldsymbol{\phi},\boldsymbol{\theta})+\mathbf{g}^T(\mathbf{w}-\mathbf{a})+1/2(\mathbf{w}-\mathbf{a})^T\mathbf{H}(\mathbf{w}-\mathbf{a})} d\mathbf{w},$$

Now, let $\mathbf{a}$ be the value $\ell(\mathbf{a};\mathbf{y},\boldsymbol{\phi},\boldsymbol{\theta})$ such that $\mathbf{g} = \mathbf{0}$. Then,

$$\int_{\mathbf{w}} e^{\ell(\mathbf{w};\mathbf{y},\boldsymbol{\phi},\boldsymbol{\theta})} d\mathbf{w} \approx e^{\ell(\mathbf{a};\mathbf{y},\boldsymbol{\phi},\boldsymbol{\theta})} \int_{\mathbf{w}} e^{-1/2(\mathbf{w}-\mathbf{a})^T\mathbf{H}(\mathbf{w}-\mathbf{a})} d\mathbf{w}.$$

where $\mathbf{H_a}$ indicates $\mathbf{H}$ evaluated at $\mathbf{a}$. We know from the normalizing constant of a multivariate Gaussian distribution that this integral is $(2\pi)^{N/2}|-\mathbf{H}_{\mathbf{a}}^{-1}|^{1/2}$, so

$$\int_{\mathbf{w}} e^{\ell(\mathbf{w})} d\mathbf{w} \approx e^{\ell(\mathbf{a})}(2\pi)^{N/2}|-\mathbf{H_a}|^{-1/2} = [\mathbf{y}|\mathbf{a},\boldsymbol{\phi}][\mathbf{a}|\boldsymbol{\theta}](2\pi)^{N/2}|-\mathbf{H_a}|^{-1/2}.$$

A marginal maximum likelihood estimator for $\boldsymbol{\phi},\boldsymbol{\theta}$, given $\mathbf{a}$, is

$$\{\hat{\boldsymbol{\phi}},\hat{\boldsymbol{\theta}}\} = \underset{\boldsymbol{\phi},\boldsymbol{\theta}}{\arg\max}\,[\log[\mathbf{y}|\mathbf{a},\phi] + \log[\mathbf{a}|\theta] - (1/2)\log|-\mathbf{H_a}(\boldsymbol{\phi},\boldsymbol{\theta})|] \qquad (12.3)$$

where we drop terms that do not contain $\hat{\boldsymbol{\phi}}$ or $\hat{\boldsymbol{\theta}}$, but show the dependence of $\mathbf{H_a}$ on $\hat{\boldsymbol{\phi}}$ and $\hat{\boldsymbol{\theta}}$. These results depend on $\mathbf{g} = \mathbf{0}$. We use Newton-Raphson steps to attain this next, conditional on $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$.

Note that, assuming conditional independence of $\mathbf{y}$ on $\mathbf{w}$,

$$\log([\mathbf{y}|\mathbf{w},\boldsymbol{\phi}][\mathbf{w}|\boldsymbol{\lambda}]) = \sum_{i=1}^{N}\log(f(y_i;w_i,\boldsymbol{\phi})) - \frac{1}{2}(\mathbf{w}-\mathbf{X}\hat{\boldsymbol{\beta}})^T\boldsymbol{\Sigma}_{\boldsymbol{\lambda}}^{-1}(\mathbf{w}-\mathbf{X}\hat{\boldsymbol{\beta}}) + C$$

where $C$ are terms that do not contain $\mathbf{w}$. Let $\mathbf{d}$ be the vector with $i$th component,

$$\mathbf{d}[i] \equiv \frac{\partial\log(f(y_i;w_i,\boldsymbol{\phi}))}{\partial w_i}.$$

Next, note that

$$\frac{\partial[-\frac{1}{2}(\mathbf{w}-\mathbf{X}\hat{\boldsymbol{\beta}})^T\boldsymbol{\Sigma}_{\boldsymbol{\lambda}}^{-1}(\mathbf{w}-\mathbf{X}\hat{\boldsymbol{\beta}})]}{\partial\mathbf{w}} = -\boldsymbol{\Sigma}_{\boldsymbol{\lambda}}^{-1}\mathbf{w} + \boldsymbol{\Sigma}_{\boldsymbol{\lambda}}^{-1}\boldsymbol{X}\hat{\boldsymbol{\beta}}$$

so the gradient will be,

$$\mathbf{g} = \mathbf{d} - \boldsymbol{\Sigma}_{\boldsymbol{\lambda}}^{-1}\mathbf{w} + \boldsymbol{\Sigma}_{\boldsymbol{\lambda}}^{-1}\mathbf{X}\hat{\boldsymbol{\beta}}$$

For example, if $f(y_i;w_i,\boldsymbol{\phi})$ is binomial with logit link function where the expected probability is $\mu_i = \exp(w_i)/(1+\exp(w_i))$, then $d[i] = y_i - \mu_i n_i$. For the Hessian, let $\mathbf{D}$ be a diagonal matrix with $i$th component,

$$\mathbf{D}[i,i] \equiv \frac{\partial^2\log(f(y_i;w_i,\boldsymbol{\phi}))}{\partial w_i^2}.$$

because all second partials are 0 when $i \neq j$ (due to conditional independence). Next, notice that

$$\frac{\partial^2[-\frac{1}{2}(\mathbf{w}-\mathbf{X}\hat{\boldsymbol{\beta}})^T\boldsymbol{\Sigma}_{\boldsymbol{\lambda}}^{-1}(\mathbf{w}-\mathbf{X}\hat{\boldsymbol{\beta}})]}{\partial\mathbf{w}\partial\mathbf{w}^T} = -\boldsymbol{\Sigma}_{\boldsymbol{\lambda}}^{-1} + \boldsymbol{\Sigma}_{\boldsymbol{\lambda}}^{-1}\mathbf{X}(\mathbf{X}^T\boldsymbol{\Sigma}_{\boldsymbol{\lambda}}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\Sigma}_{\boldsymbol{\lambda}}^{-1}$$

and so

$$\mathbf{H} = \mathbf{D} - \boldsymbol{\Sigma}_{\boldsymbol{\lambda}}^{-1} + \boldsymbol{\Sigma}_{\boldsymbol{\lambda}}^{-1}\mathbf{X}(\mathbf{X}^T\boldsymbol{\Sigma}_{\boldsymbol{\lambda}}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\Sigma}_{\boldsymbol{\lambda}}^{-1} \tag{12.4}$$

For example, if $f(y_i;w_i,\boldsymbol{\phi})$ is binomial with the logit link function, then $\mathbf{D}[i,i] = -\mu_i n_i/(1+\exp(w_i))$.

Conditional on $\boldsymbol{\phi}$ and $\boldsymbol{\lambda}$, a Newton-Raphson update is,

$$\mathbf{w}^{[k+1]} = \mathbf{w}^{[k]} - \mathbf{H}^{-1}\mathbf{g}$$

and upon convergence we set $\mathbf{a} = \mathbf{w}$ in eqrefeq:m2LLmargMLE for any update of the likelihood. Notice that this makes the marginally MLE doubly iterative, as we solve for $\mathbf{a}$ while optimizing for $\boldsymbol{\phi}$ and $\boldsymbol{\lambda}$.

## 12.x.6 Gradients and Hessians

## 12.x.7 Estimation of Fixed Effects and Prediction

To obtain variances, use the following result. For some linear combination $\mathbf{B}\hat{\mathbf{w}}$, where the weights are contained in the matrix $\mathbf{B}$ for predicted random effects $\hat{\mathbf{w}}$,

$$\text{var}(\mathbf{B}\hat{\mathbf{w}}) = \text{E}_{\mathbf{w}}[\text{var}(\mathbf{B}\hat{\mathbf{w}}|\mathbf{w})] + \text{var}_{\mathbf{w}}[\text{E}(\mathbf{B}\hat{\mathbf{w}}|\mathbf{w})]$$

For our models, note that $\text{var}(\mathbf{w}) = \boldsymbol{\Sigma}_{\boldsymbol{\theta}}$, we assume unbiasedness, $\text{E}(\hat{\mathbf{w}}|\mathbf{w}) = \mathbf{w}$, and, marginally, $\text{var}(\hat{\mathbf{w}}) = -\mathbf{H}^{-1}$, which does not depend on $\mathbf{w}$. Hence,

$$\text{var}(\mathbf{B}\hat{\mathbf{w}}) = \mathbf{B}(-\mathbf{H}^{-1})\mathbf{B}^T + \mathbf{B}\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\mathbf{B}^T$$

For the estimation of fixed effects, then, $\hat{\boldsymbol{\beta}} = \mathbf{B}\hat{\mathbf{w}}$, where $\mathbf{B} = (\mathbf{X}^T\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}$ we have

$$\text{var}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}^T\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}(-\mathbf{H}^{-1})\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\mathbf{X}(\mathbf{X}^T\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\mathbf{X})^{-1} + (\mathbf{X}^T\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\mathbf{X})^{-1}$$

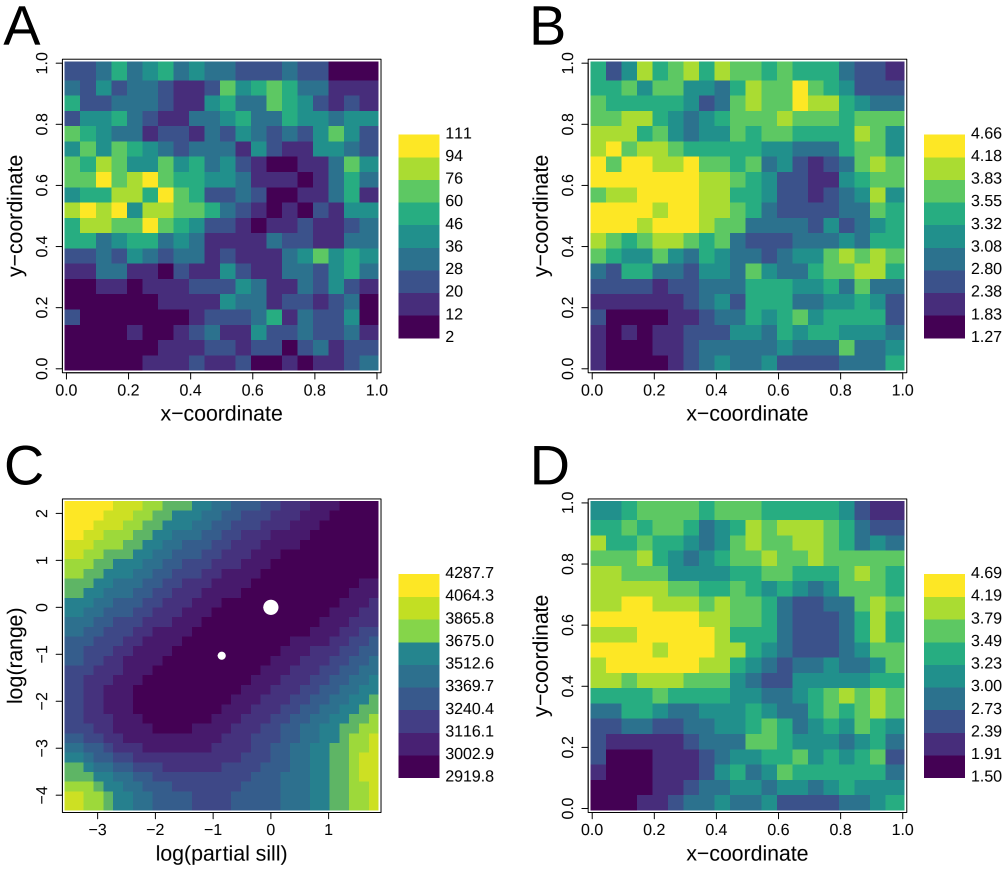

Figure 1: la te da

Recall from (5.1) that for observed $\mathbf{y}$, the generalized least squares estimator of $\boldsymbol{\beta}$ is $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{y}$ Here, we use $\hat{\mathbf{w}}$ in place of $\mathbf{y}$, $\tilde{\boldsymbol{\beta}} = (\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\hat{\mathbf{w}}$. In (5.2) the variance of $\hat{\boldsymbol{\beta}}$ is $(\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}$. Here, however, the $\mathbf{w}$ are latent and unobserved, so we need to account for estimation of the $\mathbf{w}$. It is clear from earlier that an estimator of the variance of $\mathbf{w}$ can be obtained from the likelihood as

Table 1: La te da

| effect | bias | CI90$_c$ | CI90$_u$ |
|---|---|---|---|
| $\beta_0$ | 0.111 | 0.447 | 0.704 |
| $\beta_1$ | -0.042 | 0.892 | 0.222 |
| $\beta_2$ | 0.077 | 0.909 | 0.201 |
| $\beta_3$ | -0.064 | 0.905 | 0.233 |

# References

Besag, J. (1974), "Spatial interaction and the statistical analysis of lattice systems (with discussion)," *Journal of the Royal Statistical Society, Series B*, 36, 192–236.

Breslow, N. E. and Clayton, D. G. (1993), "Approximate inference in generalized linear mixed models," *Journal of the American statistical Association*, 88, 9–25, publisher: Taylor & Francis.

Clayton, D. and Kaldor, J. (1987), "Empirical Bayes Estimates of Age-Standardized Relative Risks for Use in Disease Mapping," *Biometrics*, 43, 671–681.

Cressie, N. A. C. (1993), *Statistics for Spatial Data, Revised Edition*, New York: John Wiley & Sons.

Diggle, P. J., Tawn, J. A., and Moyeed, R. A. (1998), "Model-Based Geostatistics (Disc: P326-350)," *Journal of the Royal Statistical Society, Series C: Applied Statistics*, 47, 299–326.

Gotway, C. A. and Stroup, W. W. (1997), "A Generalized Linear Model Approach to Spatial Data Analysis and Prediction," *Journal of Agricultural, Biological, and Environmental Statistics*, 2, 157–178.

Lunn, D. J., Thomas, A., Best, N., and Spiegelhalter, D. (2000), "WinBUGS-a Bayesian modelling framework: concepts, structure, and extensibility," *Statistics and computing*, 10, 325–337, publisher: Springer.

McCullagh, P. and Nelder, J. A. (1989), *Generalized Linear Models, 2nd Edition*, Chapman & Hall Ltd.

Nelder, J. A. and Wedderburn, R. W. M. (1972), "Generalized Linear Models," *Journal of the Royal Statistical Society, Series A: General*, 135, 370–384.

Wedderburn, R. W. M. (1974), "Quasi-Likelihood Functions, Generalized Linear Models, and the Gauss-Newton Method," *Biometrika*, 61, 439–447, publisher: [Oxford University Press, Biometrika Trust].

Wolfinger, R. and O'Connell, M. (1993), "Generalized Linear Mixed Models: A Pseudo-Likelihood Approach," *Journal of Statistical Computation and Simulation*, 48, 233–243.