**Spatial Modelling**

Consider the problem of modelling spatial count data $y_{ij}$, where $i = 1, ..., N$ denote the $N$ areas. Typically, a Poisson random variable $Y_i$ is assumed to model the observed data $y_i$:

$$Y_i \sim \text{Pois}(E_i e^{\eta_i})$$

where $E_i$ is the expected number of counts to account for differences in population sizes between the areas, usually computed using internal standardisation [11], e.g. setting

$$E_i = \frac{\sum_{j \neq i} Y_j}{\sum_{j \neq i} p_j}$$

where $p_j$ is the population of the $j$th area. In disease modelling, the term $e^{\eta_i}$ is the relative risk of contracting the disease in the $i$th area. The log-relative risk is typically expressed as a regression equation where fixed, random, and covariate effects are assumed to be additive [12]. For example,

$$\eta_i = \alpha + \gamma_i + x_i,$$

where $\alpha$ denotes an intercept (fixed effect), $\gamma_i$ denotes a random effect for space, and the regression coefficient $\beta$ is the effect of covariate $x_i$. For the purpose of illustration, we generated fictitious count data $y_i$ and population sizes $p_i$ for 25 areas, as well as some covariate values $\{x_1, ..., x_25\}$. These values, together with the expected values given by Equation (2), are shown in Figure 1.

**Bayesian Smoothing**

Note that the choropleth maps in Figure 1 exhibit large spatial variations. For example, in Figure 1 (B), areas with high counts may border areas with low counts. If the aim is to map the underlying relative risks rather than the observed counts, then it is preferable to smooth the estimates of the relative risks over the areas. In a Bayesian framework, this is typically achieved by using a conditional autoregressive prior (CAR) [10] on the spatial random effects $\{\gamma_1, ..., \gamma_N\}$, which has the effect of smoothing the estimates towards a local mean [13]. The CAR prior is a joint distribution for the random effects, with

$$\gamma_i | \mathbf{y} \sim N\left(\frac{1}{\sigma^2} \sum_j w_{ij} y_j, \sigma^2\right)$$

where $\mathbf{w}$ denotes all areas excluding $i$, and $w_{ij}$ is the weight denoting the influence of area $i$ over area $j$. Collectively, these weights comprise an $N \times N$ symmetric weight matrix $W$. It is very common to use the weights defined on the basis of first order neighbours [13–14], namely

$$w_{ij} = \begin{cases} 1 & \text{if the } i^{th} \text{ and } j^{th} \text{ areas are } 1^{st} \text{ order neighbours}, \\ 0 & \text{otherwise} \end{cases}$$

Under this specification, only neighbours which are adjacent have non-zero weights. This specification may not be appropriate if the disease risk varies drastically over short distances, there are large variations in the sizes of the areas or population counts, and/or natural boundaries such as mountain ranges or rivers which alter the influence of neighbours. If the dependency between areas is assumed to be more global rather than local, then one alternative is to consider second and third order neighbours too, for example, such as the specification given by Equation (6).

$$w_{ij} = \begin{cases} w_{ij} & \text{if the } i^{th} \text{ and } j^{th} \text{ areas are } k^{th} \text{ order neighbours}, k \leq 3, \\ 0 & \text{otherwise} \end{cases}$$

Alternative specifications of the weights have also been proposed, such as deriving them from functions of geographic distance [13, 19]. Two examples are

$$w_{ij} = \begin{cases} \max(d_{ij} - d_{ij}^0) & \text{if } i \neq j \\ 0 & \text{otherwise} \end{cases}$$

and

$$w_{ij} = \exp(-0.5d_{ij})^0 & \text{if } i \neq j \\ 0 & \text{otherwise}$$

where $d_{ij}$ is the Euclidean distance between the geometric centres of the $i^{th}$ and $j^{th}$ areas. Each of these specifications achieves spatial smoothing on the parameter space. However, the lack of any obvious spatial patterns in Figure 1, and statistics like Gey’s C (the smallest p-value being 0.242 when the weights are defined using Equation (5)), indicate a lack of local spatial autocorrelation. In this situation, it may be better to perform smoothing on the covariate space.

**Smoothing on the Covariate Space**

Smoothing on the covariate space is easily achieved by defining the weights as a function of the covariate(s). The CAR prior is still assigned to the spatial random effect, but this specification smooths the risk surface over areas which have similar covariate values, which may not necessarily be close geographically. If $\delta_i = |x_i - x_j|$ is the difference in covariate values between the $i^{th}$ and $j^{th}$ areas, then two examples analogous to Equations (7) and (8) are

$$w_{ij} = \begin{cases} \max(\delta_i - \delta_j) & \text{if } i \neq j \\ 0 & \text{otherwise} \end{cases}$$

and

$$w_{ij} = \exp(-0.5\delta_i) & \text{if } i \neq j \\ 0 & \text{otherwise}$$

A comparison of these alternate specifications of the weights is shown in Figure 2.

Each of these specifications were applied to the model comprised by Equations (1) through (4). For this toy example, the result of smoothing on the covariate space rather than the parameter space seems trivial, although small improvements to performance indicators like the mean squared prediction error were observed (results not shown). Despite the vast literature on spatial modelling and alternative specifications of the weights, it appears this approach has thus far been overlooked. The full benefits of this approach to smoothing are still unclear, and further testing is required.

**Conclusions**

Smoothing is an important aspect of spatial modelling, but when the observed data lack strong spatial autocorrelation, it may not be sensible or helpful to smooth on the parameter space. This work presents an alternative approach to smoothing which addresses this problem.

**Acknowledgements**

This work was supported by the ARC Centre of Excellence for Mathematical and Statistical Frontiers.

**References**