

1 Model: logistic regression with conditional random field

1.1 Description

Let the probability of rainfall at cell be p_i , $i = 1, \dots, n$, and the occurrence of rainfall be Z_i distributed $Bernoulli(p_i)$. In this model we only consider a single explanatory variable, the temperature at each pixel, T_i . The standard logistic regression (RL) on p_i is

$$\text{logit}(p_i) = \beta_0 + \beta_t * T_i$$

where $\text{logit}(x) = \log(x/(1-x))$. Rainfall has strong spatial correlation, and so to modify the LR model to incorporate a Gaussian conditional random field (CRF). The CRF component, W describes the strength of the spatial correlation of rainfall. Let W have a multivariate Gaussian distribution with n components, $Normal(0, \Sigma^{-1})$. The covariance matrix Σ is parameterized by a sparse precision matrix Σ^{-1} in which the entries are nonzero only if two pixels are adjacent in the image. The model now becomes

$$\text{logit}(p_i) = \beta_0 + \beta_t * T_i + w_i.$$

In geostatistics this is called a Conditional Autoregressive (CAR) model, and the W is interpreted as a type of random effect on the intercept. Then (reference to Besag) the joint distribution on the CRF W is equivalent to defining the conditional distributions. Specifically,

$$w_i | w_{i \neq j} \text{distas } Normal \left(\frac{1}{m_i} \sum_{j \in n(i)} w_j, \frac{\tau^2}{m_i} \right),$$

where $n(i)$ is the set of neighbors of pixel i , and $m_i = |n(i)|$. When the precision matrix is sparse, which is the case for the gridded rainfall data, sparse matrix operations on the precision matrix can drastically reduce the computational burden of inference in CAR models.

1.2 Implementation

A Bayesian analysis with weakly informative priors on β was performed using the Stan probabilistic programming language. Only *maximum a posteriori* estimates were found.