A spatial model for rare binary events

July 15, 2016

1 Introduction

The goal in binary regression is to relate a latent variable to a response using a link function. Two common examples of binary regression include logistic regression and probit regression. The link functions for logistic and probit regression are symmetric, so they may not be well-suited for asymmetric data. An asymmetric alternative to these link functions is the complementary log-log (cloglog) link function. More recently, Wang and Dey (2010) introduced the generalized extreme value (GEV) link function for rare binary data. The GEV link function introduces a new shape parameter to the link function that controls the degree of asymmetry. The cloglog link is a special case of the GEV link function when the shape parameter is 0.

Spatial logistic and probit models are commonly presented using a hierarchical model citation. In the hierarchical framework, spatial dependence is typically modeled with an underlying latent Gaussian process, and conditioned on this process, observations are independent. However, if the latent variable is assumed to follow a GEV marginally, then a Gaussian process may not be appropriate to describe the dependence due to the fact that Gaussian processes do not demonstrate asymptotic dependence regardless of the strength of the dependence in the bulk of the data.

We propose using a latent max-stable process (de Haan, 1984) because it allows for asymptotic dependence. The max-stable process arises as the limit of the location-wise maximum of infinitely many spatial processes. Max-stable processes are extremely flexible, but are often challenging to work with in high dimensions (Wadsworth and Tawn, 2014; Thibaud and Opitz, 2015). To address this challenge, methods have been proposed that implement composite likelihood techniques for max-stable processes (Padoan et al., 2010; Genton et al., 2011; Huser and Davison, 2014). As an alternative to these composite approaches,

Reich and Shaby (2012) present a hierarchical model that implements a low-rank representation for a max-stable process. Although composite likelihoods have been used to model binary spatial data (Heagerty and Lele, 1998), we chose to use the low-rank representation of a max-stable process given by Reich and Shaby (2012).

Paragraph outlining the structure of the paper

2 Binary regression using the GEV link

Here, we provide a brief review of the the GEV link of Wang and Dey (2010). Let $Y_i \in \{0, 1\}, i = 1, \ldots, n$ be a collection of i.i.d. binary responses. It is assumed that $Y_i = I(z_i > 0)$ where $I(\cdot)$ is an indicator function, $z_i = [1 - \xi \mathbf{X}_i \boldsymbol{\beta}]^{1/\xi}$ is a latent variable following a GEV$(1, 1, 1)$ distribution, $\mathbf{X}_i$ is the associated p-vector of covariates with first element equal to one for the intercept, and $\boldsymbol{\beta}$ is a p-vector of regression coefficients. When observations

$$\pi_i = 1 - \exp\left(-\frac{1}{z_i}\right). \tag{1}$$

Although this link was selected by Wang and Dey based on its ability to handle asymmetry, the GEV distribution is one of the primary distributions used for modeling extremes. Traditionally, analysis of extreme events is done using block maxima or occurrences over a suitably high threshold. Because extreme events are rare, it is therefore reasonable to use similar methods when analyzing rare binary data.

3 Spatial dependence for binary regression

In many binary regression applications, spatial dependence is handled using a hierarchical model assuming a latent spatial process. Let $Y(\mathbf{s})$ be the observation at spatial location $\mathbf{s}$ in a spatial domain of interest

$\mathcal{D} \in \mathcal{R}^2$. We assume $Y(\mathbf{s}) = I[Z(\mathbf{s}) > 0]$ where $Z(\mathbf{s})$ is a latent spatial process. In spatial logistic and probit regression, the latent spatial process is assumed to be a Gaussian process. A Gaussian process may not be appropriate when describing dependence in the tails of the distribution because it always exhibits asymptotic independence, except in the case of perfect dependence. Because we use extreme values analysis as the foundation for rare binary analysis, we propose using a max-stable process to model the latent spatial process

The proposed max-stable process has generalized extreme value marginal distributions with location $\mathbf{X}^T(\mathbf{s})\boldsymbol{\beta}$, scale 1, and shape ξ. In our model, the scale is fixed due to identifiability concerns. Although $\boldsymbol{\beta}$ and ξ could be permitted to vary across space, we assume that they are constant across $\mathcal{D}$.

For a finite collection of locations $\mathbf{s}_1, \ldots, \mathbf{s}_n$, denote the vector of observations $\mathbf{Y} = [Y(\mathbf{s}_1), \ldots, Y(\mathbf{s}_n)]^T$. The spatial dependence is determined by the joint distribution of $\mathbf{Z} = [Z(\mathbf{s}_1), \ldots, Z(\mathbf{s}_n)]^T$, given by

$$G(\mathbf{z}) = \mathrm{P}[Z(\mathbf{s}_1) < z(\mathbf{s}_1), \ldots, Z(\mathbf{s}_n) < z(\mathbf{s}_n)] = \exp\left\{ -\sum_{l=1}^{L} \left[\sum_{i=1}^{n} \left(\frac{w_l(\mathbf{s}_i)}{z(\mathbf{s}_i)} \right)^{1/\alpha} \right]^{\alpha} \right\}, \tag{2}$$

where $G(\cdot)$ is the CDF of a multivariate GEV distribution, $w_l(\mathbf{s}_i)$ are a set of L weights that determine the spatial dependence structure, and $\alpha \in (0, 1)$ determines the strength of dependence, with α near zero giving strong dependence and $\alpha = 1$ giving joint independence. This is a special case of the multivariate GEV distribution with asymmetric Laplace dependence function (Tawn, 1990). The weights $w_l(\mathbf{s}_i)$ in (2) vary smoothly across space to induce spatial dependence. Many weight functions are possible, but the weights must be constrained so that $\sum_{l=1}^{L} w_l(\mathbf{s}_i) = 1$ for all $i = 1, \ldots, n$ to preserve the marginal GEV distribution. For example, Reich and Shaby (2012) take the weights to be scaled Gaussian kernels with knots $\mathbf{v}_l$,

$$w_l(\mathbf{s}_i) = \frac{\exp\left[-0.5 \left(||\mathbf{s}_i - \mathbf{v}_l|| / \rho \right)^2 \right]}{\sum_{j=1}^{L} \exp\left[-0.5 \left(||\mathbf{s}_i - \mathbf{v}_j|| / \rho \right)^2 \right]} \tag{3}$$

where $||\mathbf{s}_i - \mathbf{v}_l||$ is the distance between site $\mathbf{s}_i$ and knot $\mathbf{v}_l$, and the kernel bandwidth $\rho > 0$ determines the spatial range of the dependence, with large ρ giving long-range dependence and vice versa. One nice feature to this representation for the max-stable process is that at a single site i, the marginal distribution gives $P[Y(\mathbf{s}_i) = 1] = 1 - \exp\left[-\frac{1}{z(\mathbf{s}_i)}\right]$ which is the same as the marginal distribution given by Wang and Dey (2010).

The joint likelihood of $\mathbf{Y}$ is computationally challenging to obtain. Therefore, to incorporate spatial dependence into the model, we consider the hierarchical max-stable process of Reich and Shaby (2012). Consider a set of $A_1, \ldots, A_L \overset{iid}{\sim}$ Positive Stable(α) random effects associated with spatial knots $\mathbf{v}_1, \ldots, \mathbf{v}_L$. The hierarchical model is given by

$$\mathbf{Z}|A_1, \ldots, A_L \overset{indep}{\sim} \text{GEV}[\theta(\mathbf{s}), \alpha\theta(\mathbf{s}), \alpha] \tag{4}$$

$$A_l \overset{iid}{\sim} \text{PS}(\alpha)$$

where $\theta(\mathbf{s}) = \left[\sum_{l=1}^{L} A_l w_l(\mathbf{s})^{1/\alpha}\right]^{\alpha}$. Because conditional on the positive stable random effects, the $\mathbf{Z}$ are independent, the joint likelihood of $\mathbf{Y}$ is now the product of independent Bernoulli random variables.

4 Joint distribution

In Section 4.1, we give an exact expression in the case where there are only two spatial locations which is useful for constructing a pairwise composite likelihood and studying spatial dependence. For more than two locations, we are also able to compute the exact likelihood when the number of locations is large but the number of events is small, as might be expected for very rare events (see A.1).

4.1 Bivariate distribution

In a bivariate setting, the probability mass function is given by

$$\mathrm{P}[Y(\mathbf{s}_i), Y(\mathbf{s}_j)] = \begin{cases} \varphi(\mathbf{z}) & Y(\mathbf{s}_i) = 0, Y(\mathbf{s}_j) = 0 \\ \exp\left\{-\frac{1}{z(\mathbf{s}_i)}\right\} - \varphi(\mathbf{z}), & Y(\mathbf{s}_i) = 1, Y(\mathbf{s}_j) = 0 \\ 1 - \exp\left\{-\frac{1}{z(\mathbf{s}_i)}\right\} - \exp\left\{-\frac{1}{z(\mathbf{s}_j)}\right\} + \varphi(\mathbf{z}), & Y(\mathbf{s}_i) = 1, Y(\mathbf{s}_j) = 1 \end{cases} \tag{5}$$

where $\varphi(\mathbf{z}) = \exp\left\{-\sum_{l=1}^{L}\left[\left(\frac{w_l(\mathbf{s}_i)}{z(\mathbf{s}_i)}\right)^{1/\alpha} + \left(\frac{w_l(\mathbf{s}_j)}{z(\mathbf{s}_j)}\right)^{1/\alpha}\right]^{\alpha}\right\}$.

5 Quantifying spatial dependence

I still need to incorporate Brian's suggestions here In the literature on extremes, one common metric to describe the bivariate dependence is the χ statistic of Coles et al. (1999). The χ statistic between two observations z_1 and z_2 is given by

$$\chi(\mathbf{s}_1, \mathbf{s}_2) = \lim_{c \to \infty} = P(Z_1 > c | Z_2 > c). \tag{6}$$

However, in this latent variable approach, $\lim_{c\to\infty}$ may not be the most reasonable metric because the observed data are a series of zeros and ones. Therefore, we chose the κ statistic of Cohen (1960) defined by

$$\kappa = \frac{P(A) - P(E)}{1 - P(E)} \tag{7}$$

where $P(A)$ is the joint probability of agreement and $P(E)$ is the joint probability of agreement under an assumption of independence. We believe this measure of dependence to be reasonable because,

$$\lim_{\beta_0 \to \infty} \kappa(h) = \chi(h) = 2 - \vartheta(\mathbf{s}_i, \mathbf{s}_j) \tag{8}$$

where β_0 is the intercept from $\mathbf{X}^T\boldsymbol{\beta}$ and $\vartheta(\mathbf{s}_i, \mathbf{s}_j) = \sum_{l=1}^{L} \left[w_l(\mathbf{s}_i)^{1/\alpha} + w_l(\mathbf{s}_j)^{1/\alpha}\right]^{\alpha}$ is the pairwise extremal coefficient given by Reich and Shaby (2012) (see A.2). In the case of complete dependence, $\kappa = 1$, and in the case of complete independence, $\kappa = 0$.

6 Computation

For small K we can evaluate the likelihood directly. When K is large, we use MCMC methods with the random effects model to explore the posterior distribution. This is possible because the expression for the joint density, conditional on $A_1, \ldots, A_L$, is given by

$$P[Y(\mathbf{s}_1) = y(\mathbf{s}_1), \ldots, Y(\mathbf{s}_n) = y(\mathbf{s}_n)] = \prod_{i=1}^{n} \pi(\mathbf{s}_i)^{1-Y_i}[1 - \pi(\mathbf{s}_i)]^{Y_i}. \tag{9}$$

where

$$\pi(\mathbf{s}_i) = \exp\left\{-\sum_{l=1}^{L} A_l \left(\frac{w_l(\mathbf{s}_i)}{z(\mathbf{s}_i)}\right)^{1/\alpha}\right\}. \tag{10}$$

The model parameters and random effects are updated using a combination of Metropolis Hastings (MH) and Hamiltonian Monte Carlo (HMC) update steps. To overcome challenges with evaluating the positive stable density, we follow Reich and Shaby (2012) and incorporate the auxiliary variable technique of Stephenson (2009).

7 Simulation study

For our simulation study, we generate $n_m = 50$ datasets under 6 different settings to explore the impact of sample size and misspecification of link function. We generate data assuming three possible types of underlying process. For each process, we consider two sample sizes $n_s = 650$ and $n_s = 1300$.

The first of these processes is a max-stable process that uses the GEV link described in (4) with knots on a 21×21 grid on $[0, 1] \times [0, 1]$. For this process, we set $\alpha = 0.3$, $\rho = 0.025$, $\xi = 0$ for idenfiability purposes, and β_0 is set for each dataset to give 5% rarity. We then set $Y(\mathbf{s}) = I[z(\mathbf{s}) > 0]$ where $I[\cdot]$ is an indicator function.

For the second process, we generate a latent variable from a spatial Gaussian process with a mean of $\text{logit}(0.05) \approx -2.9444$ and an exponential covariance given by

$$\text{cov}(\mathbf{s}_1, \mathbf{s}_2) = \tau^2_{\text{Gau}} * \exp\left\{-\frac{||\mathbf{s}_1 - \mathbf{s}_2||}{\rho_{\text{Gau}}}\right\} \tag{11}$$

where $\tau_{\text{Gau}} = 7$ and $\rho_{\text{Gau}} = 0.10$. The mean of the Gaussian process is set to give approximately 5% rarity. Finally, we generate $Y(\mathbf{s}) \overset{ind}{\sim} \text{Bern}[\pi(\mathbf{s})]$ where $\pi(\mathbf{s}) = \exp\left\{\frac{z(\mathbf{s})}{1+z(\mathbf{s})}\right\}$

For the third process, we generate data using a hotspot method. For this process, we first generate hotspots throughout the space, and then set the probability of occurrence to be higher when a site is within a circle of radius $\rho = 0.05$ from a hotspot location. More specifically, generate $K \sim \text{Poisson}(9)$ hotspot locations, $\mathbf{v}^*_1, \ldots, \mathbf{v}^*_K$, from a uniform distribution over $[0, 1] \times [0, 1]$. If $||\mathbf{s}_i - \mathbf{v}^*_k|| < 0.05$ for any k, then $\pi(\mathbf{s}_i) = 0.70$ otherwise, $\pi(\mathbf{s}_i) = 0.01$. We then generate $Y(\mathbf{s}_i) \overset{ind}{\sim} \text{Bern}[\pi(\mathbf{s}_i)]$.

For each dataset, we fit the model using three different methods, spatial logistic regression, spatial probit regression, and the proposed spatial GEV method.

7.1 Spatial logistic and probit methods

Because logistic and probit methods represent two of the more common spatial techniques for binary data, we chose to compare our method to them. One way these methods differ from our proposed method is that they assume the underlying process is Gaussian. In this case, we assume that $Z(\mathbf{s})$ follows a Gaussian process with mean $\mathbf{X}(s)^T\boldsymbol{\beta}$ and exponential covariance function. The marginal distributions are given by

$$P(Y=1) = \begin{cases} \frac{\exp[\mathbf{X}^T\boldsymbol{\beta}+\mathbf{W}\boldsymbol{\alpha}]}{1+\exp[\mathbf{X}^T\boldsymbol{\beta}+\mathbf{W}\boldsymbol{\alpha}]}, & \text{logistic} \\ \phi(\mathbf{X}^T\boldsymbol{\beta}+\mathbf{W}\boldsymbol{\alpha}), & \text{probit} \end{cases} \tag{12}$$

where $\boldsymbol{\alpha}$ are Gaussian random effects at the knot locations, and the $\mathbf{W}$ are basis functions to recreate the Gaussian process at all sites.

7.2 Cross validation

For each dataset, we fit the model using 500 of the observations as a training set, and the remaining observations are used as a validation set to assess the model's predictive power. Because our goal is to predict a the occurrence of an event, we use Brier scores to compare the models (Gneiting and Raftery, 2007). The Brier score for predicting an occurrence at site $\mathbf{s}$ is given by $\{I[Y(\mathbf{s})=1]-P[Y(\mathbf{s})=1]\}^2$ where $I[Y(\mathbf{s})=1]$ is an indicator function indicating that an event occurred at site $\mathbf{s}$, and $P[Y(\mathbf{s})=1]$ is obtained by taking the median of the posterior distribution. We average the Brier scores over all test sites, and a lower score indicates a better fit.

We also consider the receiver operating characteristic (ROC) curve and look at the area under the curve (AUC) for the different methods and settings. AUC is commonly used as a metric to determine how well a model can classify data. When AUC = 1, the model perfectly classifies all observations in the test set, and when AUC = 0, the model misclassifies all observations in the test set. AUC is computed using the `roc`

Table 1: Relative Brier scores for GEV and Probit methods

	GEV	Probit
Setting 1	0.9047	0.9754
Setting 2	0.7885	0.9804
Setting 3	1.0275	1.0018
Setting 4	1.0264	1.0089
Setting 5	1.0458	0.9963
Setting 6	1.0565	0.9945

Table 2: Relative AUC for GEV and Probit methods

	GEV	Probit	Logit
Setting 1	0.8998	0.8973	0.8897
Setting 2	0.9458	0.9399	0.9356
Setting 3	0.7288	0.7371	0.7157
Setting 4	0.7906	0.8056	0.8115
Setting 5	0.8426	0.8458	0.8388
Setting 6	0.8756	0.8686	0.8765

function in the `pROC` package of `R` using the median of the posterior predictive distribution at the testing locations. We then average AUCs across all datasets for each method and setting to obtain a single AUC for each combination of method and setting.

7.3 Results

Table Table 1 gives the Brier score relative to the Brier score for the spatial logistic method calculated as

$$\text{BS}_{\text{rel}} = \frac{\text{BS}_{\text{method}}}{\text{BS}_{\text{logistic}}}. \tag{13}$$

Table 2 gives the AUC relative to the AUC for the spatial logistic method calculated in similarly to the relative Brier score.

We analyzed the results for this simulation study using a Friedman test at $\alpha = 0.05$ to see if at least one method had a significantly different Brier score or AUC. For any setting that yielded a significant p-value, we conducted a Wilcoxon-Nemenyi-McDonald-Thompson test to see which of the methods had

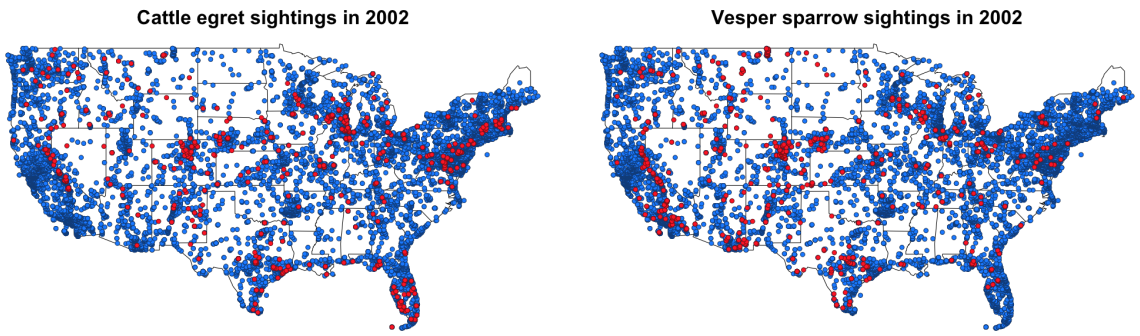

Figure 1: Reported sighting for Cattle egret (left) and Vesper sparrow (right) in 2002

different results. The full results for the Wilcoxon-Nemenyi-McDonald-Thompson tests are given in A. For all settings, we find significant results for the Friedman test comparing the Brier scores for the methods. Specifcally, we see a statistically significant reduction in Brier score using the GEV compared to logit for settings 1 and 2 and compared to probit for setting 2. However, in the other settings, the logit and probit methods tend to perform better than the GEV method.

The results using AUC are much less conclusive with only settings 1 and 4 demonstrating significant differences between the methods at $\alpha = 0.05$. As with the Brier scores, the GEV method shows a statistically significant increase in AUC over the logit method for setting 1, and for setting 4, the both the probit and logit methods show a statistically significant improvement in AUC over the GEV method.

8 Data analysis

For the data analysis, we consider data from the eBirds dataset, a citizen-based observation network of bird sitings in the United States (Sullivan et al., 2009). The data are publicly available from `http://ebird.org`. We use data from 2002, and focus on 10 different species. Figure 1 shows the sighting data for cattle egrets and vesper sparrows

9 Conclusions

Acknowledgments

A Appendices

A.1 Derivation of the likelihood

We use the hierarchical max-stable spatial model given by Reich and Shaby (2012). If at each margin, $Z_i \sim \text{GEV}(1,1,1)$, then $Z_i|\theta_i \overset{indep}{\sim} \text{GEV}(\theta, \alpha\theta, \alpha)$. We reorder the data such that $Y_1 = \ldots = Y_K = 1$, and $Y_{K+1} = \ldots = Y_n = 0$. Then the joint likelihood conditional on the random effect θ is

$$\begin{aligned} P(Y_1 = y_1, \ldots, Y_n = y_n) = &\prod_{i \leq K} \left\{ 1 - \exp\left[-\left(\frac{\theta_i}{z_i}\right)^{1/\alpha} \right] \right\} \prod_{i>K} \exp\left[-\left(\frac{\theta_i}{z_i}\right)^{1/\alpha} \right] \\ &= \exp\left[-\sum_{i=K+1}^{n} \left(\frac{\theta_i}{z_i}\right)^{1/\alpha} \right] - \exp\left[-\sum_{i=K+1}^{n} \left(\frac{\theta_i}{z_i}\right)^{1/\alpha} \right] \sum_{i=1}^{K} \exp\left[-\left(\frac{\theta_i}{z_i}\right)^{1/\alpha} \right] \\ &\quad + \exp\left[-\sum_{i=K+1}^{n} \left(\frac{\theta_i}{z_i}\right)^{1/\alpha} \right] \sum_{1<i<j\leq K} \left\{ \exp\left[-\left(\frac{\theta_i}{z_i}\right)^{1/\alpha} - \left(\frac{\theta_j}{z_j}\right)^{1/\alpha} \right] \right\} \\ &\quad + \cdots + (-1)^K \exp\left[-\sum_{i=1}^{n} \left(\frac{\theta_i}{z_i}\right)^{1/\alpha} \right] \end{aligned} \tag{14}$$

Finally marginalizing over the random effect, we obtain

$$P(Y_1 = y_1, \ldots, Y_n = y_n) = \int G(\mathbf{z}|\mathbf{A}) p(\mathbf{A}|\alpha) d\mathbf{A}.$$

$$\begin{aligned} &= \int \exp\left[-\sum_{i=K+1}^{n} \left(\frac{\theta_i}{z_i}\right)^{1/\alpha} \right] - \exp\left[-\sum_{i=K+1}^{n} \left(\frac{\theta_i}{z_i}\right)^{1/\alpha} \right] \sum_{i=1}^{K} \exp\left[-\left(\frac{\theta_i}{z_i}\right)^{1/\alpha} \right] \\ &\quad + \exp\left[-\sum_{i=K+1}^{n} \left(\frac{\theta_i}{z_i}\right)^{1/\alpha} \right] \sum_{1<i<j\leq K} \left\{ \exp\left[-\left(\frac{\theta_i}{z_i}\right)^{1/\alpha} - \left(\frac{\theta_j}{z_j}\right)^{1/\alpha} \right] \right\} \\ &\quad + \cdots + (-1)^K \exp\left[-\sum_{i=1}^{n} \left(\frac{\theta_i}{z_i}\right)^{1/\alpha} \right] p(\mathbf{A}|\alpha) d\mathbf{A}. \end{aligned} \tag{15}$$

168 Consider the first term in the summation,

$$\begin{aligned} \int \exp\left\{ -\sum_{i=K+1}^{n} \left(\frac{\theta_i}{z_i}\right)^{1/\alpha} \right\} p(\mathbf{A}|\alpha) d\mathbf{A} &= \int \exp\left\{ -\sum_{i=K+1}^{n} \left(\frac{\left[\sum_{l=1}^{L} A_l w_l(\mathbf{s}_i)^{1/\alpha}\right)^{\alpha}}{z_i} \right] \right\}^{1/\alpha} p(\mathbf{A}|\alpha) d\mathbf{A} \\ &= \int \exp\left\{ -\sum_{i=K+1}^{n} \sum_{l=1}^{L} A_l \left(\frac{w_l(\mathbf{s}_i)}{z_i}\right)^{1/\alpha} \right\} p(\mathbf{A}|\alpha) d\mathbf{A} \\ &= \exp\left\{ -\sum_{l=1}^{L} \left[\sum_{i=K+1}^{n} \left(\frac{w_l(\mathbf{s}_i)}{z_i}\right)^{1/\alpha} \right]^{\alpha} \right\}. \end{aligned} \tag{16}$$

169 The remaining terms in equation (15) are straightforward to obtain, and after integrating out the random

170 effect, the joint density for $K = 0, 1, 2$ is given by

$$P(Y_1 = y_1, \ldots, Y_n = y_n) = \begin{cases} G(\mathbf{z}) & K = 0 \\ G(\mathbf{z}_{(1)}) - G(\mathbf{z}) & K = 1 \\ G(\mathbf{z}_{(12)}) - G(\mathbf{z}_{(1)}) - G(\mathbf{z}_{(2)}) + G(\mathbf{z}) & K = 2 \end{cases} \tag{17}$$

where

$$G[\mathbf{z}_{(1)}] = P[Z(\mathbf{s}_2) < z(\mathbf{s}_2), \ldots, Z(\mathbf{s}_n) < z(\mathbf{s}_n)]$$

$$G[\mathbf{z}_{(2)}] = P[Z(\mathbf{s}_1) < z(\mathbf{s}_1), Z(\mathbf{s}_3) < z(\mathbf{s}_3), \ldots, Z(\mathbf{s}_n) < z(\mathbf{s}_n)]$$

$$G[\mathbf{z}_{(12)}] = P[Z(\mathbf{s}_3) < z(\mathbf{s}_3), \ldots, Z(\mathbf{s}_n) < z(\mathbf{s}_n)].$$

Similar expressions can be derived for all K, but become cumbersome for large K.

A.2 Derivation of the χ statistic

$$\begin{aligned}
\chi &= \lim_{p\to 0} \mathrm{P}(Y_i = 1 | Y_j = 1) \\
&= \lim_{p\to\infty} \frac{p + p - \left(1 - \exp\left\{-\sum_{l=1}^{L}\left[(-\log(1-p)w_l(\mathbf{s}_i))^{1/\alpha} + (-\log(1-p)w_l(\mathbf{s}_j))^{1/\alpha}\right]^{\alpha}\right\}\right)}{p} \\
&= \lim_{p\to 0} \frac{2p - \left(1 - \exp\left\{\log(1-p)\sum_{l=1}^{L}\left[w_l(\mathbf{s}_i)^{1/\alpha} + w_l(\mathbf{s}_j)^{1/\alpha}\right]^{\alpha}\right\}\right)}{p} \\
&= \lim_{p\to 0} \frac{2p - \left(1 - (1-p)^{\sum_{l=1}^{L}\left[(w_l(\mathbf{s}_i))^{1/\alpha} + (w_l(\mathbf{s}_j))^{1/\alpha}\right]^{\alpha}}\right)}{p} \\
&= \lim_{p\to 0} 2 - \sum_{l=1}^{L}\left[w_l(\mathbf{s}_i)^{1/\alpha} + w_l(\mathbf{s}_j)^{1/\alpha}\right]^{\alpha}(1-p)^{-1+\sum_{l=1}^{L}\left[w_l(\mathbf{s}_i)^{1/\alpha} + w_l(\mathbf{s}_j)^{1/\alpha}\right]^{\alpha}} \\
&= 2 - \sum_{l=1}^{L}\left[w_l(\mathbf{s}_i)^{1/\alpha} + w_l(\mathbf{s}_j)^{1/\alpha}\right]^{\alpha}. \qquad (18)
\end{aligned}$$

A Simulation study pairwise difference results

The following tables show the methods that have significantly different Brier scores when using a Wilcoxon-Nemenyi-McDonald-Thompson test. In each column, different letters signify that the methods have signifi-

cantly different Brier scores.

Table 3: Pairwise BS comparisons

	Setting 1	Setting 2	Setting 3	Setting 4	Setting 5	Setting 6
Method 1	A	A	A	C	B	B
Method 2	A B	B	A	B	A	A
Method 3	B	B	A	A	A B	A

References

Cohen, J. (1960) A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, **20**, 37–46.

Coles, S., Heffernan, J. and Tawn, J. (1999) Dependence Measures for Extreme Value Analyses. *Extremes*, **2**, 339–365.

Genton, M. G., Ma, Y. and Sang, H. (2011) On the likelihood function of Gaussian max-stable processes. *Biometrika*, **98**, 481–488.

Gneiting, T. and Raftery, A. E. (2007) Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of the American Statistical Association*, **102**, 359–378.

de Haan, L. (1984) A Spectral Representation for Max-stable Processes. *The Annals of Probability*, **12**, 1194–1204.

Heagerty, P. and Lele, S. (1998) A Composite Likelihood Approach to Binary Spatial Data. *Journal of the American Statistical Association*, **1459**, 1099–1111.

Huser, R. and Davison, A. C. (2014) Space-time modelling of extreme events. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **76**, 439–461.

Padoan, S. A., Ribatet, M. and Sisson, S. A. (2010) Likelihood-Based Inference for Max-Stable Processes. *Journal of the American Statistical Association*, **105**, 263–277.

Reich, B. J. and Shaby, B. A. (2012) A hierarchical max-stable spatial model for extreme precipitation. *The Annals of Applied Statistics*, **6**, 1430–1451.

Stephenson, A. G. (2009) High-Dimensional Parametric Modelling of Multivariate Extreme Events. *Australian & New Zealand Journal of Statistics*, **51**, 77–88.

Sullivan, B. L., Wood, C. L., Iliff, M. J., Bonney, R. E., Fink, D. and Kelling, S. (2009) eBird: A citizen-based bird observation network in the biological sciences. *Biological Conservation*, **142**, 2282–2292.

Tawn, J. A. (1990) Modelling multivariate extreme value distributions. *Biometrika*, **77**, 245–253.

Thibaud, E. and Opitz, T. (2015) Efficient inference and simulation for elliptical Pareto processes. *Biometrika*, **102**, 855–870.

Wadsworth, J. L. and Tawn, J. A. (2014) Efficient inference for spatial extreme value processes associated to log-Gaussian random functions. *Biometrika*, **101**, 1–15.

Wang, X. and Dey, D. K. (2010) Generalized extreme value regression for binary response data: An application to B2B electronic payments system adoption. *The Annals of Applied Statistics*, **4**, 2000–2023.