

A NEW CLASS OF FLEXIBLE LINK FUNCTIONS WITH APPLICATION TO SPECIES CO-OCCURRENCE IN CAPE FLORISTIC REGION

BY XUN JIANG^{*}, DIPAK K. DEY^{*}, RACHEL PRUNIER[†],
ADAM M. WILSON[‡] AND KENT E. HOLSINGER^{*}

University of Connecticut^{}, Western Connecticut State University[†] and
Yale University[‡]*

Understanding the mechanisms that allow biological species to co-occur is of great interest to ecologists. Here we investigate the factors that influence co-occurrence of members of the genus *Protea* in the Cape Floristic Region of southwestern Africa, a global hot spot of biodiversity. Due to the binomial nature of our response, a critical issue is to choose appropriate link functions for the regression model. In this paper we propose a new family of flexible link functions for modeling binomial response data. By introducing a power parameter into the cumulative distribution function (c.d.f.) corresponding to a symmetric link function and its mirror reflection, greater flexibility in skewness can be achieved in both positive and negative directions. Through simulated data sets and analysis of the *Protea* co-occurrence data, we show that the proposed link function is quite flexible and performs better against link misspecification than standard link functions.

1. Introduction. Understanding the underlying processes that govern the assembly of biological communities has long been of great interest to ecologists. Obviously, in the absence of species interactions and species habitat preferences, the probability that two species co-occur in a site would simply be the product of the site occupancy probabilities for each of the species. In most biological communities, however, competition [Elton (1946)] and individual response to the environment [Weiher and Keddy (1995)] are likely to play important roles in determining the species composition of local communities. Since phenotypic traits of species and environmental factors could mediate both competition and individual response, the probability of co-occurrence could also be influenced by both the traits of the species and the specific environmental conditions associated with a site. In this study we investigate the processes of community assembly in a well-defined clade, the genus *Protea* in the Cape Floristic Region (CFR) of southwestern Africa. The response variable, the number of co-occurrences of a certain pair of *Protea* species, arises naturally as a binomial variable when we define co-occurrence as the number of sites in which two species co-occur within naturally

Received December 2012; revised June 2013.

Key words and phrases. Bayesian method, community ecology, generalized linear model, MCMC, model selection, symmetric power link function.

nested watersheds. We take into consideration the spatial association among the co-occurrence of *Protea* species since it is natural to suspect areas close by would tend to have similar number of co-occurrences as a result of a latent spatial effect. Our primary interest in this study is to build a comprehensive model that could identify important factors influencing the assembly of *Protea* communities, while adjusting for both spatial association and prevalence of *Protea* in CFR.

The usual way to model the binomial response is to use a Generalized Linear Model (GLM), where we model the latent probability of “success” by a linear function of covariates through a link function [McCullagh and Nelder (1989)]. The logit, probit and Student t link functions are three of the common links used in GLM. However, the link functions mentioned above are “symmetric” links in the sense that they assume that the latent probability of a given binomial response approaches 0 with the same rate as it approaches 1. Equivalently, the probability density function (p.d.f.) that corresponds to the inverse cumulative distribution function (c.d.f.) of the link function is symmetric. However, this may not be a reasonable assumption in many cases. A commonly adopted asymmetric link function is the complementary loglog (cloglog) link function. However, the cloglog link has a fixed negative skewness. As a result, it lacks both the flexibility to let the data tell how much skewness should be incorporated and the ability to allow for positive skewness. In short, binomial data might often be better modeled with flexible link functions that allow for both positive and negative skew and that allow the data to determine the amount of skewness required.

Much work has been done to introduce flexibility into the link functions. Aranda-Ordaz (1981) proposed two separate one-parameter models for additional flexibility in the logistic model. Guerrero and Johnson (1982) used Box–Cox transformation on the odds ratio to form a more flexible class of model. Jones (2004) proposed a family of flexible distributions based on the distribution of order statistics. Stukel (1988) proposed a two-parameter class of generalized logistic models. Stukel’s model approximates many standard symmetric and asymmetric link functions quite well, but in a Bayesian framework, it may result in improper posteriors when the usual improper uniform prior is used in regressions [Chen, Dey and Shao (1999)]. Kim, Chen and Dey (2008) proposed a class of generalized skewed t -link models using a latent variable approach, which achieves proper posteriors for regression coefficients under uniform priors. Unfortunately, the range of the skewness for generalized skewed t -link is limited due to a constraint on the shape parameter required for identifiability of the model. More recently, Wang and Dey (2010) propose the generalized extreme value link function to allow more flexible skewness controlled by the shape parameter, but the standard logistic and probit links are not among the special cases of this family.

Several authors have proposed an additional power parameter on the c.d.f. corresponding to standard link functions. Nagler (1994) introduces the Scobit model, which is a generalization of the logistic model by introduction of a power

parameter. In the psychology literature, Samejima (2000) proposes the Logistic Positive Exponent Family using similar ideas. These models are part of the asymmetric parametric family proposed by Aranda-Ordaz (1981) under some reparameterizations. Gupta and Gupta (2008) propose the power normal distribution to accommodate skewness and discuss its advantages over the skew normal distribution. However, even though those link functions with power parameters can accommodate flexible skewness in one direction (e.g., positive skewness in the Scobit link), the skewness in the opposite direction can be asymmetrically limited.

In this paper we propose a new class of symmetric power link functions to model binary and binomial data, and apply it to the *Protea* species co-occurrence data. The rest of the paper is organized as follows. We introduce the *Protea* species co-occurrence data in Section 2. In Section 3 we propose a general class of power link functions based on the c.d.f. corresponding to a symmetric baseline link function and its mirror reflection. Section 4 discusses the prior specification and posterior proprieties of the parameter in the proposed model under a fully Bayesian framework. In Section 5 we introduce spatial random effects in the model to account for the spatial association in the co-occurrence data. Section 6 clarifies some computational issues in the model as well as the criteria for model comparisons. Several comprehensive simulation studies are reported in Section 7 with detailed discussions. Finally, in Section 8 we fit the proposed model on the *Protea* species co-occurrence data. We conclude our paper in Section 9 and all the proofs of the theorems are deferred to the Appendices.

2. The *Protea* species co-occurrence data. The Cape Floristic Region (CFR) is a region with remarkable biological diversity. The *Protea* species co-occurrence data we study here is derived from the Protea Atlas data set (<http://www.proteaatlas.org.za>), which includes 96,253 occurrence records for the 71 species in the genus *Protea* from 44,415 sites. Wilson and Prunier (unpublished data) constructed a series of nested watersheds covering the CFR using the 3 arc-second (90 m) research-grade digital elevation model collected by the Shuttle Radar Topography Mission (available at <http://seamless.usgs.gov>) using the `r.watersheds` function in GRASS [Grass Development Team (2008)]. The co-occurrence data used here correspond to species co-occurrences within watersheds having a mean area of approximately 55 km² (± 40 km²) nested within larger watersheds with a mean area of approximately 540 km² (± 425 km²) [see Figure 1(a) for an illustration of nested watersheds in CFR]. The smaller watersheds are considerably larger than those usually considered in community assembly studies [Vamosi et al. (2009)]. As a result, factors that are associated with a reduced probability of co-occurrence in this analysis may reflect either the consequences of competitive interactions or of habitat segregation among different watersheds.

The data are binomial because for each pair of *Protea* species, we record the number of smaller watersheds at which a particular pair co-occurs out of the total number of smaller watersheds contained within each larger watershed. The data are

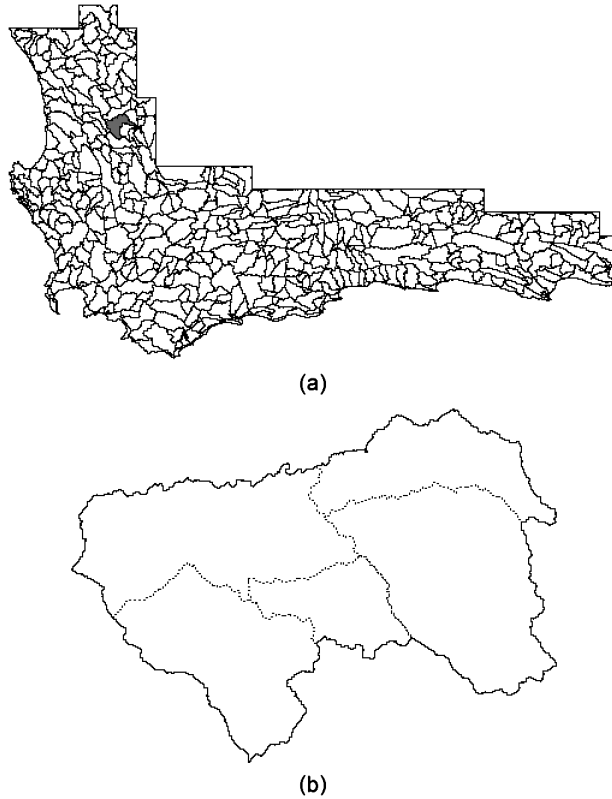


FIG. 1. Illustration of the watersheds scale in CFR. (a) The CFR showing the boundaries between the large watersheds within which smaller watersheds are nested. (b) Smaller watersheds nested within one particular parent watershed. This parent watershed is highlighted in grey in part (a).

then aggregated across the larger watersheds to cover the entire CFR. In this study we record 10,256 observations involving $\binom{71}{2}$ pairs of *Protea* species. In Figure 2

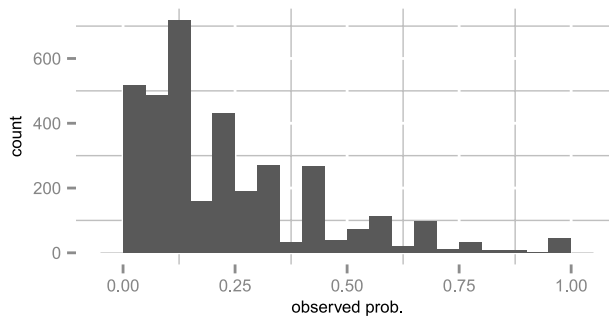


FIG. 2. Histogram of empirical frequencies of *Protea* species co-occurrence.

we plot the histogram of observed probability of co-occurrence (number of *Protea* observed co-occurrences between species pairs divided by the total number of co-occurrences possible). Although this histogram is not excessively skewed, it is reasonable to suspect that an asymmetric link function will be more appropriate for these data, because the histograms for individual species pairs can be very skewed. Our observational unit is species pairs, and for each pair of species we record seven traits that could affect the probability that they co-occur. They are as follows: (1) phylogenetic distance, which is proportional to the time since the two species diverged from a common ancestor, (2) whether or not they differ in fire survival strategy, (3) the difference in plant height, (4) the difference in month of maximum flowering, (5) whether or not they share a pollination syndrome, (6) the difference in leaf area, and (7) the difference in leaf length:width ratio. The difference is measured as either 1:0 for binary data or Euclidean distance for the continuous traits.

Our estimate of phylogenetic distance is derived from a rate-smoothed version of the phylogenetic tree presented in Valente et al. (2010). Specifically, using the topology presented in Valente et al. (2010), we estimated branch lengths under a maximum-likelihood model in PAUP* using the data used to generate the tree: DNA sequences from four chloroplast markers (*trnL* intron, *trnL-trnF* spacer, *rps16*, *atpB-rbcL* spacer), two nuclear regions (ITS and *ncpGS*), and 138 AFLP loci. We smoothed the branch lengths using NPRS in r8s [Sanderson (2003)] and calculated pairwise phylogenetic distances using cophenetic.phylo from APE [Paradis, Claude and Strimmer (2004)].

3. The symmetric power link family. Let us first specify the notation used throughout this paper. Suppose $y_i \sim \text{Binomial}(p_i, N_i)$, where p_i is the probability of success for the i th observation. Let the design matrix be \mathbf{X} with $x_i = (1, x_{i1}, x_{i2}, \dots, x_{ik})'$ the i th row of \mathbf{X} and $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \dots, \beta_k)'$ the corresponding regression coefficients. We associate p_i and x_i through a c.d.f. F as follows:

$$(3.1) \quad p_i = F(x_i' \boldsymbol{\beta}),$$

where we call F^{-1} the corresponding link function. The logit, probit, Student t link as well as the cloglog link functions are common links adopted for the binomial regression models.

Here we propose a general class of flexible link functions based on a symmetric baseline link function and its mirror reflection in the following manner. If F_0^{-1} is a baseline link function with corresponding c.d.f. F_0 for which the p.d.f. is symmetric about zero, we propose the symmetric power link family based on F as

$$(3.2) \quad F(x, r) = F_0^r \left(\frac{x}{r} \right) \mathbf{I}_{(0,1]}(r) + [1 - F_0^{1/r}(-rx)] \mathbf{I}_{(1,+\infty)}(r).$$

The intuition for the development of (3.2) is to utilize the fact that $F_0^r(x)$ is a valid c.d.f. and it achieves flexible left skewness when $r < 1$, while the same property holds for its mirror reflection $1 - F_0^{1/r}(-x)$ with skewness being in the opposite direction. By combining the two in one single family of link functions, we could achieve flexibility in positive as well as negative skewness symmetrically with respect to the baseline link. Also, we are scaling x by the same parameter r in the formulation to prevent the mode of the p.d.f. to be too far away from zero. Clearly, by introducing an additional parameter r in (3.2), the skewness of the symmetric power link family can be adjusted from its baseline to achieve more flexibility in modeling asymmetric data.

One immediate observation in (3.2) is that $F(x, 1) = F_0(x)$, so the proposed family includes the baseline c.d.f. F_0 as a special case. Also, considering the fact that F_0 is symmetric, the proposed symmetric power link family is continuous at the break point $r = 1$, since

$$(3.3) \quad \lim_{r \rightarrow 1^+} F(x, r) = 1 - F_0(-x) = F_0(x) = F(x, 1).$$

As we will be dealing with introduction of flexible skewness into the link function, we specify our measurement of skewness here. We adopt Arnold and Groeneveld's (1995) skewness measure with respect to the mode here. Under certain conditions, the skewness of a random variable X is defined as $\gamma_M = 1 - 2F(M_x)$, where $F(\cdot)$ is the c.d.f. of X with corresponding mode M_x . By definition, the skewness is between -1 and 1 , with 0 indicating symmetry. In (3.2), it follows directly that $F(x, r) = 1 - F(-x, \frac{1}{r})$. In other words, the p.d.f. of the symmetric power family with power parameter r is the mirror image of the p.d.f. with power parameter $\frac{1}{r}$. This implies that if the skewness of $F(x, r)$ is ξ , then the skewness of $F(x, \frac{1}{r})$ will be $-\xi$. Here, by combining the power of a standard symmetric link distribution function and its reflection in one single link, we can accommodate flexible skewness in both directions simultaneously, while retaining the desirable property of having the standard baseline link function as a special case. We propose three symmetric power link function families based on different baseline link functions as follows.

3.1. *The symmetric power logit (splogit) link family.* If we choose F_0 to follow the logistic distribution with location 0 and scale 1 , then we call $F(x, r)$ defined in (3.2) the symmetric power logit (splogit) family, and we call the corresponding link function the splogit link. The skewness of the splogit distribution can be found analytically as $\gamma_M = 1 - 2(\frac{r}{r+1})^r$ for $0 < r < 1$, and $\gamma_M = 2(\frac{1}{r+1})^{1/r} - 1$ for $r > 1$. As a result, it is negatively skewed when $0 < r < 1$, positively skewed when $r > 1$, and reduces to the symmetric logit link when $r = 1$. Figure 3(a) and (d) shows the p.d.f. and c.d.f. corresponding to the splogit link with $r = 0.2, 1, 5$, respectively. It is clear that as the power parameter r varies, so does the approaching rate to 0 and 1 for the splogit link. The range of skewness provided by the splogit

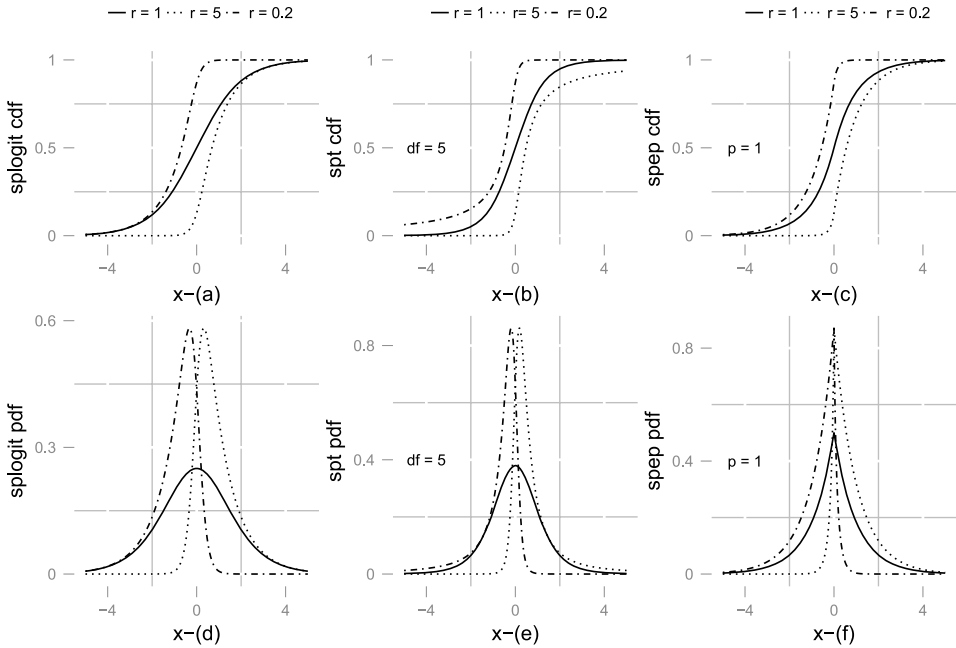


FIG. 3. Symmetric power link c.d.f. and p.d.f. for different value of r under logit, Student t and exponential power baseline c.d.f. functions.

family is unlimited, reaching -1 and 1 , respectively, as r tends to 0 and $+\infty$ [see Figure 4(a)].

3.2. *The symmetric power t (spt) link family.* Many authors have suggested using a Student t link (degrees of freedom denoted as ν) as an alternative to the logit and probit links. Mudholkar and George (1978) show that the Student t link with 9 degrees of freedom has the same kurtosis as the logistic distribution. Albert

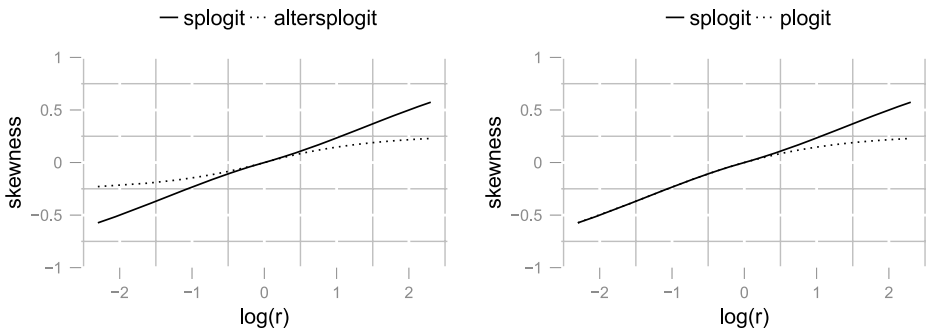


FIG. 4. Skewness range of splogit against plogit and altersplogit as $\log(r)$ varies. The possible skewness ranges from -1 to 1 under the definition of Arnold and Groeneveld (1995).

and Chib (1993) suggest using the t distribution with 8 degrees of freedom and provide a full implementation in a Bayesian framework. Liu (2004) proposes the robit model which uses the Student t distribution with known or unknown degrees of freedom as the link function and shows that it is a robust alternative to the logit and probit model. It is widely known that the Student t link with large degrees of freedom approximates the probit link. Now, by bringing in the power parameter r in the sense of (3.2), we can add more flexible skewness in the class of Student t link functions. We call this new class of link functions the symmetric power t (spt) family. Similarly, in Figure 3(b) and (e) we see the p.d.f. and c.d.f. for the spt link with 5 degrees of freedom for different values of r . Clearly, the spt link family allows us to adjust both the skewness of the distribution and the heaviness of the tails by varying r and ν , therefore accounting for an extremely rich class of link functions. In Arnold and Groeneveld's (1995) sense, the closed form expression of the spt link skewness is not available, but numerically we can show the skewness is quite flexible with small to medium ν but becomes more restricted as ν increases. Notice that when r is fixed to be positive, the symmetric power probit distribution is very similar to the power normal distribution proposed by Gupta and Gupta (2008).

3.3. *The symmetric power exponential power (spep) link family.* The exponential power (ep) distribution was first introduced by Subbotin (1923). The ep distribution is symmetric with density

$$(3.4) \quad f(x; \mu, \sigma, p) = c^{-1} \exp\left(-\frac{|z|^p}{p}\right),$$

where $-\infty < x < +\infty$, $\sigma > 0$, $p \geq 1$, $z = (x - \mu)/\sigma$, and $c = 2\sigma p^{1/p-1}\Gamma(1/p)$. Clearly, normal distribution is a special case with $p = 2$, and heavier tail distribution can be obtained as we set p to be less than 2. Also, the Laplace distribution is another special case when $p = 1$. If we set $\mu = 0$ and $\sigma = 1$, the ep distribution becomes symmetric about zero and with flexible tail properties as p varies. If we set the ep as our baseline link function F_0 , we end up with the symmetric power exponential power (spep) link family [see Figure 3(c) and (f) for the corresponding p.d.f. and c.d.f. for $p = 1$ at different values of r]. Here we restrict p to be within the range of $[1, 2]$ for our proposed spep link family since the skewness of the p.d.f. becomes restricted for $p > 2$, that is, with a thinner tail than normal distributions. However, even with this restriction, the spep link family still provides extremely flexible range of skewness and adjustment of tail behavior in one single family of link functions.

3.4. *Comparison with other power link.* As discussed in Section 1, many authors have proposed to bring in a power parameter to allow more flexible skewness

in the link function. Again, let F_0^{-1} be the baseline link; the traditional power link family is defined by

$$(3.5) \quad H(x, r) = F_0^r(x).$$

Also, by choosing a different tail in (3.2), there is an alternative way of constructing the symmetric power link given by choosing the other side of the tail as follows:

$$(3.6) \quad F^*(x, r) = [1 - F_0^{1/r}(-rx)]\mathbf{I}_{(0,1]}(r) + F_0^r\left(\frac{x}{r}\right)\mathbf{I}_{(1,+\infty)}(r).$$

Here, adopting the logit baseline for all three, we compare model (3.2) (splogit) with model (3.5) (plogit) and (3.6) (altersplogit) to illustrate the advantage of our proposed symmetric power link in terms of skewness range. Adopting the other baseline link discussed above will lead to similar results.

The advantage of skewness range of splogit is illustrated in Figure 4. Comparing with plogit, when $\log(r)$ is negative, the skewness of splogit and plogit is exactly the same, which is due to the fact that the formulation of splogit follows closely as plogit when $0 < r \leq 1$. However, on the other side, the skewness of splogit reaches 1 as $\log(r)$ goes to infinity, which has a clear advantage over plogit with a skewness limit of 0.264. Similar comparison reveals that by choosing the appropriate side, splogit has skewness advantage on both tails over the altersplogit link. A more flexible skewness means that the probability of success under the splogit link can approach 0 (or 1) in a rate that can never be achieved under the plogit or altersplogit link, which makes it more flexible in dealing with skewed data, as we will show later in the simulation study.

4. The prior and posterior proprieties. We adopt the following class of prior distributions on our proposed symmetric power link family and investigate its posterior proprieties. For regression coefficient β , we adopt the usual uniform prior, that is, $\pi(\beta) \propto 1$. For the power parameter r , we adopt a proper gamma prior $\pi(r)$ with mean one and reasonably large variance. If we denote the likelihood of the model to be $\mathbf{L}(\beta, r|\mathbf{y})$, consequently, the joint posterior density of our regression model becomes

$$(4.1) \quad \begin{aligned} \pi(\beta, r|\mathbf{y}) &\propto \mathbf{L}(\beta, r|\mathbf{y})\pi(\beta)\pi(r) \\ &\propto \prod_{i=1}^n [F(x'_i\beta, r)]^{y_i} [1 - F(x'_i\beta, r)]^{1-y_i} \pi(r). \end{aligned}$$

Clearly, the posterior distribution is proper if and only if

$$(4.2) \quad \int_{\mathbb{R}^+} \int_{\mathbb{R}^k} \mathbf{L}(\beta, r|\mathbf{y})\pi(r) d\beta dr < \infty.$$

Notice that when r is a point mass at 1, $F(x, r)$ becomes the baseline link function F_0 . Our goal is to investigate whether the introduction of a power parameter r

would change the posterior propriety compared to merely adopting the symmetric baseline function F_0 . Here we let $\mathbf{L}_0(\boldsymbol{\beta}|y)$ denote the likelihood under the corresponding baseline link F_0 .

THEOREM 1. *If under the baseline link F_0 we have*

$$(4.3) \quad \int_{\mathfrak{R}^k} \mathbf{L}_0(\boldsymbol{\beta}|y) d\boldsymbol{\beta} < \infty,$$

then the posterior under the corresponding power link $\pi(\boldsymbol{\beta}, r|y)$ is also proper, that is,

$$(4.4) \quad \int_{\mathfrak{R}^+} \int_{\mathfrak{R}^k} \mathbf{L}(\boldsymbol{\beta}, r|y)\pi(r) d\boldsymbol{\beta} dr < \infty.$$

Theorem 1 states that by introducing an additional power parameter r in the sense of (3.2), the posterior propriety under the uniform $\boldsymbol{\beta}$ prior is unchanged with a proper prior for r . Chen and Shao (2001) studied the conditions for the propriety of the posterior distribution under general link functions. The following theorem resolves the posterior proprieties of the three proposed symmetric power link families under uniform $\boldsymbol{\beta}$ priors. For the spt and spep link families, throughout this paper we adopt proper priors on ν and p .

THEOREM 2. *Let $\tau_i = 1$ if $y_i = 0$ and $\tau_i = -1$ if $y_i = 1$, and define \mathbf{X}^* to be the matrix with the i th row $\tau_i x_i'$. Suppose the following conditions hold:*

- (i) *The design matrix \mathbf{X} is of full column rank.*
- (ii) *There is a positive vector $\mathbf{a} = (a_1, a_2, \dots, a_n)' \in \mathfrak{R}^n$ such that $\mathbf{a}'\mathbf{X}^* = 0$.*

Then the proposed splogit and spep links lead to proper posteriors under the above prior setup, while the same result also holds for the spt family with degrees of freedom $\nu > k$, where k is the number of columns of \mathbf{X} .

5. Spatial random effects. Geological and climatic features vary greatly across the CFR. For example, the climate in the western part of CFR is Mediterranean with rainfall concentrated in the winter months, while the climate in the eastern CFR is more aseasonal. While such climate features could affect the pattern of *Protea* co-occurrence, we are primarily interested in identifying biological features that influence co-occurrence. Thus, we add spatial random effects in the model to account for all latent, unmeasured environmental effects that are spatially structured. Here we use the intrinsically conditionally autoregressive (ICAR) model [Besag and Green (1993)] to capture these spatial effects. The ICAR model has gained increasing usage in the past two decades due to its convenient implementation in the context of Gibbs sampling for fitting hierarchical spatial models [Banerjee, Carlin and Gelfand (2004)].

In the *Protea* co-occurrence context we suppose that there are K parent watersheds and that the proximity matrix A is defined by $A_{ij} = 1$ if watersheds i and j are adjacent and $A_{ij} = 0$ otherwise. Following notation in Section 3, the number of co-occurrence between species l and m within watershed k is modeled as

$$(5.1) \quad y_{lmk} \sim \text{Binomial}(p_{lmk}, N_k),$$

$$(5.2) \quad p_{lm} = F(x'_{lm}\boldsymbol{\beta} + w_k),$$

where N_k is the number of smaller watersheds within parent watershed k , and w_k is the spatial random effect associated with watershed k , where $k = 1, 2, \dots, K$. At the next stage, the spatial random effects (w_1, w_2, \dots, w_K) follow the ICAR model, that is,

$$(5.3) \quad p(w_1, w_2, \dots, w_K) \propto \exp\left[-\frac{1}{2\tau^2} \sum_{i \neq j} A_{ij}(w_i - w_j)^2\right].$$

Clearly, (5.3) is not a proper distribution, so it cannot be used to model data directly. However, here we use it as a prior distribution on the second stage of the hierarchical model which avoids this problem. In addition, to make \mathbf{w} fully identifiable, we impose the constraint $\sum_k w_k = 0$.

6. Computational issues.

6.1. *Markov chain Monte Carlo (MCMC) sampling.* The posterior distribution given in (4.1) is relatively easy to sample given the standard baseline link c.d.f. F_0 . To run the Gibbs sampler, we subsequently sample from the complete conditional distributions $[\boldsymbol{\beta}|r, \mathbf{y}]$, $[r|\boldsymbol{\beta}, \mathbf{y}]$ (also $[v|\boldsymbol{\beta}, r, \mathbf{y}]$ if under *spt* and $[p|\boldsymbol{\beta}, r, \mathbf{y}]$ if under *spep*). Each draw can be done using the Adaptive Rejection Metropolis algorithm [Gilks, Best and Tan (1995)] which is implemented in JAGS [Plummer (2003)]. Due to the conditional nature of the ICAR distribution, the Gibbs sampler of spatial random effects is conveniently constructed and the details are discussed in Banerjee, Carlin and Gelfand (2004). All the computations in this paper are done in JAGS or geoBUGS.

6.2. *Covariate effects.* Czado and Santner (1992) pointed out that it is more appropriate to compare the covariate effects under different link functions with the estimated probabilities since the estimates of $\boldsymbol{\beta}$ will depend on the choice of link functions. In view of this, we use the method suggested by Chib and Jeliazkov (2006) to calculate the average effect of the covariates estimated probabilities. Here we denote the set of all parameters in our model to be $\boldsymbol{\theta}$. For example, if we want to estimate the effect of covariate x_i , we integrate out parameters $\boldsymbol{\theta}$ by its MCMC posterior samples, and marginalize out other covariates \mathbf{x}_{-i} by their empirical distributions to get an estimate of the predictive distribution

$$(6.1) \quad [p|x_i, \mathbf{y}] = \int [p|x_i, \mathbf{x}_{-i}, \boldsymbol{\theta}, \mathbf{y}]\pi(\boldsymbol{\theta}|\mathbf{y})\pi(\mathbf{x}_{-i}) d\boldsymbol{\theta} d\mathbf{x}_{-i}.$$

Then if we compute this estimated probability under two specific values of x_i , for example, under 0 and 1, the difference in the computed probabilities gives an estimated effect of covariates x_i as it changes from 0 to 1.

6.3. Bayesian model comparison. To compare the performance of models under different link functions, we calculate two summary measures. The first one is the Deviance Information Criterion (DIC), which balances the fit of a model to the data with its complexity. The DIC measure is calculated with the posterior mean of deviance penalized by the effective number of parameters under the Bayesian framework [Spiegelhalter et al. (2002)]. The other measure we consider here is the logarithm of the pseudo-marginal likelihood (LPML), which measures the accuracy of prediction based on leave-one-out cross-validation ideas. The LPML measure [Ibrahim, Chen and Sinha (2005)] is a summary statistic of the conditional predictive ordinate (CPO) criterion [Gelfand, Dey and Chang (1992)]. The model with the larger LPML indicates better fit of competing models.

7. Simulation study. Here we conduct three sets of simulation studies. The first one compares the performance of our proposed symmetric power link function against some other standard or flexible link functions. The second one focuses on the performance of splogit link versus plogit link when the data is generated by a skewed distribution. The third one investigates specifically how our proposed model performs against other flexible link functions on a larger scale simulation. To simplify our simulation, here we consider Bernoulli responses instead of binomial. Before we go any further, we introduce two other flexible link functions for comparison purposes.

Stukel (1988) proposed the generalized logistic link family with parameter $\alpha = (\alpha_1, \alpha_2)'$ as follows:

$$p_i = G(h_\alpha(x_i'\beta)),$$

where G is the c.d.f. of the logistic distribution. When $x_i'\beta \geq 0$,

$$h_\alpha(x_i'\beta) = \begin{cases} \alpha_1^{-1}(\exp(\alpha_1 x_i'\beta) - 1), & \alpha_1 > 0, \\ x_i'\beta, & \alpha_1 = 0, \\ -\alpha_1^{-1} \log(1 - \alpha_1 x_i'\beta), & \alpha_1 < 0, \end{cases}$$

and for $x_i'\beta \leq 0$,

$$h_\alpha(x_i'\beta) = \begin{cases} -\alpha_2^{-1}(\exp(\alpha_2 |x_i'\beta|) - 1), & \alpha_2 > 0, \\ x_i'\beta, & \alpha_2 = 0, \\ \alpha_2^{-1} \log(1 - \alpha_2 |x_i'\beta|), & \alpha_2 < 0. \end{cases}$$

Also, Czado (1994) proposed another two parameters family link functions given by specifying

$$h_{\alpha}(x_i'\beta) = \begin{cases} \frac{(x_i'\beta + 1)^{\alpha_1} - 1}{\alpha_1}, & x_i'\beta \geq 0, \\ -\frac{(-x_i'\beta + 1)^{\alpha_2} - 1}{\alpha_2}, & x_i'\beta < 0. \end{cases}$$

First we compare splogit link with other link functions with detailed simulation. We generate 2 covariates for our simulation study. We independently generate one binary covariate x_1 with 0 and 1 randomly chosen, and the second covariate x_2 is generated independently from $N(0, 3)$. Our vector of covariates is denoted as $X = (1, x_1, x_2)'$. The true regression coefficient $\beta = (\beta_0, \beta_1, \beta_2)'$ is set to be $(0, 1, 1)'$ for all simulations. With the same value of $X'\beta$, we carry out our studies under three scenarios based on three true models as follows:

SCENARIO 1. The binary data are generated from the symmetric logistic link model with $F^{-1}(p_i) = \log(p_i/(1 - p_i))$.

SCENARIO 2. The binary data are generated from the complementary loglog (cloglog) link with $F^{-1}(p_i) = \log(-\log(1 - p_i))$. It is easily calculated that the skewness of the corresponding F is -0.264 , under the definition of Arnold and Groeneveld (1995).

SCENARIO 3. The binary data are generated from the loglog link with $F^{-1}(p_i) = -\log(-\log(p_i))$. The corresponding F is the mirror reflection of the c.d.f. corresponding to the cloglog link, and therefore with skewness 0.264 .

As described in Section 6, we conduct a fully Bayesian analysis on the above three simulated data sets. The prior of β is chosen to be $N(0, 10^4)$ and the prior for r is set to be exponential with parameter 1. In the spt model the prior for v is chosen to be Gamma(8, 1) and for the spep model the prior for p is chosen to be unif(1, 2). In the ‘‘Stukel’’ and ‘‘Czado’’ models, the priors for parameters a_1 and a_2 are set to be $N(0, 10^2)$. For each scenario, we repeat the same setting for two different sample sizes $N = 500$ and $N = 2000$ to see how sample size would affect our inference. After 2000 burn-ins, the models mix pretty well and we obtain 4000 posterior samples for each parameter. We summarize DIC and LPML measures in order to make model comparisons. Our simulation results are summarized in Table 1 and in Figure 5.

We notice from Figure 5 that under sample sizes of both 500 and 2000, the posterior mean of r in the splogit link is close to 1 when the true model is logit (symmetric), significantly less than 1 when the true model is cloglog (left-skewed), and significantly greater than 1 when the true model is loglog (right skewed). Not

TABLE 1

Posterior covariate effects for β_1 and β_2 , posterior median for v and p in *spt* and *spep* models as well as model comparisons under sample sizes 500 and 2000. Bold numbers indicate the corresponding fit is under the true model. Covariate effects are measured between value 0 and 1. Larger LPML and smaller DIC indicate a better fit

Fitted model	True model														
	logit					cloglog					loglog				
	β_1	β_2	v/p	LPML	DIC	β_1	β_2	v/p	LPML	DIC	β_1	β_2	v/p	LPML	DIC
Sample size = 500															
logit	0.12	0.11		-180.6	361.1	0.15	0.06		-137.6	274.7	0.13	0.17		-136.9	273.6
cloglog	0.10	0.15		-184.3	367.4	0.15	0.05		-131.1	262.1	0.12	0.23		-145.1	289.5
loglog	0.15	0.09		-189.2	375.4	0.14	0.06		-161.2	317.4	0.14	0.14		-134.4	268.6
Stukel	0.11	0.09		-183.4	365.8	0.16	0.04		-131.7	261.5	0.13	0.14		-137.2	270.5
Czado	0.11	0.10		-182.4	363.9	0.16	0.04		-132.0	262.3	0.13	0.14		-136.7	269.7
splogit	0.12	0.11		-181.0	359.7	0.15	0.05		-131.5	262.1	0.13	0.16		-135.9	270.9
spt	0.12	0.11	7.23	-181.1	360.5	0.15	0.04	6.02	-131.7	261.9	0.13	0.16	7.43	-136.4	271.7
spep	0.11	0.11	1.41	-181.2	360.6	0.15	0.05	1.23	-132.6	264.0	0.13	0.16	1.38	-135.9	270.9
Sample size = 2000															
logit	0.13	0.10		-696.6	1393.1	0.09	0.04		-491.3	982.5	0.11	0.16		-507.0	1013.9
cloglog	0.13	0.14		-711.6	1422.0	0.09	0.02		-477.7	955.6	0.11	0.22		-544.7	1088.0
loglog	0.12	0.08		-730.3	1457.4	0.09	0.05		-525.8	1050.0	0.11	0.13		-500.6	1000.9
Stukel	0.13	0.10		-698.5	1396.7	0.10	0.02		-478.9	956.6	0.10	0.14		-501.8	1003.5
Czado	0.13	0.10		-698.3	1396.2	0.10	0.02		-479.0	957.1	0.10	0.13		-501.4	1002.8
splogit	0.13	0.10		-696.9	1392.9	0.09	0.02		-481.2	962.4	0.11	0.13		-501.4	1002.2
spt	0.13	0.10	7.21	-697.5	1394.2	0.09	0.03	9.50	-483.1	966.4	0.11	0.13	8.54	-502.3	1004.1
spep	0.12	0.10	1.40	-697.5	1394.1	0.09	0.03	1.43	-478.9	956.9	0.10	0.15	1.34	-501.7	1002.4

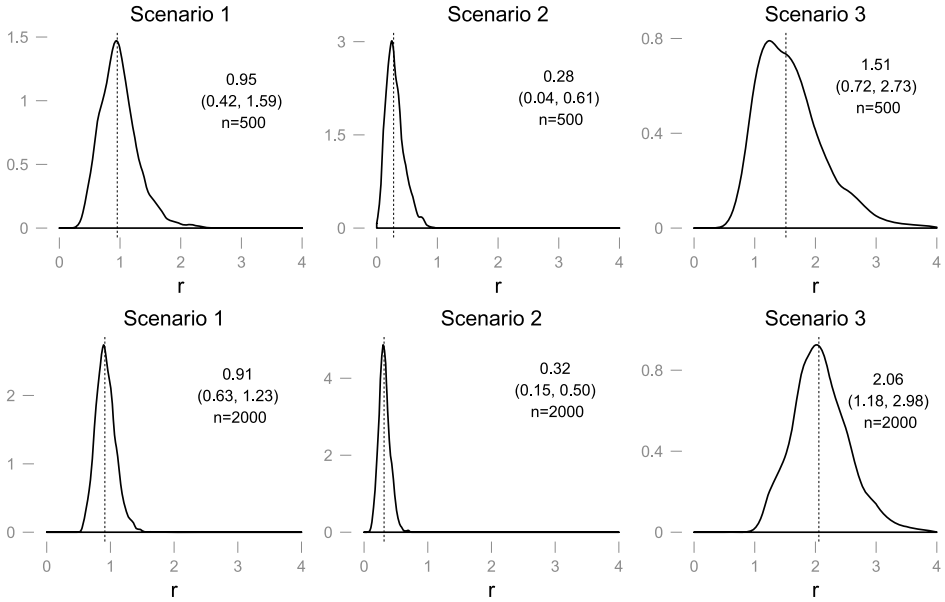


FIG. 5. Posterior density plot for power parameter r under the splogit model with sample sizes 500 and 2000. The posterior median and HPD interval of r are also reported in the plot.

surprisingly, the posterior standard deviation of r becomes considerably smaller as the sample size increases from 500 to 2000. Also, the performance of r in the spt and spep links is quite similar to the splogit link shown in Figure 5. In conclusion, the power parameter r captures the skewness of the true model very well.

Table 1 summarizes some other simulation results of the study. Comparing with standard links (logit, cloglog, loglog), calculated average covariate effect of β_1 and β_2 for the symmetric power links tends to be much closer to the value under the true model (bold) than other standard models (logit, cloglog, loglog when it is not the true model). We also observed that the symmetric power link provides estimates of covariate effects that are extremely close to the true model and significantly better than other standard link functions in terms of LPML and DIC. On the other hand, the symmetric power link performs extremely close to “Stukel” and “Czado” models in terms of both covariate effects and model comparisons. Overall, our proposed model performs well and proves to be robust enough to handle various scenarios under simulated data with symmetric, positive and negative skewness.

We conduct the second simulation study to examine the performance of the splogit link against the plogit link (3.5) and altersplogit link (3.6), when the data is generated from distributions with various skewness. We simulate data from the generalized extreme value (gev) distribution with c.d.f. as follows:

$$(7.1) \quad G(x) = \exp \left[- \left(1 + \xi \frac{x - \mu}{\sigma} \right)_+^{-1/\xi} \right].$$

Wang and Dey (2010) propose the gev model as another flexible link function to model binary response data. The skewness of (7.1) is controlled by the shape parameter ξ , which is calculated as $1 - 2 \exp\{-(1 + \xi)\}$ under Arnold and Groeneveld's (1995) definition. Notice that $\xi = -0.3$ indicates the skewness is zero. To simulate data from the gev distribution, we set F to be G with $\mu = 0, \sigma = 1$, and adopt the same covariates setup as in the first study. We choose $\xi = -3.3, -0.3, 2.7$, respectively, to represent left-skewed, symmetric and right-skewed data. For each value of ξ , we generate 100 data sets and fit splogit against plogit and altersplogit, respectively, to compare the average performance of the two. Again, we obtain 4000 posterior samples for each simulation after 2000 burn-in periods.

Figure 6 summarizes the difference of DIC between the model fits under splogit and plogit (a), splogit and altersplogit (b) of 100 replicates for each simulation.

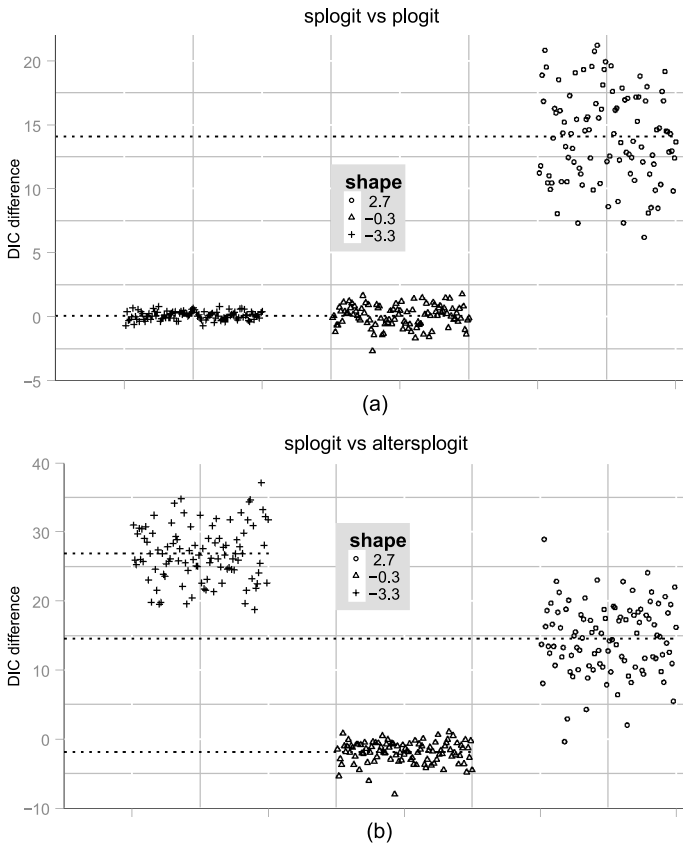


FIG. 6. Difference of DIC comparing (a) splogit and plogit, (b) splogit and altersplogit when shape parameter $\xi = -3.3, -0.3, 2.7$, respectively. Positive DIC difference indicates splogit has a better fit. In each case the simulation is repeated 100 times. The dotted line indicates the mean of DIC for the repetitions.

TABLE 2
Percentage of best performance among splogit, gev, “Stukel” and “Czado” out of 200 simulations. The best performance is determined each time by the lowest DIC value

	% Lowest DIC			
True model	splogit	gev	Stukel	Czado
logit	49.5%	31.0%	19.5%	0%
cloglog	26.5%	62.5%	8.5%	2.5%
loglog	63.0%	9.0%	24.0%	4.0%

For (a), the advantage of using the average DIC of splogit is not obvious when $\xi = -3.3, -0.3$, but becomes positive at $\xi = 2.7$. For (b), the average DIC advantage of splogit is positive at $\xi = -3.3, 2.7$, but ignorable at $\xi = -0.3$. This is exactly what we expected by looking at Figure 4 in that the splogit has skewness advantage over plogit when the data is left skewed, and has skewness advantage over altersplogit in both skewness directions.

In the third study we compare the performance of the splogit model against gev, “Stukel” and “Czado” models on a larger scale, while only focusing on model comparisons. Here, by larger scale we mean repeating the fitting of all 4 models 200 times under different true models, and record the best performance each time according to the DIC measure. To simplify the comparison, we pick sample size $N = 200$ and generate one continuous covariate from standard normal distribution. For each of the 200 simulations, the true value of $\beta = (\beta_0, \beta_1)'$ is generated from $N(1, 0.1^2)$. The prior setups of splogit, “Stukel” and “Czado” are the same as before, while the prior of ξ in the gev model is set to be uniform $(-1, 1)$. The simulation scheme is also similar to the first study, as we set the true model to be from logit, cloglog and loglog, then we fit splogit, gev, “Stukel” and “Czado” models, respectively, to find the percentage of best performance in terms of DIC.

In Table 2 we clearly see the advantage of the proposed splogit link over other link functions. The splogit link model performs the best when the true model is logit and loglog at 49.5% and 63.0%, respectively, where the gev model and “Stukel” model come as distant second places with 31.0% and 24.0%. The gev model outperforms splogit when the true model is cloglog, however, this is expected since cloglog is a special case of a gev model when $\xi = 0$. Nevertheless, in the cloglog case, the splogit model performs better than “Stukel” and “Czado” models at 26.5%. Overall, we see the robustness of the proposed splogit link against other flexible link functions.

8. Data analysis. Here we apply the model described in Section 3 on the *Protea* species co-occurrence data. The data is provided as supplementary material [Jiang et al. (2013)]. As discussed earlier, we include phylogenetic distance (GD),

fire survival strategy (FSS), plant height, month of maximum flowering (MMF), pollination syndrome, specific leaf area (SLA) and leaf length width ratio (LWR) as factors in the model and prevalence probability (with a logit transformation) as a covariate. Among them, FSS, MMF and pollination are binary and the rest are continuous. Notice that in order to model the species co-occurrence, Palmgren (1989) proposed a method of running two logistic regressions on two species separately and related the two with a regression on odds ratio. While it is no problem to integrate our sp link in the Palmgren model to replace the logit link, it is not suitable for our particular data since the co-occurrence data has the traits difference between two species, but not the traits of two species separately. For simplicity, we only adopt splogit as a representative of the symmetric power link family, however, adopting the spt and spep links would lead to similar results. After 10,000 burn-ins the model parameters and spatial random effects mix pretty well and then another 10,000 samples have been obtained. Table 3 summarizes the results under different links with ICAR prior on the spatial random effects. The priors of the regression coefficients are set to be normal with mean 0 and variance 10^4 . The spatial model is realized utilizing GeoBUGS, an add-on to WinBUGS [Lunn et al. (2000)].

First, let us look at the first half of Table 3. When we model the probability of co-occurrence, that is, $P(y = 1)$, we see that for the splogit and plogit models the estimate of the power parameter is around 0.38, corresponding to a left-skewed link. Recall that splogit is equivalent to plogit when $r \leq 1$. As a result, parameter estimates and model comparison criteria for the splogit and plogit are roughly the same. The gev model with an estimate of $\xi = 0.042$ also corresponds to a left-skewed model with a DIC slightly worse than splogit and plogit. Finally, all three flexible links perform much better than logit in terms of DIC.

In order to show the advantage of the splogit link, the second half of Table 3 considers the probability of not co-occurring, $P(y = 0)$, instead. We notice that due to the symmetric construction of splogit, modeling $P(y = 0)$ and $P(y = 1)$ are essentially the same, the only change being in the sign of the parameters. Here we can see that the parameter estimates and model comparison criterion value of splogit and plogit are different and splogit has a clear edge in terms of model comparison. These results are consistent with analytical expectations and our simulations: splogit and plogit are equivalent when $r \leq 1$, but when $r > 1$, splogit performs better than plogit. In the gev model, the estimate of $\xi = -0.639$ now corresponds to a right-skewed model, and with a DIC of 23,101.5, the gev model fits worse than splogit and plogit. Again, all three of them fit better than the standard logit link. Figure 7 plots probability curves under splogit, plogit, logit and gev links as different covariates vary in modeling $P(y = 0)$. We see that the curve under splogit has a more flexible tail behavior that results in a better fit.

In one sense, it is not surprising that “Prevalence” is the predominant influence on the probability of co-occurrence. Let n_{ik} be the number of small watersheds in which species i is found in watershed k and n_k be the number of small watersheds in watershed k . Then $p_{ik} = (n_{ik} + 1)/(n_k + 2)$ is the probability that species i is

TABLE 3

Posterior median, HPD interval and DIC measure under splogit, plogit, gev, logit models of the Protea co-occurrence data. The first half of the table models $P(y = 1)$, while the second half models $P(y = 0)$. Bold numbers indicate significant factors. Smaller DIC indicates better fit

Variables	splogit		plogit		logit		gev	
	Median	HPD interval	Median	HPD interval	Median	HPD interval	Median	HPD interval
Modeling $P(y = 1)$								
Intercept	-0.294	(-0.354, -0.235)	-0.300	(-0.349, -0.248)	0.457	(0.407, 0.509)	0.019	(-0.023, 0.063)
GD	0.019	(0.005, 0.033)	0.019	(0.004, 0.032)	0.032	(0.013, 0.052)	0.020	(0.002, 0.036)
FSS	0.025	(-0.004, 0.059)	0.025	(-0.002, 0.050)	0.053	(0.010, 0.098)	0.027	(-0.007, 0.059)
Height	-0.023	(-0.038, -0.009)	-0.023	(-0.038, -0.009)	-0.032	(-0.052, -0.012)	-0.026	(-0.043, -0.008)
MMF	0.013	(-0.026, 0.052)	0.014	(-0.026, 0.050)	0.037	(-0.014, 0.092)	0.014	(-0.031, 0.058)
Pollination	-0.029	(-0.060, 0.004)	-0.029	(-0.057, 0.001)	-0.030	(-0.073, 0.011)	-0.038	(-0.074, -0.003)
SLA	-0.003	(-0.017, 0.011)	-0.003	(-0.018, 0.010)	-0.006	(-0.024, 0.012)	-0.006	(-0.022, 0.010)
LWR	-0.027	(-0.044, -0.012)	-0.027	(-0.043, -0.011)	-0.037	(-0.059, -0.017)	-0.031	(-0.051, -0.013)
Prevalence	0.844	(0.815, 0.873)	0.842	(0.818, 0.866)	1.146	(1.125, 1.164)	0.991	(0.963, 1.016)
r/ξ	0.380	(0.331, 0.0431)	0.376	(0.336, 0.423)			0.042	(0.020, 0.062)
DIC		23,002.0		23,001.0		23,335.2		23,006.3
Modeling $P(y = 0)$								
Intercept	0.295	(0.249, 0.333)	0.313	(0.293, 0.327)	-0.456	(-0.508, -0.406)	-0.579	(-0.609, -0.547)
GD	-0.019	(-0.033, -0.003)	-0.008	(-0.014, -0.003)	-0.032	(-0.051, -0.012)	-0.015	(-0.026, -0.003)
FSS	-0.024	(-0.051, 0.005)	-0.013	(-0.026, -0.002)	-0.054	(-0.097, -0.011)	-0.026	(-0.047, -0.003)
Height	0.023	(0.009, 0.037)	0.009	(0.003, 0.015)	0.031	(0.010, 0.051)	0.017	(0.005, 0.029)
MMF	-0.013	(-0.053, 0.023)	-0.008	(-0.024, 0.008)	-0.037	(-0.091, 0.016)	0.006	(-0.024, 0.038)
Pollination	0.030	(-0.001, 0.059)	0.009	(-0.004, 0.021)	0.029	(-0.014, 0.072)	0.022	(-0.002, 0.045)
SLA	0.003	(-0.010, 0.018)	0.002	(-0.004, 0.008)	0.006	(-0.013, 0.024)	0.006	(-0.005, 0.017)
LWR	0.028	(0.012, 0.045)	0.011	(0.004, 0.017)	0.037	(0.015, 0.058)	0.022	(0.009, 0.035)
Prevalence	-0.845	(-0.866, -0.818)	-0.340	(-0.380, -0.294)	-1.146	(-1.166, -1.127)	-0.677	(-0.686, -0.669)
r/ξ	2.624	(2.344, 2.948)	2.998	(2.693, 3.441)			-0.639	(-0.661, -0.619)
DIC		23,001.8		23,069.1		23,335.0		23,101.5

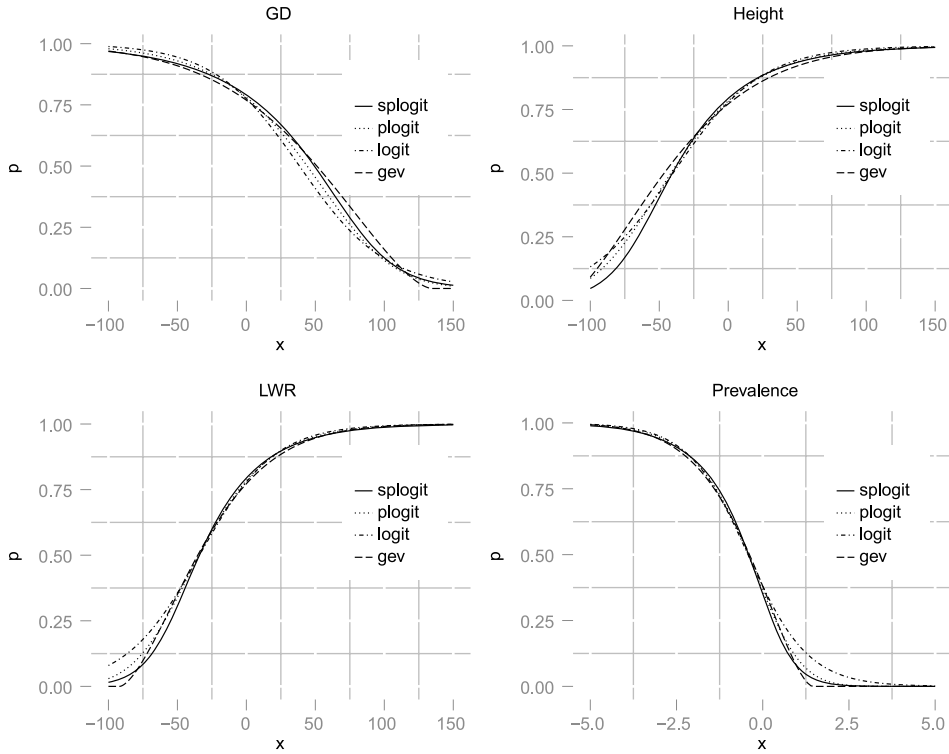


FIG. 7. Probability curves under *splogit*, *plogit*, *logit* and *gev* links as different covariates vary in modeling $P(y = 0)$. *GD*, *Height*, *LWR* and *Prevalence* are chosen since they are significant in all three models. Other covariates are fixed to be at the mean and coefficients are fixed to be at the posterior median.

found in watershed k , and the prevalence of the species pair i and j in watershed k is $p_{ik}p_{jk}$. In short, if two species are both common within a watershed, they are likely to co-occur, and if they are both uncommon, they are unlikely to co-occur.

In another sense, however, the importance of “Prevalence” may be surprising. It indicates that to a large extent species co-occur or not as if they were randomly assigned to small watersheds, suggesting that the biotic factors included in our analysis have relatively little effect on whether or not they co-occur. Competitive effects, if they existed, would lead similar species to co-occur less often than expected [corresponding to negative regression coefficients when modeling $P(y = 1)$]. It may not be surprising that competitive effects are small in this analysis, since the small watershed scale is much larger than the scale at which individual plants would compete, but habitat partitioning among watersheds would lead to the same pattern. Thus, the small negative coefficients associated with “Height” and “LWR” (leaf length-width ratio) suggest not only that competitive effects on the structure of *Protea* communities at this scale, if any, are small, but also that habitat partitioning has a similarly small influence on the probability of co-occurrence.

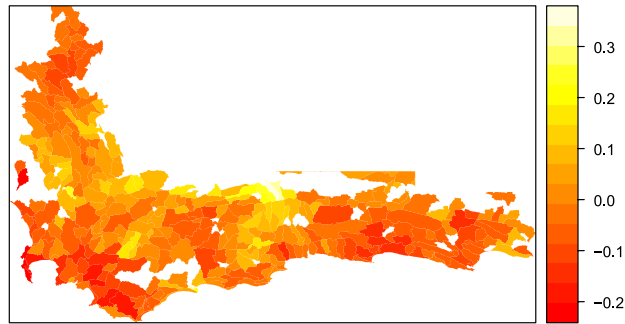


FIG. 8. Plot of spatial random effects in CFR under splogit fit.

We regard “GD” (phylogenetic distance) as a proxy for unmeasured traits that influence co-occurrence, and the positive coefficient on it may be surprising. Its magnitude is similar to that of the negative coefficients on “Height” and “LWR,” and it is small relative to “Prevalence,” but the positive sign indicates that closely related taxa occurring in the same large watershed are more likely to co-occur within small watersheds than expected by chance. Perhaps this association reflects some degree of habitat filtering [Shipley, Vile and Garnier (2006), Weiher and Keddy (1995)] on traits that we did not measure in this study.

The spatial clustering effects on co-occurrence probabilities are also obvious. In Figure 8 we see negative effects clustered in the east part and southwest corner of CFR, while positive effects are clustered in the middle part of the region. This clustering could reflect the interaction of rainfall (winter rainfall in the west, aseasonal in the east) and elevation (highest elevations inland, lowest along the coast). In future studies, we will explore both the patterns of co-occurrence at different spatial scales and the extent to which climate or other environmental features are associated with residual spatial variation in this analysis.

9. Discussion. In this paper we introduced a new family of flexible link functions. The proposed power link family can accommodate flexible skewness in both positive as well as negative directions, while retaining the baseline standard link as a special case. Simulation results show the proposed link performs well under various skewness scenarios. Also, the proposed link is computationally straightforward and efficient to implement. In addition, the power parameter idea illustrated here may be used to construct new link functions. For example, we could use an asymmetric link function c.d.f. as our baseline link. Using a power parameter might make a difference in bringing in desirable flexibility.

One potential problem with the proposed power link is that the power parameter r influences both the skewness and the mode of the link function p.d.f. Although this effect has been greatly reduced by scaling x by r in our model as defined in (3.2) and discussed in Section 3, the effect still exists with relatively large r values.

One solution might be to adjust the effect out with calculated mode values, yet it is computationally expensive especially under pt and pep links when there is no analytical solution for the mode of the c.d.f. function.

APPENDIX A: PROOF OF THEOREM 1

Observing that in our definition the power link functions F_r naturally split with respect to $r = 1$, we have

$$\begin{aligned} & \int_{\mathfrak{R}^+} \int_{\mathfrak{R}^k} \mathbf{L}(\boldsymbol{\beta}, r|\mathbf{y})\pi(r) d\boldsymbol{\beta} dr \\ &= \int_0^1 \int_{\mathfrak{R}^k} \mathbf{L}(\boldsymbol{\beta}, r|\mathbf{y})\pi(r) d\boldsymbol{\beta} dr + \int_1^\infty \int_{\mathfrak{R}^k} \mathbf{L}(\boldsymbol{\beta}, r|\mathbf{y})\pi(r) d\boldsymbol{\beta} dr. \end{aligned}$$

Clearly, the link function in the latter part is the mirror reflection of the first part, in other words, $F(x, r) = 1 - F(-x, \frac{1}{r})$, therefore, we only need to prove the first part of the integration is finite. For the baseline link F_0 we have

$$\begin{aligned} F_0(x) &= \int_{\mathfrak{R}} I(u \geq -x) d(-F_0(-u)), \\ 1 - F_0(x) &= \int_{\mathfrak{R}} I(u > x) dF_0(u). \end{aligned}$$

Since F_0 is continuous, by Fubini’s theorem we have

$$\begin{aligned} & \int_{\mathfrak{R}^k} \mathbf{L}_0(\boldsymbol{\beta}|\mathbf{y}) d\boldsymbol{\beta} \\ &= \int_{\mathfrak{R}^k} \prod_{i=1}^n F_0(x'_i \boldsymbol{\beta})^{y_i} [1 - F_0(x'_i \boldsymbol{\beta})]^{1-y_i} d\boldsymbol{\beta} \\ &= \int_{\mathfrak{R}^k} \int_{\mathfrak{R}^n} I(u_i \geq \tau_i x'_i \boldsymbol{\beta}, 1 \leq i \leq n) d(\tau_i F_0(\tau_i u_i)) d\boldsymbol{\beta} \\ &= \int_{\mathfrak{R}^n} \int_{\mathfrak{R}^k} I(\mathbf{X}^* \boldsymbol{\beta} \leq \mathbf{u}) d\boldsymbol{\beta} d\mathbf{F}_0(\mathbf{u}), \\ &< \infty, \end{aligned}$$

where $\mathbf{F}_0(\mathbf{u}) = (\tau_1 F_0(\tau_1 u_1), \tau_2 F_0(\tau_2 u_2), \dots, \tau_n F_0(\tau_n u_n))$.

Then under the power link, since $|F_0| \leq 1$ and $\pi(r)$ is a proper density, we have

$$\begin{aligned} & \int_0^1 \int_{\mathfrak{R}^k} \mathbf{L}(\boldsymbol{\beta}, r|\mathbf{y}) d\boldsymbol{\beta} \pi(r) dr \\ &= \int_0^1 \int_{\mathfrak{R}^n} \int_{\mathfrak{R}^k} I(\mathbf{X}^* \boldsymbol{\beta} \leq \mathbf{u}) d\boldsymbol{\beta} d\mathbf{F}(\mathbf{u}, r) \pi(r) dr \\ &= \int_0^1 \int_{\mathfrak{R}^n} \int_{\mathfrak{R}^k} I(\mathbf{X}^* \boldsymbol{\beta} \leq \mathbf{u}) \mathbf{F}_0^{r-1}(\mathbf{u}) d\boldsymbol{\beta} d\mathbf{F}_0(\mathbf{u}) \pi(r) dr \end{aligned}$$

$$\begin{aligned} &\leq \int_0^1 \int_{\mathbb{R}^n} \int_{\mathbb{R}^k} I(\mathbf{X}^{*'} \boldsymbol{\beta} \leq \mathbf{u}) d\mathbf{F}_0(\mathbf{u}) \pi(r) dr \\ &< \infty. \end{aligned}$$

APPENDIX B: PROOF OF THEOREM 2

By Theorem 1, we only need to prove the theorem for $r = 1$. Let $\mathbf{u} = (u_1, u_2, \dots, u_n)'$ be i.i.d. random variables with distribution function F_0 . Now, under the condition listed in Theorem 2, it follows directly from Lemma 4.1 of Chen and Shao (2001) that there exists a constant K such that

$$\|\boldsymbol{\beta}\| \leq K \max_{1 \leq i \leq n} |u_i|,$$

whenever

$$\mathbf{X}^{*'} \boldsymbol{\beta} \leq \|\mathbf{u}\|.$$

Therefore, following the derivation in Theorem 1 under the baseline link F_0 , we have

$$\begin{aligned} &\int_{\mathbb{R}^k} \mathbf{L}_0(\boldsymbol{\beta}|\mathbf{y}) d\boldsymbol{\beta} \\ &= \int_{\mathbb{R}^n} \int_{\mathbb{R}^k} I(\mathbf{X}^{*'} \boldsymbol{\beta} \leq \mathbf{u}) d\boldsymbol{\beta} d\mathbf{F}_0(\mathbf{u}) \\ &\leq K \int_{\mathbb{R}^n} \max_{1 \leq i \leq n} |u_i| d\mathbf{F}_0(\mathbf{u}) \\ &\leq K \sum_{1 \leq i \leq n} E|u_i|^k. \end{aligned}$$

Clearly, in the logistic and exponential power cases we have $E|u_i|^k < \infty$, while in the Student t case the same condition will hold as long as the degrees of freedom $\nu > k$.

Acknowledgments. The authors thank Dr. Tony Rebelo and the Protea Atlas project for providing the *Protea* occurrence data in this paper.

SUPPLEMENTARY MATERIAL

Protea species co-occurrence data set (DOI: 10.1214/13-AOAS663SUPP;.zip). We provide the *Protea* species co-occurrence data set used in the data analysis section.

REFERENCES

- ALBERT, J. H. and CHIB, S. (1993). Bayesian analysis of binary and polychotomous response data. *J. Amer. Statist. Assoc.* **88** 669–679. [MR1224394](#)
- ARANDA-ORDAZ, F. J. (1981). On two families of transformations to additivity for binary response data. *Biometrika* **68** 357–363. [MR0626394](#)
- ARNOLD, B. C. and GROENEVELD, R. A. (1995). Measuring skewness with respect to the mode. *Amer. Statist.* **49** 34–38. [MR1341197](#)
- BANERJEE, S., CARLIN, B. P. and GELFAND, A. E. (2004). *Hierarchical Modeling and Analysis for Spatial Data* **101**. Chapman & Hall/CRC, Boca Raton, FL.
- BESAG, J. and GREEN, P. J. (1993). Spatial statistics and Bayesian computation. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **55** 25–37. [MR1210422](#)
- CHEN, M.-H., DEY, D. K. and SHAO, Q.-M. (1999). A new skewed link model for dichotomous quantal response data. *J. Amer. Statist. Assoc.* **94** 1172–1186. [MR1731481](#)
- CHEN, M.-H. and SHAO, Q.-M. (2001). Propriety of posterior distribution for dichotomous quantal response models. *Proc. Amer. Math. Soc.* **129** 293–302. [MR1694452](#)
- CHIB, S. and JELIAZKOV, I. (2006). Inference in semiparametric dynamic models for binary longitudinal data. *J. Amer. Statist. Assoc.* **101** 685–700. [MR2256181](#)
- CZADO, C. (1994). Parametric link modification of both tails in binary regression. *Statist. Papers* **35** 189–201.
- CZADO, C. and SANTNER, T. J. (1992). The effect of link misspecification on binary regression inference. *J. Statist. Plann. Inference* **33** 213–231. [MR1190622](#)
- ELTON, C. (1946). Competition and the structure of ecological communities. *The Journal of Animal Ecology* **15** 54–68.
- GELFAND, A. E., DEY, D. K. and CHANG, H. (1992). Model determination using predictive distributions with implementation via sampling-based methods. Technical report, DTIC.
- GILKS, W. R., BEST, N. G. and TAN, K. K. C. (1995). Adaptive rejection Metropolis sampling within Gibbs sampling (Corr: 97V46 P541–542 with R. M. Neal). *J. Appl. Stat.* **44** 455–472.
- GRASS DEVELOPMENT TEAM (2008). Geographic resources analysis support system (GRASS). Available at <http://grass.osgeo.org>.
- GUERRERO, V. M. and JOHNSON, R. A. (1982). Use of the Box–Cox transformation with binary response models. *Biometrika* **69** 309–314. [MR0671968](#)
- GUPTA, R. D. and GUPTA, R. C. (2008). Analyzing skewed data by power normal model. *TEST* **17** 197–210. [MR2393359](#)
- IBRAHIM, J. G., CHEN, M. H. and SINHA, D. (2005). *Bayesian Survival Analysis*. Wiley, New York.
- JIANG, X., DEY, D. K., PRUNIER, R., WILSON, A. M. and HOLSINGER, K. E. (2013). Supplement to “A new class of flexible link functions with application to species co-occurrence in cape floristic region.” DOI:[10.1214/13-AOAS663SUPP](https://doi.org/10.1214/13-AOAS663SUPP).
- JONES, M. C. (2004). Reply to Comments on “Families of distributions arising from distributions of order statistics.” *TEST* **13** 1–43.
- KIM, S., CHEN, M.-H. and DEY, D. K. (2008). Flexible generalized t -link models for binary response data. *Biometrika* **95** 93–106. [MR2409717](#)
- LIU, C. (2004). Robit regression: A simple robust alternative to logistic and probit regression. In *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives* (A. Gelman and X.-L. Meng, eds.) 227–238. Wiley, New York.
- LUNN, D. J., THOMAS, A., BEST, N. and SPIEGELHALTER, D. (2000). WinBUGS—A Bayesian modelling framework: Concepts, structure, and extensibility. *Statist. Comput.* **10** 325–337.
- MCCULLAGH, P. and NELDER, J. A. (1989). *Generalized Linear Models*, 2nd ed. Chapman & Hall/CRC, Boca Raton, FL.

- MUDHOLKAR, G. S. and GEORGE, E. O. (1978). A remark on the shape of the logistic distribution. *Biometrika* **65** 667–668.
- NAGLER, J. (1994). Scobit: An alternative estimator to logit and probit. *American Journal of Political Science* **38** 230–255.
- PALMGREN, J. (1989). Regression models for bivariate binary responses. Working Paper 101, UW Biostatistics Working Paper Series.
- PARADIS, E., CLAUDE, J. and STRIMMER, K. (2004). APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics* **20** 289–290.
- PLUMMER, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In *Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003)* 20–22.
- SAMEJIMA, F. (2000). Logistic positive exponent family of models: Virtue of asymmetric item characteristic curves. *Psychometrika* **65** 319–335.
- SANDERSON, M. J. (2003). r8s: Inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics* **19** 301–302.
- SHIPLEY, B., VILE, D. and GARNIER, É. (2006). From plant traits to plant communities: A statistical mechanistic approach to biodiversity. *Science* **314** 812–814. [MR2264265](#)
- SPIEGELHALTER, D. J., BEST, N. G., CARLIN, B. P. and VAN DER LINDE, A. (2002). Bayesian measures of model complexity and fit. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **64** 583–639. [MR1979380](#)
- STUKEL, T. A. (1988). Generalized logistic models. *J. Amer. Statist. Assoc.* **83** 426–431. [MR0971368](#)
- SUBBOTIN, M. T. (1923). On the law of frequency of error. *Mat. Sb.* **31** 296–301.
- VALENTE, L. M., REEVES, G., SCHNITZLER, J., MASON, I. P., FAY, M. F., REBELO, T. G., CHASE, M. W. and BARRACLOUGH, T. G. (2010). Diversification of the African genus *Protea* (Proteaceae) in the Cape biodiversity hotspot and beyond: Equal rates in different biomes. *Evolution* **64** 745–760.
- VAMOSI, S. M., HEARD, S. B. and WEBB, C. O. (2009). Emerging patterns in the comparative analysis of phylogenetic community structure. *Mol. Ecol.* **18** 572–592.
- WANG, X. and DEY, D. K. (2010). Generalized extreme value regression for binary response data: An application to B2B electronic payments system adoption. *Ann. Appl. Stat.* **4** 2000–2023. [MR2829944](#)
- WEIHER, E. and KEDDY, P. A. (1995). The assembly of experimental wetland plant communities. *Oikos* **73** 323–335.

X. JIANG
D. K. DEY
DEPARTMENT OF STATISTICS
UNIVERSITY OF CONNECTICUT
215 GLENBROOK RD. UNIT 4120
STORRS, CONNECTICUT 06269
USA
E-MAIL: xun.jiang@uconn.edu
dipak.dey@uconn.edu

A. M. WILSON
DEPARTMENT OF ECOLOGY
AND EVOLUTIONARY BIOLOGY
YALE UNIVERSITY
165 PROSPECT STREET
NEW HAVEN, CONNECTICUT 06520
USA
E-MAIL: adam.wilson@yale.edu

R. PRUNIER
DEPARTMENT OF BIOLOGICAL
AND ENVIRONMENTAL SCIENCES
WESTERN CONNECTICUT STATE UNIVERSITY
181 WHITE STREET
DANBURY, CONNECTICUT 06810
USA
E-MAIL: prunierr@wcsu.edu

K. E. HOLSINGER
DEPARTMENT OF ECOLOGY
AND EVOLUTIONARY BIOLOGY
UNIVERSITY OF CONNECTICUT
75 N. EAGLEVILLE ROAD, UNIT 3043
STORRS, CONNECTICUT 06269
USA
E-MAIL: kent.holsinger@uconn.edu